

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Hyytinen, Ari; Lundberg, Sofia; Toivanen, Otto

Title: Design of public procurement auctions : evidence from cleaning contracts

Year: 2018

Version: Accepted version (Final draft)

Copyright: © 2018, The RAND Corporation.

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Hyytinen, A., Lundberg, S., & Toivanen, O. (2018). Design of public procurement auctions : evidence from cleaning contracts. *RAND Journal of Economics*, 49(2), 398-426.

<https://doi.org/10.1111/1756-2171.12232>

DESIGN OF PUBLIC PROCUREMENT AUCTIONS: EVIDENCE FROM CLEANING CONTRACTS

Ari Hyytinen, Sofia Lundberg and Otto Toivanen *

This draft version:

November 26, 2017

Abstract

We analyze a regime change from beauty contests to first-price sealed-bid and scoring auctions, using Swedish data on public procurement of cleaning services. In beauty contests, the lowest bid often lost, leaving substantial money on the table. The procurement costs were similar before and after the regime change: i) Entry strongly decreases the procurement cost, but did not change. Entry would have decreased had the municipalities not adjusted the objects of auctions. ii) Municipalities favored inhouse suppliers in the old regime, leading to more aggressive bidding by others. With favoritism reduced, these changes balanced each other out.

KEYWORDS: public procurement, auction design, entry, beauty contests, scoring

* Ari Hyytinen, School of Business and Economics, University of Jyväskylä; ari.hyytinen@jyu.fi. Sofia Lundberg: Department of Economics, Umeå School of Business and Economics at Umeå University; sofia.lundberg@umu.se. Otto Toivanen: Aalto University School of Business and KU Leuven; Otto.Toivanen@aalto.fi. This article supersedes the earlier procurement analyses of ours that circulated under various titles, such as “Favoritism in public procurement: Evidence from Sweden” and “Politics and procurement: Evidence from cleaning contracts”. Those earlier versions had a different focus and used part of the same contract data used in this article. We would like to thank a number of colleagues and seminar participants for useful comments on those earlier versions, and Marc Rysman (Editor), three anonymous referees, Rob Porter, Tuomas Takalo, Janne Tukiainen, Roope Uusitalo and seminar participants at EARIE 2016 (Lisbon) and VATT (Helsinki 2017) for comments and discussions on the current one. The usual caveat applies.

JEL codes: H57, D44, P16

1 Introduction

Public procurement constitutes a large part of economic activity in developed and developing economies.ⁱ Unfortunately, inappropriate and inefficient practices are thought to be widespread in public procurement (e.g., Bandiera, Prat and Valletti, 2009; Cai, Zhang and Henderson, 2013). Recommendations for greater integrity and adoption of rule-based practices are therefore common and the benefits of using formal competitive auctions are often strongly advocated (Klemperer, 2002; Tadelis, 2012; Bajari, Houghton, & Tadelis, 2014). Though increasing buyers' discretion may be beneficial when quality is only partly contractible (e.g., Coviello, Guglielmo and Spagnolo, 2017), rule-based competitive auctions are in general believed to be less prone to inappropriate and inefficient practices. One reason for this belief is that in such auctions there are explicit criteria for selecting the winner.ⁱⁱ This article provides an analysis of the consequences of a regime change that induced a shift away from discretionary beauty contests to a more rule-based procurement environment, in which only first price sealed bid or scoring auctions were allowed.

Our data come from Swedish public procurement auctions of a clearly defined low-tech product, internal cleaning service contracts. The data cover two regimes that differ in how the law regulated public procurement. In the old regime, the Swedish procurement law allowed the municipalities exceptionally high degrees of freedom in choosing how to procure the services and in how to pick the winner.ⁱⁱⁱ The municipalities ended up using beauty contest auctions. In such an auction, the auctioneer does not commit to an award (allocation) rule (Klemperer, 2002; Yoganarasimhan, 2016). During the new regime, a much stricter procurement law was in place and the municipalities had to commit to a first price sealed

bid auction or to a scoring auction (Che, 1993; Asker & Cantillon, 2008; 2010) with an explicit award rule.

The stricter procurement law disciplined the Swedish procurement bureaucrats by forcing them to abandon the practice of using flexible beauty contests. A flipside of this change was that rule-based procurement regulation and more formal procurement processes, augmented by more detailed product requirements and bidder qualifications, may make procurement burdensome for potential bidders and thereby increase entry costs (Bandiera, Prat, and Valletti, 2009). Our primary goal is to understand what the regime change meant for municipalities' procurement costs by analyzing the determinants of winning bid and entry (participation) as well as the behavior of procurement bureaucrats, including potential favoritism.

The cleaning contracts procured by Swedish municipalities provide a good testing ground for us, because cleaning services have a very simple production process, are simple to contract on and, as we will demonstrate, do not allow significant quality differentiation *ex ante* or *ex post*. The product we study is thus much simpler and more homogenous than most of the products that have been considered in prior work. This has three implications for applicable theory: First, it is hard to believe that the Swedish municipalities could have obtained significant welfare gains by *ex ante* tilting the award rule in some non-price dimension via a scoring rule. This is in contrast to, e.g., the US highway procurements studied by Lewis and Bajari (2011). Second, *ex post* negotiations of significant adaptations to the scope of work or of cost overruns are unlikely and indeed were rare, unlike what has been found for some of the more complicated contracts and products (Bajari, Houghton, and Tadelis, 2014; Decarolis, 2013; Decarolis, 2014; Decarolis and Palumbo, 2015). Neither is there much scope for provision of non-

contractible quality (e.g., Cameron 2000; Spagnolo, 2012; Tadelis, 2012; Coviello, Guglielmo and Spagnolo, 2017). Finally, there is less uncertainty in the costs of performing the cleaning job over the duration of the contract than there is, say, in completing a major construction work (Spulber, 1990).

These considerations suggest that there should be relatively few reasons to depart from the policy of granting the contract to the lowest bidder (for a similar argument in a different context, see Di Tella and Schargrotsky 2003, Bandiera, Prat and Valletti 2009). This implies that holding other things constant, the procurement costs ought to be lower in the new regime, which only allowed the use of more rule-based and arguably more competitive first price or scoring auctions.

Despite the simplicity of the procured service, our findings portray a complex picture of the mechanisms at work. In the beauty contests of the old regime, the lowest bidder won rarely, leaving a large amount of money on the table. And yet, procurement costs did not decrease in the new regime.^{iv} We find two reasons for this: *First*, we uncover a strong inverse relation between actual entry and winning bids. Securing entry was therefore essential to competitive outcomes.^v The new regime's more rule-based procurement made auctions potentially more competitive and participation more burdensome for potential bidders. As a consequence, entry would have decreased, had the municipalities not adjusted the objects (contracts) of the procurement auctions after the regime change in a way that attracted further participation.

Second, in the old regime the Swedish municipalities were less price-sensitive and favored inhouse units, which are publicly-owned organizations of the municipalities that participated in the auctions as self-standing operative units. Although the favored bidder type was not declared in procurement documents, other bidders apparently learnt about it and reacted by bidding more aggressively

when an inhouse unit participated. This is consistent with what the prior work on favoritism (Myerson, 1981; McAfee and McMillan, 1989) and subsidies (preference programs) in auctions suggest (Marion, 2007; Krasnokutskaya and Seim, 2011; Athey, Coey, and Levin, 2013). When the scope for favoritism was reduced because of the regime change, the price sensitivity of the municipalities increased. The distribution of the (lowest) submitted bids however also changed as the other (than inhouse) bidders bid less aggressively. Had the bid distribution not changed, the shift from the beauty contests to rule-based auctions would have reduced procurement costs.

We also show that favoritism likely reflected the municipalities' willingness to sustain the viability of the inhouse units and their employment. The rule-based environment curbed this practice and decreased entry of inefficient suppliers, suggesting gains in allocative efficiency.

Our analysis is related to two strands of the literature. First, in contrast to much of the earlier literature, the beauty contests that we study were not based on an explicit price-preference rule. We therefore augment the literature which has explored how bidding and participation adjusts to changes in auction design and environment. Using data on US highway procurement auctions, Marion (2007) shows that an explicit bid preference program for small businesses increased the procurement costs a little, possibly due to reduced entry by larger, lower cost suppliers. Krasnokutskaya and Seim (2011) find that the small business preference program changed firms' incentives to participate in the highway procurement auctions. Lewis and Bajari (2011) analyze US highway and roadwork contracts and find that the procurement costs were higher in the scoring auctions, but this was counterbalanced by faster completion. Using data US timber sale auctions, Athey, Coey and Levin (2013) show that restricting entry decreases the sales revenue

(and efficiency) and that subsidizing smaller bidders would increase the revenue. Branzoli and Decarolis (2015) use data on Italian road construction and maintenance contracts to document that the use of first price auctions resulted in reduced entry and subcontracting. Decarolis (2013) shows that a switch to using first price auctions lowered the level of winning bids but did so at the cost of worse ex post performance (cost overruns, delays in completion) in Italy. Other relevant work include Cameron (2000), Eklöf (2005), Bajari, McMillan and Tadelis (2009), Athey, Levin and Seira (2011) and Coviello and Mariniello (2014). For example, Bajari, McMillan and Tadelis (2009) compare auctions and negotiations in procurement and stress tradeoffs between hard-to-observe quality and price when objects are complex and contractual design incomplete. In our case, exactly the opposite holds: Objects are simple and contractual design complete, at least when compared to the procurement of large construction projects, aircrafts and the like.

Our analysis is also related to endogenous participation and selective entry in auctions (e.g. Li and Zheng, 2009; 2012; Li and Zhang, 2010; Marmer, Shneyerov and Xu, 2013; Xu, 2013; Roberts and Sweeting, 2013; Gentry and Li, 2014). Cai, Zhang and Henderson (2013) show that compared to English auctions, entry was deterred and prices were lower in corruption-prone two-stage auctions of the lease contracts of urban land in China. Bhattacharya, Roberts and Sweeting (2014) analyze how the pre-auction allocation of entry-rights compares to unregulated entry in the context of first-price auctions. Coviello, Guglielmo and Spagnolo (2017) show that buyer discretion in restricting entry did not lead to worsening of ex ante or ex post procurement outcomes in Italian procurement auctions of relatively complex public works in the construction sector.

In the next section, we describe the institutional environment and the public procurements. In Section 3, we outline what the available auction theory pre-

dicts we should expect to see in our data. Section 4 is devoted to describing our bid and auctions data as well as auxiliary interviews of industry practitioners. Guided by theory, we then explore the importance of endogenous bidder participation in Section 5. There we also demonstrate the role that the characteristics of the object being auctioned played. In Section 6, we analyze a number of questions related to how the procurements worked: We start by studying the price sensitivity of municipalities, followed by an analysis of who was favored and how other bidders reacted to favoritism. We then proceed to analyze how favoritism may have affected procurement costs and whether there was active or passive waste in the sense of Bandiera, Pratt and Valletti (2009). In Section 7 we discuss efficiency. We offer brief conclusions in section eight. We report a number of auxiliary analyses in an online supplement.

2 Public procurement of cleaning services in Sweden

The institutional environment

Our procurement data cover two regimes (the law changed in 2007/2008), and due to availability, come from two separate time periods: The data from the old regime runs from 1990 to 1998 and that of the new regime from 2009 to 2010. Conditional on a municipality having decided to procure, the two regimes differed in the types of auctions that the municipalities were, by law, required to organize.^{vi}

During the old regime, the law on public procurements in Sweden stated that the lowest bidder should in general, have won (see Appendix A). However, due to an exception to this rule, municipalities had the freedom to deem that some other bid than the lowest was most advantageous economically when quality, environmental aspects, service and maintenance etc. were also taken into account.

The old procurement law thus allowed the municipalities exceptionally high degrees of freedom in picking the winner *without* committing to an award rule (i.e., to organize *Beauty contests*).

During the new regime, the law was more stringent and imposed more constraints. The new law stated that the municipalities were allowed to use either the lowest-price-principle, which corresponds to a sealed bid first price auction (*First price*), or the so-called economically most advantageous tender principle, which corresponds to a sealed bid first scoring auction (*Scoring*). The law required that the call for tender explicitly specify how bids will be evaluated and ranked after the bids have been screened against mandatory qualification and exclusion criteria. This scoring rule was to be posted in advance. Explicit weighing of different criteria was the guiding principle with scoring (for a more detailed account, see Bergman and Lundberg 2013).

The two regimes differed in how entry was regulated. During both regimes, the law allowed the municipalities to decide whether to allow free entry or not (see Appendix A). Conditional on restricting entry, the law allowed the municipalities to freely decide how many and which suppliers to invite. Municipalities often regulated entry in the old regime, but very often allowed free entry in the new regime. Unlike in the old regime, the process of awarding a cleaning contract in the new regime followed a two-step procedure after the bids had been submitted: In the first step, bids were screened against mandatory qualification and exclusion criteria. The problematic bids that did not meet these criteria were disqualified. In the second step, qualified bids were evaluated according to the pre-specified award rule (i.e., *First price* or *Scoring*).

Procurement practices

In our data, a procurement event is an instance where a municipality purchases cleaning services for one or more of its buildings and premises. Such a procurement event can consist of one or more sealed-bid auctions, each allocating a single cleaning service contract to the winning bidder. One contract can cover the cleaning of more than one building. However, combinatorial bidding was not applied: During both regimes, the bidders were allowed to submit only one bid per contract and the municipalities should have made decisions contract-by-contract.

The timing of events in the procurement process was roughly as follows: First, the municipal procurement agent chose the auction format and entry mode (free or restricted entry). The municipal procurement bureaucrats also had to make decisions about the object of each separate auction, such as which buildings a single contract covered, what the cleaning frequency was, and how long the contract was. As far as we know, reserve prices were not set. The procurements were openly and centrally advertised during both regimes, although the media changed over time. After seeing the call for tenders, a two-stage auction followed: In the first stage, the (invited) suppliers made the participation decision. The firms were not informed about who were invited when entry was restricted, but this does not mean that they could not have learnt it somehow. In the second stage, the participating suppliers submitted their bids. The municipal procurement bureaucrats then chose the winner according to the award criteria to which they had committed (if any).

Like in the US timber auctions studied by Athey, Coey and Levin (2013), there are different types of bidders in our data: Large nationally active firms, regional firms, local firms operating in a single municipality, and “inhouse units” of the municipalities. It is unlikely that these bidders were symmetric. In particular,

the inhouse units differed from the other bidders. They are publicly owned organizations that share features of a public firm and that participated in the auctions as self-standing operative units. It is hard to pin down exactly what the objective function of these publicly-controlled units was; Boycko, Shleifer and Vishny (1996) have for example argued that public firms of this type may be under pressure to increase local employment. The presence of such a bidder introduces an asymmetry and an aspect of mixed oligopoly to competition (de Fraja & Delbono, 1989).

The difference in entry regulation between the two regimes affected participation through two channels: First, the introduction of the two-step procedure changed how municipalities could exclude problematic bidders. In the old regime, one instrument for that was to not have free entry and to invite only “desirable” bidders, i.e., bidders who fulfilled (unannounced) prequalification requirements. In the new regime, free entry became the norm, but the mandatory qualification and exclusion criteria were then used to disqualify problematic bidders. Second, the regime change increased bid preparation costs, as the content of the calls for tender and the preparation of a bid became more detailed and involved (red tape).^{vii} As e.g. Bandiera, Prat and Valletti (2009) have emphasized, rule-based procurement regulation and more formal procurement processes, characterized by detailed product requirements and bidder qualifications, make procurement burdensome for potential bidders and increase the entry costs. In the new regime, suppliers had to pay more attention to the mandatory qualification and exclusion criteria in order to reduce the risk that their bid would be disqualified.

As to the choice of the auction format, a look at the procurement documents of the old regime reveals that *no* municipality committed to an award rule. The procurement bureaucrats chose to use beauty contests, where the rule, if any, to

choose the winner was not disclosed to the bidders. It is illustrative of the atmosphere of the time that the freedom allowed by the law to deviate from choosing the lowest bid was seen as beneficial. The following quote from a book by a public sector lawyer testifies to this:

“The tender having the lowest price offered should be accepted. If it has been stated in the advertisement that the most economically advantageous tender will be accepted, factors specified therein can be taken into consideration in the assessment of tenders. The factors can be stated according to a degree of priority (LOU 1 ch. 22§), however this is not a requirement. On the contrary, it can be advantageous to state in the advertisement that such factors are non-prioritized, as this increases the possibility of being able to choose the contractor.” (Löfving, 1994) p.65 (our translation, and italics).

Who benefits and why is open to debate, as this kind of flexibility opens the door also for inefficient and undesirable practices. For example, Tadelis (2012, p. 301-302) argues that *“allowing for greater discretion in contractor selection increases the possibility of favoritism, kick-backs and political corruption”*.

The situation was quite different in the new regime as the municipalities had to choose either a first price auction or a scoring auction. A perusal of the procurement documents of the new regime shows that all municipalities specified a scoring rule, unless they used the lowest-price principle (first price auction). The specific design of scoring rules differ (Bergman & Lundberg, 2013), but it were described in detail either in words or in combination with a mathematical expres-

sion (see Appendix A for a summary of the mandatory qualification and exclusion criteria and scoring rules).^{viii}

The object of procurements: Cleaning service contracts

Delivery of internal regular cleaning services is a comparatively simple production process. Cleaning of floors of various types of rooms and corridors, desktop surfaces and lavatories requires relatively unskilled labor and moderate amounts of capital (e.g., cleaning equipment, washers, and vehicles) and intermediate inputs (e.g., cleaning substances and tools).

The extensive documentation available to us on the technical specifications of the procurements and the specifics of the bids suggest that there was limited room for quality differences in the cleaning services. Nor was there much scope for non-contractible quality (Spagnolo, 2012). The most compelling support for these claims is provided by the technical specifications of the procurement instructions. The procurement instructions were in general quite detailed already in the old regime. Besides including a detailed description of the premises to be cleaned, the frequency of cleaning, cleaning method, cleaning substances that are preferred and cleaning equipment that is to be employed, they also went into much more minute detail. For example, it was common to state requirements as to the professional education of cleaning staff to be used. As if this wasn't enough, in several instances the procurement instructions went into great detail as to how each space (e.g., classroom, toilet) was to be cleaned. The instructions for the preparation of a bid were quite similar in the new regime, but they were even more detailed and took the form of mandatory qualification and exclusion criteria.

The monitoring of *delivered* cleaning was often specified, too. It was standard to require the supplier to inform the municipality on several features of the

working process, to provide records of hours of work, workforce and machinery employed etc. Importantly for us, according to our auxiliary interview evidence, the industry practitioners seem to agree with the view that delivered quality did not differ between the two regimes.

3 Theoretical framework

The above description of the institutional environment and procurement practices suggests that we need a theoretical framework that encompasses i) the regime change from the beauty contests to using rule-based auction formats; ii) the changing entry costs and endogenous participation by asymmetric bidders; and iii) the governance of entry by the organizers of the procurements.^{ix}

Discretionary vs. rule-based auction formats

A shift away from using the discretionary beauty contests to only using more rule-based competitive auctions is frequently believed to improve efficiency and reduce inappropriate practices (Klemperer, 2002; Tadelis, 2012; Bajari, Houghton, and Tadelis, 2014). The first price sealed bid auction is optimal and cost-minimizing for a fixed number of bidders, when the bidders are symmetric and have independent private costs (Myerson, 1981; Riley and Samuelson, 1981). Asker and Cantillon (2008) have shown that when differentiated goods are purchased, simple linear scoring auctions dominate beauty contests and price-only auctions with minimum quality thresholds in the sense that the latter two generate lower (or equal) expected utility for the organizer of the procurement (Che, 1993; Asker and Cantillon, 2010).

Due to the simplicity of the product we study, *ex ante* tilting the award rule in some non-price dimension via a scoring rule is unlikely to lead to significant

welfare gains (in contrast to, e.g., the US highway procurements studied by Lewis and Bajari 2011). Moreover, unlike what has been found for some of the more complicated contracts (Bajari, Houghton, and Tadelis, 2014; Decarolis, 2014; Decarolis and Palumbo, 2015), ex post negotiations of significant adaptations to the scope of work or of cost overruns are unlikely. There also is less scope for provision of non-contractible quality (Spagnolo, 2012; Coviello, Guglielmo, and Spagnolo, 2017) and less cost uncertainty over the duration of the contract than there is, say, in completing a major construction work (Spulber, 1990). These considerations suggest that the procurement auctions of cleaning contracts are quite similar to auctions for standardized goods. Theory therefore suggests that holding other things constant, the procurement costs ought to be lower on average in the more rule-based new regime.

Favoritism and endogenous entry

Theoretical guidance gets more involved when bid preferences (favoritism) and endogenous entry are allowed. When some types of bidders are preferred either explicitly (Myerson, 1981; McAfee and McMillan, 1989; Vagstad, 1995) or implicitly, as may happen in beauty contests (Klemperer, 2002; Yoganarasimhan, 2016) the basic theoretical prediction is that favoring high-cost suppliers may reduce procurement costs due to the increased aggressiveness of the un-favored suppliers (McAfee and McMillan, 1989). This prediction however holds in special contexts only. For example, if endogenous participation is allowed (Marion, 2007; Krasnokutskaya and Seim, 2011; Athey, Levin, and Seira, 2011), there are at least three mechanisms that may increase procurement costs: Favored suppliers can increase their bids; un-favored low-cost suppliers may win less often; and the willingness of different supplier types to participate may change. The last of these

mechanisms means that when the procurement agent prefers a certain bidder type, endogenous participation can reduce overall entry and/or lead to a compositional change towards high cost bidders. The prediction for the cost effect is ambiguous, but the empirical results of Marion (2007) and Krasnokutskaya and Seim (2011) suggest that the participation margin matters a lot and that if high-cost bidders are favored, we should expect higher procurement costs in the beauty contests of the old regime.

Favoritism in public procurement may also take the form of outright corruption (Compte, Lambert-Mogiliansky, and Verdier, 2005; Menezes and Monteiro, 2006; Tran, 2008; Bandiera, Prat, and Valletti, 2009; Cai, Zhang, and Henderson, 2013). The precise mechanism of how bribes are paid and which types of favors the procurement agent provides depend on the context and how much discretion the procurement agent has when allocating the contracts. For example, Tran (2008) analyzes how the bids and bribes paid varied when the law governing the procurements changed. Using data from a bribe-paying firm, Tran compares a period when no auctions were required to periods when mandatory best-value auctions or best-price auctions were used. In Tran's data, the key mechanism through which corruption worked were secret auctions that were run before the formal auction. The tender requirements of the formal auction were then designed so that they favored the winner of the secret (bribery) auction.

The mechanical effect that results from a supplier expecting to pay a bribe is to increase its bid price, especially if the arrangement means that it can expect to win the contract at a higher price. Corruption may also lessen competition because it increases the risk of collusive coordination by bidders (Compte, Lambert-Mogiliansky, and Verdier, 2005). It may in addition lead to a choice of auction format that results in higher prices and allows corruption to continue (Cai, Zhang,

and Henderson, 2013). These considerations suggest that the procurement costs ought to be higher in the beauty contests of the old regime.

One likely consequence of more rule-based procurement processes and more detailed product requirements and bidder qualifications was that it increased entry costs of potential bidders. The theory of two-stage auctions with costly entry – see e.g. Ye (2007), Li and Zheng (2009; 2012), Marmer, Shneyerov and Xu (2013) Bhattacharya, Roberts and Sweeting (2014) – suggest that a number of (offsetting) effects may be at work. If potential bidders have no private information about their production costs as in Levin and Smith (1994), there is no compositional (selection) effect, but entry decreases when entry costs increase. In the more plausible scenario in which the potential bidders have at least some information about their costs, free entry results in the more cost-efficient suppliers entering. This means that when entry costs increase, there also is a compositional effect. Li and Zheng (2009) identify an entry and a competition effect of potential entry. They show that the positive entry effect may dominate the standard (negative) competition effect even in a standard independent private value setting. The literature on entry in auctions also suggests that regulating entry can be efficient, because the entry costs are indirectly born by the buyer (Ye, 2007) and because it can lead to a more efficient selection of bidders (if properly implemented; see Bhattacharya, Roberts and Sweeting 2014).

The above-listed theoretical insights do not provide clear-cut predictions for us: On one hand, we expect decreased *actual* entry into free entry auctions in the new regime. Holding other things constant, this ought to increase procurement costs. There may also be an offsetting effect, because the more cost-efficient suppliers became more likely to participate in the price only and scoring auctions of the new regime. On the other hand, the possibility of favoritism further compli-

cates the predictions: Less favoritism may lead to less aggressive bidding by the un-favored firms and also change the composition of participants.

4 Data

Data sources

Our bidding and procurement data come from two surveys. The first survey was administered to all Swedish municipalities asking them for procurement documents regarding internal cleaning services (Lundberg, 2005). The documents are call for tenders or contract notice, technical specification, list of bidders, bids, and the decision protocol stating the winner of the contract. The first survey is the source of the data for the first of our regimes (1990-1998).^x For the second survey the sample of procurement auctions was identified using a procurement database that keeps track of calls for tenders in Sweden. For more details see Lundberg et al. (2015).^{xi} Additional information was collected for the municipalities that according to the procurement database had organized at least one procurement auction. The second survey, carried out in 2011, is the source of the data for the new regime (2009-2010).

Our third source of data is a set of semi-structured interviews that we conducted in summer 2015. We interviewed 22 individuals who had been involved in Swedish public procurements either during the old or the new regime, or during both. The respondents were chosen for their current or former employment at one of the municipalities that our procurement auction data cover. These interviews supplement our econometric analyses.^{xii}

Descriptive statistics

Table 1 gives basic descriptive statistics, conditioning on the regime (and auction type). Our data include 1075 auctioned cleaning contracts. This total consists of 720 beauty contests that were organized in the old regime, and 240 scoring auctions and 115 first price auctions that were run during the new regime. The total number of bids in the data is 7427.

Panel A of Table 1 describes procurements, premises and contracts. In our data, the number of contracts auctioned in a procurement event decreased over time. Each contract was the object of a separate sealed bid auction, but often covered more than one premise. The premises to be cleaned are schools, offices, day care centers, other types of municipal buildings and their combinations. We take their distribution to be exogenous for our purposes. The total size of the premises that a single contract covered (measured in square meters, m²) was larger in the new regime. Municipalities also used longer contracts and extension periods in the new regime.^{xiii} The average cleaning frequency did not change much over time and was 231 ($= 5 \times 52 \times 0.89$) days per year on average. Changing the features of the cleaning contracts (e.g., their coverage in terms of size) is a decision taken by the municipalities that affects costs of organizing procurements and supplier participation. As we later argue, the Swedish municipalities may have deliberately modified the *objects* of procurement auctions – and thus an aspect of overall procurement auction design – so as to induce entry.

Panel B of Table 1 describes the mode of entry and participation. Free entry was used much more often in the scoring and first price auctions of the new regime than in the beauty contests of the old regime. Actual entry (n) refers to the number of bids submitted in an auction. The raw data reported in Table 1 appears

to suggest that there was less actual entry in the new regime (this difference disappears when observables are controlled for). In the new regime, a small number of bids were disqualified. Potential entry (N) refers to the total number of suppliers that submitted at least one bid in a given municipality in our data during the old and new regimes (for a similar approach, see, e.g. Li and Zhang, 2010; Athey, Levin and Seira, 2011; Li and Zheng, 2012; Athey, Coey and Levin, 2013). To allow for the possibility that potential entry to the largest auctions was different (e.g., due to the smallest local firms not participating in them), we calculate potential entry separately for large and small contracts in each municipality, with the threshold for a large contract set at the 80th percentile of the distribution of the (total) size of the premises included in the contract.^{xiv} On average, the number of potential bidders was about 19.

[TABLE 1 HERE]

There are four main types of bidders: Large nationally active firms (“national”), inhouse units of the municipalities that participated in the auctions as self-standing operative units (“inhouse”), small local firms (“local”), which we define as those firms that are not national or inhouse units and that participated in auctions of only one municipality, and regional firms (“regional”), which we define as those firms that are not national or inhouse units and that participated in auctions of more than one municipality.

As Table 1 shows, the rate of participation by inhouse units (and, to some extent, local firms) was lower in the new than in the old regime. Conditional on participation, the probability of an inhouse unit winning decreased moving from the old to the new regime. Conversely, the probability of national and regional firms winning increased in the new regime.

Panel C of Table 1 describes the bids. To achieve comparability across auctions, the bid is measured using the deflated price of the cleaning service per square meter and day (frequency).^{xv} In our analysis, we also control for the type of the premises and other characteristics of the contracts and allow for scale economies. As Panel C of Table 1 shows, the average and median bids, calculated over all submitted bids (and not just winning bids), are somewhat lower in the beauty contests than in the scoring and first price auctions of the new regime. The same applies to the minimum bids. However, the average winning bids are almost the same in the two regimes and in the three auction formats. This finding means that, unconditionally, the procurement costs were not lower in the new regime. This is a striking observation, because we can also see from Table 1 that the lowest bidder won only 42% of the time in the beauty contests. Moreover, conditional on not choosing the lowest bidder, the municipalities ended up paying on average 43% more than the lowest bid in those auctions. In the new regime, the lowest bidder won much more often: When calculated over both the scoring and first price auctions, it won 76% of the auctions.

Descriptive regressions

The raw data showed that the procurement costs were not lower in the new regime. We have run a set of exploratory reduced-form regressions to check whether that also holds when we condition on observable auction and municipality characteristics. In these regressions, the dependent variable is the logarithm of the winning bid. The regressions confirm that the winning bids were not lower in the new regime (see Appendix B).

The law change in Sweden was a product of EU wide developments, which resulted in stricter procurement regulation and induced a quasi-exogenous switch

away from using the beauty contests. The obvious question is: Why didn't the shift from the discretionary beauty contests to rule-based auctions reduce procurement costs?

5 Winning bids and entry

Our review of the available theory and prior empirical work suggested that possible determinants of the competitiveness of the procurement outcomes are endogenous bidder participation (number of bidders) and changing scope for not choosing the lowest bidder (favoritism). We explore the former in this section, and the latter in the next section.

The econometric approach

To explore how the regime change and bidder participation are related to procurement costs, we model the conditional mean of the winning bids (per unit of square meter and day) using the following specification:

$$E(Bid|R, W, X, T, \mu, n) = \exp[\beta R + \alpha \ln(n) + W'\lambda + X'\theta + \gamma T + \mu] \quad (1)$$

where R is a dummy equal to one if the auction was organized in the new regime and zero if it was organized in the old regime and n measures actual entry. The parameter α measures the entry elasticity of procurement costs and tells us how important entry is for the competitiveness of the procurement.

Because the raw data suggest that object characteristics changed over time, we use two groups of explanatory variables. The first group, denoted W , includes the number of contracts (*Auctions per procurement*) awarded in the procurement event, the size (in 10 000 square meters) of the premises covered by the contract (*Size of premises*) and its square, the length (in years) of the contract (*Contract*

length), the number of years over which the contract can optionally be extended if the municipality so decides when the initial contracting period expires (*Extension*), and the scaled number of days during which the cleaning takes place (*Frequency*) and its square.^{xvi} The second group, denoted X , consists of dummies for types of premises covered by a contract (*School*, *Office*, *Day-care center*, *Other*), municipal unemployment rate (*Unemployment*), population density (*Population density*) in thousands of inhabitants per square kilometer, share of inhabitants having a higher education (*Education*) and a binary indicator for those municipalities where leftwing parties have more than 50 percent of the seats in the municipal council (*Red majority*).^{xvii} We take X as exogenous, but consider the possibility that W is potentially determined by the municipalities (so as to govern entry).

We also include a vector of municipal fixed effects, μ ; one of the fixed effects refers to the group all those municipalities that organize less than five auctions. We include a linear within-sample trend, T , because the regime change may have coincided with an underlying unobserved trend in cleaning costs (due to e.g. technological development).^{xviii} We control for this possibility also using the other time-varying (municipal) covariates.

Identification of the effect (α) of actual entry on the winning bid in (1) may be hampered by the endogeneity of n . For example, suppliers may be more willing to pay the entry cost and enter a procurement auction if they expect the winning bid to be higher. Consequently, we model actual entry using the following conditional mean specification as our first stage:

$$E(n|R, W, X, T, \mu, F) = \exp\left[\tilde{\beta}R + W' \tilde{\lambda} + X' \tilde{\theta} + \tilde{\gamma}T + \tilde{\mu} + F' \tilde{\kappa}\right] \quad (2)$$

where $F = \{Free\ entry, Free\ entry \times R, Ln(N)\}$ is the vector of instruments and $\tilde{\kappa}$ is the associated vector of parameters.

The identifying assumption in using *Free entry* as an instrument is that restricting entry into an auction affects winning bids only through its effect on actual participation. It is our understanding that in the old regime, municipalities restricted entry partly in order to exclude problematic bidders. Our interpretation of the industry practice is that restricted entry was a means to reduce hassle costs of dealing with unreliable bidders. In contrast, the two step procedure described earlier was applied in the new regime to screen the submitted bids against the mandatory qualification criteria. This means that in the new regime, not allowing for free entry likely played a much smaller role in deterring participation by problematic bidders.^{xix} Our exclusion restriction would be violated if *Free entry* affected not only the amount of entry, but also its composition. One of our robustness tests is therefore to control for the types of bidders.

We allow for regime-specific differential effects by using the *Free entry* - indicator and its interaction with the regime -indicator (*Free entry* \times *R*) as instruments. If entry restrictions were used to deal with the hassle costs caused by problematic bidders especially in the old regime, we expect *Free entry* to get a positive coefficient.^{xx} The coefficient of *Free entry* \times *R* should imply a smaller effect of free entry on actual entry in the new regime.

We use potential entry ($\ln(N)$) as an additional instrument for actual entry $\ln(n)$. Here, the exclusion restriction is that the number of potential bidders has no effect on the winning bid beyond its effect on the number of actual bidders. Also this assumption could be violated through a composition effect that we address in our robustness tests. The benefit of using potential entry as an instrument is that it allows us to also explore whether the effect of entry on winning bids depends on the type of the entering bidder(s).^{xxi}

Our instruments $F = \{Free\ entry, Free\ entry \times R, Ln(N)\}$ predict actual entry well and our main IV results are also robust to using estimation techniques which allow for weak instruments. Moreover, we use the method of Conley et al. (2012) to explore how strongly our results depend on the identification assumption that the decision to restrict participation was not systematically related to the expected competitiveness of an auction. The method allows us to show how, if at all, our baseline results change if we allow $\{Free\ entry, Free\ entry \times R\}$ to be mildly positively correlated with the unobserved component of the winning bids in the second stage, violating our exclusion restriction. This case would correspond to a situation in which municipalities were more likely to allow free entry into auctions in which the expected winning bid was higher.

Our focus on the regime change means that we compare the beauty contests with what happened on average in the new regime's first price and scoring auctions. That is, we group the scoring auctions together with the first price auctions and contrast how the new, more rule-based regime differs from the old discretionary regime. When estimating (1) and (2), we nevertheless at times allow separate effects for the first price and scoring auctions of the new regime.^{xxii} We do so to illustrate the robustness of our findings, *not* to estimate the (causal) effect on procurement costs of using a first price auction instead of a scoring auction.^{xxiii}

Instrumental variable estimations

First-stage results for actual entry

We report the first-stage regressions for actual entry in Table 2. The dependent variable is the logarithm of the number of submitted bids ($Ln(n)$).^{xxiv} Besides the regime dummy (or its variants; see below), the vector of explanatory variables includes W , X , T and the municipal fixed effects.

We report results from three different specifications. Model 1 is our preferred first stage specification: In this model, we include R and the vector of instruments, is $F = \{Free\ Entry, Free\ Entry \times R, Ln(N)\}$. In Model 2, we keep the instrument vector unchanged but substitute *Scoring* and *First price* -indicators for R . In Model 3, we allow for a slightly extended instrument set of by substituting $Free\ entry \times Scoring$ and $Free\ entry \times First\ price$ for $Free\ entry \times R$. We acknowledge that the choice by a municipality to use *Scoring* or *First price* auction may be endogenous: We report the results for Models 2 and 3 only to show that our findings are *not* due to the way we treat the regime / auction format dummies.

[TABLE 2 HERE]

Table 2 shows that the coefficient of *Free entry* is positive and highly significant in our preferred Model 1: The use of restricted entry was associated with fewer actual entrants in the beauty contests. However, the coefficient of the interaction term between the free entry and regime indicator is negative and significant and almost of equal absolute size with the coefficient of the *Free entry* -indicator. In the new regime, allowing *Free entry* became the norm (recall Table 1) and, as the sum of the coefficients show, it had no effect on actual entry. This finding is consistent with the view that there was less need in the new regime to screen problematic bidders out by restricting entry.^{xxv} Table 2 also shows that the elasticity of actual entry w.r.t. potential entry is always highly significant (and close in size to the elasticity estimate of 0.4 reported in Li and Zheng 2009). The F-test reported for Model 1 in the table suggests that the instruments are not weak. The results for Models 2 and 3 confirm the findings made from Model 1.

Figure 1 illustrates the economic meaning of the first stage estimates. Panel A reports the sample means of the logarithm of actual entry, $Ln(n)$, for the old and

new regime. The raw data suggest that there was less actual entry in the new regime. The mechanisms at work are however more complicated, as shown in Panels B, C and D.

In Panel B we ask how changes in the instruments, F , and regime, R , affect predicted entry. The difference in the height of the bars in Panel B shows that, keeping other things constant, the predicted entry decreases by 23% when R and F change from their values in the old regime (left bar) to the values they obtain in the new regime (right bar).^{xxvi} Going beyond what is shown in Panel B, consider a free entry auction, holding all other things constant: If we now change the regime indicator from zero to one, the predicted actual entry decreases by 42% (p-value < 0.01). This says that there was much less entry into a free entry auction of the new regime than into a similar free entry auction of the old regime. These findings suggest that had other things remained constant, entry would have reduced after the regime change.

In Panel C, we explore how changes in contract characteristics, W , affect entry, holding other things constant. The left bar displays predicted entry when W is set to the means of the old regime and the right bar displays predicted entry when W is set to the means of the new regime. The heights of the bars differ, because the municipalities increased in the new regime the length and extension periods of the contracts and the total size of the premises that each contract covered (see Table 1). The first stage estimates imply that such changes are associated with more entry. These contract changes *increased* entry by 19.5% moving from the old to the new regime. The difference in the heights of the bars is statistically significant (p-value < 0.01).

Panel D of Figure 1 displays the *combined* effect of regime changes (shown in Panel B) and contract changes (shown in Panel C) on entry: As the panel shows, the net effect of all of these changes on entry is very close to zero.

[FIGURE 1 HERE]

To sum up, we find that, overall, entry *did not* change from the old to the new regime. This net effect masks however many mechanisms at work. Holding other things constant, predicted entry was lower in a free entry auction of the new regime, when compared to a similar free entry auction of the old regime. A likely reason for the lower entry in the free entry auctions of the new regime is that red tape and thus participation costs increased in the more rule-based procurement environment; increased red tape also came up in our interviews of the industry practitioners. Another possible reason for it is that potential entrants anticipated the increased emphasis on price (see the next section) and thus a more competitive auction. Indeed, entry would have decreased, had the municipalities not adjusted the objects of the auctions by making the contracts “larger” e.g. by increasing the total size of the premises that each contract covered as well as the length and extension periods. This interpretation of the first stage estimates is consistent with our informal discussions with practitioners, who mentioned that longer extension periods and larger auctions are a means to influence the amount of entry.

Second stage results for winning bids

We report in Table 3 an OLS regression and a set of IV (2SLS) regressions in which the dependent variable is the logarithm of the winning bid. The first column reports the OLS estimates and the remaining columns refer to the IV estimates: Models 1 to 3 correspond to the second stage of the first stage regressions reported in Table 2, with Model 1 that uses the regime dummy being our preferred spec-

ification. . Again, we report the results for Models 2 and 3 only to show that our findings are *not* due to the way we treat the regime/auction format dummies.

The first column shows that the OLS estimate of the entry elasticity of the winning bid is -0.11 and significant, suggesting that more entry is associated with lower procurement costs. However, our preferred IV model (Model 1) produces a much larger and highly significant elasticity estimate of -0.55. The upward OLS bias is intuitive, indicating that there was more entry into auctions where the (potential) bidders expected the winning bid to be high. The last two columns of the table show that the result for the entry elasticity does not depend on how we treat the regime indicator (or the auction type indicators) in the model.^{xxvii} The estimated entry elasticity of the winning bid shows that securing entry is very important for competitive procurement outcomes in our data.

[TABLE 3 HERE]

Figure 2 consists of four panels, each illustrating the economic implications of the second stage estimates of our preferred model (Model 1 of Table 3). Panel A reports the sample means of the logarithm of the winning bids, $\ln(\text{winning bid})$, for the old and new regimes. This panel confirms that the unconditional procurement costs were very similar in the two regimes. Panels B, C and D present decomposition similar to the one we used in Figure 1 and illustrate the complexity of the mechanism at work. Panel B shows that, keeping other things except the regime constant, the predicted bids were slightly lower in the new regime.^{xxviii} Panel C displays how contract changes (changes in W) directly affect the winning bids, going from the old to new regime: They resulted in slightly higher winning bids in the new regime.^{xxix} Finally, Panel D shows the total effect of these changes (shown in Panel B and in Panel C) on the winning bids. For Panel D we additionally allow entry to adjust and to take its regime-specific predicted values, as im-

plied by the first-stage estimates. The net effect of all the changes on the winning bid is very small: The two bars are almost of equal height.

To sum up Figure 2, we conclude that for a typical municipality and premise (e.g. school) that was to be cleaned, procurement costs *did not* change from the old to new regime. Our IV findings and decompositions are hence consistent with what the raw data and reduced form regressions already suggested. However, as Figures 1 and 2 show, a number of counter-balancing adjustments took place simultaneously.

[FIGURE 2 HERE]

The decompositions presented in Figure 1 and 2 showed that the procurement costs would *not* have been the same in the two regimes if the municipalities had not induced more entry in the new regime by adjusting the objects of auctions. To see how important securing entry was, we can check what the effect of the documented 23% reduction in entry (see Panel B of Figure 1) would have been on the winning bids had the municipalities not neutralized it. The second stage estimates imply that holding other things constant, the effect of such a decrease in entry would be to increase bids by 12.5% (p-value < 0.01).^{xxx}

Robustness

We have considered the robustness of our baseline IV results in a number of ways (see Appendix C): i) We checked that the estimated entry effect is not due to a compositional effect, i.e., due to different types of suppliers winning in auctions with more entry; ii) we considered in a number of ways the possibility that our instruments are weak; iii) we evaluated the plausibility of our exclusion restrictions using the approach developed by Conley, Hansen and Rossi (2012); iv) we re-ran our analyses using an alternative measure of potential entry and an al-

ternative estimation sample; and v) we checked that outlier bids are not driving our findings.

6 Favoritism

If reduced entry does not explain why the winning bids were not lower in the more rule-based new regime, then what does? In this section we turn to the other possible determinant of the competitiveness of the procurement outcomes, i.e., the change in the scope for not choosing the lowest bidder.

Price sensitivity of municipalities

We first show that *decreased* price sensitivity of municipalities cannot explain why there is no difference in the average winning bids of the two regimes. The reason we pay attention to price sensitivity is that in our data, roughly two thirds of the auctions in the new regime were scoring auctions (see Table 1). The scoring rules could have made the municipalities less price sensitive in the new regime than in the old. If they did, it would explain why the winning bids did not change when the regime changed.

To obtain a benchmark, we use auxiliary data on the scoring rules used in the new regime (see Appendix D) to calculate the average weight attached to the bids in the auctions of the new regime. The average weight was about 0.7. To estimate the change in price sensitivity, we study the procurement bureaucrats' choice of the winning bid using a random utility model (McFadden, 1974), i.e., the Conditional logit model.^{xxxii} The model is estimated using submitted bids and not just the winning bids. The more negative the coefficient of the bid is in these estimations, the more weight the price received in the choice of the winner.

Table 4 reports the estimation results for six different Conditional logit models (Models 1 to 6). In the first four models, data for *First price* auctions are excluded from the estimations, because in them the bids completely determine the choice of the winner. Models 1 to 4 therefore focus on the difference in price sensitivity between *Beauty contests* and *Scoring* auctions. The key explanatory variables are *Bid* (in krona per square meter and per day), which captures the price sensitivity in the *Beauty contests* of the old regime, and its interaction with the scoring indicator, $Bid \times Scoring$. The coefficient of the interaction term tells us how the price sensitivity differs in the scoring auctions relative to the beauty contests. In Model 1, there are no additional explanatory variables. In Model 2, we include the number of bids that each supplier submitted in the particular procurement and the total number of bids that each supplier submitted during the entire observation period. We use the former variable to control for the possibility that the municipalities take in the account how active the suppliers are in the particular procurement event (i.e., the potential multi-unit nature of the procurement). We use the latter variable as a crude proxy for the reputation of the suppliers in the market. In Model 3, we replicate the estimations of Model 1, except that we now add the supplier type dummies (*Local*, *Regional*, *National*; omitted: *In-house*) and their interactions with the indicator for the scoring auctions. Model 4 replicates Model 3, except that we now also add the two firm-level controls from Model 2. Finally, Models 5 and 6 are equivalent to Models 3 and 4, respectively, except for two things: First, they are estimated using a sample which also includes *First price* auctions and second, in these models $Bid \times R$ has been substituted for $Bid \times Scoring$. The new interaction tells us how much more (or less) price sensitive the auctions were in the new regime on average.

[TABLE 4 HERE]

Models 1 to 4 reported in Table 4 yield two main findings. First, the coefficient of *Bid* is negative and highly significant, implying that a higher bid reduced the probability of winning in the old regime. Second, the coefficient of the interaction term, *Bid* \times *Scoring*, is negative and highly significant. This finding means that the procurement bureaucrats were less price sensitive in the beauty contests than in the scoring auctions.^{xxxii} The results from Models 5 and 6 provide further support for the lower price sensitivity of the municipals in the old regime. In particular, the coefficients of *Bid* and *Bid* \times *R* in Models 5 and 6 show that the weight attached to price in the beauty contents of the old regime was about a quarter ($0.26 \approx -4/(-4-11.3)$) of the weight that was attached to the bid in the new regime.

These results have two implications: First, given that the average weight attached to the price in the new regime was 0.7 (see above), the weight implicitly attached to the price in the beauty contests of the old regime was about 0.18 ($= 0.26 \times 0.7$). Second, *decreased* price sensitivity of municipalities clearly cannot explain why there is no difference in the average winning bids of the two regimes.

Our interview respondents provided additional perspective on these findings. First, a clear majority of the respondents indicated that besides inexperience in using formal scoring rules during the old regime, major reasons for not using them were convenience (86% of respondents; see Appendix D) and the freedom to pick the desired winner (77% of the respondents). Second, the respondents gave a number of reasons for not choosing the lowest bid in the old regime. Besides seeing it as a means to secure quality, the most often mentioned reason (91% of the respondents) was the need to be able to choose a particular preferred bidder, at least once in a while. Corroborating this, 80% of the respondents agreed with the view that certain types of suppliers were treated favorably in the old regime.

Who was favored? The special role of inhouse units

A useful by-product of the estimation results of Table 4 is that they allow us to take a closer look at *which* suppliers were treated favorably. As we show next, the estimations allow us to establish that, conditional on the submitted bid, it was the inhouse units that were more likely than the other types of suppliers to win in the beauty contests.

We start by zooming at Models 3 to 6 of Table 4, which include firm type dummies and in which *inhouse* units are the base category. In these models, the other suppliers' type dummies obtain statistically significant *negative* coefficients. This finding shows that the *inhouse* units were more likely than the other types of suppliers to win in the beauty contests of the old regime *conditional on the submitted bids*. This finding is in line with the raw data (Table 1), which already suggested that *inhouse* units were particularly likely to participate *and win* auctions in the old regime: They submitted a bid to 60% of the auctions of the old regime, and won half of them. Table 4 shows moreover that this pattern disappeared in the new regime: When the firm type dummies are interacted with the indicator for the scoring auctions (Models 3 and 4) or the regime indicator (Models 5 and 6), the interactions obtain positive coefficients that essentially counterbalance the direct negative effects.

The economic magnitudes are not negligible: Our estimates imply that in the beauty contests, the willingness of municipalities to pay for the cleaning services of the *inhouse* units appears to have been roughly 0.4-0.5 krona (per square meter and per day) higher than for the services of the other types of suppliers.^{xxxiii} This is a substantive amount when compared to the mean of the winning bids (0.64 krona; see Table 1). This finding motivates us to take a closer look at the special role of *inhouse* units.

The interviews provide additional support for these findings (see Appendix D). The responses of the interviewed industry practitioners suggest that although the favored supplier likely varied across municipalities, both local firms and in-house units were treated favorably. The respondents also believed that inhouse units were very rarely treated unfavorably, whereas other firms were. When specifically asked, 71% of the respondents agreed with the view that inhouse units got a favorable treatment in the old regime.

How did un-favored suppliers react to favoritism?

Theory and prior empirical work suggest that the presence of a favored bidder may affect the procurement costs and lead to more aggressive bidding by un-favored bidders firms (McAfee and McMillan, 1989; Athey, Coey, and Levin, 2013). Such behavior might explain why the more competitive rule-based auctions did not result in lower procurement costs. To explore how the *presence* of an inhouse unit in an auction affected bidding, we proceed in two steps. In the first step, we study the winning bids, using a modified version of equation (1). We then explore in the second step the bid *distribution* more thoroughly, using all submitted bids (and not just winning bids).

We start by estimating the following version of equation (1) by GMM:

$$E(Bid|\bullet) = \exp[\beta R + W'\lambda + X'\theta + \gamma T + \mu + \alpha \text{Ln}(\chi D + n - D)] \quad (3)$$

where D is an indicator variable which takes value of one if an inhouse unit *participated* in an auction, and is zero otherwise. In this model, parameter χ explicitly measures whether entry into an auction by an inhouse-unit has a *differential impact* from the entry of other types of suppliers. If $\chi = 1$, there is no difference in the entry effects; if so, we are back to our baseline model (1). Because α is likely negative, a positive χ below unity implies that the impact of entry on the winning

bid is smaller when an inhouse unit participated in the auctions. Finally, if χ is positive and larger than one, the impact of entry on the winning bid is greater when an inhouse unit participated.

We treat D as endogenous and instrument it with a measure S of an inhouse unit's relative importance among all potential bidders. If an inhouse unit ever submitted a bid in a municipality (in our data), S is one divided by the number of potential entrants, calculated separately for the large and small auctions of each municipality. Otherwise, S obtains a value of zero.^{xxxiv} The exclusion restriction is that the composition of the potential bidder pool affects the winning bid through the composition of the actual bidders only. The other instruments are as they were in our preferred IV estimation of Model 1, Table 3.^{xxxv}

Table 5 displays the results of estimating (3) by GMM. We estimate the model in log-log transformed form so that the dependent variable is the logarithm of the winning bid. The reported model specifications correspond to those used earlier in Table 4. The estimates of the first column of the table show two things. First, the elasticity of the winning bid with respect to the number of bidders is about -0.4; this compares to the -0.5 we obtained in our main specification. Second, the estimate of χ is negative, and the null hypothesis $\chi = 1$ is clearly rejected. This implies that entry by an inhouse unit *reduced* the negative impact of entry on the winning bid; it is as if there were *fewer* bidders than there actually were. The estimates from the other two models confirm these findings.

[TABLE 5 HERE]

The analysis reported in Table 5 does *not* allow us to explore the hypothesis that favoritism leads to more aggressive bidding by other firms (McAfee and McMillan, 1989; Athey, Coey, and Levin, 2013). The reason is that the winning bids are a function of both suppliers' bidding behavior and the municipalities'

choice behavior. A sole focus on the winning bids may therefore mask important features of our data. To uncover them, we proceed to take a look at the *distribution of all submitted* bids. We do so to explore how submitted bids at different points of the distribution are affected by the presence of an inhouse unit. The question we are interested in is whether the submitted bids are adjusted differentially at the different points of the distribution in response to inhouse participation. Aggressive bidding might be especially prominent in the lower tail of the bid distribution, which motivates a quantile regression analysis.

Table 6 displays the results of a reduced form quantile regression (Koenker and Bassett Jr, 1978), estimated using the sample that includes all submitted bids. The dependent variable is the logarithm of the submitted bids (inflation-adjusted price of the cleaning service per square meter and day). To focus on the more relevant left tail of the bid distribution, we report the estimates for the 5th, 10th, 15th, 25th, 50th and 75th percentiles. We use the same set of control variables (W , X , T) as before (see Appendix D for alternative quantile regression specifications). In addition, we include the indicator variable D , which takes value of one if an inhouse unit participated in an auction, as well as its interaction with the regime indicator. The coefficient of D measures how the submitted bids differed in the presence of an inhouse unit in the beauty contests of the old regime from the beauty contests in which such a unit did not participate. The reported standard errors are obtained via bootstrapping (100 replications).

[TABLE 6 HERE]

The key finding is that the indicator D obtains a negative and significant coefficient at each percentile that we report, with very little variation in the absolute value across quantiles. This means that the presence of an inhouse unit in the beauty contests was associated with *lower* bids throughout the bid distribution.

Although we cannot conclusively rule out selection, it seems implausible that the inhouse units would have intentionally participated only in the most competitive auctions of the old regime. The findings reported in Table 6 are therefore consistent with – if not suggestive of – favoritism leading to more aggressive bidding by other firms (McAfee and McMillan, 1989; Athey, Coey, and Levin, 2013). The more aggressive bidding counterbalanced the lower price sensitivity of the municipalities in the beauty contests. This counterbalancing provides a further reason for the procurement costs being similar in the two regimes.

How did favoritism affect procurement costs?

We now take a more detailed look at what the counterbalancing meant for the procurement costs. We illustrate the relative importance of supply (bidding) and demand (winner choice) behavior for the procurement costs by exploiting within-auction rankings of the submitted bids. This analysis allows us to demonstrate how the distribution of submitted bids changed and how this change was associated with the increased price sensitivity of the municipalities.

To start with, we rank the bids of each auction in our data based on their monetary component (price). We then calculate the probability of the municipalities choosing a bid of a given rank, as well as the average bid and the average winning bid for each rank.

Figure 3 reports the probability of a bid being the winning bid for the six lowest bid ranks, separately for the old (left) and new (right) regime. The figure shows that conditional on a bid being in ranked the lowest or the second lowest, the probability of it being the winning bid is much higher in the new regime than in the old regime. Reflecting this difference, about 60% of the winning bids of the old regime and about 90% of the winning bids of the new regime were the lowest

or the second lowest bid of the auction. The difference illustrates how the greater price sensitivity of the municipalities in the new regime affected the distribution of *winning* bids over the ranks.

[FIGURE 3 HERE]

Figure 4 consists of Panels A to D, which display the average of the logarithm of all submitted bids and the average of the logarithm of winning bids for each bid rank, separately for the new and old regime. A comparison of Panel A to B and of Panel C to D shows that, in all the lowest ranks, the averages of the submitted and winning bids were quite a bit lower in the old regime. For example, the submitted bids in the lowest rank were on average 27% lower in the old regime than in the new regime (p-value = 0.036). Although some of the lowest bids of the old regime may not have been credible and may have been submitted by firms with poor reputation, these numbers illustrate how the left tail of the bid rank distribution changed, as predicted by McAfee and McMillan (1989).

[FIGURE 4 HERE]

Although acknowledging that it is difficult to distinguish between supply and demand side behavior without using a structural model, we use the ranks to answer a simple counterfactual question: What would the average of the logarithm of the winning bid have been in the new regime, had the average submitted bid of each rank remained as they were in the *old* regime and if the municipalities had drawn the winning bids in the same proportion from the ranks as they actually did in the *new* regime? The average of the logarithm of the winning bid would in this case have been -0.67. This compares to -0.51, which is the mean of the logarithm of the winning bid in the new regime. The difference between the two is -0.16 (p-value < 0.01). This decomposition shows that had the distribution of the submitted

bids not changed, the shift from the favoritism-prone beauty contests to rule-based auctions would have reduced procurement costs.

The ranks also highlight how the composition of the lowest ranked bidders and winners changed. In the old regime, the inhouse units submitted roughly a third of the bids in each of the five lowest ranks. In the new regime almost all of the bids in the five lowest ranks were submitted by suppliers other than the inhouse units. What's more interesting is that conditional on an inhouse unit submitting the lowest, second lowest or third lowest bid in the old regime, 78%, 49% and 45% of its bids won the auction, respectively. The corresponding percentages for the other suppliers were much lower at 35%, 15%, and 11%, respectively. The roles were almost completely reversed in the new regime.

Was favoritism active or passive waste?

Even though favoritism may have been associated with more aggressive bidding (as also the prior work suggests), it apparently did not lead to lower procurement costs. Why, then, were the municipalities so price insensitive in the old regime?

We can address this question by making use of the approach of Bandiera, Prat and Valletti (2009). Their approach allows us to explore whether or not choosing the lowest bid in the old regime is indicative of “active waste”, which means that the procurement agent directly benefited from the inflated bids. Alternatively, not choosing the lowest bid may mirror “passive waste”, which refers to X-inefficiency. Passive waste takes place when for instance poor practices and decisions lead to unintentional, inefficient spending of public resources. Application of the approach of Bandiera, Prat and Valletti (2009) to our data show (see Appendix D for details) that those municipalities that were *less* likely to choose the lowest bid in the old regime were also *more* likely to choose a scoring auction

in the new regime. This revealed preference is consistent with municipalities “actively wasting” public resources in the old regime.

What could active waste mean in the context of municipal public procurements in Sweden? In particular, was the practice of not choosing the lowest bid associated with outright corruption (Tran 2008; Cai, Zhang and Henderson 2013), or did such behavior generate some other kinds of benefits? We think that in our context, active waste is *unlikely* to mean outright corruption for three reasons: First, Sweden is one of the least corrupted countries in the world, so the institutional environment is quite different from, e.g., that of Cai, Zhang and Henderson (2013). Our understanding of the Swedish procurement practices and our discussions with the industry practitioners do not lend support for outright corruption either. Second, active waste may mirror municipalities’ intentional behavior to support employment in the favored inhouse units (McAfee and McMillan 1989; Vagstad 1995), as such public firms are arguably under pressure to increase local employment (Boycko, Shleifer, & Vishny, 1996). Some of our econometric findings are consistent with this view (again, see Appendix D for details).

Finally, the interviewed respondents provide four main reasons for the favorable treatment of the inhouse units in the old regime: Local employment, political involvement, labor union involvement, and ease of doing business with them (see Appendix D). Other reasons may also have mattered, but these reasons were mentioned by more than four out five respondents.

In sum, albeit the documented favoritism was associated with more aggressive bidding by other firms, it also seems to mirror municipalities’ intentional behavior to support the viability and employment of the inhouse units.

7 Efficiency

We have established four key findings so far: *First*, the lowest bid rarely won in the beauty contests, leaving a relatively large amount of money on the table (section 4.2). The procurement costs were nevertheless about the same in the two regimes (Section 4 and 5). *Second*, neither entry nor the procurement costs would have been the same in the two regimes if the municipalities had not induced more entry in the new regime by auctioning larger units and offering longer contracts (Section 5). Securing entry in this fashion was very important for improving the competitiveness of the auctions in the new regime. *Third*, the weight attached to price in the beauty contests was about a quarter of the weight that was on average attached to the bids in the new regime (Section 6). This was related to favoritism, as the inhouse units were often able to win the beauty contests even when they submitted higher bids than other suppliers (Section 6). Favoritism led to more aggressive bidding by other suppliers, shifting the distribution of submitted bids (Section 6). The effect of the more aggressive bidding was counterbalanced by the lower price sensitivity of the municipalities in the old regime. Had the bid distribution not changed, the shift from the favoritism-prone beauty contests to rule-based auctions would have reduced procurement costs (Section 6). *Fourth*, the documented favoritism likely mirrors municipalities' intentional behavior to support the viability and employment of the inhouse units (Section 6).

What do these findings imply for efficiency? This is a hard question to answer conclusively, but on balance, we are inclined to argue that efficiency improved in the new regime.

Quality of service and ex post renegotiations: We do not have hard data to compare the two regimes in these dimensions, but the nature of the product we

study suggests that there was relatively little scope for either the quality or ex post negotiations of adaptations to change from the old to new regime.

Our interview respondents reported experiencing some problems with quality in both regimes (e.g., some sort of inadequate delivery or non-performance; see Appendix E) but we have no evidence of quality differences across regimes.^{xxxvi} Thus increased quality is unlikely to explain why the more rule-based auctions did not decrease prices. In addition, problems in *delivered* quality or ex post adjustments rarely lead to renegotiations (see Appendix E).^{xxxvii} This is in contrast to what the prior work focusing on more complex products and contracts suggest (Bajari, Houghton, & Tadelis, 2014; Decarolis, 2014; Decarolis and Palumbo, 2015).

Cost of production: We argue that the new regime increased allocative efficiency through at least the following mechanisms: First, the price sensitivity of the municipalities increased and the supplier-level (fixed) entry costs likely increased (red tape); both of these lead to inefficient suppliers more often not participating in the new regime. Our interview evidence supports this view: Our respondents clearly indicate that the inhouse units had a cost disadvantage (Appendix E) and they participated and won much less often in the new regime. Second, our econometric and interview evidence shows that favoritism was reduced in the new regime. This change likely increased allocative efficiency. Third, it seems plausible that the transparency of the procurement and award mechanisms increased from the old to new regime. In the new regime less efficient suppliers had a more accurate estimate of their chance of winning, allowing them to opt out if the prospect of winning was poor.

In sum, we do not find systematic evidence on reduced quality, but find arguments for reduced costs. We are therefore inclined to say that allocative efficiency improved.

8 Conclusions

Conventional wisdom suggests that, absent complications, the shift to the more rule-based procurement environment of the new regime should have decreased procurement costs. It did not in the case of Swedish procurement of cleaning services. Instead, the picture that emerged from our analysis showed a complex reality, involving favoritism that targeted high cost suppliers in the old regime as well as increased red tape and redesign of auction objects in the new regime. Our findings augment the prior literature on how different auction formats and particularly bid preference programs work (McAfee and McMillan, 1989; Marion, 2007; Krasnokutskaya and Seim, 2011; Athey et al., 2013), because the beauty contests that we have studied were not based on an explicit price-preference rule.

A more rule-based procurement environment indeed curbs favoritism. The expected cost savings of reduced favoritism do not necessarily come about unless one pays attention to the importance of endogenous entry for procurement costs and, simultaneously, to changes in the bidding behavior of suppliers.

In Sweden, allocative efficiency seems to have improved going from the old, favoritism-prone regime to the more rule-based new regime. However, the numerous changes and mechanisms at work prevent us from pinning down how welfare changed: A central complicating factor is how to evaluate welfare implications of aggregate entry costs over bidders and auctions in the two regimes, as well as the lost (local) welfare due to the municipalities not being able to support

local employment, especially in the public sector, by awarding contracts to in-house units in the new regime.

References

Aronsson, T., Lundberg, J. and Wikström, M. "The impact of regional public expenditures on the local decision to spend." *Regional Science and Urban Economics*, Vol. 30 (2000), pp. 185-202.

Asker, J. and Cantillon, E. "Procurement when price and quality matter." *RAND Journal of Economics*, Vol. 41 (2010), pp. 1-34.

Asker, J. and Cantillon, E. "Properties of scoring auctions." *RAND Journal of Economics*, Vol. 39 (2008), pp. 69-85.

Athey, S., Coey, D. and Levin, J. "Set-asides and subsidies in auctions." *American Economic Journals: Microeconomics*, Vol. 5 (2013), pp. 1-27.

Athey, S., Levin, J. and Seira, E. "Comparing open and sealed bid auctions: Evidence from timber auctions." *Quarterly Journal of Economics*, Vol. 126 (2011), pp. 207-257.

Bajari, P., McMillan, R. S. and Tadelis, S. "Auctions versus negotiations in procurement: An empirical analysis." *Journal of Law, Economics and Organization*, Vol. 25 (2009), pp. 372-399.

Bajari, P., Houghton, S. and Tadelis, S. 2014. "Bidding for incomplete contracts: An empirical analysis of adaptation costs." *American Economic Review*, Vol. 104 (2014), pp. 1288-1319.

Bandiera, O., Prat, A. and Valletti, T. "Active and passive waste in government spending: Evidence from a policy experiment." *American Economic Review*, Vol. 99 (2009), pp. 1278-1308.

Bergman, M. A. and Lundberg, S. "Tender evaluation and supplier selection methods in public procurement: An empirical analysis." *Journal of Purchasing and Supply Management*, Vol. 19 (2013), pp. 73-83.

Bhattacharya, V., Roberts, J. W. and Sweeting, A. "Regulating bidder participation in auctions." *RAND Journal of Economics*, Vol. 45 (2014), pp. 675–704.

Boycko, M., Shleifer, A. and Vishny, R. W. "A theory of privatisation." *The Economic Journal*, Vol. 106 (1996), pp. 309-319.

Branzoli, N. and Decarolis, F. 2015. "Entry and subcontracting in public procurement auctions." *Management Science*, Vol. 61 (2015), pp. 2945 - 2962.

Cai, H., Zhang, Q. and Henderson, V. J. "China's land market auctions: evidence of corruption?" *RAND Journal of Economics*, Vol. 44 (2013), pp. 488-521.

Cameron, C. A., and Trivedi, P. K. *Microeconometrics: Methods and applications*. New York: Cambridge University Press, 2005.

Cameron, L. J. "Limiting buyer discretion: Effects of performance and price in long-term contracting." *American Economic Review*, Vol. 90 (2000), pp. 265-281.

Che, Y. K. "Design competition through multidimensional auctions." *RAND Journal of Economics*, Vol. 24 (1993), pp. 668-680.

Compte, O., Lambert-Mogiliansky, A. and Verdier, T. A. "Corruption and competition in procurement auctions." *RAND Journal of Economics*, Vol. 36 (2005), pp. 1-15.

Conley, T. G., Hansen, C. B. and Rossi, P. E. "Plausibly exogenous." *Review of Economics and Statistics*, Vol. 94 (2012), pp. 260-272.

Coviello, D. and Mariniello, M. "Publicity requirements in public procurement: Evidence from a regression discontinuity design." *Journal of Public Economics*, Vol. 109 (2014), pp. 76-100.

Coviello, D., Guglielmo, A. and Spagnolo, G. "The effect of discretion on procurement performance." Forthcoming in *Management Science*. <https://doi.org/10.1287/mnsc.2016.2628>, 2017.

de Fraja, G. and Delbono, F. "Alternative strategies of a public enterprise in oligopoly." *Oxford Economic Journal*, Vol. 41 (1989), pp. 302-311.

Decarolis, F. "Awarding price, contract performance, and bids screening: Evidence from procurement auctions." *American Economic Journal: Applied Economics*, Vol. 6 (2014), pp. 108-132.

Decarolis, F. "Comparing public procurement auctions." Manuscript, 2013.

Decarolis, F. and Palumbo, G. "Renegotiation of Public Contracts: An Empirical." *Economic Letters*, Vol. 132 (2015), pp. 77-81.

Di Tella, R. and Schargrodsy, E. "The role of wages and auditing during a crackdown on corruption in the City of Buenos Aires." *Journal of Law and Economics*, Vol. 46 (2003), pp. 269-292.

Eklöf, M. "Assessing social costs of inefficient procurement design." *Journal of the European Economic Association*, Vol. 3 (2005), pp. 826-850.

European Commission. Annual public procurement implementation review 2012. Brussels: 9.10.2012 SWD(2012) 342 final, 2012.

Gentry, M. and Li, T. "Identification in auctions with selective entry." *Econometrica*, Vol. 82 (2014), pp. 315-344.

Jehiel, P. and Moldovanu, B. "A note of revenue maximization and efficiency in multi-object auctions." *Economics Bulletin*, Vol. 3 (2001), pp. 1-5.

Jehiel, P. and Moldovanu, B. "An economic perspective on auctions." *Economic Policy*, Vol. 18 (2003), pp. 269-308.

Klemperer, P. "What really matters in auction design." *Journal of Economic Perspectives*, Vol. 16 (2002), pp. 169-189.

Koenker, R. and Bassett Jr, G. "Regression Quantiles." *Econometrica*, Vol. 46 (1978), pp. 33-50.

Krasnokutskaya, E. and Seim, K. "Bid preference programs and participation in highway procurement." *American Economic Review*, Vol. 101 (2011), pp. 2653-2686.

Levin, D. and Smith, J. L. "Equilibrium auctions with entry." *American Economic Review*, Vol. 84 (1994), pp. 585-599.

Lewis, G. and Bajari, P. "Procurement contracting with time incentives: Theory and evidence." *The Quarterly Journal of Economics*, Vol. 126 (2011), pp. 1173-1211.

Li, T. and Zhang, B. "Testing for affiliation in first-price auctions using entry behavior." *International Economic Review*, Vol. 51 (2010), pp. 837-850.

Li, T. and Zheng, X. "Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions." *Review of Economic Studies*, Vol. 76 (2009), pp. 1397-1429.

Li, T. and Zheng, X. "Information acquisition and/or bid preparation: A structural analysis of entry and bidding in timber sale auctions." *Journal of Econometrics*, Vol. 168 (2012), pp. 29-46.

Lundberg, S. "Restrictions on competition in municipal competitive procurement in Sweden." *International Advances in Economic Research*, Vol. 11 (2005), pp. 353-366.

Lundberg, S., Marklund, P-O., Strömbäck, E. and Sundström, D. "Using public procurement to implement environmental policy: an empirical analysis." *Environmental Economics and Policy Studies*, Vol. 17 (2015), pp. 487-520.

Löfving, C. *Kommunal upphandling: upphandling och näringsverksamhet efter EES-avtalet*. 2. Stockholm: Fritzes, 1994.

Marion, J. "Are bid preferences benign? The effect of small business subsidies in highway procurement auctions." *Journal of Public Economics*, Vol. 91 (2007), pp. 1591-1624.

Marmer, V., Shneyerov, A. and Xu, P. "What model for entry in first-price auctions? A nonparametric approach." *Journal of Econometrics*, Vol. 176 (2013), pp. 46-58.

McAfee, P. R. and McMillan, J. "Government procurement and international trade." *Journal of International Economics*, Vol. 26 (1989), pp. 291-308.

McFadden, D. L. "Conditional logit analysis of qualitative choice behavior." In *Frontier in Econometrics*, by Paul Zarembka, 105-142. New York: Academic Press, 1974.

Menezes, F. and Monteiro, P. "Corruption and auctions." *Journal of Mathematical Economics*, Vol. 42 (2006), pp. 97-108.

Moreira, M. J. "A conditional likelihood ratio test." *Econometrica*, Vol. 71 (2003), pp. 1027-1048.

Myerson, R. B. "Optimal auction design." *Mathematics of Operations Research*, Vol. 6 (1981), pp. 58-73.

OECD. *Government at a Glance 2015*. Paris: OECD Publishing. http://dx.doi.org/10.1787/gov_glance-2015-en, 2015.

Riley, J. G. and Samuelson, W. F. "Optimal auctions." *American Economic Review*, Vol. 71 (1981), pp. 381-392.

Spagnolo, G. "Reputation, competition, and entry in procurement." *International Journal of Industrial Organization*, Vol. 30 (2012), pp. 291-296.

Spulber, D. F. "Auctions and contract enforcement." *Journal of Law, Economics and Organization*, Vol. 6 (1990), pp. 325-344.

Tadelis, S. "Public procurement design: Lessons from the private sector." *International Journal of Industrial Organization*, Vol. 30 (2012), pp. 297-302.

Tran, A. "Can procurement auctions reduce corruption? Evidence from the internal records of a bribe-paying firm." Manuscript, 2008.

Vagstad, S. "Promoting fair competition in public procurement." *Journal of Public Economics*, Vol. 58 (1995), pp. 283-307.

Xu, P. "Nonparametric estimation of entry cost in first-price procurement auctions." *Journal of Applied Econometrics*, Vol. 176 (2013), pp. 1046-1065.

Ye, L. "Indicative bidding and a theory of two-stage auctions." *Games and Economic Behavior*, Vol. 58 (2007), pp. 181-207.

Yoganarasimhan, H. "Estimation of beauty contest auctions." *Marketing Science*, Vol. 35 (2016), pp. 27-54.

Directives and Laws:

COUNCIL DIRECTIVE 92/50/EEC of 18 June 1992 relating to the coordination of procedures for the award of public service contract

Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts.

The Public Procurement Act (LOU 2007:1091).

The Public Procurement Act (LOU 1992:1528)

Table 1: Descriptive statistics

	Old regime		New regime	
	Beauty contest	Scoring	First price	All
Number of auctions	720	240	115	1075
Total number of bids	5359	1435	633	7427
<i>Panel A: Characteristics of procurements, premises and contracts (sample: 1075 auctions)</i>				
<i>Contract characteristics (W)</i>				
Auctions per procurement (# of contracts auctioned)	22.55	14.23	3.84	18.69
Size (per contract, m2)	2462.7	5182.7	9774.7	3852.1
Contract length (in years)	1.99	2.53	2.55	2.17
Extension (option for extension; in years)	0.80	1.81	1.40	1.09
Frequency (number of cleaning days in a year / 260)	0.89	0.91	0.84	0.89
<i>Characteristics of premises</i>				
School (dummy)	0.42	0.14	0.21	0.34
Office (dummy)	0.08	0.10	0.12	0.09
Day-care center (dummy)	0.41	0.32	0.04	0.35
Other (incl. multitype and unknown; dummy)	0.09	0.44	0.63	0.23
<i>Panel B: Entry and participation (sample: 1075 auctions)</i>				
<i>Actual and potential entry</i>				
Free entry (dummy)	0.57	0.96	0.97	0.70
Actual entry (# of submitted bids, n)	7.44	5.97	5.50	6.91
# of qualified bids	7.44	5.92	5.19	6.86
Potential entry (# of potential bidders, N)	18.62	18.70	18.10	18.58
<i>Participation by firm type (share of auctions)</i>				
At least one national	0.96	0.98	0.92	0.96
At least one inhouse	0.60	0.03	0.09	0.41
At least one regional	0.86	0.93	0.91	0.88
At least one local	0.71	0.59	0.24	0.63
<i>Winning by firm type (cond. on participation)</i>				
National	0.38	0.44	0.47	0.40
Inhouse	0.51	0.14	0.10	0.49
Regional	0.28	0.46	0.59	0.36
Local	0.13	0.23	0.07	0.15
<i>Panel C: Minimum and winning bids</i>				
<i>Bid data (krona, per m2 and day)</i>				
Minimum bid	0.53	0.63	0.65	0.57
Winning bid	0.64	0.67	0.65	0.65
Average bid (of all winning and non-winning bids)	0.78	0.86	0.81	0.80
Median bid (of all winning and non-winning bids)	0.76	0.83	0.79	0.78
<i>Winning. vs. minimum bids</i>				
Lowest bid wins (fraction)	0.42	0.65	1.00	0.53
(Winning - lowest bid) / Lowest bid (³ 0)	0.25	0.07	0.00	0.18
(Winning - lowest bid) / Lowest bid (> 0)	0.43	0.20	-	-

Notes: Data come from two surveys and refer to two periods (Old regime = 1990-1998, and New regime = 2009-2010). During the old regime, the Swedish municipalities were allowed to run Beauty contests. During the new regime, they were forced by law to use either First-price or Scoring auctions.

Table 2: Actual entry -regressions (first-stage)

	Model 1	Model 2	Model 3
	OLS	OLS	OLS
Free entry	0.494*** (0.122)	0.466*** (0.113)	0.466*** (0.113)
Free entry ´ R	-0.481** (0.212)	-0.397* (0.203)	-
Free entry ´ Scoring	-	-	-0.401* (0.236)
Free entry ´ First price	-	-	-0.388 (0.321)
Ln(N)	0.329*** (0.093)	0.325*** (0.089)	0.325*** (0.092)
Municipal FE	Yes	Yes	Yes
Control variables (W, X)	Yes	Yes	Yes
R	Yes	No	No
Scoring	No	Yes	Yes
First price	No	Yes	Yes
F-test (of instruments)	13,56	13,62	10,21
Observations	1,067	1,067	1,067

Notes: The standard errors reported in parentheses are clustered by procurements. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The dependent variable is Ln(Number of submitted bids). Only the coefficients and standard errors of the instruments are displayed. In Model 1, the coefficients of the direct effects of the auction formats (First price and Scoring, not reported) were restricted to be equal (i.e., the model includes the regime dummy, R). In Model 2, the coefficients of Scoring and First price -dummies (not reported) are allowed to differ. In Model (3), an extended instrument set is used.

Table 3: Log(winning bid)

	2SLS			
	OLS	Model 1	Model 2	Model 3
Ln(n)	-0.112** (0.050)	-0.545*** (0.144)	-0.585*** (0.150)	-0.585*** (0.150)
Regime	0.066 (0.158)	-0.086 (0.185)	-	-
Scoring	-	-	-0.068 (0.182)	-0.068 (0.182)
First price	-	-	-0.306 (0.227)	-0.306 (0.227)
Control variables (W, X)	Yes	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes	Yes
Observations	1,067	1,067	1,067	1,067

Notes: The standard errors reported in parentheses are clustered by procurements. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The corresponding first stage regressions are reported in Table 2: In Models 1 and 2, the instruments are {Free entry, Free entry x R, and Ln(N)}. In Model 3, the instruments are {Free entry, Free entry x Scoring, Free entry x First price, and Ln(N)}. In the model estimated by OLS and in Model 1, the coefficients of Scoring and First price auctions were restricted to be the same.

Table 4: Choice of winner (Conditional logit)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Bid	-4.058*** (0.278)	-3.934*** (0.278)	-3.948*** (0.285)	-4.049*** (0.288)	-3.948*** (0.285)	-4.035*** (0.287)
Bid ´ Scoring	-6.905*** (1.054)	-6.664*** (1.058)	-7.142*** (1.069)	-6.725*** (1.072)	-	-
Bid ´ R	-	-	-	-	-11.382*** (1.209)	-11.253*** (1.223)
Firm-level controls:						
# of bids in the procurement	-	0.198 (0.823)	-	-1.551* (0.823)	-	-1.560* (0.822)
# of submitted bids in the data	-	0.082*** (0.012)	-	0.168*** (0.028)	-	0.147*** (0.026)
Firm type -dummies:						
National	-	-	-1.443*** (0.111)	-1.927*** (0.141)	-1.443*** (0.111)	-1.862*** (0.136)
Regional	-	-	-2.016*** (0.134)	-1.541*** (0.163)	-2.016*** (0.134)	-1.609*** (0.159)
Local	-	-	-2.148*** (0.167)	-1.546*** (0.202)	-2.148*** (0.167)	-1.630*** (0.198)
National ´ Scoring (or: ´ R)	-	-	1.132 (1.216)	0.666 (1.228)	2.111** (0.835)	1.616* (0.834)
Regional ´ Scoring (or: ´ R)	-	-	1.658 (1.216)	1.108 (1.229)	2.643*** (0.838)	2.135** (0.838)
Local ´ Scoring (or: ´ R)	-	-	2.112* (1.238)	1.530 (1.251)	3.045*** (0.872)	2.495*** (0.874)
Firm type dummies:						
Omitted firm type	No	No	Yes	Yes	No	Yes
First price -auctions	Excluded	Excluded	Excluded	Excluded	Included	Included
Observations	6,783	6,783	6,783	6,783	7,413	7,413

Notes: These estimations use data on submitted bids, not just winning bids. The standard errors reported in parentheses are clustered by auctions. *** p < 0.01, ** p < 0.05, * p < 0.10. In Models 3 and 4, firm type dummies are interacted with Scoring indicator, and in Models 5 and 6, they are interacted with the regime indicator, R.

Table 5: Implications of the presence of inhouse units for bidding
Y= Log(winning bid), Sample: Winning bids, Method: GMM)

	GMM Model 1	GMM Model 2	GMM Model 3
Ln(n)	-0.404** (0.164)	-0.446*** (0.171)	-0.446*** (0.171)
D (Inhouse unit): c	-0.771 (0.591)	-0.648 (0.689)	-0.644 (0.692)
Control variables (W, X)	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes
Test: c = 1 (p-value)	0,003	0,017	0,018
Observations	1,066	1,066	1,066

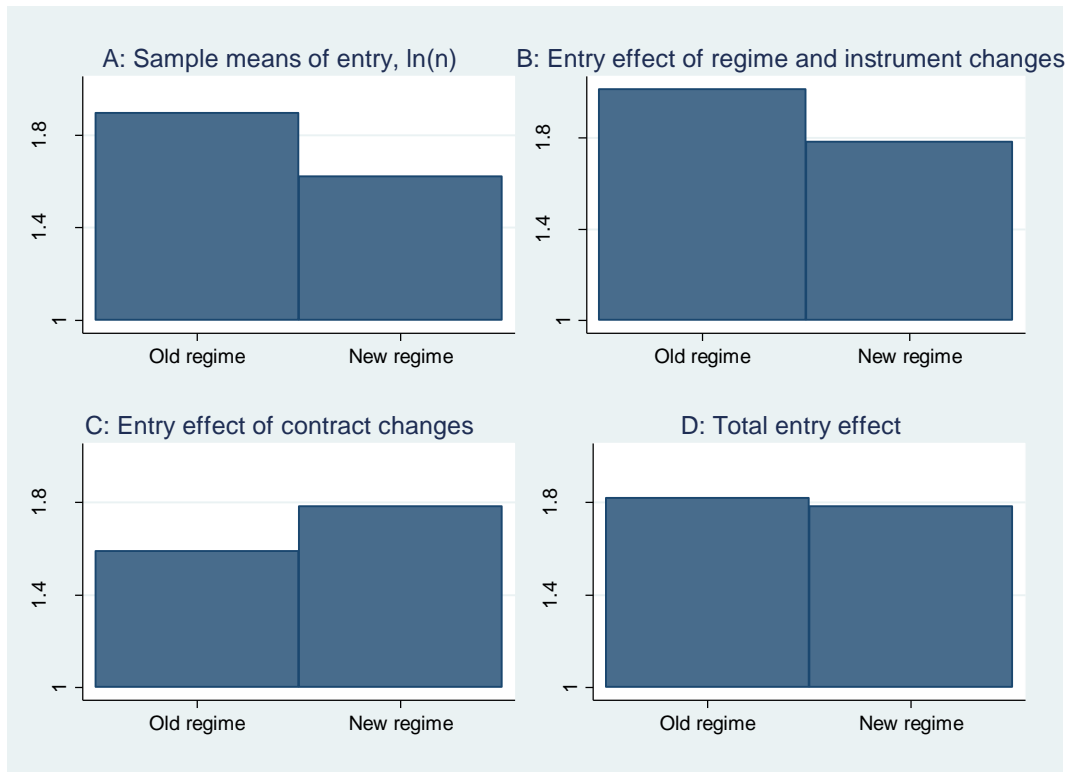
Notes: The standard errors reported in parentheses are clustered by procurements. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. GMM estimates of equation (3). The models reported here are augmented versions of those reported in Table 3. In Models 1 and 2, the instruments are {Free entry, Free entry x R, Ln(N) and S}, where S is one divided by the number of potential entrants, as calculated separately for large and small auctions of each municipality, if an in-house unit ever submitted a bid those auctions; otherwise, S obtains a value of zero. In Model 3, the instruments are {Free entry, Free entry x Scoring, Free entry x First price, Ln(N) and S}.

Table 6: Characterizing the bid distribution
Method: Quantile regression

	Percentiles				
	p5%	p10%	p25%	p50%	p75%
Regime (R)	-0.246 (0.154)	0.055 (0.099)	0.115 (0.073)	0.130** (0.056)	0.109** (0.047)
D_inhouse	-0.209*** (0.064)	-0.182*** (0.044)	-0.193*** (0.036)	-0.185*** (0.031)	-0.220*** (0.027)
D_inhouse x R	0.455*** (0.141)	0.390*** (0.087)	0.328*** (0.063)	0.285*** (0.048)	0.263*** (0.064)
Control variables (W, X, T)	Yes	Yes	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes	Yes	Yes
Observations	7,364	7,364	7,364	7,364	7,364

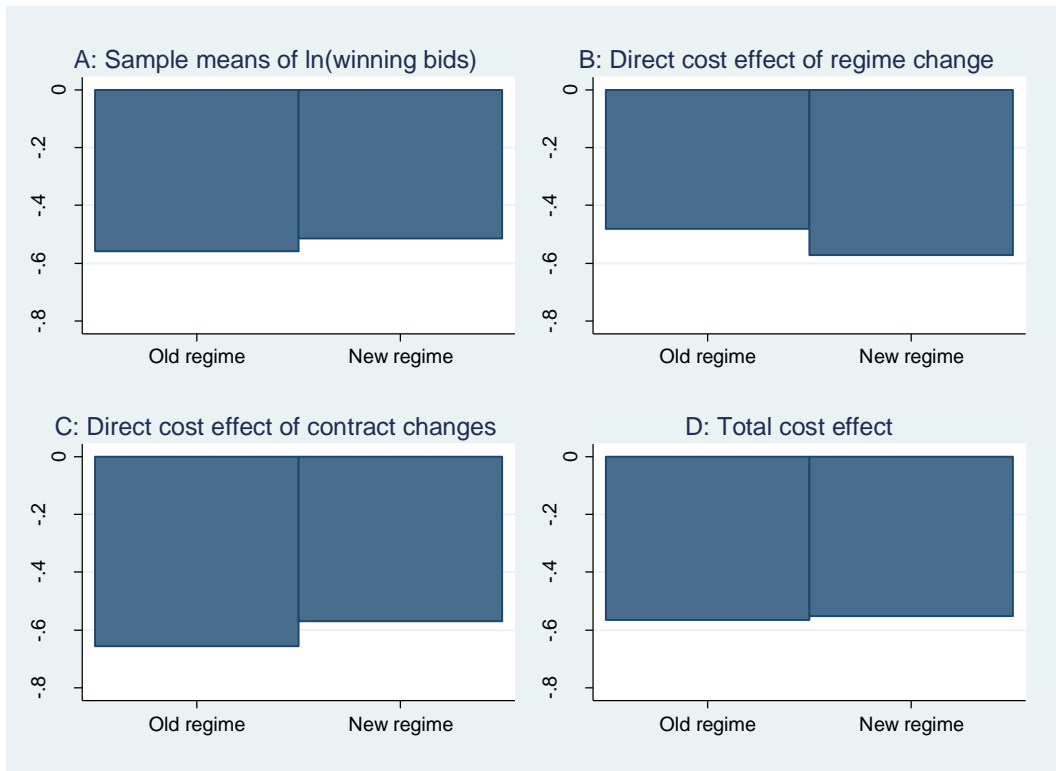
Notes: The standard errors reported in parentheses are bootstrapped (100 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The dependent variables is Ln(bid) and the estimating sample includes all submitted bids.

Figure 1: Entry and its decomposition



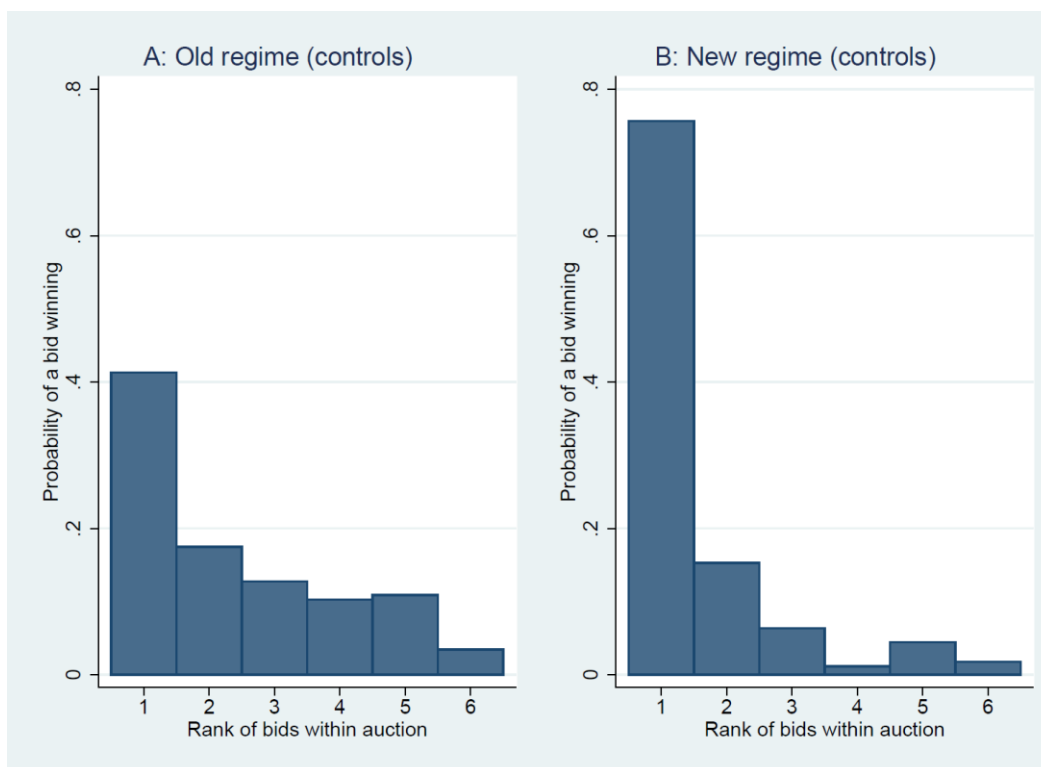
Notes: Actual entry is measured by the logarithm of number of submitted bids, $\ln(n)$. Panel A reports the sample means of the actual entry for the old and new regime. In Panel B-D, $\{X, T, \mu\}$ are set at their overall sample mean calculated over both regimes. Panel B shows how the predicted entry changes when, holding other things constant, the instruments $F = \{Free\ entry, Free\ entry \times R, \ln(N)\}$, and the regime indicator, R , change from the values they obtain in the old regime to their values in the new regime. In Panel B, contract characteristics W are fixed at the sample mean calculated for the new regime. Panel C shows how the predicted entry changes when, holding other things constant, the contract characteristics, W , change from the values they obtain in the old to their values in the new regime. In Panel C, R and F are fixed at their sample means calculated for the new regime. Panel D displays the combined effect of the two effects displayed in Panel B and C. In Panel D, R , F and W assume their subsample values in the old and new regimes, respectively.

Figure 2: Winning bid and its decomposition



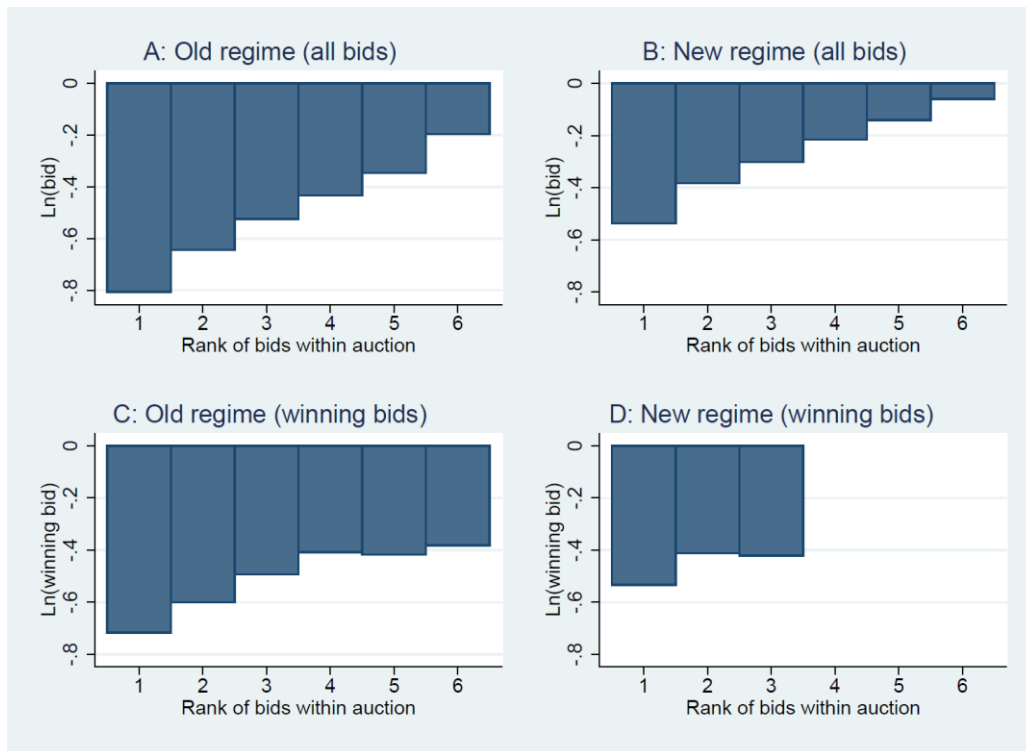
Notes: Winning bid (cost) is measured by the logarithm of winning bid. Panel A reports the sample means of the winning bids for the old and new regime. In Panel B-D, $\{X, T, \mu\}$ are set at their overall sample mean calculated over both regimes. Panel B shows how the predicted winning bid changes when, holding other things constant, the regime indicator, R , changes from the value it obtains in the old regime to its value in the new regime. In Panel B, W and n are fixed at the sample means calculated for the new regime. Panel C shows how the predicted winning bid changes when, holding other things constant, the contract characteristics, W , change from the values they obtain in the old to their values in the new regime. In Panel C, R and n are fixed at their sample means calculated for the new regime. Panel D displays the combined effect of the two effects displayed in Panel B and C. In Panel D, R , W and F assume their subsample values in the old and new regimes, respectively. The effect of F comes through entry only, as predicted by the estimated first stage.

Figure 3: Probability of a bid winning by the rank of the bid



Notes: The x-axis measures the rank of a bid in an auction, based on the monetary component (price) of the bid. The lowest bid obtains the lowest rank. Only the six lowest ranks are displayed, because it was untypical for a winning bid to have a higher rank. The 6th rank includes those few cases in which the winning bid had a higher rank. The y-axis measures the probability of municipalities choosing a bid from a given rank. The figure on the left refers to the old regime and the one on the right to the new regime. The reported probabilities are adjusted for $\{W, X, T, \mu\}$, but the raw probabilities look very similar.

Figure 4: Average bid and average winning bid by the rank of the bids



Notes: The x-axis measures the rank of a bid in an auction, based on the monetary component (price) of the bid. The lowest bid obtains the lowest rank. The y-axis measures the average of the logarithm of the bids in each rank. The reported means are adjusted for $\{W, X, T, \mu\}$. Panel A (old regime) and Panel B (new regime) use data on all submitted bids. Panel C (old regime) and Panel D (new regime) use data on the winning bids. In Panel D, the means of the winning bids are reported only for the three lowest rank categories, because in the new regime, there were only a very small number of winning bids in the higher ranked categories (as can be seen from Panel B of Figure 3).

Endnotes

ⁱ Public procurement amounted to about 12% of the GDP among the OECD member countries in 2013 (OECD, 2015). The EU Commission has estimated that each year, different levels of government spend about 20% of EU's GDP to procure goods, works and services (European Commission, 2012).

ⁱⁱ Using internal records of a bribe-paying firm from an Asian developing country, Tran (2008) finds that mandating the use of auctions curbs corruption only if the procurement auctions are open and if buyers' discretion to pick the winner is limited.

ⁱⁱⁱ Sweden applied the EU procurement law already in the early 1990s and also subsequently. The Swedish rules were relatively lax in the 1990s, because the European rules of that time allowed high degrees of freedom in organizing procurements.

^{iv} Our auxiliary interview evidence suggests that neither improved quality nor increased ex post renegotiations explains this finding.

^v Our IV estimates suggest that the entry elasticity of winning bids is about -0.55, i.e., that when the entry decreases by 10%, procurement costs go up by 5.5%.

^{vi} The municipalities were allowed to freely choose whether to procure or to produce in-house during both regimes. We take in this article as given municipalities' decisions to clean certain premises using procured cleaning services.

^{vii} For example, calls for tender in the new regime included a comprehensive set of mandatory qualification and exclusion criteria and detailed instructions of how the bidders were expected to show that they meet these criteria. However, the service in itself, and the technical specifications remained by and large the same.

^{viii} It is likely that the costs of preparing for and organizing procurement auctions went up in the new regime due to the more detailed mandatory qualification and exclusion criteria and the explicit use of scoring.

^{ix} The available theoretical models do not account for all of these features simultaneously. A structural approach is therefore not likely to be a fruitful direction to proceed.

^x In the first survey, the response rate was 79.5 % and ¼ of the respondents organized at least one procurement auction during 1990-98. >90% of the data of the first survey is from 1995-1998. We have supplemented the survey data with municipality characteristics from Statistics Sweden.

^{xi} The data base is maintained by Visma Commerce AB. This is the largest data base in Sweden and it covers approximately 90 percent of all procurements. It didn't exist at the time of the first survey.

^{xii} In addition to these interviews, one of the authors has had numerous informal discussions with practitioners about Swedish procurement practices.

^{xiii} A typical extension clause means that the contract continues for a specified (extra) period if there were no problems in delivery during the initial contract period. Our data do not allow us to determine which contracts were not extended. Discussions with practitioners suggest that the extension was typically granted.

^{xiv} This means that all auctions in the same size group of a given municipality have the same number of potential bidders. Our results are robust to using the median instead of the 80th percentile as the threshold for large contracts.

^{xv} The bids are expressed in 2013 Swedish krona (1 € ≈ 8.96 krona at the end of 2013). In the regressions where we model bids as outcomes, the dependent variable is the logarithm of the bids.

^{xvi} Many suppliers bid for more than one object in multi-object procurements. Combinatorial bidding was forbidden, thus the winner should have been picked object-by-object. Such choice behavior is a source of inflexibility, e.g., Jehiel and Moldovanu (2001; 2003). We control for the multi-object feature by the number of contracts awarded.

^{xvii} Although not perfect, this is a parsimonious way to capture the main division in Swedish politics, e.g. Aronsson, Lundberg and Wikström (2000).

^{xviii} Our results are robust to excluding T from the model, or to adding its square.

^{xix} Decaloris (2014) argues that a switch to excluding problematic bidders *after* the bids have been submitted may lessen competition among bidders who anticipate being able to later negotiate cost overruns. Although we cannot completely exclude this possibility, such behavior does not appear characterize our data.

^{xx} We would *not* expect a positive coefficient for *Free entry*, if municipalities were more likely to allow free entry into auctions in which *insufficient* entry was expected.

^{xxi} All our results go through if we include potential entry in the 2nd stage of the IV. When we do so, potential entry usually obtains a positive but insignificant coefficient in the 2nd stage.

^{xxii} We replace the *R*-dummy with a (2 x 1) vector of indicator variables for the two auction formats. Having the regime indicator, *R*, in the model is equivalent to restricting the coefficients of the *First price* and *Scoring* -dummies to be equal (leaving *Beauty contest* as the omitted category).

^{xxiii} We cannot estimate such a causal effect because there are no good instruments for the choice of the auction format during the new regime.

^{xxiv} We cluster the standard errors at the procurement level. We have re-run all of these models as count regressions (Poisson models). The results did not change.

^{xxv} In Model 3, we cannot reject the null hypothesis that the coefficient of *Scoring* × *Free entry* is equal that of *First price* × *Free entry* (p-value = 0.47).

^{xxvi} Although large, this difference is not significant at conventional levels (p-value = 0.17). This is mostly due to the larger standard error of the coefficient of *R*. Keeping *R* fixed and changing *F* only, the drop in $\ln(n)$ is -0.29 and the p-value is 0.08.

^{xxvii} In Model 2 and 3, we cannot reject at the 5% significance level the joint null hypothesis that the coefficients of *Scoring* and *First price* are zero (p-values = 0.06 and 0.06, respectively). The coefficient of *First price* is different from *Scoring* in Models 2 and 3 (p-values = 0.02 and 0.02, respectively), but this difference could be due to selection, i.e., due to municipalities using first price auctions for objects for which bidding can be expected to be competitive.

^{xxviii} The difference in the predicted bids is 8.6% (but not statistically significant), i.e., equal to the coefficient of *R* in the second stage (see Table 3, Model 1).

^{xxix} The difference in the height of the bars is 8.4% but is not statistically significant.

^{xxx} This raises the question why the municipalities did not try to induce even more entry. A partial answer is that inducing more entry by adjusting the objects was not without costs. Our second stage estimates reveal the trade-off that the municipalities faced. Holding entry constant, the *direct* effect of making contracts larger and longer was to increase the procurement costs (see Panel C of Figure 2).

^{xxxi} The model allows us to condition out all additively separable effects that are related to observable and unobservable characteristics of the municipals, (un)observable features of the auction format, and to (un)observable contract characteristics. For a detailed explanation of the conditional logit, see Appendix D or Cameron and Trivedi (2005).

^{xxxii} This increase in the price sensitivity is robust across the specifications in the table. We also estimated a model that includes supplier-specific fixed effects, with all suppliers submitting <20 bids grouped into a single category.

^{xxxiii} We ask: by how much should a supplier of type *j* have lowered its bid compared to an inhouse unit to have the same probability of winning, keeping all else the same? We get the answers by dividing the supplier type × *Beauty contest* -coefficient with the Bid × *Beauty contest*-coefficient.

^{xxxiv} E.g., if there were five potential entrants to a large auction and ten potential entrants to a small auction, and if an in-house unit was in both of them, the instrument takes value of 0.2 for the large and 0.1 for the small auctions in that municipality.

^{xxxv} A pseudo first stage of the 2SLS estimation shows that our additional instrument, *S*, predicts *D* well. The t-static of the coefficient of *S* is 5.84 in a regression in which *D* is regressed on the control variables (*W*, *X*, *T*, μ), other instruments (*F*) and *S*.

^{xxxvi} One change was that the municipalities required more proof of quality in the new regime at the bidding stage.

^{xxxvii} In those rare cases they did, our discussions with the industry practitioners suggest that the result seems often to have been a reduction in price, and not additional payments to suppliers.

Online supplement - not for publication

This online supplement consists of five separate appendices (Appendix A - E).

Each Appendix provides auxiliary material for one of the main sections of our article, Hyytinen, Lundberg and Toivanen “Design of public procurement auctions: Evidence from cleaning contracts”.

Appendix A: Auxiliary material to Section 2

This Appendix has two parts. In Part A we provide additional information on the regulation of the choice of the supplier in the Swedish procurement auctions. In Part B we offer a summary of the mandatory qualification criteria and scoring rules used in the new regime.

Part A: Description of procurement legislation

Regulation during the old regime

The Public Procurement Act (“Lag 1992:1528 om offentlig upphandling, LOU”) was not yet in force in 1990-1993, but the rules that applied then were essentially the same as under the act. The law was based on the EU rules that prevailed at the time: The Swedish law followed the COUNCIL DIRECTIVE 92/50/EEC of 18 June 1992 relating to the coordination of procedures for the award of public service contract.

What follows is an excerpt from COUNCIL DIRECTIVE 92/50/EEC of 18 June 1992 relating to the coordination of procedures for the award of public service contract: CHAPTER 3, Criteria for the award of contracts, Article 36:

- 1. Without prejudice to national laws, regulations or administrative provisions on the remuneration of certain services, the criteria on which the contracting authority shall base the award of contracts may be:*

- a. where the award is made to the economically most advantageous tender, various criteria relating to the contract: for example, quality, technical merit, aesthetic and functional characteristics, technical assistance and after-sales service, delivery date, delivery period or period of completion, price; or*
 - b. the lowest price only.*
- 2. Where the contract is to be awarded to the economically most advantageous tender, the contracting authority shall state in the contract documents or in the tender notice the award criteria which it intends to apply, where possible in descending order of importance.¹*

Regulation during the new regime

The new Public Procurement Act (“Lag 2007:1091 om offentlig upphandling, LOU”) became effective as of January 1, 2008. It was a direct consequence of the 2004 EU procurement directive, 2004/18/EC (Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts).

What follows is an excerpt from Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts, Section 3, Award of the contract, Article 53, Contract award criteria:

- 1. Without prejudice to national laws, regulations or administrative provisions concerning the remuneration of certain services, the criteria on which the contracting authorities shall base the award of public contracts shall be either:*

¹ This clause was effectively neither followed nor enforced, enabling the municipalities to run beauty contests.

- a. *when the award is made to the tender most economically advantageous from the point of view of the contracting authority, various criteria linked to the subject-matter of the public contract in question, for example, quality, price, technical merit, aesthetic and functional characteristics, environmental characteristics, running costs, cost-effectiveness, after-sales service and technical assistance, delivery date and delivery period or period of completion, or*
- b. *the lowest price only.*

2. *Without prejudice to the provisions of the third subparagraph, in the case referred to in paragraph 1(a) the contracting authority shall specify in the contract notice or in the contract documents or, in the case of a competitive dialogue, in the descriptive document, the relative weighting which it gives to each of the criteria chosen to determine the most economically advantageous tender.*

Those weightings can be expressed by providing for a range with an appropriate maximum spread. Where, in the opinion of the contracting authority, weighting is not possible for demonstrable reasons, the contracting authority shall indicate in the contract notice or contract documents or, in the case of a competitive dialogue, in the descriptive document, the criteria in descending order of importance.

Part B: Description of entry modes, the mandatory qualification and exclusion criteria and scoring rules

Entry modes

Both the old and new procurement laws allowed for four types of entry modes, called Simplified, Open, Restricted, Negotiated. The main difference between these four modes is that Simplified and Open allowed free entry, although the other two (Restricted, Negotiated)

did not. Although negotiations were allowed in some types of procurements, they were not used in the procurements that we study.

Our indicator for free entry takes value of one when the entry mode of an auction was Simplified or Open, and is zero otherwise (i.e., when entry was restricted).

Mandatory qualification and exclusion criteria and scoring rules

In the old regime, there were no formal mandatory qualification and exclusion criteria or explicitly announced scoring rules. In the new regime, the law allowed only first price or scoring auctions, and formal mandatory qualification and exclusion criteria were used widely.

We do not have comparable data on the scoring rules for all the scoring auctions of the new regime. There are three reasons for this: i) For some scoring auctions the scoring rule information was partly missing or the rule was incompletely defined; ii) the scoring rules can take different forms and are not entirely comparable across the scoring auctions (see also below); and iii) we cannot always reliably distinguish whether a particular clause was related to the scoring rule (award criterion) or whether it was one of the mandatory qualification and exclusion criteria.

We have analyzed the available (somewhat scattered) data on mandatory qualification and exclusion criteria and scoring rules as follows: Building on Lundberg et al. (2015), we first grouped the criteria into two main categories: Non-environmental quality criteria (“Quality criteria”) and Environmental criteria (“Environmental criteria”). In the original data there were 26 different kinds of quality criteria and 28 environmental quality criteria. On average, the cleaning service auctions of the new regime used about ten quality criteria and six environmental criteria. Many of these were related to the mandatory qualification and exclusion criteria (see below), but for a number of cases we cannot say exactly which. Moreover, some of the criteria were very similar and had similar kind of content. Therefore,

following Lundberg et al. (2015), we aggregated both the various environmental criteria and the various quality criteria into six main variables, respectively. The six quality variables are “Financial status, FIN” (e.g., data on supplier’s financial condition); “Insurance, INS (e.g. documentation of the supplier having required insurance); “Experience, EXP (e.g., documentation of relevant experience in similar assignments); “Performance plan, PER” (e.g., a description of how the cleaning service contract will be carried out); “Social criteria, SOC” (e.g. documentation of supplier having collective labor agreements with the union, etc); and ”Staffing, STAF” (e.g., documentation describing the qualifications of the employees). The six environmental variables are “Environmental management system, EMS” (e.g., having environmental certificates and/or fulfilling different ISO 14000 standards), “Eco labeling ECO” (e.g., having/using the EU Ecolabel for the cleaning products); “Vehicle, VEH” (e.g. meeting of certain emission standards for cars); “Meeting various chemical-usage and environmental codes, CHEM” (e.g., meeting the Swedish Chemicals Agency Code of Statutes 2008 and similar regulations); “Eco monitoring, MON” (Intention of the authority to monitor that the supplier meets the required environmental standards); and “Other eco demands, OTHER “ (other types of environmental or allergy-related criteria).

Table A1.1 provides summary statistics for the quality and environmental variables. It shows that except for SOC, most of the quality criteria were used in more than 90% of the auctions (incl. first price auctions) of the new regime. This finding means that many of them were mandatory qualification and exclusion criteria.

Table A1.1: Quality criteria

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A: Quality criteria					
FIN	355	0.98	0.14	0	1
INS	355	0.95	0.21	0	1
EXP	355	0.97	0.18	0	1
PER	355	0.99	0.07	0	1
SOC	355	0.54	0.50	0	1
STAF	355	0.94	0.25	0	1
Panel B: Environmental					

criteria					
MON	355	0.14	0.34	0	1
ECO	355	0.79	0.40	0	1
EMS	355	0.69	0.46	0	1
VEH	355	0.25	0.43	0	1
CHEM	355	0.87	0.34	0	1
OTHER	355	0.55	0.50	0	1

Notes: This table summarizes quality and environmental criteria used either as the mandatory qualification and exclusion criteria or as a part of the scoring rule. The sample includes both scoring auctions and first price auctions. The exact data on the scoring rule formulae are not available for all the scoring auctions.

A key finding from the above description is that due to the level of detail of the criteria imposed by the municipalities, there was effectively very little room for non-contractible quality. It also seems that many of the listed criteria were mandatory qualification and exclusion criteria, and not parts of the scoring rules. Moreover, compared to more complex products (e.g. construction projects), these procurement criteria are relatively simple. Many of them were related to the supplier meeting a certain environmental standard or to (easy-to-verify) aspects of its business operations, such as its financial condition or basic employee qualifications.

Bergman and Lundberg (2013) provide a detailed analysis of the various types of scoring rules used in Sweden. We briefly discuss them in Appendix D.

References used in this Appendix:

Bergman, M. A. and Lundberg, S. "Tender evaluation and supplier selection methods in public procurement." *Journal of Purchasing and Supply Management*, Vol.19 (2013), pp. 73-83.

Lundberg, S., Marklund, P-O. Strömbäck, E. and Sundström, D. "Using public procurement to implement environmental policy: an empirical analysis." *Environmental Economics and Policy Studies*, Vol. 17 (2015), pp. 487-520.

Appendix B: Auxiliary material to Section 4

In this Appendix, we report a set of reduced-form regressions to which we refer in Section 4 of the main article. We use these reduced-form regressions to explore how the regime change is associated with the winning bids.

Table A2.1 shows how procurement costs vary between the two regimes. It displays regressions in which the dependent variable is the logarithm of the winning bid and which include the regime dummy ($R = 1$ if regime is new; and $= 0$ otherwise), or auction format dummies, as the key regressor(s). The table reports results for six different model specifications. Model 1 includes only the regime dummy. In Model 2 we add the two groups of explanatory variables. The first group of variables, W , includes the number of contracts (*Auctions per procurement*) awarded in the procurement event, the size (in 10 000 square meters) of the premises covered by the contract (*Size of premises*) and its square, the length (in years) of the contract (*Contract length*), the number of years over which the contract can optionally be extended if the municipality so decides when the initial contracting period expires (*Extension*), and the scaled number of days during which the cleaning takes place (*Frequency*) and its square. The second group of control variables, X , consists of dummies for types of premises covered by a contract (four categories: *School*, *Office*, *Day-care center*, *Other*), municipal unemployment rate (*Unemployment*), population density (*Population density*) in thousands of inhabitants per the geographical size of the municipality, share of inhabitants having a higher education (*Education*) and a binary indicator that obtains a value of one for those municipalities where leftwing parties have more than 50 percent of the seats in the municipal council (*Red majority*). We also include a linear within-sample trend (T). Model 3 is equivalent to Model 2, except that it also includes municipal fixed effects.

As the table shows, the winning bids do not differ across the two regimes in Models 1-3. The coefficient of the regime dummy is positive, but has a large standard error and is thus not significant.

Table A2.1: Reduced form regressions, dependent variable = Ln(winning bid), OLS

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Regime-dummy, R	0.043 (0.063)	0.111 (0.172)	0.105 (0.159)	-	-	-
Scoring	-	-	-	0.052 (0.074)	0.111 (0.170)	0.114 (0.158)
First price	-	-	-	0.026 (0.069)	0.112 (0.188)	0.051 (0.179)
Control variables (W, X, T)	No	Yes	Yes	No	Yes	Yes
Municipal FE	No	No	Yes	No	No	Yes
R2	0.002	0.321	0.479	0.002	0.321	0.479
Observations	1,075	1,067	1,067	1,075	1,067	1,067

Notes: The standard errors reported in parentheses are clustered by procurements. *** p < 0.01, ** p < 0.05, * p < 0.10.

We also report in the table results from Models 4 to 6: These models are equivalent to Models 1 to 3, except that they do not force the coefficient of *Scoring* auction be equal to the coefficient of *First price* auction. When we substitute the indicators of the two auction formats for the regime indicator, the null hypothesis that the procurement costs were equal across the three auction formats cannot be rejected in any of Models 4 to 6. For example, in Model 6, p-value of the test statistic is 0.52. We cannot reject the null hypothesis that the coefficients of the *First price* and *Scoring* -dummies are equal either. For example, the p-value of the test statistic is 0.40 in Model 6.

In sum, the reduced-form models reported here provide no evidence that the winning bids were different across the old and new regimes. We have re-run these reduced form-models also in two alternative ways: First, we included our instruments (see main text) as additional regressors. Second, instead of controlling for a linear within-sample time trend, we either dropped the trend variable altogether or added its square to the models. These alternative regressions confirm that procurement costs were not lower in the new regime.

Appendix C: Auxiliary material to Section 5

In this Appendix we report a number of robustness tests for IV estimations of equations (1) and (2), reported in Section 5 of the main text. The Appendix consists of three parts: Part A: explores in detail whether our instruments are weak. Part B studies winner type and entry effect (composition). Finally, in Part C we report details of the other robustness tests that we have implemented.

Part A: Weak instruments

As we reported in Table 2 of the main body of the article, the F-test for the joint significance of our instruments is 13.56 in our preferred IV model (Model 1 of Table 3). This exceeds ten, which is the threshold often used to detect weak instruments (Staiger and Stock, 1997; Stock, Wright and Yogo, 2002). This finding supports the view that, when considered as a bundle, there is variation in the instruments that allows predicting the endogenous variable, even after conditioning on the other observables and the municipal fixed effects.

To err on the conservative side, we have explored further the possibility that our instruments are weak. We have implemented the following analyses:

First, we applied Moreira's (2003) conditional likelihood-ratio test (CLR-test) to our preferred IV model. The CLR-test allows us to test the null hypothesis that the coefficient of the number of submitted bids is zero in the second stage, *without assuming that the instruments are strong*. The CLR-test rejects the null hypothesis firmly (p -value < 0.01). The weak-instrument-robust 95% confidence intervals associated with the statistic is [-1.09, -0.17]. Roughly put, this is the set of parameter values for α that are consistent with the data when we allow the instruments to be weak. More formally, the 95% confidence interval refers to those coefficient values for which the rejection probability is below 95%. We conclude that

the documented negative effect of actual entry on the winning bid is not an artifact of weak instruments.

Second, we have checked the partial R2 statistic in the first-stage of our IV (see, e.g., Bound, Jaeger, and Baker 1995). In our set up, the statistic measures the correlation between the logarithm of actual entry and the instruments after partialling out the effect of the included exogenous variables. The benefit of using the partial R2 statistic is that it accounts for the possibility that the included exogenous variables predict actual entry strongly. Standard R2 or adjusted R2 would in such a case give a too promising picture of the capability of the first-stage to explain variation in the endogenous variable. In our preferred IV specification, the partial R2 statistic is equal to 0.114. This compares quite favorably to the adjusted R2 from the first-stage, which suggests that overall, we explain about 64% of the variation of the logarithm of actual entry. We point out that a large fraction of the explanatory power in the first-stage comes from the municipal fixed effects: Were the municipal fixed effects not included in the first-stage, the partial R2 statistic would be equal to 0.20 and the adjusted R2 about 40%. To our best knowledge, there are no clear-cut criteria in the literature on how large the partial R2 statistic ought to be, but our number appears not to be particularly low.

Third, a standard, yet somewhat informal way to check how IV works is to explore the reduced form of IV, i.e., the regression of the ultimate outcome variable (here: $\ln(\text{winning bid})$) on the instruments and the controls. We found that the three instruments are jointly highly significant ($F = 7.19$, p-value of the F-test = 0.0001) in this reduced-form regression. The instruments also obtained the expected signs: Free entry obtains a negative coefficient, as does the number of potential bidders. These patterns are consistent with our IV results.

Fourth, we have checked our (over-identified) 2SLS estimates against Limited information maximum likelihood (LIML) -estimates, using our preferred baseline model from Table 3 (i.e. Model 1). The motivation to do this is that the LIML-estimator has better small

sample properties than 2SLS when instruments are weak (see, e.g., Stock, Wright, and Yogo 2002). The LIML estimates of the effect of entry on the winning bids look very similar and the standard errors are not larger (coeff. = -0.60, p-value < 0.001). These findings suggest that our IV findings are not driven by a small sample bias, if the instruments happened to be weak.

Sixth, the biases of IV-estimators may increase with the number of instruments (Hahn and Hausman 2003). It has been argued that just-identified LIML IV is approximately median-unbiased even with weak instruments. Moreover, in the just-identified case, LIML and 2SLS are equivalent. Therefore, running the IV using the best single instrument is yet another robustness test that we can implement. When we use the Free entry-indicator (but not the interaction, nor $\ln(N)$) as the instrument, the free entry is highly significant and positive in the first stage of the IV (coeff = 0.41, p-value < 0.0001). Moreover, the coefficient of the submitted number of bids in the second stage is in line with our baseline findings, and in fact more negative (coeff = -0.84, p-value = 0.0001). This suggests that if anything, our baseline IV findings may be biased towards zero and thus conservative.

Finally, as Stock and Yogo (2005) explain, an alternative consequence of weak instruments (i.e., poor predictive ability of the instruments) is that when a parameter is estimated by IV-estimators, hypothesis tests concerning the estimated parameter suffer from size distortions. Using this as an insight, they develop an explicit test of weak instruments: Instruments are weak if a Wald test at the 5% level can have an actual rejection rate of no more than a given threshold (e.g., 10%, or 15%). To implement the test, suppose that we test the null hypothesis that actual entry has no effect on the winning bids and that we accept a rejection rate of 10% (which is inflated relative to the nominal 5% rate of the Wald test). The tabulations in Stock and Yogo (2005) show that we can reject the null hypothesis of weak instruments if a minimum eigenvalue statistic obtains a value larger than 22.30. In our case,

the statistic is 43.20. We thus reject the null hypothesis of our instruments being weak also using this approach.

To conclude, instruments are weak if, conditional on the included regressors, they don't have sufficient variation to explain variation in the endogenous variable. The analyses reported above show that such lack of variation is not a problem in our empirical set up.

Part B: Winner type and entry effect (composition)

Could the negative effect of actual entry on bids be driven by a compositional change among the winners? To explore this, we have re-run Models 1 to 3 of Table 3 of the main text after including supplier type dummies as additional controls. Following Marion (2009), these estimations are meant to uncover how much of the entry effect is due to within firm-type effects (i.e., lowering the bids of particular supplier type) and how much is due to a compositional effect (i.e., changing which type of supplier wins the auction). These estimations (see Table A3.1) show that the elasticity estimate does not change when supplier type dummies are included. This finding provides support for the view that the estimated entry effect is due to additional entry inducing more aggressive bidding by suppliers and is *not* due to different *types* of suppliers winning in auctions with more entry.

Table A3.1: Log(winning bid), firm type effects included

	2SLS		
	Model 1	Model 2	Model 3
Ln(n)	-0.515*** (0.134)	-0.558*** (0.140)	-0.558*** (0.140)
Scoring	-0.083 (0.180)	-0.070 (0.177)	-0.070 (0.177)
First price	-0.083 (0.180)	-0.298 (0.223)	-0.297 (0.223)
Control variables (W, X)	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes
Firm type dummies	Yes	Yes	Yes
Observations	1,067	1,067	1,067

Notes: Cluster-robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The baseline estimations for these regressions are those reported in Table 3 of the main text. Variable $\text{Ln}(n)$ refers to actual entry, i.e., the logarithm of the number of submitted bids.

Part C: Other robustness tests

We have collected the results from the rest of our robustness analyses (#1-#5) to Table A3.2.

Taking each of them in turn:

Exclusion restriction (robustness test #1): We evaluate the *plausibility* of our exclusion restrictions using the approach developed by Conley et al. (2012). This approach can be applied if one suspects that the exclusion restriction of the instruments does not hold *exactly*. In our case, this would amount to arguing that one or more of our instruments $F = \{Free\ entry, Free\ entry \times R, Ln(N)\}$ are correlated with the unobservables influencing the winning bid.

To explain the method briefly, let γ be a (3×1) vector of parameters that measures whether the instruments, F , can be excluded from the second stage of 2SLS. If $\gamma = (0, 0, 0)$ exactly, the variables have no direct effect on the winning bids. If we allow a ‘prior’ distribution for γ , we can discuss how close the exclusion restriction is to being satisfied. If it is likely that γ is close to zero and if the probability that it is far away from zero decreases sufficiently rapidly, we can say that the instruments are plausibly exogenous. We can then explore how robust the results are to small violations of the exclusion restriction. We have implemented the local-to-zero method of Conley, Hansen and Rossi (2012), as follows. First, our prior is that γ is normally distributed. We allow for the following means of the prior distribution: $\{(0.1, 0.1, 0.1), (-0.1, -0.1, -0.1), (0.1, 0.1, 0.0), (0.0, 0.0, -0.1)\}$. For example, $(0.1, 0.1, 0.0)$ is equivalent to assuming that the use of free entry was positively correlated

with the unobservable determinants of the procurement costs, but potential entry was not. The variance of γ is set to $(0.1^2, 0.1^2, 0.1^2)$.

Table A3.2 displays the results for various implementations of the local-to-zero method of Conley et al. (2012). It shows that the coefficient of actual entry varies a little, stays in the range $[-0.6, -0.5]$ and remains significant at better than 5% level in each case. Our IV results are hence robust to correlation of our instruments with the unobservables influencing the winning bids. For example, robustness test #1-c shows that our results hold even if municipalities had been more likely to allow free entry into auctions in which the expected procurement costs were higher.

We have also checked that our results are robust to the possibility that restricted entry was used in the old regime to deter unreliable bidders who were expected to bid *too low* (see Decaloris 2014). This amounts to checking whether the use of free entry in the old regime was positively associated with the unobserved determinants of the winning bids. We therefore set the means of the prior distribution of γ to $(0.2, 0.0, 0.0)$. When we do so, we find that the coefficient of $\text{Ln}(n)$ is -0.57 (p-value = 0.011). Our IV results are robust to the possibility that restricted entry was used to deter unreliable bidders in the old regime.

Table A3.2: Robustness checks

	Coefficient of $\text{Ln}(n)$ (i.e., a)
#1: Exclusion restriction (Conley et al. 2012)	
a: Local deviation: $\gamma = (0.1, 0.1, 0.1)$	-0,569 ** (0,224)
b: Local deviation: $\gamma = (-0.1, -0.1, -0.1)$	-0,521 ** (0,224)
c: Local deviation: $\gamma = (0.1, 0.1, 0.0)$	-0,559 ** (0,224)
d: Local deviation: $\gamma = (0.0, 0.0, -0.1)$	-0,534 ** (0,224)
#2: Alternative measure of potential entry	-0,696 *** (0.193)

#3: Alternative estimation sample (623 auctions)	-0,579 *** (0.292)
#4: Outliers (winzORIZED at +/- 3 std)	-0,331 *** (0.105)
Municipal FE	Yes
Control variables	Yes

Notes: The standard errors reported in parentheses are clustered by procurements. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The baseline for these robustness checks is Model 1, reported in Table 3 of the main text.

Alternative measure of potential entry (robustness test #2): We have in our baseline analysis measured potential entry using the total number of suppliers that submitted at least one bid in a given municipality in our data, calculated separately for large and small contracts. The threshold for a large contract has so far been the 80th percentile of the distribution of the total size of the premises, calculated separately for each municipality. If we use the median instead, we find that in the first stage, the coefficient of potential entry decreases a little but remains significant. The IV estimate from the second stage, in contrast, becomes a bit more negative (increases in absolute value) and remains significantly different from zero.

Alternative estimation sample (robustness test #3): In our baseline estimations we include municipal fixed effects for the municipalities that organize procurements in the old or in the new regime. If we restrict our estimation sample to only those municipalities that organized procurements during both periods, the size of our estimation sample decreases considerably, to 623 auctions. When we repeat the 2SLS analyses reported in Table 3, we find that our results are robust to using this smaller sample. For example, as Table A3.2 shows, the entry elasticity of the winning bids is -0.58 for our preferred IV model. Despite the smaller sample, the estimate is still significant at better than 5% level.²

² We would like to point out two things about the smaller sub-sample. First, the municipalities which remain in this sub-sample organized price only auctions in the new regime less often than in the full sample. In the full sample, the share of price only auctions is about one third. In the sub-sample of the municipalities that organized

Outliers (robustness test #4): We have also checked that very large or small winning bids are not driving our findings. There are many ways to identify outliers and to explore their effects, but we identified them as those bids that were larger (or smaller) than three standard deviations from the sample mean. We then dropped those observations from the estimation sample. The results from estimating our preferred IV model using this trimmed sample are reported in the table above: As can be seen, our IV estimate decreases, but remains negative and significant. In an alternative analysis, we kept the large and small observations in the sample, but set them to be equal to the value they would obtain had they been exactly three standard deviations from the sample mean. In this (unreported) estimation, the IV estimate of the coefficient of $\ln(n)$ is -0.48 (p-value < 0.001).

Different trend specification: Finally, in our baseline analysis, we control for a within-sample (linear) time trend. Our IV estimations (reported in Tables 2 and 3 of the main text) are also robust to dropping the trend variable altogether; and to adding its square to the model.

References used in this Appendix:

- Bound, J., Jaeger, D. A. and Baker, R. M. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak". *Journal of the American Statistical Association*, Vol. 90 (1995), pp. 443–450.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. "Plausibly exogenous." *Review of Economics and Statistics*, Vol. 94 (2012), pp 260-272.
- Decarolis, F. 2014. "Awarding price, contract performance, and bids screening: Evidence from procurement auctions." *American Economic Journal: Applied Economics*, Vol. 6 (2014), pp. 108-132.

procurements during both periods of our data, the share is about one fifth. Second, potential entry no longer predicts actual entry in this (much) smaller sample. However, *Free entry* works as in the full sample.

- Hahn, J. and Hausman, J. A. "Weak instruments: Diagnosis and cures in empirical econometrics". *American Economic Review Papers and Proceedings*, Vol. 93 (2003), pp. 118–125.
- Li, T. and Zheng, X. "Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions." *Review of Economic Studies*, Vol. 76 (2009), pp. 1397-1429.
- Marion, J. "Are bid preferences benign? The effect of small business subsidies in highway procurement auctions." *Journal of Public Economics*, Vol. 91 (2007), pp. 1591-1624
- Moreira, M. J. "A conditional likelihood ratio test." *Econometrica*, Vol. 71 (2003), pp. 1027-1048.
- Staiger, D. and Stock, J. H. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, Vol. 65 (1997), pp. 557 – 586.
- Stock, J. H. and Yogo, M. "Testing for weak instruments in linear IV regression". In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. H. Stock, New York: Cambridge University Press, pp. 80–108, 2005.
- Stock, J. H., Wright, J. H. and Yogo, M. "A survey of weak instruments and weak identification in generalized method of moments" *Journal of Business and Economic Statistics* Vol. 20 (2002), pp. 518–529.

Appendix D: Auxiliary material to Section 6

This Appendix consists of five parts: In Part A, we provide a brief summary of how we calculated the price weights. In Part B, we provide a summary table of the interview questions to which we refer in Section 6 of the main text. In Part C, we describe the random utility model (Conditional logit model), which we use in the main text to analyze the price sensitivity of the municipalities. In Part D we report further analyses using quantile regressions with the aim to explore 1) whether the bid distributions were different in the two regimes and 2) how other bidders reacted to the presence of an *inhouse* bidder. Finally, in Part E we detail our investigation based on Bandiera, Pratt and Valletti (2009) on passive versus active waste.

Part A: Calculation of price weights

We have also analyzed the exact formulae of the scoring rules.³ These data were obtained from the calls of tenders. A scoring rule can require a scaling of the price (bid) to match the quality measure or, alternatively, transforming of the quality measure to monetary units (scaling to price).⁴ These rules have in practice different designs and mathematical formulae. For example, in our data a commonly used scoring rule was the one in which the “quality-equivalent of the monetary bid” was obtained by multiplying the maximum obtainable quality score by the ratio of the lowest submitted bid to bidder i 's bid (i.e., $S_i = S_{max} \times (Bid_{Lowest} / Bid_i)$). This score measures the monetary value of supplier i 's bid in units of the quality measure. This scaled bid and the actual quality score of supplier i were then given weights

³ Our article's data for the new regime are a subset of the data analyzed by Bergman and Lundberg (2013), who explore the various forms and designs of scoring rules in more detail. Our analysis here builds on this prior work.

⁴ A third type of scoring auction is “quality only scoring”. In this case suppliers compete only in the quality dimension and the contracting authority announces a fixed payment in the call for tender. We have only one such example in our data.

(summing to one) when the final total score was calculated. This was used in 40.3% of the scoring auctions in our data.

The weight that a scoring rule attached to the price is not available for all auctions. Nor are the weight data entirely comparable across the auctions, because the scoring rules differ between auctions. We have comparable weight data for about 73% of the scoring auctions of the new regime. Using these data, we find that the average weight in the scoring auctions was 0.49. In an auction that chooses the winner solely based on price, the weight on the monetary bid is (normalized) to one. Taking into account missing data on the scoring rules, the average weight attached to the price was then $0.49 \times (2/3) \times 0.73 + 1 \times (1/3) \times (1/0.73) \approx 0.70$. If we assume that the average weight in the scoring auctions for which the weights are not observed was the same as in the other scoring auctions, the weight attached to the price in the new regime would be $0.49 \times (2/3) + 1 \times (1/3) \approx 0.66$.

Part B: Summary of the interview data

In this part of the Appendix, we provide a summary table of the auxiliary interview questions used in Section 6.

Table A4.1 Interview evidence (referred to in Section 6)

Q1	Reasons for not using formal scoring rules in the old regime (share of respondents indicating the reason, many choices allowed)	
	Didn't have experience to use formal scoring rules	0.91
	Convenience	0.86
	Lobbying by bidders	0.18
	Decision by municipal politicians	0.18
	It gave the freedom to pick the winner	0.77
	Other, please explain	0.09
Q2	Reasons for not choosing the lowest bid in the old regime (share of respondents indicating the reason, many choices allowed)	
	The difficulty in writing the calls for tenders so that every bidder would provide adequate quality	0.77
	The need to take quality of service into account	0.73
	The need to avoid bids by "fly-by-night" operators with a high probability of default winning	0.64
	The need to be able to choose small/local/etc firms, at least once in a while	0.90

Other, please explain	0.27
Q3 Agree with the view that certain types of suppliers were treated favorably in the old regime (share)	0.80
Q4 Agree with the view that in-house cleaning services got a favorable treatment in the old regime (share)	0.71
Q5 Type of supplier that was favored in the old regime (share of respondents indicating the type, many choices allowed)	
In-house	0.44
Large national/international firms	0.06
Regional firms	0.25
Local firms	0.63
Favored supplier type varied across municipalities	0.94
Q6 Agree with the view that certain types of suppliers were treated unfavorably in the old regime (share)	0.75
Q7 Type of supplier that was disfavored in the old regime (share of respondents indicating the type, many choices allowed)	
In-house	0.00
Large national/international firms	0.73
Regional firms	0.27
Local firms	0.07
Un-favored firm type varied across municipalities	0.93
Q8 Reasons for favoring in-house cleaning units (share of respondents indicating the reason, many choices allowed)	
Local employment	0.92
Labor union involvement	0.83
Political involvement	0.90
Ease of doing business with them	0.92
Low quality of private firms' service	0.44
Financial unreliability of private firms	0.22
Other, please specify	0.70

Notes: The number of respondents was 22, but not all respondents answered to all questions. The shares reported take the non-response into account. The responses are derived from semi-structured interviews that followed a pre-set protocol and had a common core based on a set of survey questions that were presented to all respondents. The interviewer was however allowed to discuss freely about the auctions with the respondents during the interviews, so the interview protocol was not exactly the same across the interviews.

Part C: Details on the discrete choice model (Conditional Logit model)

As we explain in the main text, we use the random utility model of McFadden (1974) to study the choice of the winning bid and to estimate how the price sensitivity of the municipalities changed from the old to new regime.

To describe the model, let the municipalities be indexed by m , $m = 1, \dots, M$, premises to be cleaned by i , $i = 1, \dots, I_m$, and bidders (suppliers) by j , $j = 1, \dots, J_{mi}$. The indirect utility of municipality m from choosing bidder j to clean building i is:

$$U_{mij} = \psi_{mi} + (\eta_1 + \eta_2 \times \text{Scoring}_{mi}) \times \text{bid}_{mij} + \beta' F_j + q_{mij} + \varepsilon_{mij}, \quad (\text{A_eq_4.1})$$

where ψ_{mi} refers to the additively separable effects of municipal/procurement/object characteristics (including the format of the auction), bid_{mij} to the bid (price) of supplier j for object i in municipality m (in krona per square meter per day), Scoring_{mi} is an indicator that takes value one if the auction format used in auction i of municipality m was scoring, F_j to supplier attributes, q_{mij} to ‘quality’, and ε_{mij} to an error term.

The municipal/procurement/object characteristics, ψ_{mi} , reflect the mean utility that municipality m obtains when it has its premises cleaned and the object-specific deviations from the mean. It thus captures all additively separable effects of observable and unobservable municipal characteristics on municipal utility, e.g., regional structure, demographics, income distribution, voter preferences, and propensity to procure services. The term also refers to (un)observable object characteristics, such as the type, size, location, etc. of the object. It captures differences in the indirect utility derived, e.g., from having a clean health center as compared to having clean sports facilities. The assumed additive separability of these effects and the distributional (logit) assumption on the error term (see below) allow us to condition all these effects out in the estimation. The term controls in addition for the additively separable effects on the utility of those characteristics of the procurement event that do not vary over the bidding suppliers, such as whether or not entry to the auction was open, which auction format was used, and whether or not the object was auctioned as a part of a multi-object procurement.

The second term in (A_eq_4.1) specifies the effect of a submitted bid on the choice, with the coefficient measuring the weight given to the bid. The interaction terms allow us to explore whether the weight attached to price is different between the two auction formats (or, in some specifications that we estimate, between the two regimes).

The third term in (A_eq_4.1), F_j , allow us to capture the possibility that there are *supplier-specific*, as opposed to object-specific, quality differences (i.e., ex ante corporate-level quality differences). For example, a piece of information in the bids through which the suppliers are able to ‘differentiate themselves’ (besides the price) is the corporate identity of the bidder. This may e.g. convey information about the bidder’s experience. To capture this, we use supplier-type dummies or fixed effects in some specifications. We can also alternatively include the number of bids a supplier submits during the entire sample period (to proxy reputation) and the number of bids a supplier submits in a given procurement event (to capture elements of “combinatorial bidding”, if any).

The fourth term in (A_eq_4.1), q_{mij} , refers to non-price attributes. It allows for the possibility that municipalities care about the quality of cleaning of *a particular object* for which suppliers are bidding (i.e., ex ante object-level quality differences). The extensive documentation available to us on the technical specifications of the procurements, calls of tenders and the specifics of the bids however suggest that it is likely that there are *no* major ex ante quality differences *at the object-level*. That is, conditional on the corporate identity of the bidders, it is not likely that there are large, ex ante discernible quality differences *between the bids for a specific object*. There is little room for a supplier to differentiate one-self quality-wise, conditional on F_j , suggesting $q_{mij} \approx 0$.

The last term in (A_eq_4.1), ε_{mij} , is a stochastic error term that captures intrinsic randomness in municipality decision making. Given ψ_{mi} , the error term only contains bidder-object specific unobservables. It therefore allows for idiosyncrasies in the decision-making of the procurement bureaucrats. We assume that ε_{mij} was unobservable to bidders and distributed i.i.d. type I extreme value.

Given the above assumptions, and imposing approximation $q_{mij} \approx 0$, the probability that bidder w wins in a procurement auction for object i organized by municipality m is (McFadden 1974):

$$\Pr[y_{mi} = w] = \frac{\exp\{\tilde{U}_{miw}\}}{\sum_{j=1}^{J_{mi}} \exp\{\tilde{U}_{mij}\}} \quad (\text{A_eq_4.2})$$

where

$$\tilde{U}_{miw} = \psi_{mi} + (\eta_1 + \eta_2 \times \text{Scoring}_{mi}) \times \text{bid}_{miw} + \beta' F_w$$

$$\tilde{U}_{mij} = \psi_{mi} + (\eta_1 + \eta_2 \times \text{Scoring}_{mi}) \times \text{bid}_{mij} + \beta' F_j.$$

As specified, the model corresponds to the standard Conditional logit model and can be estimated by maximum likelihood (ML).

Part D: Quantile regressions

Table A.4.2 displays the results for four sets (Panels A-C) of reduced form quantile regressions (Koenker and Bassett 1978), estimated using the sample that includes all submitted bids. The dependent variable is, as before, the logarithm of the price of the cleaning service per square meter and day (frequency), adjusted for inflation. To focus on the more relevant left tail of the bid distribution, we report the estimates for the 5th, 10th, 15th, 25th, 50th and 75th percentiles. The reported standard errors are obtained via bootstrapping (100 replications).

Panel A and B show how the bid distributions differ between the old and new regimes, unconditionally and conditionally. In Panel A, we only include the regime indicator. Consistent with the raw data (see Table 1), Panel A shows that the bids are throughout the entire distribution a little higher in the new regime. In Panel B, we use a reduced form specification which includes as controls $W = \{\text{Auctions per procurement}, \text{Size of premises}, \text{Size of premises}^2, \text{Contract length}, \text{Extension}, \text{Frequency}, \text{Frequency}^2\}$, $X = \{\text{School}, \text{Office},$

Day-care center, Unemployment, Population density, Education, Red majority}, Trend (T) and the municipal fixed effects (μ). Panel B shows that conditioning on $\{W, X, T, \mu\}$ matters. The bids of the new regime now appear to be lower in the lower tail and higher in the upper tail of the distribution.

Table A4.2: Characterizing the bid distribution

Panel A: $Y=\text{Log}(\text{bid})$, Sample: All submitted bids, Method: Quantile regression

	Percentiles				
	p5%	p10%	p25%	p50%	p75%
Regime (R)	0.060** (0.030)	0.041* (0.022)	0.044*** (0.015)	0.047*** (0.018)	0.032* (0.017)
Control variables (W, X, T)	No	No	No	No	No
Municipal FE	No	No	No	No	No
Observations	7,427	7,427	7,427	7,427	7,427

Panel B: $Y=\text{Log}(\text{bid})$, Sample: All submitted bids, Method: Quantile regression

	Percentiles				
	p5%	p10%	p25%	p50%	p75%
Regime (R)	-0.202 (0.131)	0.022 (0.109)	0.133 (0.085)	0.158*** (0.057)	0.176*** (0.044)
Control variables (W, X, T)	Yes	Yes	Yes	Yes	Yes
Municipal FE	Yes	Yes	Yes	Yes	Yes
Observations	7,364	7,364	7,364	7,364	7,364

Panel C: $Y=\text{Log}(\text{bid})$, Sample: All submitted bids, Method: Quantile regression

	Percentiles				
	p5%	p10%	p25%	p50%	p75%
Regime (R)	-0.113*** (0.036)	-0.076*** (0.027)	-0.054*** (0.021)	-0.036* (0.021)	-0.047*** (0.017)
D_inhouse	-0.212*** (0.031)	-0.185*** (0.021)	-0.160*** (0.017)	-0.130*** (0.015)	-0.132*** (0.017)
D_inhouse x R	0.509*** (0.072)	0.421*** (0.065)	0.289*** (0.046)	0.253*** (0.081)	0.253*** (0.063)
Control variables (W, X, T)	No	No	No	No	No
Municipal FE	No	No	No	No	No
Observations	7,364	7,364	7,364	7,364	7,364

Notes: The standard errors reported in parentheses are bootstrapped (100 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

To take into account the special role of the inhouse units, we turn to the quantile regressions reported in Panel C. These models include the indicator variable (D), which takes value of one

if an inhouse unit participated in an auction, as well as its interaction with the regime indicator. The coefficient of D measures how the submitted bids differed in the presence of an inhouse unit in the beauty contests of the old regime from the beauty contests in which such a unit did not participate. The key finding is that the indicator D obtains a negative and significant coefficient at each percentile that we report. Additional results are 1) that the coefficient of the indicator D reduces in absolute value when we move up in the distribution, suggesting that those firms bidding low reacted more strongly to the presence of an inhouse bidder than those bidding high; and 2) that the interaction between the indicator D and the regime indicator R always obtains a positive and significant coefficient larger in absolute value than the direct effect of D .

Part E: Characterizing the nature of favoritism

Here we explore why the municipalities were price insensitive in the old regime. To this end, we follow the approach of Bandiera, Prat and Valletti (2009). Applied to our context, their approach suggests that not choosing the lowest bid in the old regime may be evidence of “active waste”, which means that the procurement agent directly benefits from the inflated bids. Alternatively, such choice behaviour may mirror “passive waste”. Passive waste takes place when for instance poor practices and decisions lead to unintentional, inefficient spending of public resources (Bandiera, Prat, and Valletti, 2009).

We examine, in particular, whether those municipalities that were *less* likely to choose the lowest bid in the old regime were *less* likely to choose a first price auction in the new regime. This revealed preference would be consistent with municipalities actively wasting public resources (Bandiera, Prat, and Valletti, 2009) in the old regime. The opposite would be consistent with passive waste, as it would indicate that when forced to choose between a *First*

price and *Scoring* auction, a municipality that in the past often did not award a contract to the lowest bidder (i.e., was price insensitive) now wants to use a competitive first price auction.

We implement this analysis in two steps. In the first step, we form a dummy that takes value one for those auctions of the old regime in which the lowest bid won, and zero otherwise. We use a linear probability model (OLS) and regress this dummy on W (object characteristics) and X (municipal characteristics, other controls) and the municipal fixed effects, using data from the old regime only. The estimated municipal fixed effects of this regression provide us with information on which municipalities were particularly price sensitive and likely in the old regime to award the contract to the lowest bidder. The municipal fixed effects are jointly highly significant (F-test = 213.02, p-value < 0.001). In the second step, we use data from the new regime and form an indicator which takes the value one for first price auctions and is zero for scoring auctions. We then regress this indicator on the estimated municipal fixed effects from step one. This regression is also a linear probability model. We display the results from this second step in Table A4.3.

The results show that the coefficient of the municipal fixed effects is positive and significant in Model (1) where we have no other controls and Model (2) where we add contract characteristics. This suggests that those municipalities that were *less* likely to choose the lowest bid in the old regime were also *more* likely to choose a scoring auction in the new regime. This revealed preference is consistent with municipalities “actively wasting” public resources (Bandiera, Prat, and Valletti, 2009) in the old regime.

Model (3) provides some support for the viewpoint that active waste is *unlikely* to mean outright corruption. It shows that when we add the municipal characteristics (*Unemployment*, *Population density*, *Education*, and *Red majority*), the coefficient of the municipal fixed effects remains positive but is no longer significant. This is an interesting finding, because it indicates that our municipal-level control variables are pretty good at capturing the

determinants of the choice of the auction format in the new regime. Although not shown in the table, the *Red majority* - indicator, which is one for those municipalities where leftwing parties have a controlling majority in the municipal council, obtains a negative and significant coefficient in Model 3. This is consistent with such municipalities actively not using first price auctions and thus choosing auction formats that allow non-price considerations, such as local employment or union involvement, when picking the winner. The other significant predictor in Model 3 is the unemployment rate, but it obtains a little surprisingly a positive coefficient. In the estimation sample, the pairwise correlation of *Red majority* with the estimated municipal fixed effects is -0.56 (p-value < 0.001), whereas that of *Unemployment* and the fixed effects is -0.03 (p-value = 0.59).

Table A4.3: Choice between First price and Scoring auctions (New regime)

	OLS Model (1)	OLS Model (2)	OLS Model (3)
Municipal FE	0.438*** (0.100)	0.326** (0.147)	0.094 (0.162)
Control variables: Contract characteristics	No	Yes	Yes
Control variables: Municipal characteristics	No	No	Yes
Observations	252	252	252

Notes: The standard errors reported in parentheses are clustered by procurements. *** p < 0.01, ** p < 0.05, * p < 0.10. The dependent variable is an indicator which takes value one for the first price auctions and is zero for the scoring auctions. The municipal FEs are estimated using data from the old regime.

As can be seen from Table A4.1 (see especially Q8), the interview respondents' answers provide further support for the above considerations and particularly for the view that active waste was *unlikely* to mean outright corruption. The four main reasons for the favorable treatment of inhouse units in the old regime were local employment, political involvement, labor union involvement, and ease of doing business with them. These reasons were mentioned by more than four out five respondents.

References used in this Appendix:

- Bandiera, O. Prat, A., and Valletti, T. "Active and passive waste in government spending: Evidence from a policy experiment." *American Economic Review*, Vol. 99 (2009), pp. 1278-1308.
- Bergman, M. A. and Lundberg, S. "Tender evaluation and supplier selection methods in public procurement." *Journal of Purchasing and Supply Management*, Vol. 19 (2013), pp. 73-83.
- Koenker, R. and Bassett Jr. G. "Regressions quantiles." *Econometrica*, Vol. 46 (1978), pp. 33-50.
- McFadden, D. L. "Conditional logit analysis of qualitative choice behavior." In *Frontier in Econometrics*, by Paul Zarembka, 105-142. New York: Academic Press, 1974.

Appendix E: Auxiliary material to Section 7

In this Appendix, we provide a summary table of the interview questions used in Section 7.

Table A5.1 Interview evidence (referred to in Section 7)

Q1	Delivered quality of cleaning worse or better in the new regime (1=much worse, ..., 5=much better)	3.26
Q2	Delivered ancillary quality worse or better in the new regime (1=much worse, ..., 5=much better)	3.85
Q3	Renegotiation due to non-delivery/inadequate delivery in	
	Old regime (1=very infrequent, ..., 5=very frequent)	1.94
	New regime (1=very infrequent, ..., 5=very frequent)	2.23
Q4	Renegotiation due to municipality altering the conditions in	
	Old regime (1=very infrequent, ..., 5=very frequent)	2.20
	New regime (1=very infrequent, ..., 5=very frequent)	1.82
Q5	Non-performance (any reason) more common in	
	New regime (1=much less common, ..., 5=much more common)	3.26
Q6	Early termination of contract due to non-delivery/inadequate delivery in	
	Old regime (1=very infrequent, ..., 5=very frequent)	1.36
	New regime (1=very infrequent, ..., 5=very frequent)	1.73
Q7	Main problems in delivering cleaning services (share of respondents indicating the reason, many choices allowed)	
	The supplier keeping the schedule (old regime)	0.17
	The supplier keeping the schedule (new regime)	0.18
	The quality of the cleaning (old regime)	0.44
	The quality of the cleaning (new regime)	0.77
	The quality of the cleaning equipment used (old regime)	0.28
	The quality of the cleaning equipment used (new regime)	0.05
	The quality of the personnel (old regime)	0.22
	The quality of the personnel (new regime)	0.50
	Financial issues with the supplier (old regime)	0.17
	Financial issues with the supplier (new regime)	0.23
	General unreliability of the supplier (old regime)	0.22
	General unreliability of the supplier (new regime)	0.27
	Deteriorating performance towards the end of the contract (old regime)	0.28
	Deteriorating performance towards the end of the contract (new regime)	0.32
	Other, please specify (old regime)	0.00
	Other, please specify (new regime)	0.32
Q8	Perceived cost differences between suppliers in	
	Old regime (0 = yes, 1 = no)	0.31
	New regime (0 = yes, 1 = no)	0.33
Q9	If perceived cost differences in the old regime, cost ranking of suppliers (1=lowest cost, ..., 5=highest cost)	
	In-house	2.88
	Large national/international firms	2.13
	Regional firms	2.29
	Local firms	2.00
Q10	If perceived cost differences in the new regime, cost ranking of suppliers	

(1=lowest cost, ..., 5=highest cost)	
In-house	3.36
Large national/international firms	1.77
Regional firms	1.92
Local firms	2.54

Notes: The number of respondents was 22, but not all respondents answered to all questions. The shares reported take the non-response into account. The responses are derived from semi-structured interviews that followed a pre-set protocol and had a common core based on a set of survey questions that were presented to all respondents. The interviewer was however allowed to discuss freely about the auctions with the respondents during the interviews, so the interview protocol was not exactly to the same across the interviews.