Abstract
This thesis empirically investigates important phenomena surrounding procurement auctions. The first chapter examines the effect of uncertain design evaluations on firms’ behaviour in public procurement auctions in which firms compete on both price and design proposal. The second chapter takes a closer look at the source of variation in design evaluations. The last chapter then examines the costs and benefits of two widely used auction formats in procurement of construction projects.

In chapter 1, we investigate firms’ competition over price and product design in the context of Design-Build (DB) auctions. In a DB auction, firms’ design proposals are independently evaluated by reviewers of the government. Reviewers’ evaluations contain uncertainty from a bidder’s point of view, leading luck to curtail differences in firms’ chances of winning. The estimated model predicts that uncertain evaluations exacerbate auctioneer’s uncertainty in auction outcomes. An alternative mechanism that shuts down the impact of uncertain evaluations on bidders’ behaviour would mitigate the auctioneer’s uncertainty in the price and the design quality by 22% and 33%, respectively.

In chapter 2, we study strategic interactions among expert reviewers in evaluating a set of proposals. As expert reviewers may be concerned about their own career, they may not truthfully report the signal of proposal quality. We model and estimate the effects of such an incentive on reviewers’ behaviour using a sample of public procurement auctions, so called Design-Build (DB) auctions, where potential contractors compete on price and design quality. The DB auctioneer assigns a set of expert reviewers to independently evaluate each design proposal submitted by the potential contractors. The challenge in identifying the strategic effects comes from unobserved heterogeneity in design quality. We circumvent the confounding issue by exploiting the exclusion restriction that peer’s observed characteristics are
independent of one’s quality signals. We find that (i) reviewers have incentive to conform to the experienced reviewer’s score, and also that (ii) reviewers bias his/her evaluation of favourite design upward. The empirical results suggest a need for careful mechanism design that takes into account of experts’ incentives in various competitive environments where the outcome relies on experts’ opinions.

In chapter 3, we empirically investigate the costs and benefits of two widely used procurement auction formats: Lump-Sum Auction (LSA) and Unit-Price Auction (UPA). In an LSA, each bidder submits a single price for an entire project, and receives its price bid upon completion of the project if there is no contractual change. In a UPA, each bidder submits a price for each construction component of a project, and the final pay to the contractor depends on quantity change on contracted components during the construction phase. The existing literature only provides partial predictions about the performance differences between the two auction formats. In order to deal with the procurer’s selection of auction formats, we exploit exogenous variation in (i) procurer’s capacity constraints, and (ii) expected weather disturbances conditional on its realization. We estimate that LSA would reduce the final pay to the contractor by more than 19% relative to UPA for the projects with medium level of project risk. We find no evidence of differences in the degree of cost overruns/underruns between LSA and UPA projects. The empirical results indicate a potential for saving in infrastructure expenditure, and suggests a need for a model that takes into account for both ex-ante and ex-post bidders’ behaviour in both types of auction formats.
Acknowledgements

I am indebted to Professor Victor Aguirregabiria for invaluable advice and generous guidance. I would also like to thank Professor Ettore Damiano, Professor Rahul Deb, Professor Yao Luo, Professor Robert McMillan, Professor Junichi Suzuki, Professor Yuanyuan Wan and the participants in CEPR/JIE, EARIE, and Jornadas de Economía Industrial for helpful comments and discussions.
## Contents

1 Strategic Design under Uncertain Evaluations: Theory and Evidence from Design-Build Auctions

1.1 Introduction .............................................. 1

1.2 Institutional Details and Data ............................ 5
   1.2.1 Design-Build Procurement Auction .................. 5
   1.2.2 Data .................................................. 8

1.3 A Structural Model of a DB Auction ..................... 12
   1.3.1 Model ................................................ 13
   1.3.2 Numerical Exercise .................................. 18

1.4 Identification ............................................ 20
   1.4.1 Econometric Specification of Structural Model ....... 21
   1.4.2 Separating the Sources of Variation in Equilibrium Strategies ............. 24
   1.4.3 Unidentified $\gamma$ ................................ 25
   1.4.4 Nonparametric Identification of Reduced Form Factor Model ............... 26

1.5 Estimation ................................................. 28

1.6 Results ................................................... 29

1.7 Does Evaluation Uncertainty Matter? .................... 32
   1.7.1 Simulation of DB Auction with Varying Number of Reviewers ............... 32
   1.7.2 Auction Design for Design Auction: Second-Price Auction with Design Score Contingent Transfer ......................... 33

1.8 Conclusion ............................................... 36

2 Identifying Strategic Behaviour in Expert Evaluations: Evidence from Design-Build Auctions

2.1 Introduction .............................................. 38

2.2 Institutional details and data ............................ 39
   2.2.1 Design-Build auctions ............................... 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2 Heterogeneity in reviewers’ evaluations</td>
<td>41</td>
</tr>
<tr>
<td>2.2.3 Data</td>
<td>42</td>
</tr>
<tr>
<td>2.2.4 Descriptive statistics on reviewers’ evaluations</td>
<td>43</td>
</tr>
<tr>
<td>2.2.5 Descriptive statistics on reviewers’ characteristics</td>
<td>44</td>
</tr>
<tr>
<td>2.2.6 Descriptive evidence on strategic evaluations using score distance</td>
<td>46</td>
</tr>
<tr>
<td>2.2.7 Descriptive evidence on strategic evaluations using design rankings</td>
<td>48</td>
</tr>
<tr>
<td>2.3 A structural model of strategic evaluations</td>
<td>51</td>
</tr>
<tr>
<td>2.3.1 Basic model</td>
<td>51</td>
</tr>
<tr>
<td>2.3.2 Model with reviewer’s preferences on the competition outcome</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3 Specification of reviewers’ preferences and information structure</td>
<td>55</td>
</tr>
<tr>
<td>2.4 Identification and Estimation</td>
<td>57</td>
</tr>
<tr>
<td>2.4.1 Identification</td>
<td>57</td>
</tr>
<tr>
<td>2.4.2 Estimation</td>
<td>58</td>
</tr>
<tr>
<td>2.5 Estimation Results</td>
<td>59</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>62</td>
</tr>
<tr>
<td>3 What Really Matters in Procurement of Risky Projects? Lump-Sum Auction versus Unit-Price Auction</td>
<td>63</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>63</td>
</tr>
<tr>
<td>3.2 Institutional Details and Data</td>
<td>66</td>
</tr>
<tr>
<td>3.2.1 Procurement process under LSA and UPA</td>
<td>66</td>
</tr>
<tr>
<td>3.2.2 Data</td>
<td>68</td>
</tr>
<tr>
<td>3.3 Model and Estimation</td>
<td>70</td>
</tr>
<tr>
<td>3.3.1 Empirical Model</td>
<td>70</td>
</tr>
<tr>
<td>3.3.2 Instruments and Local Average Treatment Effect</td>
<td>73</td>
</tr>
<tr>
<td>3.4 Estimation Results and Discussions</td>
<td>74</td>
</tr>
<tr>
<td>3.4.1 Estimation Results</td>
<td>74</td>
</tr>
<tr>
<td>3.4.2 Why Do We See What We See?</td>
<td>77</td>
</tr>
<tr>
<td>3.5 Conclusion</td>
<td>78</td>
</tr>
</tbody>
</table>
3.6 Appendix for Chapter 1 ................................................................. 84
  3.6.1 Definition of Observables ....................................................... 84
  3.6.2 Proof of Proposition 1 ............................................................ 84
  3.6.3 Proof of Proposition 2: Multiplicative Separability of Partially Reduced Payoff Function ....................................................... 84
  3.6.4 Proof of Proposition 3: Equilibrium Existence, Monotonicity, and Differentiability 85
  3.6.5 Proof of Proposition 4: ............................................................ 86
  3.6.6 Proof of Proposition 5: Multiplicative Separability of Bidding Strategies ....... 87
  3.6.7 Equilibrium Computation Algorithm ......................................... 87
1 Strategic Design under Uncertain Evaluations: Theory and Evidence from Design-Build Auctions

1.1 Introduction

Design competitions in the face of uncertain evaluations are common in a variety of settings. An academic researcher may face an uncertain evaluation of her grant proposal. A consulting firm may face an uncertain client’s evaluation of its unique proposal. For large infrastructure projects, public procurements that involve billions of dollars often solicit designs from consulting firms, and those competing firms are also uncertain about how government officials will evaluate their designs of the infrastructure project.1

Prompted by these examples, we study the effect of a client’s uncertain evaluation on suppliers’ design choices competing for a contract to produce a customized product. Transactions involving customized products require a design of the end-product by the supplier before the product is manufactured, but suppliers typically do not know precisely what end-product their client would like.2 Uncertain evaluations introduce an element of luck into design competitions, providing heterogeneous suppliers with heterogeneous incentives: good designers face a lower chance of winning from increased uncertainty in evaluation while bad designers face a higher chance of winning a contract.3 It is not clear how heterogeneous suppliers respond to a change in the degree of uncertainty in a client’s evaluation of their design proposals. Thus, the identification of the effect of uncertain evaluations on suppliers’ behaviour is an empirical question.

To address this question, we use hand-collected data on Design-Build (DB) procurement auctions from the Florida Department of Transportation (FDOT), which are used by many state departments of transportation in the U.S. and around the world.4 In a DB auction, bidders compete over price and design to win a contract to deliver an infrastructure project, ranging from bridge repair to building construction. Upon receiving price and design proposals, each reviewer employed by the FDOT independently evaluates and assigns a score to every design proposal. The quality score of a design proposal is then determined

---

1Public Private Partnerships, which have surged in popularity, are also an example of public procurement that involves a design competition among consulting firms.

2While suppliers may communicate with a client to reduce uncertainty, the client may be unwilling to do so since repeated interactions can be very costly. Speed of delivery is often an important consideration in procurement.

3Note that a supplier faces uncertainty in the evaluation of its rivals’ designs as well.

4As of October 2010, there are 39 state departments of transportation that use DB, including California, Delaware, Georgia, Minnesota, etc. DB auctions are also common in other developed countries, including Canada, Japan, and Sweden.
by the average reviewer’s evaluation. The bidder with the lowest price per quality score ratio (PQR) wins the project, and receives its price bid upon completing the project.\(^5\)

The data reveal a substantial amount of disagreements among reviewers for a given design proposal, and conversations with DB contractors confirm that uncertainty in design evaluation is a substantial concern among DB firms. Such uncertainty in design evaluations, which we refer to as \textit{evaluation uncertainty}, has not been considered to date in the vast auction literature. Indeed, to the best of my knowledge, there is virtually no previous empirical work that has investigated the implications of uncertain design evaluations on supplier behavior.

At the heart of the arguments, we develop an estimable model in which each bidder strategically chooses its price and design proposal in the face of uncertain design evaluations. The model allows for unobserved auction heterogeneity, and two sources of firm heterogeneity: variable costs that depend on the design quality choice and fixed costs. Most importantly, the model deals with the problem that the econometrician does not observe actual design quality of a proposal. We propose a sequential method for estimating the degree of evaluation uncertainty together with the primitives of the model.

We distinguish two sources of heterogeneity in reviewers’ evaluations: vertical reviewer heterogeneity and horizontal reviewer heterogeneity. Vertical reviewer heterogeneity comes from the fact that reviewers have different quality standards, which affect design scores of different bidders equally.\(^6\) While vertical reviewer heterogeneity is rank-neutral, horizontal reviewer heterogeneity is not. For instance, if a reviewer happens to place a higher value on the visual aspect of a bridge than other reviewers, the bidder with a fancy design proposal may unexpectedly obtain a high score from the reviewer. While we argue that both forms of reviewer heterogeneity are uncertain to bidders when they make their bids, this paper focuses on the role of horizontal reviewer heterogeneity, which we interpret as evaluation uncertainty, mainly because vertical reviewer heterogeneity does not affect the rankings of design proposals, and so is unlikely to affect bidders’ incentives in a DB auction.\(^7\)

\(^5\)PQR is a winner selection rule used by many state departments of transportation, including Alaska, Michigan, North Carolina, and South Dakota.

\(^6\)For example, reviewers’ leniency in assigning a score may be captured by vertical reviewer heterogeneity since a lenient reviewer tends to give a high score to every design proposal.

\(^7\)In my sample, most of reviewers show up only once or twice in ten years and reviewers are not experienced, which partially explains the discrepancy in reviewers’ evaluations. This observation is intriguing since the FDOT could reduce evaluation uncertainty by training reviewers through on-the-job training. Non-repeated use of reviewers could be rationalized as a precautionary measure against corruption between reviewers and bidders.
We show nonparametric identification of a structural model of a DB auction and estimate in two steps. First, we use data on price bids, reviewers’ evaluations, and observable characteristics together with a reduced-form model to recover bidders’ cost structure and evaluation uncertainty, taking into account the presence of unobserved auction heterogeneity and vertical reviewer heterogeneity. Second, we combine the estimates from the previous step with bidders’ first-order optimality conditions to identify the distribution functions of bidders’ private cost information.

The estimated model predicts that uncertain design evaluations raise uncertainty in the auction outcomes from the auctioneer’s standpoint. A large amount of uncertainty in design evaluation implies that winner selection is heavily influenced by luck. Inefficient bidders benefit from noisy evaluations since the bidders would otherwise lose in the absence of reviewers’ subjective judgments. Contrary to inefficient bidders, efficient bidders are less likely to win due to increased contribution of luck. These asymmetric effects of evaluation uncertainty on bidding incentives results in inefficient bidders becoming less competitive than efficient bidders upon increasing evaluation uncertainty. Consequently, greater bidders’ uncertainty in evaluations exacerbates the dispersion in price and design quality, resulting in greater uncertainty for the auctioneer in auction outcomes.

A large amount of uncertainty in auction outcomes may not be desirable from the auctioneer’s standpoint. A low winning price may raise a chance of bankruptcy during the implementation of the project while a high winning price may make the procurement unaffordable if the auctioneer is budget-constrained. In light of these issues, we propose a second-price auction in which a bidder’s design score determines the transfer amount, shutting down the effect of evaluation uncertainty on bidders’ behaviour. The alternative mechanism is simple and dominant strategy-implementable. Based on the estimated model, we find that the alternative mechanism reduces the auctioneer’s uncertainty in both the amount paid and the quality of the winning design by 22% and 33%, respectively.\(^8\)

The results obtained in this paper has an important economic implication in that design competition under uncertain design evaluations may affect a client adversely due to the endogenous response of suppliers. The client may ensure the attributes of the end-product by committing to select the contractor based on an objective measure, and not letting his subjective judgment distort suppliers’ behaviour.\(^9\)

\(^8\) The auctioneer’s uncertainty is measured by standard deviation of the winning price and the winning design quality.

\(^9\) An example of such objective measure is price.
We build on and contribute to two growing literatures: the literature on multi-attribute auctions, and the literature on the structural estimation of auction models. DB auctions are a particular type of a multi-attribute auction, also known as a scoring auction, in which bidders compete in attributes of an auctioned product. While first-price low-bid auctions rank bidders solely based on price, scoring auctions select the winner based on several attributes of a bid; the DB auction is a type of scoring auction where the winner is selected based on PQR.10 Krasnokutskaya, Song, and Tang (2012) empirically investigate an auction environment in which the attributes-based winner selection rule is unknown to bidders.11 In this paper, we argue that disclosing information about a buyer’s preference may not be straightforward, and uncertainty in subjective evaluation is likely to remain after disclosing the selection rule whenever a design score is used as a means of selecting the winner.12

From a technical point of view, we add to the literature of multi-attribute auctions by developing an estimation approach when some attributes are unobserved, precluding the standard inversion approach pioneered by Guerre, Perrigne, and Vuong (2000). As the econometrician observes only a few noisy quality signals for each design proposal, the actual design decision of each bidder cannot be pointwise-identified from the data. We also take into account other relevant issues in estimation. Unobserved auction heterogeneity is also particularly relevant in the context of the analysis here since the effect of evaluation uncertainty is local.13 As emphasized in Krasnokutskaya (2011), ignoring the existence of unobserved auction heterogeneity overemphasizes the dispersion in the private information of bidders, which in turn leads the effect of uncertainty in design evaluation on bidders’ behaviour to be underestimated.14 My structural model together with its implied reduced form allow for estimation of bidders’ cost and evaluation uncertainty in the presence of various types of unobserved heterogeneities.

---


11 Krasnokutskaya, Song, and Tang (2012) make a distinction between multi-attribute auctions and scoring auctions based on whether the auctioneer’s taste is observed or not. In this paper, we do not make this distinction and treat them synonymously.

12 Nakabayashi and Hirose (2013) consider public procurement auctions in Japan in which quality is verifiable. However, many state departments of transportation assign reviewers to evaluate design proposals.

13 i.e., reviewers’ preference heterogeneity can switch the rankings of design proposals only when bidders’ design proposals are close to each other in true design quality.

14 We also allow bidders to have private information in variable cost and fixed cost of production to rationalize the observed bids’ distribution as discussed in Asker and Cantillon (2008).
The rest of the paper is organized as follows. Section 1.2 provides institutional details about the DB auction process, and describes the data. Section 1.3 develops a structural model and derives comparative statics results. Section 1.4 shows identification of the structural model. Section 1.5 describes the structural estimation procedure, and Section 1.6 presents the estimation results. Section 1.7 demonstrates the economic significance of evaluation uncertainty through simulation, and also demonstrates the effect of a change in auction mechanism on the auction outcomes. Section 1.8 concludes.

1.2 Institutional Details and Data

1.2.1 Design-Build Procurement Auction

Here we describe DB procurement process in detail, and explain some institutional details that become important from structural modeling perspective. Some questions to be answered in this section include: (i) Who are the reviewers and how are they selected? (ii) What are the necessary pre-qualification requirements to become a “bidder” in a DB auction? (iii) Do bidders know who would be reviewing their designs ex-ante? (iv) Would bidders have incentives to lobby reviewers to win auctions?

DB procedure can be decomposed into two consecutive stages, a pre-selection stage and a bidding stage. In the pre-selection stage of a DB procurement, the FDOT posts an advertisement on-line which lists information about the project location, description of work, criteria for evaluating a letter of interest, and technical qualification requirements. Then, reviewers are selected from a pool of the FDOT employees by a department secretary based on qualifications and availability. Meanwhile, an interested builder and a designer match to form a DB firm. The DB firm then writes a letter of interest to the FDOT. The appointed reviewers then evaluate the letter of interest based on the criteria described in the advertisement, which include past performance grades of builders and designers, DB experience, and current capacity of builders. Those DB firms that are judged as pre-qualified by reviewers are short-listed and become a “bidder”. There is no specific rule as to how many DB firms should be short-listed, and the number of short-listed bidders ranges from 2 to 5 in the sample. The identities of these bidders are posted on-line and become common knowledge. The bidders then receive the request for proposal, which describes

---

15 A DB firm usually consists of a builder sub-contracting with a designer.
16 There was only one auction with one bidder in the original DB auction records sent by the FDOT, and this auction is not used in any part of the analyses.
detailed specification of the project and design evaluation criteria.\textsuperscript{17}

Following the pre-selection stage, the bidders now enter the bidding stage. All the bidders and the reviewers meet in a mandatory pre-proposal meeting in which the reviewers provide proposal instructions and the scope of the project. Both design and price bids are usually due 1 to 2 months after the pre-proposal meeting, and the design and price bids need be sent to the FDOT in separate envelopes. Next, the reviewers independently evaluate each design proposal, and the quality score of a design is determined by the average reviewer’s evaluation. Finally, the price bids are opened to determine the winner of the project based on PQR.

The answers to some of the questions raised at the beginning of this subsection are clear from the above description of events. First, the reviewers are all employees of the FDOT. In addition, the appointed reviewers’ compensation are salary based, and not based on each review task. Therefore, it is likely that reviewers’ incentive to exert effort in reviewing tasks is not strong, and could potentially explain some of the variation in reviewers’ evaluations. Second, past record is an important factor for a DB firm to be pre-qualified as a bidder. From a subset of DB records for which the identities of applicants are available, we find that the number of bidders significantly differs from the number of applicants. This observation is intriguing since if the FDOT simply wants to lower the winning PQR, decreasing competition through removing potential bidders would adversely affect the winning PQR. An explanation for this fact is that the FDOT’s objective is not only to lower the winning PQR, but also to avoid ex-post default after the

\textsuperscript{17}Design evaluation criteria vary across auctions. Some repeatedly observed evaluation criteria include warranty, innovative aspect of design, maintenance of traffic, construction methods, commitment to environmental protection, project schedule, etc.
contract is made. As is well known, renegotiation is prevalent in government procurement auctions.\textsuperscript{18} If the FDOT wishes to avoid costly renegotiation and/or possible ex-post default, then the FDOT may choose to moderate competition through pre-screening process, leaving some rent to the winner of a project.\textsuperscript{19} Lastly and most importantly, the bidders do observe the reviewers who evaluate their designs before bidding stage in a pre-proposal meeting. While the presence of pre-proposal meeting could imply that some of uncertainty is resolved ex-ante, it is not clear how much information a bidder possesses about the reviewers at the time of bidding. Knowing the identities of reviewers is meaningful to a bidder only if some pattern or tendency can be inferred from the reviewers’ identities. The sample of DB projects shows that the majority of reviewers are appointed only once in a decade and thus, the bidders are unlikely to make an inference about reviewers’ characteristics from their past evaluations.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Builder</td>
<td>3.61</td>
<td>3.68</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Designer</td>
<td>3.05</td>
<td>3.23</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Reviewer</td>
<td>1.68</td>
<td>1.68</td>
<td>1</td>
<td>16</td>
</tr>
</tbody>
</table>

\textsuperscript{a} The sample contains 110 DB auctions procured between years 2000 and 2011.
\textsuperscript{b} In total, there are 53 builders, 64 designers, and 250 reviewers in the sample.

The pre-proposal meeting also casts a doubt on bidders’ incentives to lobby reviewers. Table 1 shows how frequently a particular reviewer is observed in the sample. On average, a reviewer shows up in less than two auctions and most reviewers show up only once in a decade. Thus, bidders play a one-shot game rather than a repeated game if they are to connect to a particular reviewer. While the one-shot nature of the game may not preclude bidders’ incentive to lobby reviewers, its scope can be significantly limited by reducing the benefit of establishing a reviewer specific connection. Further, it is difficult for a reviewer to ensure winning of a particular bidder in a DB auction. While the reviewer could raise the chance of winning for a particular bidder by enlarging the gap in quality scores between the bidder and others, the reviewer cannot ensure that the bidder wins since he/she does not know how other reviewers will evaluate

\textsuperscript{18} Theoretical analyses of procurement auctions with ex-post bankruptcy is found in Board (2007).

\textsuperscript{19} For the analysis of incomplete contract in public procurement projects, see Bajari, Houghton, and Tadelis (2006).
the design and what the price bids are.

In short, the reviewers from the FDOT have a weak incentive to exert effort, and have little experience in evaluating design proposals in DB auctions. Past performance is an important determinant that makes a DB firm pre-qualified, and the FDOT’s objective is unlikely achieving the lowest possible PQR. While bidders meet with reviewers before deciding on their bids, knowing the identities of reviewers is unlikely to reduce uncertainty in reviewers’ evaluations.

1.2.2 Data

We investigate a sample of DB auctions that took place between years 2000 and 2011 in Florida. Although DB is also a common practice in other states (e.g., Alaska, Pennsylvania, Minnesota, etc.), scoring rules and point systems differ across these state departments of transportation. Therefore, a single department of transportation, the FDOT, is chosen for consistency and auction record availability. In particular, records on design evaluations in other state departments of transportation are often aggregated and not preserved at the individual reviewer level.

The sample of DB auctions used in the analysis here is a result of selecting a subset of original DB auctions, and the selection procedure is as follows. First, we requested the FDOT for the records of all DB auctions that have been procured between years 2000 and 2011. The FDOT provided 152 auction records. Second, we manually compiled a dataset from the provided records, and all the auction records with only 1 bidder, or missing engineer’s estimate of project cost, or with modified scoring rule, or missing reviewers’ evaluations are excluded from the dataset. An engineer’s estimate of project cost is an important control for project size heterogeneity across auctions in the data. Moreover, the original dataset contained a variant of DB auctions (DB’ auctions) in which the scoring rule involves a time incentive component\(^{20}\). We do not consider DB’ auctions here to maintain consistency in auction format.\(^{21}\) Those auctions with no individual reviewer’s evaluations are also excluded since the variation in reviewer level scores is the focus of the paper. The selected sample is complemented by bidder characteristics, which we obtained through web-scraping.

\(^{20}\) The variant of DB auctions is a combination of DB and A+B auction studied in Bajari and Lewis (2011).

\(^{21}\) It turns out that DB’ auctions have a larger horizontal reviewer heterogeneity than DB auctions on average, suggesting a selection of auction format by the FDOT.
In total, 42 auctions are excluded from the sample and not used in any part of the data analysis. Consequently, we are left with 110 DB auctions with detailed information on design evaluations. Out of the excluded records, 28 auctions are removed for having DB’ rule, 11 auctions are removed for not showing engineers’ estimates, 2 auctions are removed for not showing individual reviewer level evaluation scores, and 1 auction is removed for having only one participating bidder.

Figure 2 is an actual record of design evaluations for a bridge construction project, and is one of the DB auctions with a large spread in design evaluations across reviewers in the sample. The first and second rows of the table shows the identity of 3 bidders and 5 reviewers, respectively. The first and second columns show 10 evaluation categories and weights. Each reviewer independently reviews each quality aspect of a design proposal, and assigns a score out of the category specific maximum score. Then, these scores are summed across all categories to obtain the total score of a design proposal, which we define as a reviewer’s evaluation. These total scores are averaged across reviewers to determine the quality score of a bidder’s design proposal. The three bidders are ranked by their PQR, and the bidder with the lowest PQR wins the project. A large variation in reviewers’ evaluations can be easily verified in this auction. For example, the difference in evaluation scores assigned by JD and DK is 24 points for Cone & Graham/Jacob, which is 24% of the maximum allowable points. Also, JD ranks Cone & Graham/Jacob fifth and Johnson Bros./GAI third, while DK ranks Cone & Graham/Jacob first and Johnson Bros./GAI fourth.

---

22 The set of excluded auctions accounts for 27% of the original auction records.
Table 2: Summary Statistics of Key Variables

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winning PQR ($1,000 / score point)</td>
<td>193</td>
<td>266</td>
<td>3.125</td>
<td>1125</td>
<td>110</td>
</tr>
<tr>
<td>Winning Price ($1,000,000)</td>
<td>16.6</td>
<td>22.8</td>
<td>0.253</td>
<td>103</td>
<td>110</td>
</tr>
<tr>
<td>Winning Quality Score (score point)</td>
<td>86.3</td>
<td>5.57</td>
<td>69.7</td>
<td>95.5</td>
<td>110</td>
</tr>
<tr>
<td># Bidders / Auction</td>
<td>3.12</td>
<td>0.534</td>
<td>2</td>
<td>5</td>
<td>110</td>
</tr>
<tr>
<td># Reviewers / Auction</td>
<td>3.82</td>
<td>0.800</td>
<td>3</td>
<td>6</td>
<td>110</td>
</tr>
<tr>
<td>PQR ($1,000 / score point)</td>
<td>249</td>
<td>346</td>
<td>3.125</td>
<td>1945</td>
<td>338</td>
</tr>
<tr>
<td>Price ($1,000,000)</td>
<td>20.9</td>
<td>28.6</td>
<td>0.253</td>
<td>142</td>
<td>338</td>
</tr>
<tr>
<td>Reviewer’s Evaluation (score point)</td>
<td>84.3</td>
<td>8.19</td>
<td>38.6</td>
<td>100</td>
<td>1296</td>
</tr>
</tbody>
</table>

The summary statistics is calculated based on 110 DB auctions procured between years 2000 and 2011.

Figure 2: Summary of Evaluation Scores on E7E10 Barge Canal Bridge Design Build Project

Table 2 shows the summary statistics of the key variables. Prices are adjusted for inflation and expressed in 2011 USD. The average winning price is more than 16 million USD. Considering the fact that the average winning price in usual first-price low-bid auction is 7.4 million USD in Florida, DB auctions seem to be adopted for relatively large scale projects. A quality score is the average of quality scores.
across reviewers’ evaluations, which is the weighted sum of category level scores.\textsuperscript{23} Since the maximum quality scores vary across auctions, every quality score is standardized by its maximum possible score, and expressed out of 100 points.\textsuperscript{24}

Table 3: Distribution of Winning Price and Quality Score

<table>
<thead>
<tr>
<th></th>
<th>Lowest Price</th>
<th>Non-Lowest Price</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Quality Score</td>
<td>38</td>
<td>19</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>(34.5%)</td>
<td>(17.2%)</td>
<td>(51.8%)</td>
</tr>
<tr>
<td>Non-Highest Quality Score</td>
<td>51</td>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>(46.3%)</td>
<td>(1.8%)</td>
<td>(48.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>89</td>
<td>21</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>(80.9%)</td>
<td>(19.1%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

The above calculation is based on 110 DB auctions procured between years 2000 and 2011.

Table 3 shows how many winners received the non-highest design quality score, and how many winners bid the non-lowest price. It is readily seen that neither lowest price bidder nor highest quality score bidder always win in a DB auction. Indeed, the majority of the winners in DB auctions do not receive the highest quality score.

In order to gain a sense about how much variation exists in price bids and reviewers’ evaluations, consider the following simple decomposition of variance in the natural logarithm of price, reviewer’s evaluation, and price per reviewer’s evaluation. Table 4 shows the decomposition of variance in the above three variables into between-auction, within-auction-between-bidder, and within-bidder-between-reviewer. As prices do not vary within-bidders, within-bidder-between-reviewer variation in price per reviewer’s evaluation is entirely driven by the variation in reviewers’ evaluations. As the sample contains projects of different sizes, most of the variation in price per reviewer’s evaluation is driven by between-auction variation. Conversely, the rest of within-auction variations are comparable in size. Note here that within-bidder-between-reviewer variation in reviewers’ evaluations accounts for 69% of its total variance. Also, 17% of within-auction variation in price per reviewer’s evaluation is explained by within-bidder-between-reviewer variation in evaluations.

\textsuperscript{23}Each evaluation category has an auction specific maximum allowable score.

\textsuperscript{24}Note that rescaling of quality scores does not introduce any problem to my analysis here as the winner selection rule is price per quality score which is scale invariant.
Table 4: Variance Decomposition of Price Bids and Reviewers’ Evaluations

<table>
<thead>
<tr>
<th>Variables</th>
<th>Between-Auction</th>
<th>Within-Auction</th>
<th>Between-Bidder</th>
<th>Within-Bidder</th>
<th>Between-Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Price)</td>
<td>1.43</td>
<td>0.158</td>
<td>(0.0971)</td>
<td>(0.00739)</td>
<td></td>
</tr>
<tr>
<td>ln(Reviewer’s Evaluation)</td>
<td>0.0389</td>
<td>0.0516</td>
<td>(0.00540)</td>
<td>(0.00400)</td>
<td>0.0779</td>
</tr>
<tr>
<td>ln(Price per Reviewer’s Evaluation)</td>
<td>1.43</td>
<td>0.170</td>
<td>(0.0970)</td>
<td>(0.00842)</td>
<td>0.0779</td>
</tr>
<tr>
<td>Obs</td>
<td>110</td>
<td>338</td>
<td>1296</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Estimates of standard deviations at each level of the hierarchy are presented.
\textsuperscript{b} Standard errors in parentheses

1.3 A Structural Model of a DB Auction

In order to allow for endogenous response of bidders to a change in the auction environment, we construct a model where bidders compete over price and design quality under uncertain design evaluations and uncertain rivals’ bids. That is, the model incorporates three types of uncertainty: i) A bidder is uncertain about what its rivals’ bids are, ii) how its design proposal will be evaluated, and iii) how its rivals’ designs will be evaluated.

An efficient bidder, who submits a low PQR, experiences a lower probability of winning while an inefficient bidder, who submits a high PQR, may win with a higher probability upon an increase in evaluation uncertainty. That is, evaluation uncertainty \textit{curtails} the differences in bidders’ chances of winning. The asymmetric effects of evaluation uncertainty on bidding incentives across different types of bidders together with strategic uncertainty makes the equilibrium effects of evaluation uncertainty on bidders’ behaviour not obvious.

On top of strategic uncertainty, the model introduces multi-dimensional types in a bidder’s cost structure to account for complexity in bidding strategies. For example, a bidder consisting of a good builder and a bad designer may take a low-price-low-quality strategy while a bidder consisting of a bad builder with a good designer may take a high-price-high-quality strategy, exploiting their comparative advantages. However, these two completely different types of bidders may end up with very similar objective PQR.
bids, and generating complex bidding strategies is non-trivial without multi-dimensional types.

In the following, the multidimensional choice problem of a bidder of multidimensional types is reduced to one dimensional choice problem of a bidder of a single pseudo type as in Asker and Cantillon (2008) to gain model tractability. A Bayesian Nash Equilibrium of the model is characterized, and a numerical exercise is provided to shed a light on how bidders’ respond to a change in evaluation uncertainty.

1.3.1 Model

Consider \( N \equiv |\mathcal{N}| \) risk neutral bidders where \( \mathcal{N} \) denotes the set of bidders in a given DB auction. As a request for proposals lists all pre-qualified bidders, every bidder knows who participate in a DB auction. For the sake of notational simplicity, we suppress the auction subscript \( a \) in this section. Let \( \{p_i, q_i\} \in \mathbb{R}_+^2 \) be the price bid, and the objective quality of the design proposed by bidder \( i \in \mathcal{N} \), respectively. Also, let \( b_i \equiv p_i/q_i \) be the objective PQR bid of bidder \( i \), which is assumed to be responsive only within the support \([0, \overline{b}] \subset \mathbb{R}_+\), and the government rejects proposals outside the bounds.\(^{25}\) To capture comparative advantage aspect of a bidder in a DB auction, bidder \( i \) of the model is characterized by a variable cost type, \( v_{c_i} \), and a fixed cost type, \( f_{c_i} \).

Now, define the ex-post payoff of bidder \( i \) by:

\[
\pi_i^{\text{post}} = \begin{cases} 
    p_i - v_{c_i} C(q_i) - f_{c_i} & \text{if bidder } i \text{ wins} \\
    0 & \text{otherwise}
\end{cases}
\]

where \( v_{c_i} C(q_i) \) and \( f_{c_i} \) consist of variable and fixed cost of delivering the project at quality level \( q_i \). \( C(.) \) is increasing, convex, and differentiable (i.e., \( C_{q} > 0, C_{qq} > 0 \)). The convexity is necessary to generate a smooth substitution between price and design quality given the winner selection rule. \( C(.) \) is also common across bidders.\(^{26}\) Let \( c_i \equiv \{v_{c_i}, f_{c_i}\} \in T_i \subset \mathbb{R}_+^2 \) and \( c_{-i} \) denote the vector of bidders’ types in the auction excluding bidder \( i \). Bold cases are used to refer to a vector with \( i_{th} \) element corresponding to bidder \( i \)’s

\(^{25}\)An interpretation of the boundedness assumption is that the government does not accept a bid that does not meet its cost.

\(^{26}\)Later in specifying the econometric form of the model, we allow for \( v_{c_i} \) and \( f_{c_i} \) to be partially observed by bidder \( i \)’s rivals. However, we assume these two parameters are entirely private information of bidder \( i \) in this section without loss of generality and for expositional simplicity.
choice or type (e.g., \( b \equiv [b_1, \ldots, b_i, \ldots, b_N] \), \( c_{-i} \equiv [c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_N] \)). Also, let \( f(c_{-i} | c_i) \) denote the joint distribution of \( c_{-i} \) conditional on the realization of bidder \( i \)'s type, which is everywhere differentiable over its compact support.

Define evaluation noise as the total amount of subjectivity that a set of reviewers introduces into a quality score of bidder \( i \)'s design proposal. Further, define evaluation uncertainty as the degree of dispersion in evaluation noise. Assume that evaluation noise, \( w_i \), generates realization of a quality score by multiplicatively affecting \( q_i \), and is independently distributed from design quality. Let \( F_{w_i}(w) \) be the joint distribution function of evaluation noise \( w_i \). The probability of winning conditional on a vector of PQR bids, \( \tilde{G}_i(b) \), is given by:

\[
\tilde{G}_i(b) = \int_{w} 1\{\text{bidder } i \text{ is the winner given } b\} dF_{w_i}(w)
\]

\[
= \int_{w} 1 \left\{ \frac{p_i}{q_i w_i} < \frac{p_j}{q_j w_j}, \forall j \neq i \right\} dF_{w_i}(w)
\]

\[
= \int_{w} 1 \left\{ \ln(b_i) - \ln(w_i) < \ln(b_j) - \ln(w_j), \forall j \neq i \right\} dF_{w_i}(w)
\]

As bidder \( i \) does not observe the private information of its rivals, bidder \( i \)'s probability of winning is obtained by integrating \( \tilde{G}_i(b) \) over the distribution of bidder \( i \)'s rivals' strategies. Let \( \psi_{-i}(c_{-i}) \) denote a vector of bidder \( i \)'s rivals' (arbitrary) PQR strategies. Then,

\[
G_i(b_i, \psi_{-i}) \equiv \int_{c_{-i}} \tilde{G}_i(b_i, \psi_{-i}(c_{-i})) f(c_{-i} | c_i) dc_{-i}
\]

Finally, bidder \( i \)'s interim expected payoff conditional on participating in a DB auction is defined as \( \pi^{int}_i \equiv G_i(b_i, \psi_{-i}) \pi^{post}_i \). The problem of bidder \( i \) is then defined as:

\[
\max \left\{ \max_{p_i, q_i} \pi^{int}_i \right\} \text{ s.t. } p_i/q_i = b_i \in [0, \bar{b}], 0 \right\}
\]

Let \( \{p_{i}^{BR}(c_i), q_{i}^{BR}(c_i)\} \) be a best response correspondence of bidder \( i \) of type \( c_i \) with an arbitrary belief about its rivals' strategies. A Bayesian Nash Equilibrium is a state in which every bidder's belief is consistent with best responses of its rivals. A Bayesian Nash Equilibrium is called “pure” if every bidder’s strategy is a deterministic function of own type.
Definition 1. A Pure Strategy Bayesian Nash Equilibrium consists of a profile of best response functions \( \{p^{BR}(c), q^{BR}(c)\} \) in which every bidder \( i \in \mathcal{N} \) believes its rivals bid according to \( \{p_{-i}^{BR}(c_{-i}), q_{-i}^{BR}(c_{-i})\} \).

Before proceeding, we show that this two-dimensional decision problem can be transformed into a one-dimensional choice problem. Consider the optimization problem in (1) with an additional constraint that \( b_i = \alpha \) for some \( \alpha \in [0, \bar{b}] \). This constrained optimization problem has a unique solution, and the values of price and objective design quality that solve this problem are given by the following closed form expressions:

\[
q_i(\alpha) = C_q^{-1}(\alpha/vc_i) \tag{2}
\]
\[
p_i(\alpha) = q_i(\alpha)C_q^{-1}(\alpha/vc_i) \tag{3}
\]

where \( C_q^{-1}(\cdot) \) is the inverse of \( C_q(\cdot) \).

**Proposition 1.** For any given \( b_i = \alpha \in [0, \bar{b}] \), there is always a unique pair of \( \{p_i, q_i\} \in \mathbb{R}_+^2 \) that maximizes \( i \)'s interim expected payoff conditional on participation.

Proof in Appendix. This proposition establishes that pricing and designing decisions are uniquely determined for any given PQR. It follows that the problem of a bidder can be rewritten as one-dimensional choice problem, such that:

\[
\max_{b_i \in [0, \bar{b}]} G_i(b_i, \psi_{-i}) (p_i(b_i) - vc_i C(q_i(b_i)) - fc_i)
\]

where \( p_i(\cdot) \) and \( q_i(\cdot) \) are the functions defined in (2) and (3), respectively. Now, we make a simplifying assumption on \( C(\cdot) \) for tractability.

**Assumption 1.** The design cost function \( C(\cdot) \) is homogeneous of degree \( \gamma > 1 \).

Assumption 1 implies scale invariance and allows for sorting of bidders in terms of a single index by guaranteeing that equilibrium bidding strategies be additively log-separable in auction-bidder heterogeneities.

**Proposition 2.** Given Assumption 1, equilibrium PQR strategy of bidder \( i \) is a sole function of a single
index, \( e_i \equiv f c_i / C_q^{-1}(1/vc_i) \).

\[
p_i(b_i) - vc_i C(q_i(b_i)) - fc_i = C_q^{-1}(1/vc_i) (u(b_i) - e_i)
\]

where \( u(b_i) \equiv b_i C_q^{-1}(b_i) - vc_i C(C_q^{-1}(b_i)) \).

Proof in Appendix. The single index is an increasing function of \( vc_i \) and \( fc_i \). Suppose, for example, that the quality cost function is a power function, such that \( C(q) = q^\gamma \) with \( \gamma > 1 \). \( \gamma \) captures the weight assigned on \( vc_i \) and \( fc_i \) within \( e_i \), and \( vc_i \) and \( fc_i \) would equally contribute to \( e_i \) when \( \gamma = 2 \) (i.e., \( e_i = vc_i fc_i \)). In the extreme case where \( \gamma \) is substantially larger than 2, \( e_i \) is essentially determined by \( fc_i \).

Define \( e_i \) as the efficiency type of bidder \( i \). It is trivial to see that PQR strategy is a sole function of efficiency type \( e_i \).

\[
\psi_i(e_i) \equiv \arg \max_{b_i \in [0, \bar{b}]} C_q^{-1}(1/vc_i) G_i(b_i, \psi_{-i}) (u(b_i) - e_i)
\]

\[
= \arg \max_{b_i \in [0, \bar{b}]} G_i(b_i, \psi_{-i}) (u(b_i) - e_i) \quad \forall \ i \in \mathcal{N}
\]

The following simplifying assumption on the distribution of evaluation noise \( w_i \) is made in order to characterize equilibria.

**Assumption 2.** (Smooth Density and Independence) : Log evaluation noise, \( \ln(w_i) \), is drawn independently from a smooth density with an infinite support.

**Assumption 3.** (Profitable Participation) : Every bidder has a chance to make some profit (i.e., \( \bar{b} > u^{-1}(\bar{e}) \) where \( \bar{e} \) is the most inefficient bidder).

Assumption 2 and 3 together guarantee differentiability of the probability of winning function and participation of every bidder.

**Proposition 3.** Equilibrium Existence, Monotonicity, and Differentiability: There exists a pure strategy Bayesian Nash Equilibrium. In any equilibrium, \( \psi_i(e_i) \) is non-decreasing and differentiable in \( e_i \) \( \forall \ i \in \mathcal{N} \) in the interior of the domain.
Proof in Appendix. Let \( \{p^\psi(c), q^\psi(c)\} \) be the corresponding price and objective design quality strategy profiles in an equilibrium. If bidders’ strategies are interior, which we assume for the rest of the paper, the first order optimality condition from the above one-dimensional choice problem together with the ratio of (2) and (3) gives the following two equations:

\[
\psi_i(e_i) = vc_i C_q(q^\psi_i(c_i)) - \frac{u_\psi(\psi_i(e_i))}{u(\psi_i(e_i)) - e_i} G_i(\psi(e))
\]

where \( g_i(\psi) \equiv \partial G_i(\psi)/\partial \psi \) and \( \psi(e) \) is an equilibrium PQR strategy profile. Condition (4) is interpreted as bidder \( i \)'s offer of price per quality is set at the marginal cost of providing a unit of design quality. Condition (5) is commonly seen in an auction model with private information in which the marginal cost of raising PQR is equalized to its marginal benefit.

Note that condition (4) is independent of the distribution of evaluation noise. If a bidder becomes less competitive upon a change in the distribution of evaluation noise, then it will respond by producing a design of higher quality, and there is no way to lower its price while raising both its objective PQR and its design quality. Therefore, any change in the distribution of evaluation noise induces positive co-movements among price, design quality, and PQR bids though this effect can be either positive or negative.

**Proposition 4.** Let \( \tau \) be any parameter that affects the winner selection outcome, but does not have any effect on bidders’ exogenous costs. Then, \( \text{sign} \left( \frac{d\psi_i(e_i)}{d\tau} \right) = \text{sign} \left( \frac{dp^\psi_i(c_i)}{d\tau} \right) = \text{sign} \left( \frac{dq^\psi_i(c_i)}{d\tau} \right) \). That is, a cost-irrelevant parameter \( \tau \) induces positive co-movements in bidders’ strategies.

Proof in Appendix. \( \tau \) can be anything that influences strategic behaviour, but does not enter bidder’s cost. For instance, suppose that \( \tau \) represents the number of participating bidders in an auction. An intense competition may lower the price of the product but it may also deteriorate the quality of the customized product, invoking race to the bottom.

A particular implication of this proposition is when \( \tau \) represents a measure of uncertainty in reviewers’ evaluations. The intuition behind Proposition 4 is straightforward. Suppose that a bidder decides to become less competitive upon an increase in evaluation uncertainty for some reason. An increase in its PQR bid implies that each unit of design quality is offered at a higher price. Therefore, the bidder
strategically substitute design quality for price, producing a design of higher quality at a higher price.

Proposition 4 implies that it may be non-trivial to improve auction outcomes in both price and design quality through a marginal change in the auction rule. For example, any attempt to improve both price and design quality by changing the weight assigned on quality score in the winner selection process would fail in improving either price or design quality. This result suggests that a policy maker may be interested in a mechanism that shuts down the within-bidder substitution between price and design quality choices. We provide such a mechanism in Section 1.7.

1.3.2 Numerical Exercise

While a theoretical characterization of bidders’ equilibrium behavior is difficult, a numerical exercise may shed a light on the effects of evaluation uncertainty on bidders’ behaviour of different types. We also demonstrate the equilibrium effect of evaluation uncertainty on the distribution of price and design quality bids. All the parameter values in this numerical exercise are set equal to the estimates obtained from a structural estimation of the model.\textsuperscript{27}

Figure 3 illustrates the effect of evaluation uncertainty on the probability of winning function keeping rivals’ strategies constant. An increase in evaluation uncertainty \textit{flattens out} the probability of winning function. There are two distinct paths where increased randomization affects the winner selection process. The first path is the level effect that is heterogeneous across different types of bidders. Suppose, for example, that there is an exogenous increase in evaluation uncertainty. Inefficient bidders expect to win the project with higher probability than before since these bidders have little chance of winning in the absence of reviewers’ subjective judgments. Therefore, inefficient bidders have an incentive to shade their bids and enjoy a higher payoff upon winning. In contrast to inefficient bidders, efficient bidders experience an exogenous decrease in their chance of winning, generating incentive to become more competitive to ensure they win. Therefore, the level effect generates a greater dispersion of PQR bids and associated

\textsuperscript{27}To keep the computation of equilibrium simple, we assume that $\ln(w_i)$ follows Type 1 Extreme Value distribution. This assumption allows for a closed form expression for the probability of winning function $G_i(.)$ that resembles Tullock’s Contest Success Function.
choice of price and design quality bids.

Figure 3: The Effects of Uncertain Design Evaluation on Probability of Winning

![Figure 3: The Effects of Uncertain Design Evaluation on Probability of Winning](image)

Figure 4: The Effects of Uncertain Evaluations on Equilibrium Bidding Strategies

![Figure 4: The Effects of Uncertain Evaluations on Equilibrium Bidding Strategies](image)

The second effect is the slope effect that is common across different types of bidders. Since an increase in evaluation uncertainty lowers the marginal effect of lowering PQR bid on every bidder’s probability of winning, the incentive to invest on winning is weakened. Consequently, every bidder becomes less
competitive and therefore, increased randomization in winner selection generates a higher PQR bid (and associated choices of price and design quality bid) on average. Figure 4 illustrates the equilibrium effect of evaluation uncertainty on the distribution of price and design quality bids.

The numerical exercise shows that, on average, evaluation uncertainty is costly monetary wise, but improves design quality. In addition, a rise in evaluation uncertainty is associated with increased dispersion in both price and design quality. That is, evaluation uncertainty exacerbates the difference in price and design quality among bidders, which in turn leads to a greater amount of uncertainty in the auction outcomes from the point of view of the auctioneer.

These two findings have some important policy implications. First, evaluation uncertainty, on average, can be seen as a transfer to the contractor. As luck plays a larger role in determining the winner of a project, bidders lose incentive to provide a competitive offer, leaving a larger rent to bidders. Second, a rise in evaluation uncertainty comes with an additional cost of increased uncertainty in auction outcomes. If the auctioneer is budget constrained, an unexpectedly high winning price for a mega project may result in cancellation of the procurement itself. Thus, if the auctioneer is budget constrained, introducing a mechanism that reduces uncertainty in auction outcomes may become a valuable option.

### 1.4 Identification

Identification of the model is challenging due to unobserved design choices. That is, the econometrician only observes noisy signals of design quality through reviewers’ evaluations. Since pointwise-identification of the design quality is infeasible due to the incidental parameter problem, the probability of winning function $G_i$ and its density $g_i$ cannot be obtained directly from the data, precluding the standard inversion approach pioneered by Guerre, Perrigne, and Vuong (2000).\textsuperscript{28} Further, identification problem is complexified due to unobserved auction and reviewer heterogeneities, which may confound the degree of evaluation uncertainty.

As is demonstrated in Krasnokutskaya (2011), public procurement auctions of infrastructure projects involve a substantial amount of unobserved cost heterogeneity across auctions: auction characteristics that are observed to all participating bidders but unobserved to the econometrician. Ignoring the pres-

---

\textsuperscript{28}Estimation of design quality by the average reviewers’ evaluation faces incidental parameter problem since each auction involves only a few reviewers.
ence of unobserved auction heterogeneity exaggerates the dispersion of bids explained by bidders’ private information, which in turn underestimates the impact of evaluation uncertainty on bidding strategies. In other words, evaluation noises swap bidders’ rankings only if bidders’ are in a close competition.

Some reviewers are more lenient than others. Lenient reviewers assign a high score to every design bid, which exerts little effect in determining the rankings of bidders. We define disagreements among reviewers arising from such a reviewer specific characteristics as vertical reviewer heterogeneity. The econometrician would exaggerate the degree of evaluation uncertainty if vertical reviewer heterogeneity is not taken into account. Thus, evaluation uncertainty is measured by the amount of discrepancy in reviewers’ evaluations partialling out vertical reviewer heterogeneity, which is defined as horizontal reviewer heterogeneity.

In short, ignoring unobserved auction heterogeneity would underestimate the impact of evaluation uncertainty on bidders’ behaviour. On the other hand, ignoring vertical reviewer heterogeneity would overestimate the effect of evaluation uncertainty on bidders’ behaviour.

The rest of this section is organized as follows. First, the model from the previous section is specified in an econometric form, and the identifying assumptions are described. Second, we describe the implications of the identifying assumptions for the reduced form equations of the model, which is defined as Reduced Form Factor (RFF) model. Lastly, we elaborate the identification of the primitives of the model via RFF model.

1.4.1 Econometric Specification of Structural Model

Suppose that the econometrician observes a random sample of $A$ auctions, indexed by $a$, with the following information for each auction: price, \( \{p_{ia} : i = 1, 2, ..., N_a\} \), where $i$ is the bidder subindex, and $N_a$ is the number of bidders in auction $a$. Each individual reviewer’s evaluation of each bidder’s design proposal, \( \{q_{ria}^0 : r = 1, 2, ..., R_a\} \), where $r$ is the reviewer subindex, and $R_a$ is the number of reviewers in auction $a$. There is also information about exogenous auction characteristics, $Z_a$, which may include, for example, the engineer’s estimate of the project cost and project types (e.g., road, bridge, building construction, etc). These data are assumed to be generated from the equilibrium of the model from Section 1.3.

An auction is characterized by two types of unobserved heterogeneities: unobserved cost heterogeneity $\theta_a$ and unobserved measure heterogeneity $\eta_a$. On one hand, unobserved cost heterogeneity $\theta_a$ may capture project cost commonly observed to participating bidders but unobserved to the econometrician. On the
other hand, unobserved measure heterogeneity $\eta_a$ may capture auction heterogeneity in scoring difficulty. That is, some auctions may be more difficult for a bidder to score high than other auctions.

Reviewer heterogeneity is also decomposed into two components: vertical heterogeneity $\mu_{ra}$ and horizontal heterogeneity $\xi_{ria}$. Vertical heterogeneity captures the variation in evaluations specific to a certain reviewer, which may be interpreted as reviewers’ leniency in evaluating a design in general. On the other hand, horizontal heterogeneity is defined as an idiosyncratic noise that has nothing to do with the characteristics of reviewers. Vertical heterogeneity and measure heterogeneity both capture design rank preserving variation in evaluations. That is, swap of design rankings occur due to horizontal heterogeneity but not for the other two since vertical heterogeneity and measure heterogeneity does not vary across bidders. For this reason, we measure evaluation uncertainty as the dispersion in horizontal heterogeneity.

While the model in Section 1.3 abstracts from the information asymmetry between bidders and the econometrician, empirical results can be significantly influenced by an assumption on who observes what. I.A.1 below makes explicit who observes which component of a particular bidder’s cost.

**Identifying Assumption 1.** *Bidders’ types are given by:*

$$v_{cia} = \exp\{Z_a \beta_v + (1 - \rho) \theta_a + \varepsilon_{via}^v\} \quad (6)$$

$$e_{ia} = \exp\{Z_a \beta_f + \rho \theta_a + \varepsilon_{ia}^e\} \quad (7)$$

where $Z_a$ is a vector of auction characteristics which are observed to all bidders and the econometrician. $\theta_a$ is an unobserved cost heterogeneity, observed by all participating bidders but unobserved to the econometrician, that are independently and identically distributed across auctions. Lastly, $\{\varepsilon_{via}^v, \varepsilon_{ia}^e\}$ are private information observed only by bidder $i$, which is independently and identically distributed across bidders, but allowed to be correlated in an arbitrary manner within a bidder. The within-bidder correlation of $\varepsilon_{ia}^v$ and $\varepsilon_{ia}^e$ is natural as $\varepsilon_{ia}^e$ is a function of $\varepsilon_{ia}^v$ by construction. $\rho \in [0, 1]$ is a weight parameter assigned on unobserved auction heterogeneity across the two cost functions. To keep the model simple, I.A.2 parametrizes the variable cost function as follows.
Identifying Assumption 2. \( C(.) \) is a power function.

\[
C(q) = q^\gamma
\]

where \( \gamma > 1 \) is assumed for convexity.

Denote the logarithm of a variable by \( {}' \) (i.e., \( {}' \equiv \ln(p) \)). As the econometrician does not observe the objective design quality, we impose an assumption that links noisy reviewers evaluations and the objective design quality.

Identifying Assumption 3. A reviewer’s evaluation of a bidder’s design, \( q_{ria}^0 \), is noisy but an unbiased estimate of true quality \( q_{ia} \).

\[
q_{ria}^0 = \hat{q}_{ia} + \hat{w}_{ria}
\]

where the random variable \( \hat{w}_{ria} = \eta_a + \mu_{ra} + \xi_{ria} \) represents the sum of the measure, vertical, and horizontal heterogeneity. Note that the decomposition of evaluation noise into vertical and horizontal heterogeneity allows for affiliation in evaluations among bidders’ design proposals that arises from reviewer specific characteristics: a reviewer’s leniency in evaluating design proposals in general. A lenient (resp. stringent) reviewer may give out a high (resp. low) score to every design proposal due to his/her low (resp. high) quality standard.

Given the above assumptions, the primitives of the model to be identified are (i) the distribution of unobserved auction heterogeneities, \( F_\theta \) and \( F_\eta \), (ii) the distribution of vertical and horizontal heterogeneities, \( F_\mu \) and \( F_\xi \), and (iii) the joint distribution of variable cost and efficiency type, \( F_{s,v} \).\(^{29}\) The following subsection describes how the model together with I.A.1-I.A.3 can help identify the primitives of the model.

\(^{29}\)The curvature parameter \( \gamma \) turned out to be not identified when \( \varepsilon_{ia}^0 \) and \( \varepsilon_{ia}^v \) are correlated. This result is further discussed later in this section.
1.4.2 Separating the Sources of Variation in Equilibrium Strategies

For the moment, let us consider an equilibrium with perfectly observable design quality. That is, everyone agrees on the quality of any proposed design. I.A.1 together with I.A.2 imply that the equilibrium strategies of bidders are functions of observed heterogeneity \( Z_a \), unobserved heterogeneity \( \theta_a \), and also functions of private information \( \{ \varepsilon^v_{ia}, \varepsilon^e_{ia} \} \). Proposition 5 establishes the log additive separability of equilibrium strategies in cost components.

**Proposition 5.** Consider a monotone pure strategy equilibrium characterized by a set of first-order conditions (4) and (5), which are denoted by \( \{ p(Z, \theta, \varepsilon^v, \varepsilon^e), b(Z, \theta, \varepsilon^e) \} \). Then, the equilibrium reduced form PQR strategies of bidder \( i \) is additively log-separable in every cost component, and satisfies the structural equation (4), such that:

\[
\dot{b}(Z_a, \theta_a, \varepsilon^e_{ia}) = Z_a \tilde{\beta}_f + \tilde{\rho} \theta_a + \dot{b}(0, 0, \varepsilon^v_{ia}) \tag{8}
\]

\[
\dot{p}(Z_a, \theta_a, \varepsilon^v_{ia}, \varepsilon^e_{ia}) = Z_a \tilde{\beta}_v + \theta_a + \tilde{\gamma} \dot{b}(0, 0, \varepsilon^e_{ia}) + \varepsilon^v_{ia} \tag{9}
\]

\[\forall i \in \mathcal{N} \text{ where } \tilde{\gamma} \equiv \gamma/(\gamma - 1), \tilde{\rho} \equiv \frac{\gamma - 1}{\gamma} \rho, \tilde{\beta}_f \equiv \frac{\gamma - 1}{\gamma} \beta_f, \text{ and } \tilde{\beta}_v \equiv \beta_v + \beta_f .\]

Proof in Appendix. Note here that design quality is unobserved and so is PQR. That is, \( \dot{b}(Z_a, \theta_a, \varepsilon^e_{ia}) \) is unobserved to the econometrician. I.A.3 links an observed evaluation of a design and the actual design quality. I.A.3 links an observed evaluation of a design and the actual design quality.

With I.A.3, equations (8) and (9) can be rewritten in terms of the observed dependent variables:

\[
\dot{\varphi}_{ria} = Z_a \tilde{\beta}_f + \tilde{\rho} \theta_a + \dot{s}_{ia} + \dot{w}_{ria} \tag{10}
\]

\[
\dot{p}_{ia} = Z_a \tilde{\beta}_v + \theta_a + \tilde{\gamma} \dot{s}_{ia} + \varepsilon^v_{ia} \tag{11}
\]

where \( s_{ia} \equiv b(0, 0, \varepsilon^e_{ia}) \), which is a sole function of bidder \( i \)’s unobserved efficiency level, and \( \varphi_{ria} \equiv p_{ia}/q^0_{ria} \) (i.e., price per reviewer \( r \)’s evaluation). We define (10) and (11) as reduced form factor (RFF) model since the two equations are represented as functions of the primitives of the model, and are linked through the latent factors \( s_{ia} \) and \( \theta_a \). Note that additive separability of \( \dot{w}_{ria} \) also follows from the model, and \( \dot{w}_{ria} \) does not enter the price equation as bidders do not know how reviewers evaluate designs. To identify \( F_r \),
we simulate RFF model to obtain probability of winning and its density. Then, we back out $\varepsilon_{ia}$ for every realization of $s_{ia}$ using the first order condition (5). We elaborate the identification of the model in the following subsection.

1.4.3 Unidentified $\gamma$

This subsection demonstrates the source of partial identification of RFF model. To this end, consider a simplified version of RFF model in (10) and (11) with bidder heterogeneities only. That is, the model considered here involves neither reviewer heterogeneity (i.e., $\hat{w}_{ria} = 0$) nor unobserved auction heterogeneity (i.e., $\theta_a = 0$).

$$\hat{\varphi}_{ia} = \hat{s}_{ia} \quad (12)$$

$$\hat{p}_{ia} = \hat{\gamma} \hat{s}_{ia} + \varepsilon_{ia} \quad (13)$$

Within-auction covariance structure of the model is then given by:

$$E[\hat{p}_{ia}^2] = Var(\hat{s}_{ia}) + 2 \hat{\gamma} Cov(\hat{s}_{ia}, \varepsilon_{ia}^v) + Var(\varepsilon_{ia}^v) \quad E[\hat{p}_{ia} \hat{p}_{i'a}] = 0$$

$$E[\hat{p}_{ia} \hat{\varphi}_{ia}] = \hat{\gamma} Var(\hat{s}_{ia}) + Cov(\hat{s}_{ia}, \varepsilon_{ia}^v) \quad E[\hat{p}_{ia} \hat{\varphi}_{i'a}] = 0 \quad (14)$$

$$E[\hat{\varphi}_{ia}^2] = Var(\hat{s}_{ia})$$

$$E[\hat{\varphi}_{ia} \hat{\varphi}_{i'a}] = 0$$

In (14), there are three equations with four unknowns, and so the full identification of the model is clearly not possible. In particular, within-bidder substitution between price and design quality can be explained either through $\gamma$ or through $Cov(\hat{s}_{ia}, \varepsilon_{ia}^v)$. Cost curvature parameter $\gamma$ is the elasticity of substitution between price and design quality, and measures how costly a unit increase in design quality is. If a given increase in bidder’s design quality is comes with a large increase in price, and therefore a large increase in PQR, this variation could be explained by having a high elasticity of substitution (i.e., high $\gamma$). If inefficient bidders tend to consist of a bad designer and a bad builder, however, the co-movement in price and design quality can also be explained without having $\gamma$ large. Therefore, $\gamma$ cannot be separately identified from $Cov(\hat{s}_{ia}, \varepsilon_{ia}^v)$. In other words, having a high design cost is observationally equivalent to
having a positive assortative matching between builders and designers.

1.4.4 Nonparametric Identification of Reduced Form Factor Model

In order to identify the degree of evaluation uncertainty from a bidder’s point of view, we isolate the part of a reviewer’s evaluations that bidders knew at the time of bidding from the part they did not. RFF model exploits the fact that bidders do not observe reviewers’ evaluations of their designs at the time of bidding, and also reviewers do not observe bidders’ price bids at the time of evaluation. In particular, RFF model assumes that disagreement among reviewers on the design quality of a proposal is unknown to the bidder, but the part of design quality that is agreed among reviewers is known to the bidder at the time of bidding.

Note that the notion of design quality here is broad in the sense that the latent factor captures all the information that reviewers have about bidder $i$ at the time of evaluation. That is, if all reviewers agree that a particular design proposal is of high quality, then the design is deemed of high quality. Therefore, a bidder’s reputation or the impression that reviewers received from a particular bidder in the pre-proposal meeting can be regarded as a part of design quality as long as reviewers agree and a bidder knows what reviewers know about the bidder.

Proposition 6. \{\( F_\theta, F_\eta, F_\mu, F_\xi, F_{s,v}, F_e \)\} are all nonparametrically identified.

For simplicity, we omit the observables, \( Z_a \), in showing the nonparametric identification of RFF model below.\(^{30}\) The identification argument here closely follows Carneiro, Hansen, and Heckman (2003).

First, the measure of evaluation uncertainty (or equivalently, horizontal reviewer heterogeneity) is identified by exploiting the fact that multiple reviewers evaluate multiple designs in an auction. Intuitively, the identification comes from discrepancy in reviewers’ evaluations net of reviewer specific characteristics (e.g., leniency).

\[
(\hat{\phi}_{ria} - \hat{\phi}_{r'ia}) - (\hat{\phi}_{ri'a} - \hat{\phi}_{r'i'a}) = (\xi_{ria} - \xi_{r'ia}) - (\xi_{ri'a} - \xi_{r'i'a}) \text{ for } i' \neq i, \ r' \neq r \quad (15)
\]

As the LHS of (15) is observed, and the RHS is a sum of iid random variables, \( F_\xi \) is nonparametrically identified.\(^{30}\)

\(^{30}\) The proof trivially goes through with observables \( Z_a \).
identified through deconvolution. It follows that:

\[
(\hat{\varphi}_{ria} - \hat{\varphi}_{r'ia}) - (\xi_{ria} - \xi_{r'ia}) = \mu_r - \mu_{r'} \quad \text{for } r' \neq r
\]  

(16)

Since the distribution of the LHS of (16) is known, \( F_\mu \) can be again identified through deconvolution.

Second, bidder heterogeneities can also be identified taking into account the possible correlation between \( \varepsilon_{ia}^v \) and \( \varepsilon_{ia}^e \). As \( \gamma \) is known and \( \hat{\varphi}_{ria} - \hat{\varphi}_{r'ia} = \xi_{ria} - \xi_{r'ia} \), it follows that:

\[
(\hat{p}_{ia} - \hat{p}_{i'ia}) - \tilde{\gamma} (\hat{\varphi}_{ria} - \hat{\varphi}_{r'ia}) - \tilde{\gamma} (\xi_{ria} - \xi_{r'ia}) = \varepsilon_{ia}^v - \varepsilon_{i'ia}^v \quad \text{for } i' \neq i
\]

and so the marginal distribution of \( \varepsilon_{ia}^v \), \( F_v \), is identified from the within-auction residual variation in price bids that are not explained by the variation in PQR bids across bidders. Now, \( F_{s,v} \) can be identified again by deconvolution using the obtained \( F_v \) from:

\[
\hat{p}_{ia} - \hat{p}_{i'ia} = \tilde{\gamma} (\hat{s}_{ia} - \hat{s}_{i'ia}) - (\varepsilon_{ia}^v - \varepsilon_{i'ia}^v) \quad \text{for } i' \neq i
\]  

(17)

where the LHS of (17) is known.

Third, \( \tilde{\rho} \), \( F_\theta \) and \( F_\eta \) are identified from between-auction variation in price and PQR bids, respectively.

\[
\frac{E[\hat{p}_{ia} \hat{\varphi}_{r'ia}]}{E[\hat{p}_{ia} \hat{p}_{i'ia}]} = \tilde{\rho}
\]

\[
\hat{p}_{ia} - \tilde{\gamma} \hat{s}_{ia} - \varepsilon_{ia}^v = \theta_a
\]

\[
\hat{\varphi}_{ria} - \tilde{\rho} \theta_a - \hat{s}_{ia} - \mu_r - \xi_{ria} = \eta_a
\]

which completes the identification of RFF model.

Finally, the distribution of efficiency private information \( F_e \) can be identified using equation (5) by integrating over \( F_{s,v}, F_\mu \), and \( F_\xi \), as in Guerre, Perrigne and Vuong (2000), such that:

\[
\kappa_1(\gamma) \frac{\gamma_{-1}^{\frac{1}{\gamma}}}{\gamma_{-1}^{\frac{1}{\gamma}}} s_{ia}^{\gamma_{-1}^{\frac{1}{\gamma}}} + \kappa_2(\gamma) \frac{1}{\gamma_{ia}(s_{ia})} \frac{G_{ia}(s_{ia})}{\gamma_{ia}(s_{ia})} = \exp\{\varepsilon_{ia}^e\}
\]

(18)

where \( \kappa_1(.) \equiv \gamma^{\frac{1}{\gamma-1}} - \gamma^{\frac{1}{\gamma-1}} \) and \( \kappa_2(.) \equiv \frac{\gamma}{\gamma-1} \kappa_1(.) \). Thus, \( F_e \) can be identified from repeated auctions.
1.5 Estimation

The two-step estimation procedure proposed here closely follows the preceding identification section. While the variance covariance structure of RFF model is identified without any parametrization of the distribution functions, joint normality assumption on $\dot{s}_{ia}$ and $\varepsilon_{ia}^\epsilon$ is imposed for simplicity. We estimate the parameters of RFF model by Method of Moments with normalized curvature parameter $\gamma$.\(^{31}\) Then, we back out the distribution of efficiency private information $F_e$ via simulation.

**Step 1: Estimation of Reduced Form Factor Model.**

Let $\sigma_j$ denote the variance of unobservable $j \in \{\theta, \eta, s, \mu, \xi, v\}$ and $\delta \equiv \text{Cov}(\varepsilon_{ia}^\nu, \varepsilon_{ia}^\epsilon)$. We use $\Theta$ to denote the vector of variance covariance parameters plus $\tilde{\rho}$. First, we estimate the equilibrium bidding strategies by RFF model specified in (10) and (11) by OLS, partialing out the effects of observables. Then, we obtain OLS residuals (denoted by $\hat{p}_{ia}$, $\hat{\phi}_{ria}$) and compute the sample variance covariances specified below.

\[
M[\hat{p}_{ia}^2] = \hat{\sigma}_\theta + \gamma^2 \hat{\sigma}_s + 2 \gamma \hat{\delta} + \hat{\sigma}_\nu \\
M[\hat{p}_{ia} \hat{\phi}_{ria}] = \hat{\rho} \hat{\sigma}_\theta + \gamma \hat{\sigma}_s + \hat{\delta} \\
M[\hat{\phi}_{ria}^2] = \rho^2 \hat{\sigma}_\theta + \hat{\sigma}_\eta + \hat{\sigma}_s \\
M[\hat{\phi}_{ria} \hat{\phi}'_{ria}] = \rho^2 \hat{\sigma}_\theta + \hat{\sigma}_\eta + \hat{\sigma}_\mu
\]

where $M[.]$ denotes sample mean. Let $\Omega$ denote the vector of the above 8 sample averages, and $\Lambda(.)$ to denote the RHS function of $\Theta$. Then, $\Theta$ can be estimated by inverting $\Omega$, such that:

\[
\hat{\Theta} \equiv \Lambda^{-1}(\Omega)
\]

**Step 2: Estimation of $F_e$.**

In the second step, we draw a random variable from the distribution of normalized PQR strategy $\hat{F}_s$. Then, for a given realization of a random variable, we obtain an estimate of probability of winning and its density by numerically integrating over the distribution of evaluation noise $\hat{F}_w$ and the distribution of rivals’ PQR.

\(^{31}\)We set $\gamma = 2$ for the rest of the paper.
strategies $\hat{F}_s$. Finally, we evaluate the LHS of equation (5), and repeating this process gives estimate of $\hat{F}_e$. More specifically, (i) we draw pseudo random variables $s_{1k}$ from $\hat{F}_s$, (ii) numerically integrate over the distribution of rivals’ strategies $s_{-1} = [s_{2k}, s_{3k}, \ldots, s_{Nk}]$ and evaluation noise $w = [w_{1k1}, w_{1k2}, \ldots, w_{NkL}]$ by repeatedly drawing from $\hat{F}_{-s}$ and $\hat{F}_w$ where we set $L = 10^3$ to obtain the estimate of the probability of winning, such that:

$$\hat{G}(s_{1k}; N, R) = \int \int 1\{\hat{s}_{1k} - \hat{w}_{jkl} < \hat{s}_{jk} - \hat{w}_{jkl} \text{ for } j \neq 1 \in N\} d\hat{F}_{-s} d\hat{F}_w$$

Similarly, $\hat{g}(s_{1k}; N, R)$ can be obtained by numerically differentiating $\hat{G}(s_{1k}; N, R)$. (iii) Obtain the simulated efficiency private cost, $\hat{e}_{1k}$, by evaluating equation (18).

$$\kappa_1(\hat{\gamma}) \frac{\hat{\gamma}_{1a}}{\hat{\gamma}_{ia}} + \kappa_2(\hat{\gamma}) \frac{\hat{\gamma}_{1k}}{\hat{\gamma}_{ka}(s_{1k}; N, R)} = \exp\{\hat{e}_{1k}\}$$

Iterate (i) through (iii) $K$ times (where we set $K = 10^3$) to estimate the distribution of $\hat{e}_{1k}$ by repeatedly evaluating (20). Note here that the distribution of $\hat{e}_{1k}$, denoted by $\hat{F}_e(\cdot; N, R)$, depends on the number of bidders and reviewers. Thus, we compute $\hat{F}_e(\cdot; N, R)$ for all possible combinations of $N$ and $R$ in the data.

### 1.6 Results

Table 5 shows estimates of $\Theta$. The vector of observed auction characteristics includes engineer’s estimate of project cost, the number of bidders and reviewers in the auction to capture the effect of competition and evaluation uncertainty. Project types are classified into road, bridge, building, and others.

The first striking finding is the large estimate of horizontal reviewer heterogeneity $\sigma_\xi$. Horizontal reviewer heterogeneity is as large as 27% of the within-auction heterogeneity in PQR bids $\sigma_s$. The large estimate of $\sigma_\theta$ suggests that ignoring unobserved auction heterogeneity would exaggerate $\sigma_s$ as pointed out in Krasnokutskaya (2011) and Bajari, Houghton, and Tadelis (2006). A consistent estimation of $\sigma_s$ is particularly relevant in the context of the analysis here since horizontal reviewer heterogeneity is likely

---

32While the model does not predict that bidders’ strategies are additively log-linear in bidder level characteristics, we control them to approximate observed bidder heterogeneity when estimate. Project types are classified into road, bridge, building, and others.
to swap the ranking of bidders by chance only in a close competition. Therefore, overestimating the dispersion in private information of bidders results in underestimation of the effect of horizontal reviewer heterogeneity on bidders’ behaviour.

As is well known in the procurement auction literature, engineer’s cost estimate captures much of between-auction heterogeneity in project size. What seems puzzling at first glance here is the insignificant coefficient estimate on the number of bidders. If competition is all it captures, the insignificant estimate is intriguing. However, the number of bidders may be correlated with some unobserved auction heterogeneity. For example, if the FDOT observes project complexity, and if the FDOT tends to let more applicants participate into an auction for more complex project, then the effect of competition on pricing and designing decisions can be completely offset by project complexity. That is, bidders are on average inefficient at implementing a complex project, and therefore the effect of competition on bidding strategies is hidden by project complexity. For the same reason, correlation between the number of reviewers and bidders’ behaviour can be hidden by project complexity.

Figure 6 shows the distribution of bidders’ private cost information for varying number of bidders and reviewers. Intense competition and horizontal reviewer heterogeneity is associated with a right shift of the distribution of private cost information. That is, the more bidders or more reviewers there are, the more inefficient each bidder is on average. This finding is in line with the estimation result obtained from RFF model. After computing the equilibrium using the estimated distribution of efficiency level, we find no significant difference in mean bids across the number of bidders. That is, the competition effects on pricing and designing strategies are offset by the asymmetry in the distribution of private information. Therefore, the structural model here is consistent with the fact that the number of bidders and the number of reviewers are insignificantly correlated with both pricing and designing strategies in the RFF model.

---

33 DB auctions have fewer bidders than the standard first-price low-bid auction due to pre-screening of potential bidders. Therefore, small variation in the number of bidders can be attributing to the imprecise estimates.
Table 5: Reduced Form Factor Model: Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>log(PQR)</th>
<th>log(Price)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FDOT Engineer’s Cost Estimate)</td>
<td>0.982</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>(0.0167)</td>
<td>(0.0155)</td>
</tr>
<tr>
<td># of Bidders / Auction</td>
<td>-0.0073</td>
<td>0.0193</td>
</tr>
<tr>
<td></td>
<td>(0.0563)</td>
<td>(0.0508)</td>
</tr>
<tr>
<td># of Reviewers / Auction</td>
<td>0.0624</td>
<td>0.0598</td>
</tr>
<tr>
<td></td>
<td>(0.0331)</td>
<td>(0.0354)</td>
</tr>
<tr>
<td>$\delta$, $\rho$</td>
<td>-0.0143</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>$\sigma_{\eta}$, $\sigma_{\theta}$</td>
<td>0.0456</td>
<td>0.0494</td>
</tr>
<tr>
<td></td>
<td>(0.00957)</td>
<td>(0.0104)</td>
</tr>
<tr>
<td>$\sigma_{\mu}$, $\sigma_{s}$</td>
<td>0.00343</td>
<td>0.0266</td>
</tr>
<tr>
<td></td>
<td>(0.00055)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>$\sigma_{\xi}$, $\sigma_{v}$</td>
<td>0.00731</td>
<td>0.0103</td>
</tr>
<tr>
<td></td>
<td>(0.00289)</td>
<td>(0.0174)</td>
</tr>
</tbody>
</table>

Auction and Bidder Characteristics | Yes | Yes |
Obs | 1296 | 338 |

Bootstrapped standard errors in parentheses. Bidder characteristics include the distance between the project work site and the builder’s nearest branch, and builder’s utilization rate.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Distribution of Efficiency Private Information

![Distribution of Efficiency Private Information](image)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
1.7 Does Evaluation Uncertainty Matter?

As shown in Section 1.3, efficient (resp. inefficient) bidders respond to greater evaluation uncertainty by becoming more (resp. less) competitive, exacerbating the bids dispersion. The above direct and behavioral effects together create a significant amount of uncertainty to the auctioneer in both the monetary cost of procurement and the quality of a contractor’s design.

This section quantifies the impact of evaluation uncertainty through a change in the number of reviewers. More specifically, we first examine to what extent evaluation uncertainty can be mitigated by adding more reviewers. As demonstrated in the following, the marginal effect of additional reviewers discipates quickly since evaluation uncertainty is a convex function of the number of reviewers. Further, allocating a large number of reviewers to a review task may come with a large opportunity cost to the FDOT, which is not captured in the model.

To shut down the effect of evaluation uncertainty on bidders’ behaviour without additional reviewers, a second-price auction with design score contingent transfer is proposed as an alternative mechanism. Under the alternative mechanism, uncertainty concerning evaluation of rivals’ design proposals is completely shut down. This alternative mechanism is also dominant strategy implementable and therefore, robust to the problem of counterfactual experiment under multiple equilibria.

1.7.1 Simulation of DB Auction with Varying Number of Reviewers

Consider a symmetric average DB auction with three bidders and varying number of reviewers $R$. The degree of evaluation uncertainty is $\hat{\tau}(R)$, which is a decreasing and convex function of $R$. The simulation results are shown in Table 7.

<table>
<thead>
<tr>
<th></th>
<th>Mean Price</th>
<th>Mean Quality</th>
<th>Standard Deviation Price</th>
<th>Standard Deviation Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = 1$</td>
<td>20.6</td>
<td>81.4</td>
<td>5.58</td>
<td>14.3</td>
</tr>
<tr>
<td>$R = 4$</td>
<td>16.2</td>
<td>72.0</td>
<td>4.10</td>
<td>11.8</td>
</tr>
<tr>
<td>$R = 9$</td>
<td>15.3</td>
<td>69.8</td>
<td>3.77</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Winning prices are expressed in $1,000,000 while winning quality scores are out of 100 points.
We find that adding five reviewers to an ordinary DB auction with four reviewers would reduce the winning price by 5.8%, and the winning design quality by 3.1%. The effects of additional reviewers on the dispersion of auction outcomes are somewhat larger than those on its expected value. The standard deviations of the winning price and design quality decline by 7.8% and 4.9%, respectively.

While increased number of reviewers mitigates evaluation uncertainty, the marginal effect of an additional reviewer declines quickly due to the convexity of evaluation uncertainty in the number of reviewers. In addition, assigning many employees of the FDOT to the review task can be prohibitively costly. To mitigate evaluation uncertainty without incurring additional administrative cost of appointing reviewers, we propose an alternative auction mechanism that significantly reduces the auctioneer’s uncertainty in the auction outcomes without deteriorating the expected auction outcomes significantly.

1.7.2 Auction Design for Design Auction: Second-Price Auction with Design Score Contingent Transfer

In this counterfactual experiment, the auctioneer uses design scores to determine the transfer amount rather than determine the contractor of a project. More specifically, we consider a second-price auction with design score contingent transfer in which the winner is selected solely based on the lowest submitted price, and the winner’s design score determines the amount of transfer that a contractor receives. This auction mechanism is of interest as it shuts down the effect of evaluation uncertainty on bidders’ behavior. Neither a bidder’s winning likelihood nor its ex-post payoff is affected by uncertain design evaluations of its rivals, and thus bidders would not respond to the uncertainty in rivals’ design evaluations.\textsuperscript{34}

In this auction, bidders simultaneously submit both price and design proposals as in an ordinary DB auction. The winner is selected by the lowest price bid. Design quality score is determined by the average reviewers’ evaluation, and it determines the amount of transfer that a bidder receives upon winning the project.\textsuperscript{35} We denote the transfer function by $trans(q_iw_i)$ where $q_i$ and $w_i$ are objective design quality and evaluation noise as defined before. The winner receives the second lowest price, denoted by $p^{(min)}_{-i}$, plus $trans(q_iw_i)$ upon winning. Now, let $E_{w_i}[\cdot]$ and $E_{p^{(min)}}[\cdot]$ denote the expectation operator over the

\textsuperscript{34} A bidder would also be non-responsive to uncertainty in its own design evaluations under risk neutrality.

\textsuperscript{35} Non-winning bidders receive zero transfer.
distribution of \( w_i \) and \( p_{i}^{\text{(min)}} \), respectively. Also, denote the probability of winning function conditional on bidder \( i \)'s own price bid by \( Pr(win|p_i) \). Then, the interim expected payoff of bidder \( i \) is defined as:

\[
\pi_i^{\text{int}} = \max_{p_i, q_i^0} E_{p_i^{\text{(min)}}} \left[ E_w [ Pr(win|p_i) (p_{i}^{\text{(min)}} + \text{trans}(q_i w_i) - v_i C(q_i) - f_i) ] \right]
\]

With a log-linear transfer function, the bidder’s profit function can be rewritten as \( \text{trans}(q_i w_i) = \phi \ln(q_i) + \phi \ln(w_i) \) where \( \phi \) is a constant. The log-linear transfer function shuts down the impact of uncertainty associated with its own design evaluation on its own strategy under the risk neutrality assumption.

\[
\pi_i^{\text{int}} = \max_{p_i, q_i^0} E_{p_i^{\text{(min)}}} \left[ E_w [ Pr(win|p_i) (p_{i}^{\text{(min)}} + \phi \ln(q_i) + \phi \ln(w_i) - v_i C(q_i) - f_i) ] \right]
\]

Taking the first-order optimality condition with respect to \( q_i \) gives the equilibrium design quality choice \( q_i^{(eqm)} \), such that:

\[
\frac{\phi}{q_i^{(eqm)}} = v_i C(q_i^{(eqm)}) \tag{21}
\]

which uniquely identifies the equilibrium design quality, and is independent of evaluation uncertainty.\(^{36}\)

Define the expected total cost by \( ETC_i \equiv -\phi \ln(q_i^{(eqm)}) + v_i C(q_i^{(eqm)}) + f_i \). Note that \( ETC_i \) is independent of own price bid as optimal quality choice is independent of own price. Therefore, bidder \( i \)'s problem collapses to:

\[
\pi_i^{\text{int}} = \max_{p_i} E_{p_i^{\text{(min)}}} [ Pr(win|p_i) (p_{i}^{\text{(min)}} - ETC_i) ]
\]

which is equivalent to the ordinal second-price low-bid auction. Thus, the (weakly) dominant truth-telling strategy of bidder \( i \) is:

\[
p_i^{(eqm)} = ETC_i \tag{22}
\]

The dominant strategy equilibrium of the game is characterized in closed form by (21) and (22).

\(^{36}\)The uniqueness of \( q_i^{(eqm)} \) is apparent since LHS is decreasing in \( q_i^{(eqm)} \) while RHS is increasing in \( q_i^{(eqm)} \) in (21).
The linear transfer parameter, $\phi$, is set so that the average fixed cost bidder minimizes the unit cost of design. To be more specific, consider the following problem. Let $q_i^{(AC)}$ denote the choice of design quality that minimizes the average cost of implementing the project for any arbitrary pair of variable cost and fixed cost. It is straightforward to obtain:

$$q_i^{(AC)} = \arg \min_{q_i^0} \frac{v_i q_i^0 + f_i}{q_i}$$

$$= \left( \frac{f_i}{(\gamma - 1)v_i} \right)^{\frac{1}{\gamma}}$$

Therefore, $\phi = \frac{\gamma}{\gamma - 1} f_c$ where $f_c$ is the estimated average fixed cost from the data, induces the level of design quality that achieves the minimum average cost for the average fixed cost bidder of any variable cost type.

We simulate and obtain the distribution of winning price and design quality in this alternative auction. Table 8 compares auction outcomes from DB and the proposed auction mechanism.

<table>
<thead>
<tr>
<th>Table 8: Second Price Auction with Design Quality Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Standard Deviation</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>DB Auction</td>
</tr>
<tr>
<td>SPA with Score Contingent Transfer</td>
</tr>
</tbody>
</table>

Winning prices are expressed in $1,000,000 while winning quality scores are out of 100 points.

SPA with design quality contingent transfer reduces the auctioneer’s uncertainty in the auction outcomes without affecting the expected auction outcomes much. Standard deviation of winning price and design quality decline by 22% and 33%, respectively. While 0.6% (resp. 1%) increase (resp. decrease) in the average winning price (resp. design quality) may not be negligible, the objective of the FDOT here may not be minimizing the expected auction outcomes. If the auctioneer wishes to avoid a large uncertainty in the auction outcomes, the alternative auction mechanism here may serve as a way to mitigate such uncertainty.

\[37\]If the FDOT’s objective is to minimize the expected objective PQR, it should let more bidders participate in each auction to encourage competition.
The reduction of uncertainty in the auction outcomes partly comes from the transfer coefficient $\phi$ that implements the efficient scale in the production of design quality by the average fixed cost type of a bidder. It is clear from (21) that bidder $i$ with $f_{c_i} < \bar{f}_c$ (resp. $f_{c_i} > \bar{f}_c$) overproduces (resp. underproduce) design quality relative to its efficient scale, generating a smaller dispersion in design bids. As its price bid is a function of its design quality, a smaller dispersion in design quality leads to a smaller dispersion in price bids.

The small changes in the expected auction outcomes are intriguing at first glance. As discussed in Section 1.3, every bidder overproduces design quality in a DB auction while bidders are, on average, producing at their efficient scale under the alternative mechanism here. However, it is low fixed cost bidders, who is likely to win the project, that overproduces design quality under the alternative mechanism, and production cost is reflected in the form of a higher price bid.

The main message here is not to suggest that the alternative mechanism is superior to the DB auction mechanism, and the results obtained in the analysis does not necessarily hold when there is a large amount of evaluation uncertainty in a DB auction. However, if the auctioneer wishes to mitigate its uncertainty in the auction outcomes for the level of evaluation uncertainty present in the data without worsening the expected auction outcomes drastically, then it may consider implementing the proposed auction mechanism.

1.8 Conclusion

This paper studied the effects of uncertain design evaluations on competing suppliers’ behavior using a sample of Design-Build auctions from the Florida Department of Transportation. We document the presence of evaluation uncertainty, which affects the rankings of bidders through introducing luck into the winner selection process. A structural model that incorporates uncertainty in both design evaluations and rivals’ bids are developed and estimated taking into account potentially confounding heterogeneities, such as unobserved auction heterogeneity and vertical reviewer heterogeneity. The structural approach is consistent with the observed fact that both the number of bidders and reviewers have no effect on bidders’ behavior on the surface. The paper also provides the first attempt in the literature of multi-
attribute auctions at estimating structural parameters when some attributes of a bid are unobserved to the econometrician. The economic significance of evaluation uncertainty is demonstrated through simulation exercises. An increase in horizontal reviewer heterogeneity not only generates a higher winning price and higher winning design quality on average, but also exacerbates dispersion in auction outcomes, resulting in greater uncertainty in auction outcomes from the auctioneer’s standpoint. A simple second-price auction with design score contingent transfer, which shuts down the impact of uncertain design evaluations on bidding incentives, would reduce the auctioneer’s uncertainty in both the winning price and design quality by 22% and 33%, respectively.
Identifying Strategic Behaviour in Expert Evaluations: Evidence from Design-Build Auctions

2.1 Introduction

Evaluating the quality of a proposal by independent experts is a standard practice in many competitive environments: research grants, publication in peer reviewed journals, prizes, sports (e.g., gymnastics, figure skating, diving, etc.), and procurements. While experts might be well-informed and possess an ability to properly evaluate the quality of a proposal, institutional structure may distort incentives of expert reviewers. Figure skating judges of Olympic games who may be concerned about their own careers, for example, may give a biased score for the athletes of their own countries.\textsuperscript{39}

We study Design-Build (DB) procurement auctions used by many state departments of transportation in the U.S. and around the world.\textsuperscript{40} In a DB auction, potential contractors compete over price and design to win a contract to implement an infrastructure project, ranging from road maintenance and bridge repair to building construction. We use hand-collected data on DB auctions from the Florida Department of Transportation (FDOT) to investigate strategic interactions among expert reviewers of the FDOT. Upon receiving design proposals, each reviewer of the FDOT independently evaluates and assigns a score to every design category of each design proposal. The quality score of a design proposal is then determined by summing scores across categorical scores, and averaging the aggregated scores across reviewers. The potential contractor with the lowest price per quality score ratio wins the project, and receives its price bid upon the completion of the project.\textsuperscript{41}

We first present reduced form evidence that reviewers’ evaluations are dependent despite the procurement rule. We then build a simple structural model of strategic reviewers’ evaluations, which captures both conformity to their peers, and also the gains from their favourite designer winning the contract. The existence of unobserved quality heterogeneity is the challenge in identifying the effect of a peer’s evaluation of a design on the evaluation of a reviewer. To avoid confounding unobserved heterogeneity with the peer

\textsuperscript{39}Zitzewits (2006) show strong evidence of biases in scoring of Olympic games.

\textsuperscript{40}As of October 2010, there are 39 state departments of transportation that use DB, including California, Delaware, Georgia, Minnesota, etc. DB auctions are also common in other developed countries, including Canada, Japan, and Sweden.

\textsuperscript{41}Price per quality is a winner selection rule used by many state departments of transportation, including Alaska, Michigan, North Carolina, and South Dakota.
effect, we exploit the variation in peers’ evaluations due to their exogenous characteristics. Our argument here is that peers’ exogenous characteristics have nothing to do with the signal of design quality that a reviewer receives, but has an effect on peer’s evaluation of a design proposal. We found evidence of both conformity to peers as well as reviewers’ incentives to bias evaluation for their favourite design.

This paper is related to the emerging literature in forensic economics (or equivalently, economics of wrongdoing) where the focus of the literature is mainly to identify agents’ hidden behaviour. The paper is also related to the vast peer effect literature where agents have incentives to conform to their peers. We contribute to the above two literature by establishing a simple model that incorporates both favouritism and conformity. Further, we provide an approach that does not require a priori knowledge about who biases for who.

The rest of the paper is organized as follows. Section 2.2 describes the institutional details and the data. Section 2.3 presents the model. Section 2.4 establishes identification and describes the estimation procedure. Section 2.5 shows and discusses the estimation results. Section 2.6 concludes.

2.2 Institutional details and data

2.2.1 Design-Build auctions

Here we describe the DB procurement process with a focus on the evaluation process. The DB procedure can be decomposed into two consecutive stages, a pre-selection stage and a bidding stage. In the pre-selection stage of a DB procurement, the Florida Department of Transportation (FDOT) posts an advertisement on-line which lists information about the project location, description of work, criteria for evaluating a letter of interest, and technical qualification requirements. Then, reviewers are selected by a department secretary from a pool of FDOT employees and based on qualifications and availability. These reviewers are mostly civil engineers who themselves are involved in the design of projects for standard procurement auctions of the FDOT.43

Meanwhile, an interested DB firm writes a letter of interest to the FDOT.44 The appointed reviewers

---


43 In standard procurement auctions the design of the project is pre-specified by the auctioneer. Every firm simply submits a price bid for a given design.

44 A DB firm usually consists of a builder sub-contracting with a designer.
then evaluate the letter of interest based on the criteria described in the advertisement, which include past performance grades, DB experience, and current capacity of DB firms. Those DB firms that are judged as pre-qualified by reviewers are short-listed and become a potential contractor. The identities of these potential contractors are posted on-line and become common knowledge. The potential contractors then receive the request for proposal, which describes detailed specification of the project and design evaluation criteria. Design evaluation criteria vary across auctions. Some repeatedly observed evaluation criteria include warranty, innovative aspect of design, maintenance of traffic, construction methods, commitment to environmental protection, project schedule, etc. Following the pre-selection stage, the potential contractors now enter the bidding stage. All the potential contractors and the reviewers meet in a mandatory pre-proposal meeting in which the reviewers provide proposal instructions and the scope of the project. Both design and price bids are usually due 1 to 2 months after the pre-proposal meeting, and the design and price bids need be sent to the FDOT in separate envelopes.

Then, each reviewer “independently” evaluates each design proposal, and assigns a score to every evaluation category (or criteria equivalently) of each design proposal. Reviewers’ evaluations are summed across categories, and the quality score of a design proposal is determined by averaging the summed scores across reviewers. Finally, the envelopes with price bids are opened, and the potential contractor with the lowest price per quality ratio wins the contract to provide the project, and receives his price bid upon completion of the project.

Figure 5: Timeline of Events in a DB Auction
2.2.2 Heterogeneity in reviewers’ evaluations

Let us take a closer look at the evaluation of design proposals submitted by those potential contractors. Figure 6 presents the actual record of reviewers’ design evaluations for a bridge construction project, which illustrates how design evaluation is implemented. The first and second rows of the table show the identity of 3 potential contractors and 5 reviewers (or evaluators), respectively. The first and second columns show the names of ten evaluation categories and their respective weights. Each reviewer independently evaluates each quality aspect of a design proposal, and assigns a score out of the category specific maximum score. Then, these scores are summed across all categories to obtain the total score of a design proposal. The summed scores are averaged across reviewers to determine the quality score of a potential contractor’s design proposal.

The example in figure 6 shows a significant amount of heterogeneity in evaluation scores across reviewers for a given design proposal. For the design proposal of Cone & Graham/Jacob, there is a differential of 24% between the evaluation scores of reviewers JD and DK. Also, reviewer JD ranks the proposal by Cone & Graham/Jacob in fifth place and the design of Johnson Bros./GAI in third place, while reviewer
DK ranks Cone & Graham/Jacob first and Johnson Bros./GAI fourth. Note that this difference across reviewers in the way they rank the proposals appears not only in the aggregate score but also at the level of specific, and apparently objective, criteria. For instance, for evaluation criterion entitled “Schedule”, we see that reviewer DH gives a score of 6/9 to Cone & Graham/Jacob and a score of 8/9 to Johnson Bros./GAI, while the scores of reviewer DK to these two proposals are 8/9 and 5/9, respectively.

This kind of heterogeneity in reviewers’ scores cannot be explained by a model where: (a) each evaluation criterion represents an objective attribute of a design and it is perfectly known to all the reviewers, though reviewers may differ in the marginal utility associated to the level of the attribute, i.e., vertical differentiation; and (b) a reviewer’s scores perfectly reveal his preferences for the different designs. In this paper, we propose and estimate a model of reviewer behavior that relaxes these two assumptions. More specifically, we relax assumption (a) by taking into account that the true attributes of a design are not perfectly known to reviewers such as each reviewer receives a different signal about the attributes of a design. We relax assumption (b) by incorporating strategic interactions or peer effects in reviewers’ preferences. For instance, inexperienced reviewers may have an incentive to mimic what experienced reviewers do despite the fact that the evaluation rule specifically states every reviewer needs to evaluate independently. Alternatively, some reviewers could behave strategically to try favor their favorite proposal by discounting the scores of the other competing proposals.

2.2.3 Data

We investigate a sample of DB auctions that took place between years 2000 and 2011 in Florida. Although DB is also common practice in other states such as Alaska, Pennsylvania, or Minnesota, among others, scoring rules and point systems differ across these state departments of transportation. Therefore, a single department of transportation, the FDOT, is chosen for consistency and availability of evaluation records. In particular, evaluation records in other state departments of transportation are often aggregated and not preserved at the individual reviewer level. Indeed, our sample shows each reviewer’s evaluation record at evaluation category level.

Our dataset combines information from three different sources: (1) auction data from the FDOT; (2) firm level data; and (3) data on reviewers’ salaries, gender, race, and experience in design evaluation.

The sample of DB auctions used in the analysis here is a result of selecting a subset of original DB
auctions, and the selection procedure is the following. We obtained from the FDOT the records of all the 152 DB auctions that have been procured between years 2000 and 2011. We compiled a dataset from the provided records, excluding from our working sample all the auction records with: (a) only 1 bidder/design; or (b) missing engineer’s estimate of project cost; or (c) missing reviewers’ evaluations. The original dataset also contained a variant of scoring auctions (i.e., DB’ auctions) in which the scoring rule not only involves the price and the average quality from reviewers’ evaluations but also a time incentive component. As the focus of the paper is on reviewers’ evaluation decisions, and not bidding incentives, we include both DB and DB’ auctions in our working sample. For the rest of the analyses, we include DB’ auctions whenever we refer to DB auctions. This sample is complemented with firm and reviewer characteristics, which we obtained through web-scraping.

The main source of reviewers’ characteristics is collected from the government salary archives provided by The Tampa Tribune. The government salary archives contain information on reviewers’ salaries, gender, race, and hired date of FDOT’s workers from 2009 to 2012. As the data do not contain reviewers’ information prior to 2009, we complement the data with the salary information provided by Florida Open Government. Florida Open Government provides information on Florida state employees’ salaries from 1995 to 2012. The data on reviewer characteristics, however, do not entirely cover our whole sample of reviewers from DB evaluation records since we do not have information on gender, race, and hire date on those who quit their job before 2009. As a result, our dataset on reviewers’ characteristics covers 81.8% of reviewers from DB evaluations records.

2.2.4 Descriptive statistics on reviewers’ evaluations

We first provide evidence on substantial reviewer heterogeneity at aggregate score level. To this end, consider variance decomposition of aggregate score $AS_{rda}$ which is assigned by reviewer $r$ on design $d$ in auction $a$.

---

45 The variant of DB auctions is a combination of DB and A+B auction studied in Bajari and Lewis (2011). In a DB’ auction, a bidder submits a price bid, design proposal, and also a time bid which describes how fast a project can be delivered. The scoring rule in a DB’ auction is $price_i + g(time\_bid_i)$ where the quality score of potential contractor $i$ is determined exactly the same manner as in a DB auction procedure and the function $g(.)$ is an increasing linear function that is known by any participant in the auction. The function $g(.)$ may differ across auctions, and therefore potential contractors may strategically substitute between price and time bids depending on the weight assigned on the time bid.

46 The sources of reviewer characteristics is found at: http://tbo.com/fact-finder/government-salaries-archive/.
Consider the following hierarchical decomposition of aggregate score.

\[ AS_{rda} = \mu^{(0)} + \mu^{(1)}_a + \mu^{(2)}_{da} + \mu^{(3)}_{rda} \]

where \( \mu^{(0)} \) is a constant parameter. \( \mu^{(1)}_a, \mu^{(2)}_{da}, \) and \( \mu^{(3)}_{rda} \) are iid random variables with zero mean and constant variance. Then, we have:

\[ Var(AS_{rda}) = Var(\mu^{(1)}_a) + Var(\mu^{(2)}_{da}) + Var(\mu^{(3)}_{rda}) \]

<table>
<thead>
<tr>
<th></th>
<th>Between-Auction</th>
<th>Within-Auction</th>
<th>Within-Design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate Score ((AS_{rda}))</td>
<td>2.66</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>((.448))</td>
<td>((.309))</td>
<td>((.137))</td>
</tr>
<tr>
<td>Obs</td>
<td>1486</td>
<td>1486</td>
<td>1486</td>
</tr>
</tbody>
</table>

Table 9: Variance Decomposition of Reviewers’ Evaluations

Table 9 presents descriptive statistics on reviewers’ evaluations. First observation is that within-design between-reviewer variation turns out to be the largest among the three variance. Approximately 56% of the variance in aggregate evaluation scores is explained by between-reviewer within-design evaluations.

2.2.5 Descriptive statistics on reviewers’ characteristics

A reviewer’s characteristics that may be particularly relevant in a peer reviewer’s evaluation is the level of experience. For example, an inexperienced reviewer may be inclined to mimic what the most experienced reviewer does if he/she does not want to look bad by deviating too much from the experienced reviewer’s score.\(^{47}\) In order to construct the experience level of a reviewer at the time of his/her appointment, we first look at how many times a particular reviewer participated in DB auction in the sample. Figure 7

\(^{47}\)Later in this section, we provide evidence that this is actually the case in a reduced form regression despite the DB procurement rule states reviewers should independently review designs.
shows a large amount of heterogeneity in participation frequency across reviewers. More than 50% of the reviewers participated only once in the review task while there is only a small subset of reviewers who participated more than 5 times during the sample period.

Figure 7: Distribution of Reviewers’ Participation Frequency

![Figure 7: Distribution of Reviewers’ Participation Frequency](image)

Figure 8: Reviewer Callback and Time Elapsed

![Figure 8: Reviewer Callback and Time Elapsed](image)
Table 10: Summary Statistics of Evaluation Scores and Reviewer Characteristics

### Summary Statistics of Reviewers’ Evaluations

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical Evaluation (score/category maximum)</td>
<td>.844</td>
<td>.133</td>
<td>0</td>
<td>1</td>
<td>10861</td>
</tr>
<tr>
<td>Aggregate Evaluation (sum of categorical evaluations)</td>
<td>84.6</td>
<td>6.15</td>
<td>55.7</td>
<td>98.42</td>
<td>456</td>
</tr>
<tr>
<td># Evaluation Categories (# of evaluation categories/auction)</td>
<td>8.26</td>
<td>2.07</td>
<td>3</td>
<td>13</td>
<td>98</td>
</tr>
<tr>
<td># Reviewers (# of reviewers/auction)</td>
<td>3.87</td>
<td>.875</td>
<td>3</td>
<td>8</td>
<td>119</td>
</tr>
<tr>
<td># Design Proposals (# of design proposals/auction)</td>
<td>3.07</td>
<td>.585</td>
<td>2</td>
<td>5</td>
<td>119</td>
</tr>
</tbody>
</table>

### Summary Statistics of Reviewers’ Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Participated Auctions (at the time of review)</td>
<td>2.43</td>
<td>2.41</td>
<td>1</td>
<td>16</td>
<td>456</td>
</tr>
<tr>
<td>Wage (annual base salary, $1,000)</td>
<td>83.1</td>
<td>21.7</td>
<td>26.2</td>
<td>130.7</td>
<td>422</td>
</tr>
<tr>
<td>Tenure (years elapsed since FDOT’s hire date)</td>
<td>15.0</td>
<td>9.82</td>
<td>.265</td>
<td>48.2</td>
<td>395</td>
</tr>
<tr>
<td>Male (=0 if female, =1 if male)</td>
<td>.834</td>
<td>.371</td>
<td>0</td>
<td>1</td>
<td>448</td>
</tr>
<tr>
<td>White (=0 if non-white, =1 if white)</td>
<td>.717</td>
<td>.450</td>
<td>0</td>
<td>1</td>
<td>475</td>
</tr>
</tbody>
</table>

The summary statistics are based on 119 DB auction procured between years 2000 and 2011.

Figure 8 shows the distribution of the time elapsed from the last review task for all the reviewers in the sample. There is a clear negative relationship between the callback rate and the time from the last review task. For example, less than 1% of the reviewers are called back after 1460 days = 4 years. Therefore, we use the first 4 years of the sample (year 2000-2003) to construct reviewers’ experience level using the number of participated auctions in the past at the time of their appointment.

Table 10 presents descriptive statistics on auction/reviewer characteristics. The average evaluation tends to be quite high (i.e., 84% of the maximum possible score). There are about 8 categories, 4 reviewers, and 3 designs involved in an auction on average. The majority of the reviewers in DB auctions is well paid, male, and white. Also, these reviewers have been working at the FDOT for a number of years.

### 2.2.6 Descriptive evidence on strategic evaluations using score distance

In this subsection, we show that inexperienced reviewers’ scores agglomerate around experienced reviewers’ scores. For the purpose of this exercise, we define a *beginner* as a reviewer who has not participated in any DB auction in the past. Also, define an *experienced* reviewer as a reviewer who participated in DB auctions
most frequently in the past. Intuitively, a beginner may have an incentive to mimic the experienced reviewer’s evaluation of a design since deviating from the experienced reviewer’s evaluation would make him/her look bad, potentially affecting his/her future career.

If the data is generated through such a mechanism, then we should see in the data that the score distance between a beginner’s score and a experienced reviewer’s score shrink as the experience level of the experienced reviewer increases. However, this observation by itself is not enough to conclude that the reviewers evaluations are interdependent. The score distance between the experienced reviewer’s score and a beginner’s score can also shrink if the experienced reviewer becomes more precise in estimating the true unobserved quality of a design as he/she becomes more experienced. To circumvent this confounding issue, we also consider the score distance between two beginners’ scores, and show that this score distance also shrinks as the experience level of the experienced reviewer increases. Shorter score distance between two beginners’ scores is unlikely explained by increasing precision of the experienced.

We first define the score distance between (i) a beginner and a beginner, and (ii) a beginner and the experienced.

\[
BDIS_{cda}^{(r,r')} = |S_{rcda} - S_{r'cda}| \quad \text{where } r' \text{ and } r \text{ are both beginners}
\]

\[
EDIS_{cda}^{(r,r')} = |S_{rcda} - S_{r'cda}| \quad \text{where } r' \text{ is the experienced but } r \text{ is a beginner}
\]

where \( S_{rcda} \) is the categorical score assigned to design \( d \) by reviewer \( r \) on category \( c \) in auction \( a \). Let \( EXPER_a \) be the experience level of the leader in an auction \( a \). Now, consider the following empirical model.

\[
BDIS_{cda}^{(r,r')} = \alpha^{(B)} EXPER_a + W_a \beta^{(B)} + u_{cda}^{(r,r')}
\]

\[
EDIS_{cda}^{(r,r')} = \alpha^{(E)} EXPER_a + W_a \beta^{(E)} + v_{cda}^{(r,r')}
\]

\(^{48}\)In case of a tie, the experienced reviewer is determined using the highest salary among those most frequently participated reviewers.
where $\alpha(B), \alpha(E), \beta(B),$ and $\beta(E)$ are parameters, $W_\alpha$ is a vector of experienced reviewer’s characteristics and project characteristics in auction $a$, and $u_{cda}^{(r,r')}$ and $v_{cda}^{(r,r')}$ are errors terms. The idea here is to show that coefficients $\alpha(B)$ and $\alpha(E)$ are both significantly negative, such that the score distance between (i) a beginner and another beginner, and (ii) the experienced reviewer and a beginner both shrink as the experienced reviewer becomes more experienced.

Table 11 shows that beginners’ evaluation scores do agglomerate around the experienced reviewer’s score as the experienced becomes more experienced. The effect of an increase in the experienced reviewer’s experience on the score distance between two beginners is similar to its effect on the score distance between beginners and the experienced. In particular, an additional unit of review experience reduces the distance between two beginners by 0.5-0.8% of the maximum possible score. Alternatively, one standard deviation increase in review experience reduces the score distance between beginners by approximately 0.2 standard deviation. Another strong predictor on score agglomeration is the weight assigned on evaluation category. We find that one standard deviation increase in the categorical weight comes with 0.1 standard deviation decrease in the score distance for both the experienced and beginners.

One thing to note here though is that an increase in categorical weight is not necessarily evidence of strategic interactions among reviewers. That is, if the FDOT assigns a higher weight on those evaluation categories that are relatively straightforward to evaluate, then we observe in the data that the score distance shrink with an increase in the weight assigned on category. However, an analogous story does not hold for the use of experienced reviewers by the FDOT. More specifically, we would observe a greater score distance if the FDOT assigns more experienced reviewers to a more complex project where evaluation of a design would be more difficult, leading to a greater dispersion in scores.

The descriptive analysis above indicates strategic interactions among reviewers, which suggests reviewers communicate before or during the evaluation process despite the procurement rule. We suppose that the observed behavior of reviewers comes from the concern that a large deviation from the experienced peer make them look bad. In the following section, we capture this incentive in our structural model as a peer effect.

### 2.2.7 Descriptive evidence on strategic evaluations using design rankings

We investigate the source of differences in design rankings across reviewers. Let $rank_{cda}$ be the rank
Table 11: Do beginners evaluation scores agglomerate around the experienced reviewer’s score?

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Score Distance between Experienced and Beginner</th>
<th>Score Distance between Two Beginners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>experienced reviewer’s experience</td>
<td>$-0.0082^{***}$</td>
<td>$-0.0056^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>experienced reviewer’s wage</td>
<td>$-0.0429^{***}$</td>
<td>$-0.0041$</td>
</tr>
<tr>
<td></td>
<td>(0.0126)</td>
<td>(0.0144)</td>
</tr>
<tr>
<td>experienced reviewer’s tenure</td>
<td>0.0003</td>
<td>$-0.0001$</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>male experienced reviewer</td>
<td>0.0668***</td>
<td>0.0596***</td>
</tr>
<tr>
<td></td>
<td>(0.0053)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>white experienced reviewer</td>
<td>$-0.0166^{**}$</td>
<td>$-0.0189^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.0056)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>weight assigned on evaluation category</td>
<td>$-0.2017^{***}$</td>
<td>$-0.1994^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.0287)</td>
<td>(0.0266)</td>
</tr>
<tr>
<td>log(price)</td>
<td>0.0096</td>
<td>0.0267</td>
</tr>
<tr>
<td></td>
<td>(0.0063)</td>
<td>(0.0074)</td>
</tr>
<tr>
<td>log(engineer’s cost estimate)</td>
<td>$-0.0162^{***}$</td>
<td>$-0.0345^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.0059)</td>
<td>(0.0074)</td>
</tr>
<tr>
<td># of reviewers/auction</td>
<td>$-0.0017$</td>
<td>$-0.0019$</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td># of designs/auction</td>
<td>0.0098***</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>experienced builder</td>
<td>$-0.0043$</td>
<td>$-0.0050$</td>
</tr>
<tr>
<td></td>
<td>(0.0041)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>experienced designer</td>
<td>$-0.0042$</td>
<td>$-0.0051$</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>utilization rate of builder</td>
<td>$-0.0044$</td>
<td>$-0.0047$</td>
</tr>
<tr>
<td></td>
<td>(0.0050)</td>
<td>(0.0052)</td>
</tr>
<tr>
<td>log distance between builder’s branch and project work site</td>
<td>0.0016</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0014)</td>
</tr>
</tbody>
</table>

Year and Project Type Fixed Effects | No | No | Yes | No | No | Yes |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R-sq</td>
<td>0.0354</td>
<td>0.0436</td>
<td>0.0714</td>
<td>0.0451</td>
<td>0.0533</td>
<td>0.0807</td>
</tr>
<tr>
<td>Obs</td>
<td>4192</td>
<td>4192</td>
<td>4192</td>
<td>3703</td>
<td>3703</td>
<td>3703</td>
</tr>
</tbody>
</table>

Standard errors are bootstrapped 1000 times, and shown in parentheses. Experienced builder (designer) are dummy variable equal to one if a builder (designer) shows up in the sample more than 5 times. Utilization rate is defined as builder’s backlog per capacity where backlog is calculated as total dollar value amount of projects at the time of bidding. Capacity is the maximum of backlog for the builder within the sample.
Table 12: Disagreement in Design Rankings among Reviewers

<table>
<thead>
<tr>
<th></th>
<th>Category Level Ranking</th>
<th>Aggregate Design Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>3590 (35.4%)</td>
<td>678 (55.6%)</td>
</tr>
<tr>
<td>Disagree</td>
<td>6533 (64.6%)</td>
<td>541 (44.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>10123</td>
<td>1219</td>
</tr>
</tbody>
</table>

# of times a pair of reviewers agreed/disagreed are tabulated in the above table.

of design \(d\) assigned by reviewer \(r\) on category \(c\) in auction \(a\) based on evaluation score. We construct a variable \(1\{rank_{rcda} \neq rank_{r'cda}\}\), which indicates discrepancy in rankings between reviewer \(r \neq r' \in R\). Similarly, define aggregate score as \(AS_{rda} \equiv \sum_{c \in C} S_{rcda}\), and the ranking based on \(AS_{rda}\) by \(RANK_{rda}\). Then, the discrepancy indicator at the aggregate level is given by \(1\{RANK_{rda} \neq RANK_{r'da}\}\). We present unconditional discrepancy in reviewers’ evaluation scores at both disaggregate and aggregate level in Table 12.

Reviewers agree on rankings only 35.4% (resp. 55.6%) of the time at disaggregate level (resp. aggregate level).\(^{49}\) Note here that reviewers tend to disagree more in rankings of designs at disaggregate category level. From the previous analysis, we find that the score distance is decreasing in categorical weight. Therefore, this discrepancy in the ranking between the disaggregated and aggregated levels can be explained by the fact that reviewers tend to agree with each other for the evaluation categories with larger weights.

Now, we run probit regression of agreement dummy variables on reviewer, design, and project characteristics to see what are the determinants of the observed discrepancies in design rankings. Table 13 shows the probit estimation result where the regressors are also defined in terms of distance or absolute difference in reviewers’ characteristics. We also include project as well as designers’ characteristics in this regression. The aggregate level ranking heterogeneity reveals an interesting pattern in the behaviour of reviewers: reviewers with vertically differentiated characteristics (e.g., wage, tenure, etc) tend to agree with each other while reviewers with horizontally differentiated characteristics (i.e., race, gender, etc)

\(^{49}\)There are 2,327 agreements and 7,796 disagreements in rankings of designs at category level.
tend to disagree with each other. If an inexperienced reviewer is concerned about his career, then he may have an incentive to mimic what his senior reviewer does. If reviewers’ racial and gender diversity reflects diversity in how one views and evaluates a certain design aspect, then horizontal diversity would be associated with disagreements in design rankings. What is interesting here though is that the difference in review experience has a positive effect on design ranking at categorical level while it has a negative impact on rankings at aggregate level.

2.3 A structural model of strategic evaluations

2.3.1 Basic model

Consider a set of design proposals \( D \), indexed by \( d \in D \), and a set of reviewers \( R \), indexed by \( r \in R \). Each reviewer \( r \) evaluates each design proposal according to several evaluation criteria, indexed by \( c \in C \). In this section, we omit the auction subscript \( a \) to avoid cluttering.

Define an *experienced* reviewer, denoted by \( e \), as the reviewer with the largest number of participation in the past DB auctions within the given set of reviewers. Also, define an *intermediate*, denoted by \( i \), as a reviewer who has some review experience in DB auctions. Lastly, let a *beginner*, denoted by \( b \), be a reviewer who has never been appointed in the past. For every proposal and evaluation criteria \((d,c)\), each reviewer \( r \) of type \( j \) receives a signal of quality \( \theta^{(j)}_{rcd} > 0 \). The vector of signals \( \theta \equiv \{ \theta^{(j)}_{rcd} : r \in R, c \in C, d \in D, j \in \{e,i,b\} \} \) consists of both publicly observable and unobservable components. As there is always a unique experienced reviewer in a given set of reviewers by definition, we omit the reviewer subscript \( r \) for the experienced (i.e., \( \theta^{(e)}_{rcd} = \theta^{(e)}_{cd} \)).

As a working assumption, the experienced reviewer’s evaluation is completely foreseen by the other reviewers. This assumption can be justified on the ground that the most experienced reviewer’s behaviour can be predicted by his/her past scoring decisions by the other reviewers while those inexperienced reviewers’ behaviour are difficult to predict. Given that many reviewers are beginners in DB auctions, it is not surprising if the most experienced reviewer demonstrate how to evaluate designs to those beginners, and beginners could be influenced by how the experienced reviewer evaluates a design.

For every design \( d \in D \) of every category \( c \in C \), the experienced, intermediates, and beginners simultaneously assign a score, \( S^{(e)}_{cd}, S^{(i)}_{rcd}, \) and \( S^{(b)}_{rod} \), respectively. We denote the payoff (i.e., loss function) of the experienced, intermediate, and beginner by \( U^{(e)}, U^{(i)}_{r}, \) and \( U^{(b)}_{r} \), respectively. The summary of
Table 13: When Reviewers Disagree with Each Other?

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$1{rank_{rda} \neq rank_{r'da}}$</th>
<th>$1{RANK_{rda} \neq RANK_{r'da}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>experience difference</td>
<td>0.0130**</td>
<td>-0.0293**</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0132)</td>
</tr>
<tr>
<td>wage difference</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>tenure difference</td>
<td>-0.0007</td>
<td>-0.0124**</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td>gender difference</td>
<td>0.0198</td>
<td>0.2417**</td>
</tr>
<tr>
<td></td>
<td>(0.0311)</td>
<td>(0.0992)</td>
</tr>
<tr>
<td>racial difference</td>
<td>-0.0154</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>(0.0301)</td>
<td>(0.0814)</td>
</tr>
<tr>
<td>log(engineer’s cost estimate)</td>
<td>-0.0500***</td>
<td>-0.0435</td>
</tr>
<tr>
<td></td>
<td>(0.0133)</td>
<td>(0.0379)</td>
</tr>
<tr>
<td># of designs/auction</td>
<td>0.2729***</td>
<td>0.3778***</td>
</tr>
<tr>
<td></td>
<td>(0.0251)</td>
<td>(0.0762)</td>
</tr>
<tr>
<td># of reviewers/auction</td>
<td>0.0889***</td>
<td>0.0590</td>
</tr>
<tr>
<td></td>
<td>(0.0179)</td>
<td>(0.0581)</td>
</tr>
<tr>
<td>experienced builder</td>
<td>0.0223</td>
<td>0.0382</td>
</tr>
<tr>
<td></td>
<td>(0.0286)</td>
<td>(0.0938)</td>
</tr>
<tr>
<td>experienced designer</td>
<td>0.0556</td>
<td>0.0619</td>
</tr>
<tr>
<td></td>
<td>(0.0291)</td>
<td>(0.0890)</td>
</tr>
<tr>
<td>utilization rate of builder</td>
<td>-0.1184**</td>
<td>-0.2874**</td>
</tr>
<tr>
<td></td>
<td>(0.0399)</td>
<td>(0.1079)</td>
</tr>
<tr>
<td>distance between builder’s branch</td>
<td>0.0005</td>
<td>0.0014</td>
</tr>
<tr>
<td>and project work site</td>
<td>(0.0003)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>Obs</td>
<td>10123</td>
<td>1219</td>
</tr>
</tbody>
</table>

Bootstrapped standard errors in parentheses. experienced builder (designer) are dummy variable equal to one if a builder (designer) shows up in the sample more than 5 times. Utilization rate is defined as builder’s backlog per capacity where backlog is calculated as total dollar value amount of projects at the time of bidding. Capacity is the maximum of backlog for the builder within the sample.
assumptions are as follows.

**Assumption 1**: A set of reviewers consists of an experienced, intermediates, and beginners.

**Assumption 2**: The experienced reviewer’s signal is public knowledge.

**Assumption 3**: Intermediates’ and beginners’ signals both consist of public and private information.

With assumptions 1-3, define the payoff function of each type of reviewers by:

\[
U^{(j)}_r = -E\left[\sum_{d \in D} \sum_{c \in C} (1 - \alpha_j) \left( S^{(j)}_{rcd} - \theta^{(j)}_{rcd}\right)^2 + \alpha_j \left( S^{(j)}_{rcd} - S^{(e)}_{cd}\right)^2 \right]_{I_r}
\]  

where \(I_r\) is the information set of reviewer \(r\) at the time of evaluation, and \(\alpha_j \in [0,1]\) for \(j \in \{e,i,b\}\). That is, reviewer \(r\)'s payoff is a weighted average of the loss in deviating from his own design quality signal, and the loss in deviating from the evaluation of the experienced reviewer (i.e., taste conformity).

Taking as given the experienced reviewer’s scores, reviewer \(r\) chooses his own evaluation of each design to maximize his payoff (i.e., Nash assumption). The best response function of reviewer \(r\), denoted by \(\sigma^{(j)}_{rcd}\) for \(j \in \{e,i,b\}\), is given by:

\[
\sigma^{(e)}_{cd} = \theta^{(e)}_{cd}
\]

\[
\sigma^{(j)}_{rcd} = (1 - \alpha_j) \theta^{(j)}_{rcd} + \alpha_j \sigma^{(e)}_{rcd}
\]

Therefore, a pure strategy Bayesian Nash equilibrium is a vector of scores \(\{S^{(j)}_{rcd}^* : r \in \mathcal{R}, c \in \mathcal{C}, d \in \mathcal{D}, j \in \{e,i,b\}\}\), such that \(S^{(j)}_{rcd}^*\) satisfies the best response condition described in (24) and (25) for every \((r,c,d)\). Now, let \(\omega_{rcd}^{(j)}\) and \(\varepsilon^{(j)}_{rcd}\) denote the public and private part of reviewer \(r\)'s signal, respectively.

Now, assumption 2 and 3 imply that \(\theta^{(e)}_{cd} = \omega^{(e)}_{cd}\) and \(\theta^{(j)}_{rcd} = \omega^{(j)}_{rcd} + \varepsilon^{(j)}_{rcd}\) for \(j \in \{i,b\}\). Then, the unique reduced form equation is given by:

\[
S^{(e)}_{cd}^* = \omega^{(e)}_{cd}
\]

\[
S^{(j)}_{rcd}^* = (1 - \alpha_j)(\omega^{(j)}_{rcd} + \varepsilon^{(j)}_{rcd}) + \alpha_j \omega^{(e)}_{cd}
\]
2.3.2 Model with reviewer’s preferences on the competition outcome

So far, we have assumed that reviewers only have a taste for reporting the signal it received, and a taste for conformity to the score of the most experienced. It might be the case that reviewers have also preferences on the outcome of the competition. In particular, a reviewer may prefer that his favorite design wins the competition. There are several possible explanations for this type of preferences, such as aversion to support a loser, prestige of supporting the winner, or just corruption. We now consider a version of the model that introduces this type of preferences.

Let \( d^*(e) \) and \( d^*(j)_r \) for \( j \in \{i, b\} \) denote reviewer \( r \)'s favorite design, taking into account the conformity to the experienced reviewer’s evaluation. That is, \( d^*(e) \equiv \{d \in D : \sum_{c \in C} \omega_{cd}(e) > \sum_{c \in C} \omega_{cd}(e') \forall d' \neq d\} \), and \( d^*(j)_r \equiv \{d \in D : \sum_{c \in C} [(1 - \alpha_j)\theta_{rd}^{(j)} + \alpha_j \omega_{cd}^{(e)}] > \sum_{c \in C} [(1 - \alpha_j)\theta_{rd'}^{(j)} + \alpha_j \omega_{cd}^{(e)}] \forall d' \neq d\} \). Then, the payoff function of reviewers are defined as:

\[
V^{(e)} = U^{(e)} - \sum_{d \in D} \sum_{c \in C} \mathbb{1}\{d \neq d^{(e)*}\} \delta_c s_{cd}^{(e)} \tag{28}
\]

\[
V^{(j)}_r = U^{(j)}_r - \sum_{d \in D} \sum_{c \in C} \mathbb{1}\{d \neq d^{(j)*}_r\} \delta_j s_{rd}^{(j)} s_{cd} \tag{29}
\]

where \( U^{(e)} \) and \( U^{(j)}_r \) for \( j \in \{i, b\} \) are as defined in the previous basic model section. While both \( \delta_e > 0 \) and \( \delta_j > 0 \) capture disutility from assigning a high score on non-favorite design, these parameters have distinct interpretations. On one hand, \( \delta_e \) captures the experienced reviewer’s incentive to discount his non-favorite design. On the other hand, \( \delta_j \) for \( j \in \{i, b\} \) captures intermediates’ and beginners’ incentives to strategically reduce the score of their non-favorite design upon an increase in the experienced reviewer’s evaluation of their non-favorite design. Note that the above strategic substitution does not affect the rankings of designs for each reviewer and therefore, any reviewer’s favorite design can be identified directly from the data. Characterization of the equilibrium in this model is again trivial. The best responses of
each type of reviewers is given by:

$$\sigma_{cd}^{(e)} = \theta_{cd}^{(e)}$$  \hspace{1cm} if \( d = d_r^{(e)} \)

(30)

$$\sigma_{cd}^{(e)} = \theta_{cd}^{(e)}/(1 + 2\delta_e)$$ \hspace{1cm} if \( d \neq d_r^{(e)} \)

(31)

$$\sigma_{cd}^{(j)} = (1 - \alpha_j) \theta_{rcd}^{(j)} + \alpha_j \sigma_{cd}^{(e)}$$ \hspace{1cm} if \( d = d_r^{(j)} \)

(32)

$$\sigma_{cd}^{(j)} = (1 - \alpha_j) \theta_{rcd}^{(j)} + (\alpha_j - \delta_j) \sigma_{cd}^{(e)}$$ \hspace{1cm} if \( d \neq d_r^{(j)} \)

(33)

and the reduced form equations are:

$$S_{cd}^{(e)*} = \omega_{cd}^{(e)}$$ \hspace{1cm} if \( d = d_r^{(e)} \)

(34)

$$S_{cd}^{(e)*} = \omega_{cd}^{(e)}/(1 + 2\delta_e)$$ \hspace{1cm} if \( d \neq d_r^{(e)} \)

(35)

$$S_{rcd}^{(j)*} = (1 - \alpha_j) \theta_{rcd}^{(j)} + (\alpha_j - \delta_j 1\{d \neq d_r^{(j)}\}) \omega_{cd}^{(e)}$$ \hspace{1cm} if \( d = d_r^{(e)} \)

(36)

$$S_{rcd}^{(j)*} = (1 - \alpha_j) \theta_{rcd}^{(j)} + (\alpha_j - \delta_j 1\{d \neq d_r^{(j)}\}) \omega_{cd}^{(e)}/(1 + 2\delta_e)$$ \hspace{1cm} if \( d \neq d_r^{(e)} \)

(37)

for \( j \in \{i, b\} \). The reduced form strategies of each type of reviewers show intuitive properties. An increase in disutility parameter \( \delta_e \) discounts the scores of non-favorite design of the experienced reviewer, which exacerbates the difference in the experienced reviewer’s favorite and non-favorite design. When \( \delta_j > \alpha_j \) (i.e., incentive to get a reviewer’s favorite design is greater than the incentive to conform), an intermediate or beginner assigns a negative weight on the experienced reviewer’s signal. The magnitude of incentives to counteract the design score of the experienced by an intermediate or beginner is naturally captured in the coefficient of \( \theta_{cd}^{(e)} \) depending on whether the favorite design of the experienced coincides with that of another type of reviewers.

2.3.3 Specification of reviewers’ preferences and information structure

A design \( d \) can be described in terms of a vector of “objective” attributes that are observable to all the reviewers \( \{X_{cd} : c \in C, d \in D\} \), where \( X_{cd} \) is a \( 1 \times K \) vector of variables observable to the researcher, and unobservables to the econometrician \( \{\omega_{rcd}^{(j)}, \varepsilon_{rcd}^{(j)} : c \in C, d \in D\} \), where \( \omega_{rcd}^{(j)} \) for \( j \in \{e, i, b\} \) is the part of unobserved heterogeneity known to all the reviewers while \( \varepsilon_{rcd}^{(j)} \) for \( j \in \{e, i, b\} \) is the private information of reviewer \( r \). As the common knowledge and private information components cannot be separately identified
as is, we assume that the common knowledge component of reviewers’ signals are modeled as follows.

\[
\omega_{cd}^{(e)} = X_{cd} \Gamma Z^{(e)} + \xi_c^{(1)} + \xi_d^{(2)} + \tilde{\omega}_{cd}^{(e)} \tag{38}
\]

\[
\omega_{r_{cd}}^{(j)} = X_{cd} \Gamma Z_{r}^{(j)} + \xi_c^{(1)} + \xi_d^{(2)} \tag{39}
\]

where \(Z^{(e)}\) and \(Z_r^{(j)}\) are both \(q \times 1\) vector of reviewers’ exogenous characteristics of the experienced reviewer and the rest of reviewers, respectively. \(\Gamma\) is a \(K \times q\) matrix of parameters that capture how reviewers’ characteristics affect the marginal utilities of different design attributes. \(\xi_c^{(1)}\) and \(\xi_d^{(2)}\) are unobserved category and design heterogeneities, respectively. \(\xi_c^{(1)}\) may capture differences in evaluation difficulty across design criterion, and \(\xi_d^{(2)}\) may capture the overall design quality difference across designs. \(\tilde{\omega}_{cd}^{(e)}\) is an idiosyncratic shock that is iid over \((r, c, d)\) that is common knowledge to every single reviewer, which may capture comparative advantage of design \(d\) on category \(c\) over other categories. Combining the above equations with private information, we have:

\[
\theta_{cd}^{(e)} = \omega_{cd}^{(e)} = X_{cd} \Gamma Z^{(e)} + \xi_c^{(1)} + \xi_d^{(2)} + \tilde{\omega}_{cd}^{(e)} \tag{40}
\]

\[
\theta_{r_{cd}}^{(j)} = \omega_{r_{cd}}^{(j)} + \varepsilon_{r_{cd}}^{(j)} = X_{cd} \Gamma Z_r^{(j)} + \xi_c^{(1)} + \xi_d^{(2)} + \varepsilon_{r_{cd}}^{(j)} \tag{41}
\]

In summary, the vector of structural parameters of the model is \((\Gamma, \alpha)\). We want to use our model and data to estimate this vector of parameters. Note here that the only component that is unobserved from the point of view of reviewer \(r’\)’s peer is \(\varepsilon_{r_{cd}}^{(j)}\) for \(j \in \{e, b\}\), and \(\theta_{cd}^{(e)}\) is completely observed by every reviewer.

We provide sufficient conditions for identification for \(\alpha_j\) and \(\delta_j\), which are the focus of the paper.
2.4 Identification and Estimation

2.4.1 Identification

The first order equation of an intermediate or beginner can be written as:

\[ S_{rcd}^{(j)} = (1 - \alpha_j) \theta_{rcd}^{(j)} + (\alpha_j - \delta_j 1\{d \neq d_r^*\}) S_{cd}^{(e)} \]

\[ = (1 - \alpha_j) \left[ X_{cd} \Gamma Z_r^{(j)} + \xi_c^{(1)} + \xi_d^{(2)} + \varepsilon_{rcd}^{(j)} \right] + (\alpha_j - \delta_j 1\{d \neq d_r^*\}) S_{cd}^{(e)} \quad \text{for } j \in \{i, b\} \]

The key identification problem comes from the correlation between the unobservable \( \xi_c^{(1)} + \xi_d^{(2)} + \varepsilon_{rcd}^{(j)} \) and the score assigned by the experienced, \( S_{cd}^{(e)} \). As shown in previous section, equilibrium scoring strategies of intermediates and beginners depend on both \( \theta_{rcd}^{(j)} \) and \( \omega_{cd}^{(e)} \). Therefore, controlling for \( \xi_c^{(1)} \) and \( \xi_d^{(2)} \) with fixed effect is not enough to solve endogeneity problem. Indeed, if designs have comparative advantage in some design criteria that are not captured by the fixed effects, then this could generate a positive correlation between \( \omega_{cd}^{(e)} \) and \( \varepsilon_{rcd}^{(j)} \) (i.e., the peer effect is confounded by unobserved quality heterogeneity). Now, we provide an identification assumption.

**ASSUMPTION ID:** (a) Unobservables \( (\omega_{cd}^{(e)}, \varepsilon_{rcd}^{(j)}) \) are independent of the exogenous observables \( (X_{cd}, Z_r^{(j)}) \) for \( j \in \{e, i, b\} \). And there exists (at least) one element in the vector \( Z^{(e)} \) (observed characteristics of the experienced reviewer), say \( Z_k^{(e)} \), such that: (b) the k-th column of \( \Gamma \), associated to \( Z_k^{(e)} \), has non-zero elements; and (c) variable \( Z_k^{(e)} \) has strictly positive variance conditional on \( Z_{kr} \), i.e., no perfect correlation between \( Z_k^{(e)} \) and \( Z_{kr} \).

Under Assumption ID, we can construct moment conditions that identify the structural parameters. It is obvious that the signal of the experienced reviewer is identified trivially, and so we show identification of intermediates and beginners below. First, it is convenient to write the equilibrium conditions as follows:

\[ S_{rcd}^{(j)} = W_{rcd} \gamma + (\alpha_j - \delta_j 1\{d \neq d_r^*\}) S_{cd}^{(e)} + (1 - \alpha_j)(\xi_c^{(1)} + \xi_d^{(2)} + \varepsilon_{rcd}^{(j)}) \]

where \( W_{rcd} \equiv (vec(X_{cd} \otimes Z_r)) \), and \( \gamma \equiv (vec(\Gamma)) \). Under Assumption ID, we have that \( \mathbb{E}(\varepsilon_{rcd}^{(j)} | X_{cd}, Z, \xi_c^{(1)}, \xi_d^{(2)}) = \)
0 for \( j \in \{i, b\} \). Therefore, we have the following moment restrictions:

\[
\mathbb{E} \left( \begin{bmatrix} W_{rcd} \\ Z^{(e)} \end{bmatrix} \left[ S^{(j)}_{rcd} - W_{rcd}\gamma - (\alpha_j - \delta_j 1\{d \neq d^*_r\}) S^{(e)}_{cd} - (1 - \alpha_j)(\xi^{(1)}_c + \xi^{(2)}_d + \varepsilon_{rca}) \right] \right) = 0 \quad \text{for} \ j \in \{i, b\}
\]

### 2.4.2 Estimation

We introduce auction subscript \( a \) for clarifying the different sources of sample variation. We first show evidence that OLS specification of the model is rejected by the data.

Define the vector of sample moment conditions:

\[
m^{OLS}(\gamma, \alpha) = \sum_{a, r, c, d, j \in \{i, b\}} \left[ W_{rcda} \left[ S^{(j)}_{rcda} - W_{rcda}\gamma - (\alpha_j - \delta_j 1\{d \neq d^*_r\}) S^{(e)}_{cd} - (1 - \alpha_j)(\xi^{(1)}_{ca} + \xi^{(2)}_{da} + \varepsilon^{(j)}_{rca}) \right] \right] = 0
\]

Then, the OLS with category and design fixed effects consistently estimates \( \alpha_j \) for \( j \in \{i, b\} \). That is, we can test if the estimated residuals from the OLS specification of the model are (positively) correlated or not. Let \( \hat{\varepsilon}^{(j)}_{rca} \) and \( \hat{\omega}^{(e)}_{cda} \) denote the estimated residuals from the OLS with category and design fixed effects. Then, we consider:

\[
\hat{\varepsilon}^{(j)}_{rca} = \lambda_j \hat{\omega}^{(e)}_{cda} + u^{(j)}_{rca}
\]

with the null hypothesis, \( H_0 \), and alternative hypothesis, \( H_A \), that:

\[
H_0 : \lambda_j = 0 \\
H_A : \lambda_j \neq 0
\]

for \( j \in \{i, b\} \), and rejection of the null hypothesis implies that OLS is inconsistent.

As we will see in the estimation results, \( \lambda_j \neq 0 \) is strongly supported by the data. Thus, we proceed to take into account of endogeneity using control function approach, which exploits the exclusion restriction that directly controls for the presence of \( \omega^{(e)}_{cda} \) in the reduced form equation of intermediates as well as
beginners. Suppose that \( \varepsilon^{(j)}_{rcda} \) and \( \tilde{\omega}^{(e)}_{cda} \) are jointly normally distributed with correlation coefficient \( \rho \), such that:

\[
\varepsilon^{(j)}_{rcda} = \rho \tilde{\omega}^{(e)}_{cda} + v^{(j)}_{rcda}
\]

where \( v^{(j)}_{rcda} \) is iid error term. First, we estimate \( \tilde{\omega}^{(e)}_{cda} \) by OLS with fixed effects for \( \xi^{(1)}_{ca} \) and \( \xi^{(2)}_{da} \), and obtain predicted residual, \( \hat{\omega}^{(e)}_{cda} \). Second, we control for \( \hat{\omega}^{(e)}_{cda} \) in the structural equations of intermediates and beginners, such that we estimate the following by OLS.

\[
S^{(j)}_{rcda} = W^{(j)}_{rcda} \gamma + (\alpha_j - \delta_j 1 \{d \neq d^*_r\}) \hat{s}^{(e)}_{cda} + (1 - \alpha_j)(\xi^{(1)}_{ca} + \xi^{(2)}_{da} + \rho \hat{\omega}^{(e)}_{cda} + v^{(j)}_{rcda})
\]

### 2.5 Estimation Results

Recall that the system of best response equations are given by:

\[
S^{(j)}_{rcda} = (1 - \alpha_j)(\omega^{(j)}_{rcda} + \varepsilon^{(j)}_{rcda}) + (\alpha_j - \delta_j 1 \{d \neq d^*_r\}) \hat{s}^{(e)}_{cda} \quad \text{for } j \in \{i, b\}
\]

and

\[
\theta^{(e)}_{cda} = X^{(e)}_{cda} \Gamma Z^{(e)}_{cda} + \xi^{(1)}_{ca} + \xi^{(2)}_{da} + \tilde{\omega}^{(e)}_{cda}
\]

\[
\theta^{(j)}_{rcda} = X^{(j)}_{rcda} \Gamma Z^{(j)}_{ra} + \xi^{(1)}_{ca} + \xi^{(2)}_{da} + \varepsilon^{(j)}_{rcd}
\]

Auction characteristics includes engineer’s estimate of the project cost and project types (e.g., roads, bridge, building construction, etc). Design characteristics include builder and designers’ experience, utilization rate, distance between the nearest builder’s branch and work site. Reviewer characteristics include reviewers’ experience, indicator for being white, indicator for being male, wage (=annual base salary), and tenure.

In order to demonstrate the degree of endogeneity that comes from unobserved design heterogeneity, we present OLS estimation results with different level of fixed effects. That is, in place of \( \xi^{(1)}_{ca} + \xi^{(2)}_{da} \), we consider two different fixed effect specifications: only design fixed effect, and design fixed effect + category
Table 14: OLS Peer Effects Estimation Results with & without Favorite Design

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$S_{rcda}^{(i)}$ (intermediate)</th>
<th>$S_{rcda}^{(b)}$ (beginner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\alpha_j$</td>
<td>0.0237 (0.0062)</td>
<td>0.0186 (0.0061)</td>
</tr>
<tr>
<td>$\alpha_j - \delta_j$</td>
<td>0.0093 (0.0058)</td>
<td>0.0083 (0.0058)</td>
</tr>
<tr>
<td>$\lambda_j$</td>
<td>0.2147 (0.0339)</td>
<td>0.2139 (0.0344)</td>
</tr>
</tbody>
</table>

| Design FE          | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               |
| Category FE        | No                | Yes               | No                | Yes               | No                | Yes               | No                | Yes               |
| Favorite Design    | No                | No                | Yes               | Yes               | No                | Yes               | No                | Yes               |
| R-sq               | 0.2122 (0.0344)   | 0.2167 (0.0340)   | 0.2141 (0.0345)   | 0.2186 (0.0560)   | 0.2122 (0.0561)   | 0.2167 (0.0557)   | 0.2141 (0.0558)   | 0.2186 (0.0558)   |
| N                  | 7962              | 7962              | 7962              | 7962              | 7962              | 7962              | 7962              | 7962              |

Robust standard errors in parenthesis. Residual from the experienced reviewer's regression model is controlled when Favorite Design = Yes.

fixed effect. Note that going more flexible specification of the fixed effect (e.g., category and design fixed effect without additive separability assumption, such as $\xi_{cda}$) will face an incidental parameter problem in estimating $\tilde{\omega}_{cda}^{(e)}$ consistently, and therefore we limit ourselves to rather restrictive in controlling for unobserved heterogeneity.

The results are presented in Table 14 for intermediates and beginners separately. The OLS estimation results show evidence of endogeneity, which is depicted by positive significant estimate of $\lambda_j$ in every specification. These results are all consistent with our supposition that $\tilde{\omega}_{cda}^{(e)}$ and $\varepsilon_{rcda}^{(j)}$ are positively correlated and also that a part of this endogeneity comes from the unobserved design-category level quality heterogeneity. To circumvent the endogeneity problem, we control for the endogenous unobservable in the experienced reviewers’ evaluation in the estimation of intermediates and beginners. The results are presented in Table 15.

The control function estimation results reveal interesting patterns. The peer effect $\alpha_j$ is estimated to be negative in the absence of favourite design, but introducing the favourite design into the model flips the signs of estimated coefficients. This observation has an intuitive explanation. In the absence of favourite design, the peer effect captures the intermediates’ as well as beginners’ incentives to lower scores for their non-favourite designs, leading to seemingly puzzling negative estimate of the peer effect.
Table 15: Control Function Estimates of Peer Effects with & without Favorite Design

<table>
<thead>
<tr>
<th>Specification</th>
<th>$S_{rcda}^{(i)}$ (intermediate)</th>
<th>$S_{rcda}^{(b)}$ (beginner)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4)</td>
<td>(1) (2) (3) (4)</td>
</tr>
<tr>
<td>$\alpha_j$</td>
<td>-0.0359 (0.0047) -0.0357 (0.0046) 0.0242 (0.0056) 0.0240 (0.0056) -0.0525 (0.0053) -0.0524 (0.0063) 0.0314 (0.0063) 0.0313 (0.0063)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_j - \delta_j$</td>
<td>-0.0448 (0.0045) -0.0437 (0.0045) -0.0639 (0.0055) -0.0625 (0.0055)</td>
<td></td>
</tr>
<tr>
<td>Design FE</td>
<td>Yes Yes Yes Yes</td>
<td>Yes Yes Yes Yes</td>
</tr>
<tr>
<td>Category FE</td>
<td>No Yes No Yes</td>
<td>No Yes No Yes</td>
</tr>
<tr>
<td>Favorite Design</td>
<td>No No Yes Yes</td>
<td>No No Yes Yes</td>
</tr>
<tr>
<td>N</td>
<td>5491 5491 5491 5491</td>
<td>5491 5491 5491 5491</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses. Predicted residuals from the experienced reviewer’s regression model is controlled when Favorite Design = Yes.

Once we introduce the favourite design component, however, the peer effect restores its intuitive sign that intermediates and beginners gain some utility from conforming to the experienced reviewer’s decision.

The magnitude of the effect turned out to be somewhat weak. The estimation result implies 10% increase in the experienced reviewer’s score leads to 0.31% increase in a beginner’s score if the design under consideration is the beginner’s favourite. If the design is not the beginner’s favourite, then the same change in the experienced reviewer’s score leads to roughly 0.6% decrease in the beginner’s score. Nonetheless, all the estimates are statistically significant at 1% level.

The paper needs to be improved in many aspects. First, we shut down the feedback effect from those inexperienced reviewers’ evaluations to the experienced reviewer’s evaluation of a design. The strong asymmetry assumption on the payoff structure has to be relaxed. Second, reviewers are unlikely to have complete information on peers’ evaluations at the time of evaluation. Yet, if reviewers have private information about their own signals, then the behaviour of peers become unpredictable. Therefore, we need a way to introduce both incomplete information and common knowledge component in a more natural way that we can identify each component from the data. Lastly, as the reduced form results indicate, we also need to take into account the weight heterogeneity across categories since the incentive of reviewers to conform is stronger for those categories with a large weight.
2.6 Conclusion

This paper considered strategic interactions of expert reviewers in the context of infrastructure procurement auction. We have shown evidence that reviewers’ evaluations are not independent despite the procurement rule. We also show evidence that the reviewers value not only conformity to the experienced reviewer’s evaluation, but also the winning of their favourite design.

In the future research, we will extend our model to relax a couple of assumptions of the model. In particular, the strong asymmetry assumption about the information structure within a set of reviewers has to be relaxed.
3 What Really Matters in Procurement of Risky Projects? Lump-Sum Auction versus Unit-Price Auction

3.1 Introduction

Cost overrun is a prevalent and substantial issue in procurement of construction projects. Cost overrun is the change in procurer’s pay to the contractor as a result of changes in the project plan during the construction phase. A recent example of a large cost overrun includes Sochi 2014 Olympics, which incurred 500% cost overrun amounting to the total cost of 51 billion dollars. Panama Canal expansion project, which was contracted in 2009 with initial price of 5.25 billion dollars, went through 1.6 billion dollar cost overrun in 2014. The Central Artery/Tunnel Project, known as Big Dig, in Boston with initial contracting amount of 2.6 billion dollars ballooned to 14.8 billion dollars. All the above examples indicate that it is not sufficient to just look at contracted price to infer the final pay to the contractor in a construction procurement. Changes in a plan is a rule rather than exception in procurement of construction projects, and the final pay to the contractor could differ significantly from the contracted price.

Two predominantly adopted auction formats for procuring public infrastructure projects are Lump-Sum Auction (LSA) and Unit-Price Auction (UPA). In an LSA, firms submit a single price bid for an entire project, and the firm with the lowest price bid contracts with the government. The contractor receives its price bid amount upon completion of the project if there is no change in the contract. That is, a small adjustment on construction component during implementation is provided at the cost of a contractor under LSA. In a UPA, on the other hand, the government’s engineer first estimates how many units of each construction component would be required for the project, and each bidder submits a price for each unit of the estimated quantity. Then, the estimated quantity for each component is multiplied by bidder’s unit-price, and then summed across components to determine the total price bid for the bidder. The winner in a UPA is determined by the lowest total price bid, and the contractor is paid the total price bid amount upon the completion of the project if there is no change in the project plan. However, if there is any quantity change for any of the estimated component during the construction phase, the contractor is obliged to provide the additional quantity at its unit-price. From the point of view of incomplete contract

UPA is used, for example, highway contracting, pipeline construction, procurement in defense, internationally supported procurement by World Bank. UPA is also used in timber auctions as in Athey and Levin (2001).
literature, LSA can be best thought of as a fixed-price contract where the contractor provides the project at a fixed-price and the risk of quantity change is born by the contractor. On the other hand, UPA can be thought of as a cost-plus contract where the procurer provides the contractor with any additional cost associated with quantity change on contracted components.\textsuperscript{51}

We study LSA and UPA projects procured between year 2003 and 2012 by the Florida Department of Transportation (FDOT), which provides detailed information on both auction and construction phases of public infrastructure projects. It is the FDOT’s belief that UPA should be used for the projects with a large project uncertainty (i.e., a large potential for deviating from the FDOT’s estimate). Indeed, Bajari and Tadelis (2001) predicts that (i) fixed-price contract gives a strong incentive for cost reduction, but (ii) a fixed-price contract involves a higher cost of renegotiation than cost-plus contract. However, their theory abstracts from imperfect competition among bidders, and also abstracts from the fact that potential contractors endogenously determine their own bids in procurement auctions. For example, LSA could result in a higher winning price bid than UPA to compensate for the fact that much of the risk is transferred to the contractor under LSA. Therefore, it is important to look at the entire procurement process to evaluate the effect of auction formats on the procurement cost. To this end, we provide empirical evidence of differences in both ex-ante and ex-post auction outcomes across the two auction formats.

From a broad perspective, this paper contributes to the vast procurement auctions and incomplete contract literature by providing seemingly puzzling empirical findings that are difficult to explain with the existing theories. Further, this paper adds to an important debate about whether project risk should be primarily allocated to the procurer or the contractor in procurement of a construction project. From a context specific perspective, this paper provides a policy suggestion about whether more or less LSA should be used in practice.

In evaluating the effects of the two auction formats on project outcomes, the selection of auction formats by the FDOT becomes a concern since the FDOT explicitly states that it uses UPA for a more complex project where a contract change is likely during the construction phase. To circumvent the selection issue, we exploit exclusion restrictions derived from (i) the FDOT’s capacity constraints, and (ii) expected weather disturbances conditional on its realization. From a detailed conversation with an

\textsuperscript{51}In case of quantity change on uncontracted components, the contracting parties renegotiate the prices in both LSA and UPA. As we show later, however, approximately 75 percent of quantity change occurs on uncontracted (new) components.
FDOT’s engineer, we found that the FDOT’s engineer faces a substantial amount of transaction cost in procuring a project using UPA simply because the FDOT’s engineer needs to keep track of quantity used for each contracted component, which could vary from a few to hundreds per project. As LSA involves a lower transaction cost (or monitoring cost) than UPA, the FDOT is more likely to use LSA for the projects of medium risk when it is heavily backlogged. These projects would have otherwise been procured using UPA in the absence of backlog. We argue that the backlog level of the FDOT has nothing to do with bidders’ costs, and therefore satisfying the exclusion restriction when comparing bidding strategies across the two auction formats. In addition to the bidding strategy, we examine to what extent the degree of cost overruns and underruns differ across the two auction formats by exploiting the timing differences in the procurement process and weather disturbances data. More specifically, we use expected weather disturbances (e.g., hurricanes, rainfall, etc), which would increase project risk as instruments for the choice of auction formats. We argue that expected weather disturbances has nothing to do with cost overrun/underrun upon conditioning on realized weather disturbances.

We find a large public cost saving effect of LSA over UPA. More specifically, we estimate that LSA reduces the final pay to the contractor by more than 19% for those projects with medium level of risk. We find neither economically nor statistically significant effect of the auction format choice on cost overrun/underrun, which seem to suggest that LSA should strictly be preferred to UPA for those projects with a medium level of project uncertainty.

The empirical results here indicate that the FDOT may be too conservative in the use of LSA, and suggests more frequent use of LSA for those projects with medium level of risk as transfer of the risk to the contractor may lead to a large saving in procurement cost. The finding also indicates a need for developing a model that links ex-ante competition for the contract award process and ex-post performance of contractor under different level of risk allocation across contracting parties in order to understand the empirical results obtained in this paper.

Despite the prevalence of the problem, empirical work involving cost overruns remains scarce in the procurement auction literature. The procurement auction literature has focused on the role of ex-ante asymmetric information rather than ex-post adjustments in contracts. Empirical studies on incomplete contracts has also lagged its vast theoretical literature, and it abstracts from ex-ante asymmetric infor-
mation in contrast to the auction literature. One of the few papers that involve both ex-ante and ex-post auction behaviour of contractors in a procurement auction setting is Bajari et al. (2013). They show that ex-post change in contract comes with substantial adjustment costs by showing that bidding strategies are increasing in both cost overruns and underruns. Decarolis (2014) studies the introduction of first-price sealed-bid auctions in public procurements in Italy, and find the perverse effects of competitive bidding auction format on performance in project delivery. Bajari and Lewis (2014) investigates time-incentive contracts with a risk of contractual change where the contractor gets punished if project delivery is delayed, and show evidence of bunching in actual completion date around the planned completion date. Bajari and Tadelis (2001) theoretically characterizes the trade-off between fixed-price and cost-plus contracts, and gives an explanation for the prevalence of these two types of contracts in practice.52

The rest of the paper is organized as follows. Section 3.2 describes the procurement procedures under both LSA and UPA, and also describes the data. Section 3.3 develops an empirical model and estimation. Section 3.5 presents the estimation results and discuss its implication. Section concludes.

3.2 Institutional Details and Data

3.2.1 Procurement process under LSA and UPA

The procurement procedures for LSA and UPA can be decomposed into a design phase followed by an auction and a construction phase. Under both LSA and UPA, FDOT’s in-house engineers specify the plan/design of the infrastructure project, which includes an estimate of the project cost, a quantity estimate on each component of the designed project for the case of UPA, and initial contingency amount (ICA). ICA is the allowance given to the contractor when unexpected changes to the contract happens without going through time-consuming costly renegotiation process.53 Then, the FDOT posts an advertisement on-line which lists information about the project location, description of work, contract duration, and engineer’s estimate of the project cost. The in-house engineers decide whether to procure the project using LSA or UPA based on the degree of uncertainty in the amount of ex-post adjustment: if there is a little (resp. large) scope of change during the construction phase, the project is procured using LSA


53For those projects with an engineer’s cost estimate smaller (resp. greater) than 5,000,000 dollars, ICA cannot exceed 5 percent (resp. 1 percent) of the engineer’s cost estimate or 50,000 dollars (resp. 150,000 dollars), whichever is less.
If the project is procured using LSA, every interested firm submits a single price bid for the entire project, and the firm with the lowest price bid wins the contract. The firm is obliged to provide the project at the price it submits unless significant change in the contract occurs during the construction phase. On the other hand, if the project is procured using UPA, every interested firm submits a price bid for each unit of a construction component of which its quantity is estimated by the FDOT’s in-house engineers. For example, if the FDOT’s in-house engineer estimates that 10 units of electronic message signs need be implemented, a bidder submits a dollar amount describing how much one unit of the 10 electronic message signs will cost. Then, these unit-price bids are multiplied by the estimated quantities, and summed across all the project components to derive a single total price. The total price is then used to rank all the participating bidders, and the bidder with the lowest total price wins the contract in UPA and the contractor is obliged to provide the contracted components at its unit-price. ICA is not disclosed to the bidders at the time of bidding, but disclosed to the contractor upon winning.

Under both LSA and UPA, the above auction phase is followed by a construction phase in which the contractor provides the designed infrastructure project. If there is no change in the project plan, the contractor receives its own bid plus ICA upon delivery of the project. If the project manager of the FDOT finds a need for extra work during the construction phase, then the FDOT’s project manager files a claim for the extra work describing the associated cost, time extension required to implement the change, and the reason for the change. If the claimed amount is less than ICA, then the cost is expended from the ICA. If the claimed amount is greater than ICA, then the amount of fund is determined through renegotiation. We define the additional (resp. reduced) work through renegotiation as “cost overrun” (resp. cost underrun). Finally, the contractor receives its price bid plus ICA plus cost overrun (or cost underrun if the change is negative).

---

54 According to an in-house engineer at the FDOT, 95% of extra work is initiated by the FDOT, and not contractors.
3.2.2 Data

We investigate a sample of LSAs and UPAs that took place between years 2003 and 2012 in Florida. These two types of auction formats are widely used across U.S. departments of transportation, but detailed information on ex-post auction outcomes is often unavailable with an exception of the FDOT. In particular, records on the reasons for contract change is not readily available to public in many states department of transportation.

The data contain information on all participating bidders’ price bids (every unit-price bid for UPAs), FDOT’s engineer’s project cost estimates, quantity estimates in case of UPAs, identities of participating bidders, project location, and description of work. Ex-post adjustments through contractual changes are recorded separately based on whether the ex-post adjustment requires more fund than ICA. If the total adjustment cost in a project is below its ICA, the adjustment cost is covered by ICA and no renegotiation takes place. If the total adjustment cost in a project is above its ICA, however, renegotiation between the contracting parties may take place.

The summary statistics indicate some major differences between LSA and UPA projects, which may be partially attributed to the selection of auction formats by the FDOT. On average, a fewer number of bidders participate in LSA than in UPA. UPA is used for relatively large project and is expected to take a longer time to complete than LSA. As LSA is used for those projects with a relatively small project risks, ICA is smaller for LSA projects than for UPA projects on average. We also see that LSA projects are less susceptible to cost overruns, and the average cost overrun in UPA is six times greater than that of LSA projects. Similarly, cost underruns and reduction in work days are more than three times larger.
Table 16: Summary Statistics of Lump-Sum Auctions

<table>
<thead>
<tr>
<th>Variable</th>
<th>LSA</th>
<th>UPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td># of Bidders / Auction</td>
<td>4.10</td>
<td>2.41</td>
</tr>
<tr>
<td>Winning Price ($1,000)</td>
<td>1850</td>
<td>2310</td>
</tr>
<tr>
<td>Engineer’s Project Cost Estimate ($1,000)</td>
<td>2122</td>
<td>2624</td>
</tr>
<tr>
<td>Contract Time Length (# of Days)</td>
<td>122</td>
<td>72</td>
</tr>
<tr>
<td>ICA ($1,000)</td>
<td>29.9</td>
<td>29.1</td>
</tr>
<tr>
<td>Cost Overrun ($1,000)</td>
<td>29.4</td>
<td>84.5</td>
</tr>
<tr>
<td>Cost Underrun ($1,000)</td>
<td>-8.15</td>
<td>80.3</td>
</tr>
<tr>
<td>Extra Work Days (# of Days)</td>
<td>3.25</td>
<td>14.3</td>
</tr>
<tr>
<td>Reduction in Work Days (# of Days)</td>
<td>-0.123</td>
<td>1.64</td>
</tr>
</tbody>
</table>

The summary statistics above are calculated based on Lump-Sum auctions and Unit-Price auctions that took place between years 2003 and 2012. Price and Funds are in thousand of US dollars.

Table 17: Distribution of Projects with Cost Overruns and Underruns under LSA and UPA

<table>
<thead>
<tr>
<th>LSA</th>
<th>UPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Overrun</td>
<td>Cost Underrun</td>
</tr>
<tr>
<td># of projects</td>
<td>169</td>
</tr>
<tr>
<td>share among total # of project</td>
<td>(30%)</td>
</tr>
<tr>
<td>Total number of projects</td>
<td>554</td>
</tr>
</tbody>
</table>

# of projects with cost overruns and underruns are tabulated above.

in UPA than LSA projects.\textsuperscript{55}

Table 17 tabulates the frequency distribution of cost overrun and underrun. It is clear that cost overrun is a rule rather than exception in public construction projects as almost 50% of all the UPA projects in the sample experienced some cost overruns. Now, cost overruns can be generated not only through a change in non-contracted components. That is, FDOT’s engineers may realize during construction phase that a completely new task is required. For example, suppose that the FDOT engineer forgot to include the procedure to prevent erosion of an existing drainage pipe, and erosion occurs. The FDOT has to replace the drainage pipe at its cost, which is not described in the project plan. To see this, we look a subset of the UPA projects data where we have information on ex-ante as well as ex-post quantity on each component.

\textsuperscript{55}Time extension is only available through renegotiation.
### Table 18: Summary Statistics on Ex-Post Adjustments in Unit-Price Contracts

<table>
<thead>
<tr>
<th>Variable</th>
<th># of Component Categories</th>
<th>Total Component Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>Addition in Contracted Components</td>
<td>4.02</td>
<td>15.4</td>
</tr>
<tr>
<td>Reduction in Contracted Components</td>
<td>4.56</td>
<td>11.4</td>
</tr>
<tr>
<td>Change on Uncontracted Components</td>
<td>8.98</td>
<td>22.9</td>
</tr>
<tr>
<td>Share of Additional Contracted Components</td>
<td>0.093</td>
<td>0.193</td>
</tr>
<tr>
<td>Share of Reduced Contracted Components</td>
<td>0.168</td>
<td>0.242</td>
</tr>
<tr>
<td>Share of Change in Uncontracted Components</td>
<td>0.739</td>
<td>0.321</td>
</tr>
</tbody>
</table>

Summary statistics on component level changes in 298 Unit-Price projects are shown in both absolute and share terms. The sample period is between year 2010 and 2014. Values are expressed in $1,000 and calculated based on unit-price bids for contracted components.

This data allows us to see how frequent quantity change on contracted components and uncontracted components are, and also how significant each type of the change on the cost of procurement.

Table 17 reveals surprising facts. On average, there are four components that experience positive and negative quantity changes while nine new components are added during construction phase of a project. A surprising observation here is that additional quantity on contracted component is fairly insignificant compared to reduction and changes on uncontracted components. Extra quantity on contracted components adds only $51,000 per project while additional costs associated with new uncontracted components is $670,000. As the number of components and project size differ across projects, we consider the share of the above variables, but the picture remains the same. Only 5.3% of the total change in costs is associated with extra quantity on contracted components. Vast majority of the cost is derived from emergence of new uncontracted components during construction.

### 3.3 Model and Estimation

#### 3.3.1 Empirical Model

We look at the effect of auction formats on various project outcomes, such as winning price bid and final pay to the contractor. A final pay to the contractor is the sum of winning price, initial contingency amount, and cost overrun/underrun. To be explicit about what we intend to identify, we first consider a simple empirical model of auction outcomes with a binary endogenous regressor. We further develop a censored regression model that takes into account the fact that cost overruns and underruns are only...
observed when change in the contract involves a high cost, and cannot be entirely covered by ICA.

Let $LSA_a$ be an indicator variable which takes the value of 0 (resp. 1) when UPA (resp. LSA) is used. Further, let $Z_a$ be the vector of instruments that affect the choice of auction formats by the FDOT but do not affect the auction outcomes through any other path. We omit the vector of controls in this section to avoid cluttering. Then, the selection of auction formats can be modeled using latent utility model, such that:

$$LSA_a = \begin{cases} 1 & \{\gamma_0 + Z_a \gamma_1 > \varepsilon_a\} \end{cases}$$ (42)

where $\gamma_0$ and $\gamma_1$ are parameters and $\varepsilon_a$ is an idiosyncratic shock. As we do not observe the potential outcomes of a single project in both states, we use observed outcome, $Y_a$, to estimate the causal effect of LSA on auction outcomes, such that:

$$Y_a = \alpha + \beta LSA_a + u_a$$ (43)

where $u_a$ is unobserved heterogeneity in project outcome. To see that endogeneity is a concern here, suppose that $\varepsilon_a$ contains unobserved project complexity, and $Y_a$ is the observed cost overrun in project $a$. As a complex project tends to involve a large change in the plan, which is captured by $u_a$, we expect a positive correlation between $\varepsilon_a$ and $u_a$. Therefore, we would underestimate the causal effect of LSA on the cost overrun since a complex project is likely procured through UPA, and we would observe those projects procured using UPA tend to have a large cost overruns/underruns. Therefore, as in Imbens and Angrist (1994), we assume independence, exclusion restriction, and relevance of our instruments to identify the local average treatment effect of using LSA over UPA. In the simple case where $Z_a$ is a scalar, the local average treatment effect is given by:

$$\beta_{LATE} = E[Y_a|Z_a] - E[Y_a|Z'_a]$$

$$Pr(LSA_a = 1|Z_a) - Pr(LSA_a = 1|Z'_a)$$ (44)

where $Z'_a \neq Z_a$, and $\beta_{LATE}$ is identified using those projects that the FDOT complied to use LSA for the reasons that are not directly related to project outcomes. The IV estimator for $\beta_{LATE}$ is obtained simply
by replacing the population moments in (44) with corresponding sample analogs. As we have multiple instruments, we average across local treatment effects estimated using different instruments.

The above specification of the model makes sense when the outcome variable is continuous. However, some project outcomes, such as cost overruns/underruns, are censored at zero as the costs of changing the plan are to some extent covered by the ICA. Therefore, we take censoring into account, and estimate by Maximum Likelihood Estimator by assuming joint normal distribution of unobserved heterogeneities.

A problem arises when an outcome variable is censored at zero since the logarithmic transformation of an outcome variable is not well defined. Yet, logarithmic transformation is intuitive and preferred as it gives a natural elasticity interpretation. To this end, we consider the following transformation of variable for cost overruns and underruns.

\[
\hat{Y}_{a}^{\text{over}} = \ln(\text{project cost estimate}_a + \text{cost overrun}_a) - \ln(\text{project cost estimate}_a) \quad (45)
\]

\[
\hat{Y}_{a}^{\text{under}} = \ln(\text{project cost estimate}_a + \text{cost underrun}_a) - \ln(\text{project cost estimate}_a) \quad (46)
\]

where \(\text{project cost estimate}_a\) is an FDOT’s engineer’s estimate of the project cost in the absence of cost overruns/underruns. Note here that both of the above transformed outcomes are censored at zero, and the transformation leads to a natural percentage change interpretation.\(^{56}\)

The censored outcomes are modeled as:

\[
LSA_a = 1\{\gamma_0 + Z_a \gamma_1 > \varepsilon_a\}
\]

\[
\hat{Y}_{a}^{\text{over}} = \alpha^{\text{over}} + \beta^{\text{over}} LSA_a + u_a^{\text{over}} \quad \text{if cost overrun}_a > 0
\]

\[
\hat{Y}_{a}^{\text{under}} = \alpha^{\text{under}} + \beta^{\text{under}} LSA_a + u_a^{\text{under}} \quad \text{if cost underrun}_a > 0
\]

\[
= 0 \quad \text{otherwise}
\]

where \(\alpha^j\) and \(\beta^j\) for \(j \in \{\text{over, under}\}\) are parameters of the model. \(\varepsilon_a, u_a^{\text{over}}, \text{ and } u_a^{\text{under}}\) are assumed to be jointly normally distributed so we can estimate with Maximum Likelihood.

\(^{56}\)The interpretation is in terms of percentage out of \(\text{project cost estimate}_a\) for cost overrun model with an assumption that \(\text{cost overrun}_a\) is relatively small compared to \(\text{project cost estimate}_a\), for example.
3.3.2 Instruments and Local Average Treatment Effect

Ideally, we would like to have the FDOT’s engineers to choose auction formats randomly across projects. Given that UPA is used for those projects with a large amount of project outcome uncertainty, the choice of auction formats may be correlated with unobserved project heterogeneity. Since uncertain projects are likely to involve more adjustment on average, we expect that cost overruns and underruns be larger under UPA than LSA projects. To circumvent the selection issue, we propose two types of instruments: FDOT’s capacity constraints, and expected weather disturbances.

Our main instruments capture the FDOT’s capacity constraints in procuring infrastructure projects. As mentioned before, the UPA involves a large transaction cost during construction phase, such that the in-house engineers needs to keep track how many units of each component are used. The transaction costs could impose significant transaction costs on the FDOT if a project involves many construction components. Therefore, the FDOT may decide to procure a project using LSA when it is severely capacity constrained.

As FDOT is split into seven regional offices that independently procure infrastructure projects, we construct three variables that capture the constraints of the FDOT’s district offices in using UPA. The first instrument is the number of projects procured simultaneously from the same district office. The second instrument is the total dollar value of unfinished projects procured in the past from the same district office, which we define it as district office backlog. As district offices may differ in their capacity (e.g., the number of in-house engineers) to procure projects, we also control for district office capacity, which is defined as the maximum district office backlog in the sample. We argue that the FDOT’s capacity constraints satisfy the exclusion restriction since the FDOT’s administrative situation is unlikely correlated with bidders’ costs. That is, bidders’ bidding strategies are independent of the FDOT’s capacity constraints.

Another type of instruments that complements the analyses when looking at ex-post auction outcomes is the expected weather disturbances. As weather is a crucial determinant of construction project outcomes, we expect that the FDOT is likely to use UPA in those regions where much weather disturbances, such as heavy rainfall and hurricane, are expected because damages to construction components may result in rework. Although relevance between expected weather disturbances and the choice of auction formats is intuitive and testable, exclusion restriction is harder to argue since expected weather disturbances may
affect cost overruns and/or underruns through some other unobservables to the econometrician. To this end, we exploit the timing differences in the procurement process, and control for realization of weather when examining the effect of auction formats on cost overruns and underruns. More specifically, we argue that the expected weather disturbances has nothing to do with cost overruns/underruns conditional on realized weather. This argument is reasonable since cost overruns and underruns are functions of realized environment while the choice of auction format is made by the FDOT’s engineers before weather disturbances are realized.

Now, as in recent literature on identification of the local average treatment effect when using instrumental variable approach, our instruments identify the effect of using LSA over UPA for those projects with medium level of project risk. To be more concrete, it is helpful to look at the procurement process from the point of view of the FDOT’s engineer. Suppose that there are two projects with a low project risk/uncertainty and a moderately high project risk. In the absence of any additional project the FDOT uses LSA for the project with low uncertainty, and uses UPA for the project with moderately high uncertainty. Suppose now that there is additional project with a very high uncertainty. The FDOT wishes to use UPA for both projects of moderately and very high uncertainty, but it does not have enough personnel to do so. Therefore, the FDOT uses UPA for the project with a very high uncertainty and uses LSA for the projects with low and moderately high project uncertainty. In summary, the projects that comply due to variation in our instruments are the marginal projects that have medium level of project risk.

3.4 Estimation Results and Discussions

3.4.1 Estimation Results

Table 19 presents the results for continuous auction outcome variables, which includes winning price bids and final pay to the contractor. Recall here that a price bid is a single lump-sum bid in LSA while it is the sum of unit-price multiplied by component-wise estimated quantity. The final pay to the contractor of a project is the sum of winning price bid, initial contingency amount, and cost overruns/underruns. The first stage results of 2SLS is presented in 20 using only FDOT’s capacity constraints as instruments, and also using expected weather disturbances in addition to capacity constraint instruments.

We construct expected weather disturbances using monthly precipitation and hurricane hits data.\textsuperscript{57}

\textsuperscript{57}County level monthly precipitation and hurricane data are retrieved from http://www.ncdc.noaa.gov/cdo-
More specifically, we estimate the belief of the FDOT’s engineer on expected total rainfall assuming that the total monthly precipitation follows AR(1) process with county and month fixed effects. Using the predicted value and contract duration of a project, we construct the expected total rainfall during the contract duration at the time of project letting. Similarly, we construct the probability that the county of project implementation is hit by hurricane at least once using county and month fixed effects.

Table 19: Estimation Results for Continuous Auction Outcomes

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Winning Price Bid</th>
<th>Final Pay to Contractor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5) (6)</td>
<td></td>
</tr>
<tr>
<td>Estimation Method</td>
<td>OLS 2SLS 2SLS OLS 2SLS 2SLS</td>
<td></td>
</tr>
<tr>
<td>LSA (=0 if UPA, =1 if LSA)</td>
<td>-0.010 -0.225** -0.269** -0.012 -0.195** -0.238**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015) (0.089) (0.113) (0.015) (0.083) (0.101)</td>
<td></td>
</tr>
<tr>
<td>Hansen J Statistic</td>
<td>1.09 0.894 0.946 1.06</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$ (2) P-value</td>
<td>0.295 0.344 0.330 0.303</td>
<td></td>
</tr>
<tr>
<td>Project Characteristics</td>
<td>Yes Yes Yes Yes Yes Yes</td>
<td></td>
</tr>
<tr>
<td>Contractor Fixed Effect</td>
<td>No No Yes No No Yes</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1624 1624 1624 1624 1624 1624</td>
<td></td>
</tr>
</tbody>
</table>

Heteroskedasticity standard errors in parentheses. Only FDOT’s capacity constraints are used as instruments for the above 2SLS estimations. Project characteristics includes log(engineer’s project cost estimate), log(contract duration), log(ICA+1), log(realized weather disturbances) and project types. Project types are also controlled and defined as a linear combinations of tasks, which are extracted from work description of bid tabs.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

web/search?datasetid=GHCNDMS and http://maps.csc.noaa.gov/hurricanes/, respectively.
The first stage results in Table 20 show intuitive signs. District office backlog level and the number of projects procured simultaneously are positively and significantly correlated with the probability of using LSA over UPA. On the other hand, district office capacity, which is measured as the maximum backlog level of a district within the sample, is negatively correlated with the use of LSA. Further, we find that those rainfall and hurricane prone regions are less likely to use LSA, which agrees to the hypothesis. The estimation results imply that, for example, doubling the number of projects procured simultaneously from the same district office would increase the probability of using LSA over UPA by 13%.

The 2SLS results in Table 19 shows a large effect of using LSA over UPA, which we cannot see in OLS suggesting a large degree of endogeneity. Note that the instruments here do not include weather disturbances and we use only FDOT’s capacity constraints in the above 2SLS estimation. The 2SLS estimates indicate that both winning price and finalize pay to the contractor reduces more than 19% if the

Table 20: First Stage Results

<table>
<thead>
<tr>
<th>First Stage Dependent Variable</th>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>District Office Constraints IV</strong></td>
<td></td>
</tr>
<tr>
<td>district office backlog</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>4.81e−10***</td>
</tr>
<tr>
<td></td>
<td>(1.56e−10)</td>
</tr>
<tr>
<td>district office capacity</td>
<td>−2.10e−10*</td>
</tr>
<tr>
<td></td>
<td>(1.18e−10)</td>
</tr>
<tr>
<td># of projects procured simultaneously</td>
<td>0.1279***</td>
</tr>
<tr>
<td></td>
<td>(0.0179)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Expected Weather Disturbance IV</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>log(expected total rainfall while on contract)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>probability of hurricane hit while on contract</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project Characteristics</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-sq</td>
<td>0.215</td>
<td>0.229</td>
</tr>
<tr>
<td>Obs</td>
<td>1624</td>
<td>1624</td>
</tr>
</tbody>
</table>

Heteroskedasticity robust standard errors in parentheses. Expected total rainfall is constructed assuming total monthly rainfall follows AR(1) process with county and month fixed effects. Probability of hurricane occurrence while on contract is constructed similarly. Project characteristics includes log(engineer’s project cost estimate), log(contract duration), log(ICA+1), log(realized weather disturbances) and project types. Project types are defined as a linear combinations of tasks, which are extracted from work description of bid tabs.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
FDOT uses LSA over UPA for the projects with medium project risk. The findings here are robust to inclusion of contractor fixed effects.

Table 21 shows estimation results for the censored outcome model we developed in the previous section. We find no evidence of differences in costs between LSA and UPA.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Cost Overrun</th>
<th>Cost Underrun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Tobit MLE</td>
<td>Tobit MLE</td>
</tr>
<tr>
<td>LSA (=0 if UPA, =1 if LSA)</td>
<td>0.000 (0.005)</td>
<td>-0.001 (0.011)</td>
</tr>
<tr>
<td>Project Characteristics</td>
<td>Yes Yes Yes Yes Yes Yes</td>
<td></td>
</tr>
<tr>
<td>Contractor Fixed Effects</td>
<td>No No Yes No No Yes</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1624 1624 1624 1624 1624 1624</td>
<td></td>
</tr>
</tbody>
</table>

Heteroskedasticity robust standard errors in parentheses. We control for log(realized rainfall while on contract) at county level when instrumenting by log(expected rainfall while on contract). We also control for the number of days that were extended for rainy days at project level. Project types are also controlled and defined as a linear combinations of tasks, which are extracted from work description of bid tabs.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.4.2 Why Do We See What We See?

The finding of the paper is puzzling and intriguing. Our empirical results suggest that LSA is strictly superior to UPA for those projects with medium level of risk. Despite the fact that UPA involves a high transaction cost during construction phase, there does not seem to be any cost saving effect when cost overruns occur. Indeed, most of the cost overruns come from emergence of new components for which LSA and UPA have exactly the same procedure. Most of the reduction in the procurement cost seems to come from reduction through contracted price or equivalently winning price.

So what explains the empirical findings here? The empirical results are consistent with the prediction of Ewerhart and Fieseler (2003), which points out inefficiency of UPA relative to first-price sealed bid auction (i.e., LSA). They show in a simple framework that UPA could select the contractor of non-lowest cost among the set of bidders. The pre-fixed estimated quantities of construction components, of which some of them are not used at all in the actual implementation of the project, distorts and generates non-monotonic bidding strategies. As a result, LSA may select more efficient bidder than UPA, which is
reflected by the large difference in the winning prices between LSA and UPA.\textsuperscript{58}

While Ewerhart and Fieseler (2003) answers the question of how LSA could be more efficient than UPA, their theory does not explain why UPA is so pervasively used in practice. They point out the possibility of (i) risk sharing between the contracting parties by transferring the risk from the contractor to the procurer, and (ii) reducing the cost of renegotiation through a simple cost-plus contract mechanism. However, the data rejects these above two hypotheses as the majority of cost overruns are generated through the change in the scope of work where contracted components are irrelevant, and the renegotiation process is exactly the same across LSA and UPA.\textsuperscript{59}

3.5 Conclusion

This paper examined a causal effect of auction formats on a variety of procurement outcomes, which includes winning price, final pay to the contractor, and cost overruns/underruns. We found a large cost reducing effect of Lump-Sum Auction over Unit-Price Auction for projects with medium level of project risk. While cost overruns/underruns are sizable and norm in construction projects, we did not find any significant differences in the degree of cost overruns/underruns across the two auction formats. Despite its pervasive use in procurement of construction projects, UPA does not seem to benefit the buyer and existing theories can only partially explain our empirical result.

For the future research, we attempt to develop a theory of Lump-Sum and Unit-Price auctions that takes into account of ex-post auction outcomes.

\textsuperscript{58}Inverted U-shaped bidding strategy in UPA is found in Athey and Levin (2001).
\textsuperscript{59}Ewerhard and Fieseler (2003) also suggest the possibility of UPA being easy to work with. Through conversation with a FDOT’s engineer, we confirmed that UPA is much more costly and involves more time and effort than LSA.
References


79


81


3.6 Appendix for Chapter 1

3.6.1 Definition of Observables

- Engineer’s estimate of project cost: A proxy for the project size.
- Distance: Distance between project site and the closest branch of bidder.
- Utilization Rate: A bidder’s backlog per capacity. Backlog is defined as the total dollar value of projects ongoing at the time of bidding. Capacity of a bidder is defined as the maximum backlog during the period the sample is taken from. Backlog and capacity are calculated using all other types of auctions and DB auctions procured by the FDOT from 1999 to 2012.
- Project Type: Projects are classified into Road, Bridge, Building, Mixed Project, Monitoring System Implementation, and Others

3.6.2 Proof of Proposition 1

Proof. Let \( \{p_i(\alpha), q_i(\alpha)\} \) be the solutions to this constrained optimization problem of a bidder \( i \), such that:

\[
\begin{align*}
\{p_i(\alpha), q_i(\alpha)\} &= \arg \max_{p_i, q_i^0} \pi^i_{\text{int}} \quad \text{subject to} \quad p_i = \alpha q_i \\
\Leftrightarrow \{\alpha q_i(\alpha), q_i(\alpha)\} &= \arg \max_{q_i^0} G_i(\alpha, \psi - i) (\alpha q_i - vc_i C(q_i) - fc_i) \\
\Leftrightarrow \{\alpha q_i(\alpha), q_i(\alpha)\} &= \arg \max_{q_i^0} \alpha q_i - vc_i C(q_i) - fc_i
\end{align*}
\]

(47)

The first-order necessary and sufficient condition w.r.t. \( q_i \) gives:

\[
\begin{align*}
\alpha - vc_i C_p(q_i) &= 0 \\
\Rightarrow q_i(\alpha) &= C^{-1}_q(\alpha/vc_i) \\
\Rightarrow p_i(\alpha) &= \alpha C^{-1}_q(\alpha/vc_i)
\end{align*}
\]

(48)

(49)

3.6.3 Proof of Proposition 2: Multiplicative Separability of Partially Reduced Payoff Function

I show that \( p_i(b_i) - vc_i C(q_i(b_i)) - fc_i \) is multiplicatively separable in \( vc_i \) and \( e_i \equiv fc_i/C^{-1}_q(1/vc_i) \) given Assumption 1.
Proof. From Proposition 1, it follows that:

\[ p_i(b_i) - v_i C(q_i(b_i)) - f_i = b_i C_q^{-1}(b_i/v_i) - v_i C(C_q^{-1}(b_i/v_i)) - f_i \]

Now, Assumption 1 implies that \( C_q^{-1}(\cdot) \) is homogeneous of degree \( 1/(\gamma - 1) \). Therefore, it follows that:

\[ b_i C_q^{-1}(b_i/v_i) - v_i C(C_q^{-1}(b_i/v_i)) = \frac{v_i^{\gamma-1}}{\gamma} (b_i C_q^{-1}(b_i) - C(C_q^{-1}(b_i))) = C_q^{-1}(1/v_i) (u(b_i) - \epsilon_i) \]

where \( u(b_i; \epsilon_i) \equiv b_i C_q^{-1}(b_i) - v_i C(C_q^{-1}(b_i)) \).

3.6.4 Proof of Proposition 3: Equilibrium Existence, Monotonicity, and Differentiability

I first show monotonicity of equilibrium PQR strategy in \( e_i \) assuming that an equilibrium exists. Suppose, to the contrary to the claim, that there exists a non-monotone equilibrium, such that there exist some \( \psi_1(e_1) \equiv \psi_1 > \psi_2 \equiv \psi_2(e_2) \) with \( e_1 < e_2 \). Then,

\[
G_i(\psi_1, \psi_{-i})(u(\psi_1) - e_1) \geq G_i(\psi_2)(u(\psi_2, \psi_{-i})) - e_1
\]

\[
G_i(\psi_2, \psi_{-i})(u(\psi_2) - e_2) \geq G_i(\psi_1, \psi_{-i})(u(\psi_1) - e_2)
\]

which implies:

\[
G_i(\psi_1, \psi_{-i})(u(\psi_1) - e_1) - G_i(\psi_2, \psi_{-i})(u(\psi_2) - e_1) \geq G_i(\psi_2, \psi_{-i})(u(\psi_2) - e_2) - G_i(\psi_1, \psi_{-i})(u(\psi_1) - e_2)
\]

\[ \Rightarrow e_1 \geq e_2 \]

which is a contradiction. Therefore, PQR strategy is non-decreasing in \( e_i \) for any equilibrium if exists.

Now, I prove existence. Multiplicative separability of the partially reduced form payoff function implies that the equilibrium PQR strategy of a bidder is independent of \( v_i \) because:

\[
\max_{b_i \in [b, B]} \pi_i^{int} = \max_{b_i \in [b, B]} G_i(b_i, \psi_{-i}) (u(b_i) - \epsilon_i)
\]

\[ = \max_{b_i \in [b, B]} \ln(G_i(b_i, \psi_{-i})) + \ln(u(b_i) - \epsilon_i) \]

To show the existence of monotone equilibrium, I show log-supermodularity between own bids \( b_i \) and
private information $e_i$. Then, I apply existence theorem proposed in Athey (2001), such that:

$$\frac{\partial^2 \ln(\pi_i^{int})}{\partial b_i \partial e_i} = \frac{\partial^2 \ln(u(b_i) - e_i)}{\partial b_i \partial e_i} = \frac{u'(b_i)}{(u(b_i) - e_i)^2} > 0$$

which completes the proof.

Now, define $d_j \equiv \ln(b_j) - \ln(w_j)$ and its distribution function as $F_d$. Suppose, for simplicity, assume that $d_j$ is iid across bidders.\textsuperscript{60} By Assumption 2, $F_d$ has a smooth density with an infinite support regardless of bidder $j$’s strategy for any $j \neq i$.

$$G_i(b_i, \psi_{-i}) \equiv Pr(\ln(b_i) < d_j + \ln(w_i) \ \forall j \neq i) = \int [1 - F_d(\ln(b_i) - \ln(w_i))]^{N-1} dF_w$$

which is differentiable with respect to $b_i \in (b, \bar{b})$.

3.6.5 Proof of Proposition 4:

Proof. Consider equation (4). By implicitly differentiating with respect to $\tau$, I have:

$$\frac{d\psi_i(e_i)}{d\tau} \equiv \frac{dq_{i,\psi}^\psi(c_i)}{d\tau}C_{qq}(q_{i,\psi}^\psi(c_i)) \nu c_i$$

(50)

which implies that $\frac{dq_{i,\psi}^\psi(c_i)}{d\tau}$ and $\frac{d\psi_i(e_i)}{d\tau}$ have the same sign. Also,

$$\frac{dp_{i,\psi}^\psi(c_i)}{d\tau} = \frac{d\psi_i(e_i)}{d\tau} d_i^\psi(c_i) + \frac{dq_{i,\psi}^\psi(c_i)}{d\tau} \psi_i(c_i)$$

(51)

Therefore,

$$\text{sign} \left( \frac{d\psi_i(e_i)}{d\tau} \right) = \text{sign} \left( \frac{dp_{i,\psi}^\psi(c_i)}{d\tau} \right) = \text{sign} \left( \frac{dq_{i,\psi}^\psi(c_i)}{d\tau} \right)$$

(52)

\textsuperscript{60}It should be obvious that affiliation of types across bidders do not cause any problem here.
3.6.6 Proof of Proposition 5: Multiplicative Separability of Bidding Strategies

Let \( b(\theta_a, e_{ia}) \) be the equilibrium PQR strategy of bidder \( i \) in an auction \( a \) with unobserved heterogeneity \( \theta_a \) with efficiency private information \( e_{ia} \). Similarly, define \( s_{ia} \equiv b(0, e_{ia}) \), such that \( s_{ia} \) is the strategy of bidder \( i \) when \( \theta_a = 0 \). I show that there is a function of \( \theta_a \), say \( h(\theta_a) \), that satisfy \( b(\theta_a, e_{ia}) = h(\theta_a) s_{ia} \). Then, I show that the two first order optimality conditions are satisfied under the conjectured equilibrium strategy profile. For the sake of simplicity, I omit observed heterogeneity in the following proof, but the proof for observed heterogeneity follows exactly the same steps as the proof for unobserved heterogeneity shown below.

Proof. The probability of winning function \( G_i(.) \) is homogeneous of degree 0, and so I have \( G_i(h(\theta_a) s_{ia}, h(\theta_a) \psi_{-i}) = G_i(s_{ia}, \psi_{-i}) \). Further, its density function \( g(.) \) is homogeneous of degree -1 since \( G_i(.) \) is homogeneous of degree 0, implying that \( g(h(\theta_a) s_{ia}) = s_{ia}/h(\theta_a) \). Therefore, LHS of (18) can be factored as follows.

\[
\kappa_1(\gamma) (h(\theta_a) s_{ia})^{\gamma-1} + \kappa_2(\gamma) (h(\theta_a) s_{ia})^{\gamma-1} \frac{G_i(h(\theta_a) s_{ia}, h(\theta_a) \psi_{-i})}{g_i(h(\theta_a) s_{ia}, h(\theta_a) \psi_{-i})} = h(\theta_a) \frac{\gamma}{\gamma-1} \left( \kappa_1(\gamma) s_{ia}^{\gamma-1} + \kappa_2(\gamma) s_{ia}^{\gamma-1} \right) \frac{G_i(s_{ia}, \psi_{-i})}{g_i(s_{ia}, \psi_{-i})}
\]

Therefore, \( h(\theta_a) = \exp\{\frac{\gamma-1}{\gamma} \rho \theta_a\} \) does not affect the first order condition (18) in any way.

Now, consider pricing strategy. It is immediate that pricing strategy is log-linear in \( \theta_a \), such that:

\[
\hat{p}(Z_a, \theta_a, \epsilon_{ia}, \epsilon_{ia}) = \frac{\gamma}{\gamma-1} \left( \frac{\gamma-1}{\gamma} \rho \theta_a + \hat{s}_{ia} \right) + (1 - \rho) \theta_a + \epsilon_{ia} - \frac{1}{\gamma-1} \hat{s}_{ia}
\]

3.6.7 Equilibrium Computation Algorithm

Step 1: Set the level of horizontal reviewer heterogeneity arbitrarily large. Denote the pseudo horizontal reviewer heterogeneity by \( \tau^j \) where \( \tau^0 \) is the initial level of pseudo horizontal reviewer heterogeneity. Then, I draw 300 types from \( e_{ia} \sim F_{\epsilon}(\cdot|N, R) \).

Step 2: Guess a strategy of each bidder and denote them by \( \psi_{i}^{j,k}(e_{i}) \) for \( k = 0 \). Given the strategy of bidders, compute the equilibrium strategy by applying the following Quasi-Newton map to every bidder of every type simultaneously (Here I suppress the dependency of strategy on \( c_i \) for the sake of visual clarity).

\[
\psi_{i}^{j,k+1} = \psi_{i}^{j,k} - H_{j,k} J_{j,k}
\]

where \( J_{j,k} \) and \( B_{j,k} \) are the Jacobian and inverse Hessian of \( \pi_i^{int} \). Since inversion of Hessian matrix is
computationally very expensive, I approximate $H^{j,k}$ (say $\hat{H}^{j,k}$) by BFSG method, such that:

$$\hat{H}^{j,k+1} = \left( I - \frac{(J^{j,k+1} - J^{j,k})(\psi^{j,k+1}_i - \psi^{j,k}_i)^T}{(J^{j,k+1} - J^{j,k})^T(\psi^{j,k+1}_i - \psi^{j,k}_i)} \right)^T \hat{H}^{j,k} \left( I - \frac{(J^{j,k+1} - J^{j,k})(\psi^{j,k+1}_i - \psi^{j,k}_i)^T}{(J^{j,k+1} - J^{j,k})^T(\psi^{j,k+1}_i - \psi^{j,k}_i)} \right) + \frac{(\psi^{j,k+1}_i - \psi^{j,k}_i)(\psi^{j,k+1}_i - \psi^{j,k}_i)^T}{(J^{j,k+1} - J^{j,k})^T(\psi^{j,k+1}_i - \psi^{j,k}_i)}$$

A necessary condition for a Bayesian Nash Equilibrium is $\psi^{j,l+1}_i = \psi^{j,l}_i$ for all $l > K$ where $K$ is some arbitrarily large integer.

Step 3: Upon convergence, I reduce horizontal reviewer heterogeneity by $\kappa > 0$, such that $\tau^{j+1} = \tau^j - \kappa$. Then, use equilibrium strategy $\psi^{j,k}_i$ as an initial guess for $\psi^{j+1,k}_i$.

Step 4: Repeat step 2 and 3 till $\tau^j = \hat{\tau}$, such that pseudo horizontal reviewer heterogeneity meets estimated horizontal reviewer heterogeneity in the data.