Trust and Trustworthiness in Procurement Contracts with Retainage

Matthew J. Walker[*], Elena Katok[†] and Jason Shachat[‡]

Abstract

In complex procurement projects, it is difficult to write enforceable contracts that condition price upon quality. Supplier non-performance becomes an acute risk, particularly when there is intense competition for the contract. An established incentive mechanism used to mitigate the problem of supplier non-performance is retainage, in which the buyer sets aside a portion of the purchase price. After project completion, the buyer determines the amount of retainage that is released to the seller, considering any defects that arise. While generally a feasible contract form to implement, the practical difficulties in assessing completion introduce a moral hazard for the buyer. We develop a structurally new game and experimental design to offer managerial insights on how retainage principles mediate trust and trustworthiness in competitive procurement settings with moral hazard. The experimental results suggest that if trust in the procurement relationship is strong enough, then retainage can mitigate the seller-side moral hazard problem and substitute for reputation in a fragmented supply chain, at the cost of inflated tender prices. In high retainage structures, there is a trade-off between trade efficiency and supplier participation in request for bids. We further develop a model of fair payment norms and offer managerial insights on how to design the retainage mechanism, conditional on prevailing levels of trust and beliefs about fairness.

Keywords: trust, procurement, competition, retainage, moral hazard

[*] Newcastle University Business School, Newcastle Upon Tyne, NE1 4SE, United Kingdom; email: matt.walker@newcastle.ac.uk (corresponding author).
[†] Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080; email: ekatok@utdallas.edu.
[‡] Durham University Business School, Durham, DH1 3LB, UK; email: jason.shachat@durham.ac.uk.

1. Introduction

A tension between competition and cooperation characterizes many procurement settings. It is common practice for public and private sector entities to rely on competitive procurement to obtain goods and deliver projects. [1] For standardized goods, price competition promotes productive efficiency and cost reduction by suppliers. For non-standardized goods, such as complex construction or infrastructure projects, the benefits of competition are less straightforward. A distinguishing feature of procurement tenders, as opposed to sales auctions, is that the bidding process represents the beginning rather than the end of the relationship (Fugger, Gretschko and Pollrich 2019). Procurers often rely on *ex-post* incentives to mitigate the risk of supplier non-performance.[2] It can be difficult, however, for the procurer to write complete and enforceable agreements *ex-ante* that condition price on the quality of works delivered (Chakravarty and MacLeod 2009, Gretschko and Pollrich 2019). Supplier cost-cutting is an ever-present issue (Lo et al. 2007, Midler 2007). This may manifest itself in reduced quality materials or unethical/unsustainable production processes (Guo et al. 2015, Chen and Lee 2016). Intense competition for the contract may further increase suppliers' incentives to cut corners later on (Chaturvedi 2021), or adversely affect relationship-building (Emiliani and Stec 2005).

An understudied incentive mechanism used in procurement to mitigate supplier moral hazard issues is retainage. A retainage provision, or retention as it is known outside the United States, is a pre-agreed percentage of the contractual price withheld from a seller by the buyer. The buyer in this context might be a client, main contractor or sub-contractor withholding money from a lower tier. Retainage has its origins in nineteenth-century British railway construction, when it was set at 20% of contractual value (Bausman 2004). Today, typical provisions range from 3 to 10% and provisions are found across most standard construction contract types (Cox et al. 2011, Nabi Mohamad et al. 2021).[3] On substantial completion of the project, retainage is released back to the seller, minus deductions for any defects that arise. Measures to safeguard cash retainage vary by country and locale.[4]

---

[1] The World Bank estimates that on average, public procurement constitutes 14.5% of gross domestic product globally (Djankov et al. 2017).

[2] See, e.g., Bajari et al. (2014) in highway procurement. An alternative mechanism to overcome the tension between competition and cooperation in procurement is proposed by Chakraborty et al. (2021). In a mixed adverse selection and moral hazard model, they show that incentives to mitigate shirking by sellers can be provided either by limiting the number of bidders, or by using an inefficient auction allocation rule.

[3] In the UK, standard building contracts are produced by the Joint Contracts Tribunal, which encourages the holding of retention monies until practical completion. In the US, standard Design-Build agreements include those produced by the American Institute of Architects, ConsensusDocs and the Engineers Joint Contract Documents Committee. These contracts often specify that half of the retainage money be released immediately, while the remainder is released after the expiration of a defects liability period.

[4] The European Commission (2009, clause 41) prescribes that retainage monies "are not paid until the satisfaction of conditions specified in the contract for the payment of such amounts or until defects have

In this paper, we develop a structurally new game and experimental design to offer managerial insights on how retainage principles mediate trust and trustworthiness in procurement settings with moral hazard.[5] Specifically, we investigate how and when contractual retainage can be used to mitigate the seller moral hazard problem in a two-tier supply chain. If operating as intended, retainage circumvents the difficulties in writing a complete contract and effectively aligns project incentives (Raina and Tookey 2013). An alternative view cited among practitioners is that retainage negatively impacts contractors' cashflow, thereby acting as a financial constraint and generating a counter-productive increase in procurement costs. Recognition of a potential hidden cost of retainage has driven a downward trend in the maximum retainage provision permitted by several US states (ASA 2018). The efficiency of using retainage to mitigate moral hazard is understudied and is challenging to measure using empirical data due to the nuances and complexity of each construction project. The laboratory enables us to isolate the causal effect of retainage on bids, quality and profits, without the confounds of project-specific factors or alternative mechanisms (e.g., repeated interactions) observed in the real world.

An important consideration in the implementation of retainage is what, in practice, constitutes *substantial completion*. Legal scholars have long recognized the difficulties inherent in determining such a doctrine (Thomas et al. 1995).[6] Failure of trade parties to understand their contractual obligations is one of the leading causes of construction disputes (Arcadis 2020). Litigation is often lengthy to pursue. A costly dispute between the Californian construction contractor, FTR, and the client Rio School District, over the latter's failure to release more than half a million dollars of retainage persisted for many years before being resolved to FTR's favor in 2015.[7] Recently, a construction sector consultation commissioned by the UK Department of Business, Energy & Industrial Strategy (Pye Tait 2017) found that late and non-payment of retainage monies from clients to contractors is commonplace, especially among lower tier suppliers. Whether or not the withholding of retainage monies is justified is often unclear, precisely because of the difficulties in verifying substantial performance.[8] What is clear from the report is that a substantial fraction of clients believe their overall project costs are higher because

---

been rectified." In New Zealand, the 2015 Construction Contracts Amendment Bill provides additional protection for the payment of retainage monies to sub-contractors. In China, retainage applies to pre-specified defects liability periods and at the time of writing enjoy additional financial guarantees from the Agricultural Bank of China.

[5] We take inspiration from Özer and Zheng (2019), who argue that "developing structurally new games and experimental design … are necessary to better understand the role of trust and trustworthiness in decision-making because they are *not* abstract issues" (Section 14.4).

[6] Corbin (1919) captured the essence over a century ago: "What constitutes substantial performance must be determined with reference to the particular facts in each case. The question is always one of degree and its solution must be doubtful in many cases" (p 761).

[7] FTR International, Inc. v. Rio School District. California Court of Appeal, 2015.

[8] We thank an anonymous referee for helping us clarify on this point.

of retainage, that retainage induces the possibility of opportunistic payment behavior and that tender prices reflect this countervailing buyer moral hazard.

The buyer moral hazard problem is accentuated when suppliers make relation-specific investments before a contract is confirmed. In 2016, German automaker Volkswagen (VW) cancelled orders worth 500 million euros with two component suppliers, in the wake of the emissions scandal that forced the company to cut approximately 1 billion euros in costs. The cancellation came too late for the suppliers, however, who had already spent 58 million euros making factory alterations in preparation for the order (Rauwald 2016).

Similar situations can arise in the construction industry. One example is when a project proceeds based on a *letter of intent* (LOI). There are various types of LOI, from a "handshake" agreement stating the intention of parties to trade, to an interim contract which is replaced by a binding contract on expiration. The peril of supplying under an LOI, without a concrete payment schedule, is demonstrated by a notable English contract law case.[9] In 2005, RTS Systems won a competitive tender to supply improved food packaging for the German dairy manufacturer Müller. Work began based on an LOI and Müller paid RTS only 30% of the agreed price up front and a further 40% later on. After expiration of the LOI and repeated deferral in the execution of a binding contract, Müller alleged product defects and refused to pay RTS the remaining 30% of the tender price. A protracted and costly legal battle ensued, centered around the basis for which a contractual agreement existed. The Supreme Court Justice pronounced on judgement day that "the moral of the story is to agree first and to start work later".

Today, the use of LOIs "remains widespread in the construction industry" (Wevill 2015, p 29). The buyer, while not explicitly designating withheld monies as retainage, may withhold a high percentage of the purchase price up front. And while a well-written LOI should allocate reasonable cost estimates for all items to avoid unprotected investments, this is not a legally binding agreement and payment is normally on a *quantum meruit* basis, which can lead to dispute over monies owed (Chappell 2021, p 9).[10] It is not difficult to specify cost structures which expose suppliers to losses and potentially deter participation in requests for bids.

Motivated by the anecdotal evidence, we vary the percentage of the contract price withheld by the buyer until after delivery of the project to shift the relative burdens of trust between buyer and seller. To that end, we compare the performance of procurement contracts in which there is either (i) zero retainage – a fixed-price contract, (ii) a retainage provision set such that suppliers can adjust their bid upwards to compensate for the increased risk of non-payment of monies, or (iii) high retainage in which the buyer pays only a small percentage of the contract

---

[9] RTS Flexible Systems Limited v Molkerei Alois Müller GmbH [2010] 14 UKSC.
[10] See, e.g., EWHC 687 (TCC), [2006] CILL 2348 and BLM Vol.23 No.6.

price up front, and consequently suppliers cannot fully compensate for future production costs incurred by bidding higher. These three arrangements approximate the market conditions (rather than specific parameters) discussed above. That is, there is either a one-sided seller moral hazard problem, a two-sided buyer and seller moral hazard problem, or a one-sided buyer moral hazard problem.

We develop an analytical model that consists of a sealed-bid reverse auction (tender) followed by a bilateral trade interaction. Specifically, a single buyer seeks to procure one unit of an indivisible good (e.g., construction of a new school) from a pool of pre-qualified suppliers. There is a commitment to procure at the lowest price, if at all, and the contract contains a fixed retainage provision. After allocation of the contract, the seller takes costly action to deliver the project and the buyer realizes the project value as a function of the seller's action. The buyer then has discretion over the retainage return decision. In other words, the buyer decides how to allocate retainage monies to compensate the seller for performance delivered.

The retainage-related part of the transaction is predicated on trust, and so game-theoretic arguments based on standard preferences predict that the retainage mechanism generates no quality improvement. Trust between agents (i.e., managers) is an important driver of supply chain success (Cerić 2016). Thus, we develop a model of fair payment norms, in which the buyer may be trustworthy or untrustworthy. We adopt Özer and Zheng's (2019, p 497) definition of trustworthiness as a voluntary behavior "in a way not to take advantage of the trustor's vulnerable position when faced with a self-serving decision that conflicts with the trustor's objective". Whereas an untrustworthy buyer always withholds retainage from the seller, a trustworthy buyer distributes retainage according to some known and exogenous fairness norm. Trust is defined as the seller's belief about interacting with a trustworthy buyer (Herold 2010).

Standard preferences imply that trade efficiency will be low in our first two contracting arrangements, because of the seller moral hazard problem, and nil in our third contracting arrangement, as the market unravels. By contrast, the model of fair payment norms demonstrates that, if there is sufficient trust in the market, then high quality delivery emerges as an equilibrium outcome in anticipation of a positive reciprocal retainage return. Designed appropriately and under sufficient conditions of trust, retainage can mitigate the seller moral hazard problem.

We fit our behavioral model to the data and provide a characterization of the optimal retainage level, given beliefs about the fair reference point in the transaction. By doing so, we contribute to a growing behavioral operations literature addressing how social preferences influence supply chain contracting (Beer et al. 2018, Cui et al. 2007, Hu et al. 2017, Katok and

Pavlov 2013, Loch and Wu 2008).[11] Closely related, but in a non-competitive setting, Davis and Hyndman (2018) find that monetary incentives may complement relational incentives when efficiency of the monetary incentive is high, but crowd out relational incentives when it is low. Also related, Li et al. (2020) observe that a target-based bonus contract induces significantly higher effort by store managers in an inventory context, but that fairness concerns undermine its effectiveness.

Our experimental data offer two primary insights for practitioners in construction and related procurement settings: (i) if trust in the supply chain relationship is strong enough, then retainage is a useful mechanism to mitigate the seller moral hazard problem and substitute for reputation in a fragmented supply chain; (ii) if there exists a subset of untrustworthy buyers in the population, then retainage is liable to inflate tender prices and high retainage structures to deter participation in the contracting process. A further experimental finding is that suppliers sub-optimally adjust their bids in anticipation of uncertain retainage returns and incur losses on relationship-specific investments.[12] Based on these insights, managers and policymakers should consider measures to facilitate trust in the *ex-post* procurement relationship during the *ex-ante* tender process, while letting the price mechanism efficiently allocate the contract. In the conclusion, we return to discuss what said measures might look like.

The present study builds on a well-established experimental economics literature examining trust and trustworthiness in principal-agent settings. Our baseline environment integrates the gift-exchange game of Fehr et al. (1993) into an auction setting. In a typical gift-exchange game, participants are assigned to the role of either buyer or seller and participate in a two stage exchange. First, the buyer sets a price. Second, the seller produces a costly product quality and creates the trade surplus. In the absence of reputational considerations, if there is a preannounced and finite number of repetitions, then the seller should incur the minimum production cost possible and receive the lowest available price. In contrast, experiments of the gift exchange game without competition, typically observed a positive relationship between price and quality, yielding a Pareto improvement relative to the equilibrium prediction and lending support to Akerlof's (1982) gift-exchange hypothesis (Anderhub et al. 2002, Fehr et al. 1997). Reciprocal behavior can be rationalized with theories of social preferences (Bolton and Ockenfels 2000, Fehr and Schmidt 1999). Yet, in price-based auction models where buyers can only transact

---

[11] In related work, Hoppe and Schmitz (2011) observe in an experiment that renegotiable option contracts can help solve the hold-up problem if sellers are endowed with bargaining power. A similar framework has been applied to a supply chain setting by Davis and Leider (2018).

[12] Related auction models in which contract renegotiation leads to lower prices *ex-post* are considered by Waehrer (1995) and Wang (2000). Shachat and Tan (2015) consider an auction-bargaining mechanism in which sellers compete at an English auction to deliver an indivisible good to a buyer. They find in an experiment that re-negotiated prices are below the winning bid although, contrary to the Nash prediction, final and initial prices are positively correlated.

with the lowest-priced seller, competition forces bidding down to minimum cost. High quality is then not a profitable seller strategy. The discontinuity in quality choice as a function of price is a direct result of the incomplete contracting model (Hart 1995).

By designing different price-setting institutions, prior studies have shown that it is possible to improve trade efficiency in auction environments with moral hazard.[13] Most closely related to this paper are Fugger, Katok and Wambach (2019) and Brosig-Koch and Heinrich (2014). These studies restrict attention to the nature of the auction selection rule with fixed-price (i.e., zero retainage) contracts. In a price-based auction, the market is inefficient with low quality production. Fugger, Katok and Wambach show that simply giving buyers the option to select a seller who did not place the lowest bid (called a buyer-determined auction) significantly raises prices and quality levels. Like in our study, interactions are one-shot, no reputation information is available, and buyers must accept vulnerability to loss to incentivize high quality. The authors employ a multi-level cost and quality design, across two different valuation schedules. Their experimental data reveal that buyer-determined auctions yield a robust improvement in cooperation and efficiency and they rationalize this finding using a model of inequity-averse preferences. Brosig-Koch and Heinrich consider a buyer-determined auction in which buyers can condition procurement acceptance decisions on past seller performance. Providing reputation information in this way significantly increases buyer profits and procurement quality, relative to the price-based auction format.

Buyer-determined auctions clearly have an important trust-building role in procurement. They are not, however, always palatable from a regulatory standpoint, not least because they may facilitate bidder collusion (Fugger et al. 2016) or discrimination in the award decision (Verdeaux 2003). We believe that by understanding how and when an alternative real-world incentive mechanism can be used to overcome seller-side moral hazard – and its trade-offs with buyer-side moral hazard – enable us to make a significant contribution on this issue.

The rest of the paper proceeds as follows. In section 2, we outline the setting and derive competing analytical results from the benchmark model and from a model of fair payment norms. In section 3, we derive testable hypotheses and summarize the experimental design. In section 4, we present our experiment results and conduct a formal statistical analysis. In section 5, we estimate the parameters of our behavioral model. In section 6, we conclude by drawing

---

[13] The first experimental analysis of procurement contracts in reverse auctions with moral hazard was conducted by Cox et al. (1996). They compared fixed-price and cost-sharing contracts. Their main finding was a trade-off between budgetary expense and efficiency. Although contracts with a greater cost-sharing element involved less procurement expense, they were also less efficient due to heightened seller moral hazard. Cost-sharing arrangements tend to be more appropriate for complex projects, accompanied by low degrees of design completeness (Bajari and Tadelis 2001). For the procurement of complex goods, a buyer commitment to negotiate with one seller can outperform an auction when there is adverse selection, scope for product improvements and/or costly renegotiation (Herweg and Schmidt 2017).

implications for managerial decision-making and delineate conditions under which the retainage contract is beneficial in procurement interactions.

2. Model and Theory

Consider a one-shot interaction in which a single buyer seeks to procure one unit of an indivisible good from a group of n pre-qualified suppliers, indexed by i. A first-price sealed-bid auction determines selection of the winning supplier (henceforth the seller) and the contract price. Auction participation is voluntary for suppliers, and the buyer can choose to not purchase after observing the contract price. The seller can produce either a high- or low-quality unit, but the setting prohibits quality contingent contracts.

In this setting, the contract price is less binding than usual as purchases are made with retainage provisions. Such a provision includes a retainage proportion ρ, which is a fraction of the contract price withheld from the seller until after unit production.[14] A general interpretation of ρ is as the degree of price flexibility. After production, the buyer observes her valuation and (indirectly) the seller's action. The amount of retainage released to the seller is then at the buyer's discretion. The retainage proportion thus regulates each party's trust burden: at low levels, the buyer possesses limited insurance against low quality production; at high levels, the seller is vulnerable to financially damaging retainage return decisions. The sequence of events is displayed in Figure 1.

In the Bidding Stage, each supplier simultaneously submits his bid, b_i, or chooses not to participate in the auction process. If at least one seller submits a bid, the one submitting the lowest bid wins the auction and the contract price is the winner's bid, $p = min\{b_1, \cdots, b_n\}$. Ties are broken randomly. If no bid is submitted, all parties earn zero and we call this outcome "market unraveling". When an auction succeeds, the profile of bidding-stage actions is announced before the next stage.

In the Procurement Stage, the buyer either accepts the winning bid and pays $(1 - \rho)p$ to the winner, or she rejects it resulting in all parties earning zero. The buyer's Procurement Stage action is $a \in \{a^0, a^1\}$, where a^0 is a rejection and a^1 is an acceptance. A buyer's agreement to purchase initiates a fundamental transformation (Williamson 1985), which describes the transition from an *ex-ante* competitive market in which multiple suppliers can tender their bid, to an *ex-post* bilateral trade relationship between the buyer and seller.

In the Production Stage, the seller chooses to produce either a high- or low-quality unit, $q_i \in \{q^L, q^H\}$. The seller incurs a sunk production cost c^j for quality level j. Production cost schedules are the same across suppliers and this is common knowledge. A seller's cost of

---

[14] For $\rho = 1$, the procurement decision can be thought of as a payment promise (uninformative cheap talk).

producing high quality is strictly greater than his cost of producing low quality, $c^H > c^L > 0$. Likewise, the buyer's valuation of the unit is increasing in quality and given by the expression v^j. Trade is preferred to no trade and surplus is increasing in quality, i.e., $v^H - c^H > v^L - c^L \geq 0$.

Finally, in the Payment Stage, the buyer observes the quality and then selects a proportion, *r,* of the retainage money to return to the seller.

Figure 1. The sequence of events in our procurement model.
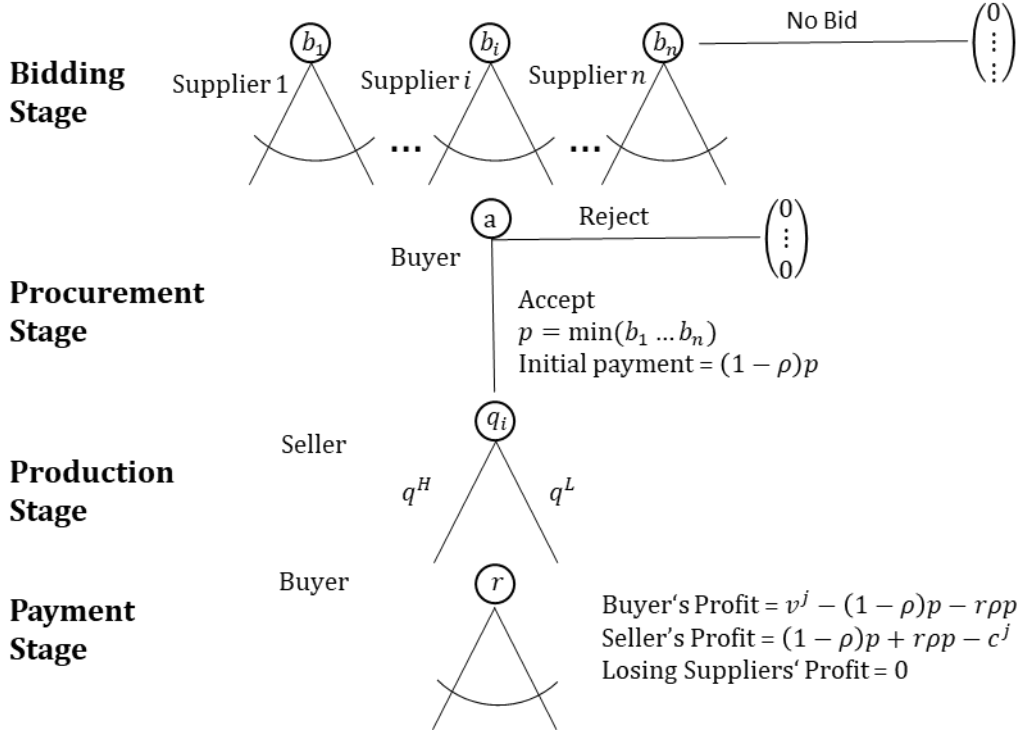
Notes: This is an extensive game tree representation of the strategic interaction. Suppliers move first and either submit a bid b_i at auction or choose not to participate. If a market forms, the buyer can either accept to trade with the lowest bidder and make a guaranteed payment equal to $(1 - \rho)p$, or refuse the transaction. The winning supplier (seller) then selects to produce a high quality, q^H, or low quality, q^L, product and incurs the production cost . The buyer is informed about the product value, v^j, and decides on a discretionary proportion *r* of the retainage money ρp to return to the seller. The seller earns a profit equal to the difference between total payment received and the cost incurred. The buyer earns a profit equal to the difference between value received and the total payment made. Non-trading parties earn zero profit.

Formally, a supplier *i*'s strategy has two components. These are a Bidding Stage action $b_i \in \{[c^L, v^H], b^0\}$ and a quality choice function $q_i(b_i | b_i = p, a = a^1)$. Bids can be submitted from a continuous interval between the seller's minimum production cost and a reserve price, which without loss we set to equal the buyer's maximum valuation for the unit, and b^0 is non-participation. A buyer's strategy also has two components: a procurement decision function, $a(p)$; and a retainage return function, $r(p, q_i | a = a^1)$.

The profits of the transacting buyer and seller are,

Buyer's profit: $\quad\quad \Pi_B = v^j - (1 - \rho)p - r\rho p$

Seller's profit: $\quad\quad \Pi_S = (1 - \rho)p + r\rho p - c^j$

Our primary interest is in whether and under what conditions there exists an equilibrium solution that implements high quality. High quality is implementable if it yields a non-negative expected payoff to the transacting parties (buyer and seller participate voluntarily) and it is incentive compatible (preferred to low quality by the seller).

2.1   Discussion of model assumptions

Before proceeding to the analysis, we provide justification for the assumptions underlying our procurement model. The model assumes: 1. a tender with multiple pre-qualified potential suppliers; 2. a discretionary purchase; 3. a non-recurring transaction; 4. a project for which it is difficult to condition price on quality; 5. a contract with retainage provision. We argue that the combination of these five assumptions strikes a reasonable balance between replicating the target setting of interest and controlling for factors that may otherwise confound the retainage channel of interest. To justify the combination of our assumptions, rather than each assumption in isolation, we consider qualitative evidence from a single locale, the UK construction sector, based on procurement regulations, impact assessments and practitioner guides (Table 1). We discuss the generalizability of our results further in the conclusion.

First, note that the use of pre-qualification questionnaires to short-list capable suppliers before the tender process is commonplace in construction and subject to regulatory mandate for large public-sector projects. Pre-qualification saves time and identifies potential tenderers who are appropriate to carry out the project. Second, most tenders include discretionary clauses that allow the buyer to reject all bids. While such provisions may also extend to the possibility of accepting the non-lowest bid (buyer-determined auction), this assumes that suppliers can be differentiated based on quality *ex-ante*. This is not the case in our model and so allowing for such a choice would confound the channel of interest by allowing price to function as a signal of quality, engendering trust that may mistakenly be attributed to the retainage mechanism (see Fugger, Katok and Wambach 2019).

Third, anecdotal evidence from impact assessments suggests that the construction supply chain is fragmented and proceeds project to project, with many one-time business relationships. Although abstracting from reputational information is restrictive from a practical perspective, it is a strength of the experimental method to control for factors that cannot be well-identified in field data. Fourth, it is difficult for third parties to verify the quality of outcomes due to the complexity of construction projects and the subjectivity around performance obligations. Finally, primary survey data collected by Pye Tait (2017) for the period 2013 to 2016 implies that

retainage provisions are contained in many UK construction contracts. While the total value of retainage held across the sector is uncertain, 75% of contractors and 80% of clients reported experiencing retainage over the last three years; the percentages of their contracts with retainage monies held were 65% and 78%, respectively.

Table 1 – Primary evidence to support the combination of model assumptions.

	Assumption	UK construction sector evidence base
1.	Multiple pre-qualified potential suppliers.	Pre-qualification questionnaires (PQQs) are mandated for public sector construction projects above EU thresholds.[a, b]
2.	Discretionary purchase.	*"Most invitations to tender contain a proviso that the employer does not guarantee to accept the lowest or any tender".*[c]
3.	Non-recurring transaction.	*"While there are some examples of large, repeat and expert clients adopting a genuinely strategic approach to their supply chain, many of the industry's customers in the private and public sector are occasional or contract for a single project. Therefore, it is usually the case that business-to-business relationships between the client and the industry, and through the supply chain, are transactional with decisions taken on an immediate or short term basis."*[d]
4.	Non-verifiability of outcomes.	"*It can be the case that quality is only apparent after construction work has been completed … it is challenging to measure the extent to which late or non-payment of retentions is for justifiable reasons. This is because opinions as to what constitutes 'justifiable' or 'unjustifiable' can differ depending on the contractor or client perspective.*"[e]
5.	Retainage provision.	Central estimates on retainage held across the construction sector in England (as of 2016):[f] • Suppliers with experience of retainage in last 3 years: 75%. • Suppliers with retainage held on current contracts: 65%. • Buyers with experience of retainage in last 3 years: 85%. • Buyers with retainage held on current contracts: 78%.

[a] Procurement policy note 8/16: Standard Selection Questionnaire (SQ) template. https:// gov.uk/government/collections/procurement-policy-notes.
[b] Thresholds laid down by European Public Contracts Directive (2014/24/EU).
[c] Chappell (2021, p 1).
[d] UK Department for Business, Energy and Industrial Strategy. Impact Assessment BEIS021(C)-16-IM, par. 4. https://gov.uk/government/consultations/retention-payments-in-the-construction-industry.
[e] Ibid., paras. 17, 25.
[f] Ibid., paras. 10, 16. The central estimates are uncertain and rely on several assumptions.

2.2 Standard Theory

In a first approach, we assume that the buyer and sellers are expected profit-maximizers. We proceed to solve for the subgame perfect equilibria (SPNE) using backward induction.[15] All proofs are contained in the supplementary materials.

Proposition 1.

A. *For $\rho \leq 1 - \frac{c^L}{v^H}$, there is a unique SPNE in which $b_i^* = \frac{c^L}{(1-\rho)} \; \forall \, i$, $a^* = a^1$, $q_i^* = q^L$ and $r^* = 0$.*

B. *For $\rho > 1 - \frac{c^L}{v^H}$, there is a unique SPNE in which $b_i^* = b^0 \; \forall \, i$, and the market unravels; off the equilibrium path, $a^* = a^1$, $q_i^* = q^L$ and $r^* = 0$.*

Proposition 1 describes how the retainage proportion regulates the burden of trust between buyer and seller. Since the buyer always returns zero retainage, the proportion only influences seller participation and bids. Below a retainage threshold determined by the ratio of a seller's minimum production cost to the buyer's maximum unit valuation, suppliers will participate in the procurement auction, submit a bid that covers their marginal cost and is acceptable to the buyer, and produce low quality; the equilibrium bid in this interval is strictly increasing in the retainage proportion. Above this threshold, no supplier wishes to participate in the auction process – no available bid at least breaks even – and so the market unravels.

Corollary 1: *Under standard preferences, (i) high quality is not an implementable outcome, (ii) equilibrium bids and the probability of market unraveling are non-decreasing in ρ.*

2.3 A Model of Fair Payment Norms

Let us consider an alternative approach in which we depart from standard assumptions on the buyer's preferences. The buyer is one of two types: with probability α, the buyer is trustworthy (T); with probability $(1 - \alpha)$, the buyer is untrustworthy (U). The prior probability α is common knowledge to all agents. We define trustworthiness with respect to the buyer's retainage return for *high quality*. Whereas both types return zero retainage in exchange for low quality, only an untrustworthy buyer returns zero retainage for high quality; a trustworthy buyer can be relied upon to make a fair retainage return for high quality, constrained by the terms of the contract. That is, whereas low quality unambiguously warrants zero retainage return by both types, the buyer moral hazard permits subjectivity in what constitutes fair compensation for high

---

[15] Other Nash equilibria exist at higher prices and associated with non-participation, but since none of these outcomes are compatible with high quality, we ignore them here.

quality. This permits us to incorporate positive reciprocity into the model.[16] Based on these definitions, we use the terms "trustworthiness" and "positive reciprocity" interchangeably.

Models of distributional fairness were originally proposed to explain individual behavior (e.g., Bolton and Ockenfels 2000, Fehr and Schmidt 1999) and typically assume that the fair reference point in a bilateral relationship is the 50-50 surplus split. While this assumption may be appropriate to describe some firm relationships, not all supply chain relationships are created equal. Thus, following Cui et al. (2007), we specify that the fair reference point for a trustworthy buyer is γ multiplied by the seller's profit, or $\gamma = \Pi_B/\Pi_S$.[17] Production costs are included in profit comparisons, based on prior evidence from hold-up experiments that buyers consider sunk costs when making decisions on final surplus divisions (Ellingsen and Johannesson 2004). The parameter γ is exogenous and represents an implicit fairness norm, which in practice may be specific to the industry, locale or transaction.

There are some notable special cases in this model of fair payment norms. The standard theory is captured by $\alpha = 0$. When $\rho = 0$, the model collapses to an extensive-form game of complete information and Proposition 1A applies. In fixed-price contracts, high quality is never an implementable outcome regardless of the degree of trust. From now on we assume that there is price flexibility ($\rho > 0$) and some probability that the buyer is trustworthy ($\alpha > 0$). The model can then be analyzed as an extensive-form game of incomplete information. We apply a perfect Bayesian equilibrium (PBE) solution concept and restrict attention to pure strategy equilibria. This requires that beliefs correspond to the objective probabilities for all equilibrium actions.

Suppose again that suppliers compete anonymously in the Bidding Stage and that in the Procurement Stage, the buyer accepts any bid from which she expects to profit. After the fundamental transformation takes place, the seller is uncertain as to the probability with which the buyer is trustworthy. Trust is defined in our setup as the seller's belief α about interacting with a trustworthy buyer. There is a direct mapping between trust beliefs and the seller's choice to produce high quality. Our approach enables the characterization of threshold trust levels at which high quality may emerge at the population level. Bolton and Ockenfels (2000) provide a similar application of their ERC theory ("the α model") to data from gift-exchange experiments.[18]

---

[16] The model generalizes to the case in which low quality is also rewarded by trustworthy types. We find little evidence to support this assumption in our experiment. The alternate specification introduces a non-monotonicity into the relationship between trust and the incentive compatibility of high quality, which we return to below.

[17] Note that whereas Cui et al. (2007) model disadvantageous inequality aversion, our paper predominantly models advantageous inequality aversion due to the market imbalance.

[18] Bolton and Ockenfels succinctly explain the rationale for such an approach as follows: "much of what we need to know has to do with the thresholds at which behavior deviates from the 'more money is preferred to less' assumption" (p 167).

In the Payment Stage, the buyer returns zero retainage with probability at least $(1 - \alpha)$. An untrustworthy buyer will always return zero ($r_U = 0$). A trustworthy buyer also returns zero on receipt of low quality. If the seller delivers high quality, then a trustworthy buyer returns the retainage proportion necessary to achieve a profit distribution as close as possible to the fair reference point. In this situation, the best-response retainage return function for a trustworthy buyer is as follows:

$$r_T(p, q^H; \gamma) = \frac{v^j + \gamma c^j + (\rho - 1)p(1 + \gamma)}{\rho p(1 + \gamma)}, \tag{1}$$

where the return in (1) is bounded in the unit interval. Thus, conditional on the winning bid, the optimal return for a trustworthy buyer produces either an interior or corner solution. For ease of notation, we define $r_T^j = r_T(p, q^j; \gamma)$. The main insights of this section would be unchanged by assuming that a trustworthy buyer offers a large enough flat reward to the seller for producing high quality. By introducing a fairness norm into the environment, we consider a variable reward and provide some foundation as to how this might be determined in practice.

Corollary 2: *The trustworthy buyer's retainage return is (i) non-decreasing in quality, and (ii) non-increasing in γ.*

In the Production Stage, the seller anticipates the retainage return decision of the buyer and chooses a quality level j to maximize his expected monetary payoff as follows:

$$E[\Pi_S] = (1 - \rho)p + E(r_T^j)\rho p - c^j \ ,$$

where a seller's expectation of the retainage return given his quality and bid is driven by his posterior belief about the buyer's trustworthiness, with $E(r_T^L) = 0$ and $E(r_T^H) = \alpha r_T^H$.

Definition 1: *The breakeven bid associated with quality level j is $\underline{b}^j = \frac{c^j}{1 - \rho(1 - E(r_T^j))}$.*

Intuitively, the higher the seller's trust and/or the trustworthy buyer's retainage return, the lower the bid that a seller can profitably submit. For correct beliefs on α, any bid accompanying q^j below the breakeven level is weakly dominated by a bid equal to $\underline{b}^j$.

Corollary 3: *Equilibrium bids and the probability of market unraveling are (i) non-decreasing in ρ, (ii) non-increasing in α, and (iii) non-decreasing in γ.*

We can now characterize the conditions under which there exists a PBE associated with the production of a high quality unit in our environment.

Proposition 2.

If $\frac{A}{1-C(1-A)} \leq \rho \leq \frac{B}{C}$, where $A = 1 - \frac{c^L}{c^H}, B = 1 - \frac{c^H}{v^H}, C = 1 - \alpha r_T^H$, there exists a PBE in pure strategies in which $b_i^* = \underline{b}^H \; \forall \; i, a^* = a^1, q_i^* = q^H, r_T^* = r_T^H$ and $r_U^* = 0$.

The proposition states that conditional on the prevailing levels of trust and beliefs about fairness, the retainage proportion can be set appropriately to incentivize high quality as an equilibrium strategy and mitigate the seller moral hazard problem. We call this the *implementable retainage interval*. The equilibrium bid is $\underline{b}^H$ and so the buyer appropriates all the gains from trade. The buyer will always find such an outcome profitable and the buyer types will separate in their final retainage return decision per the discussion above. The two inequalities that define this interval ensure that $\underline{b}^H \in [c^L, v^H]$ and $\underline{b}^H < \underline{b}^L$. In other words, the breakeven bid associated with high quality is feasible and below that of low quality.

To understand the bounds on the implementable retainage interval intuitively, note that A is a measure of seller moral hazard, B is a measure of the trade surplus generated by high quality, and C is a measure of the expected retainage return lost if the seller produces a high quality unit. As A approaches one from below, the seller has greater cost incentive to produce low quality. As B approaches one from below, there are greater potential gains from producing a high quality unit. Conditional on the winning bid, high quality is more difficult to implement using retainage (i.e., the interval narrows) the larger the seller moral hazard and the smaller the trade surplus from high quality. Similarly, high quality is easier to implement using retainage the smaller the associated expected loss of retainage.

Corollary 4: *There is a direct and positive correspondence between trust and the size of the implementable retainage interval.*

To demonstrate the trade-offs between bids, participation, and quality, consider an example. Suppose that $n \geq 2$ suppliers compete to win a procurement contract, with $c^L = 0.30$, $c^H = 0.40$, $v^L = 0.35$ and $v^H = 0.80$ (units in tens of thousands). The contract contains a retainage provision, with $\rho = 0.5$ or $\rho = 0.75$. We will test these parameter values in the experiment. Figure 2 presents the spectrum of PBE outcomes in the (α, γ) space that are supported by the model of fair payment norms for each retainage arrangement. Blue circles (green squares) in the figure indicate beliefs for which the seller delivers low quality (high quality) as part of the equilibrium bidding strategy. The number inside the shape indicates the equilibrium bid amount. The empty region in the right panel indicates market unraveling.

In the implementable regions, the equilibrium bid is (weakly) decreasing in α and increasing in γ. Fixing γ and moving horizontally from left to right, high quality outcome is

implementable above a certain trust threshold. Similarly, fixing α and moving vertically downwards, high quality is implementable above a certain fairness threshold, provided α is high enough. The right panel demonstrates that in a high retainage arrangement, the behavioral model can support high quality in addition to market unraveling as an equilibrium outcome.

Figure 2. Parametric example: equilibrium outcomes supported by the behavioral model.
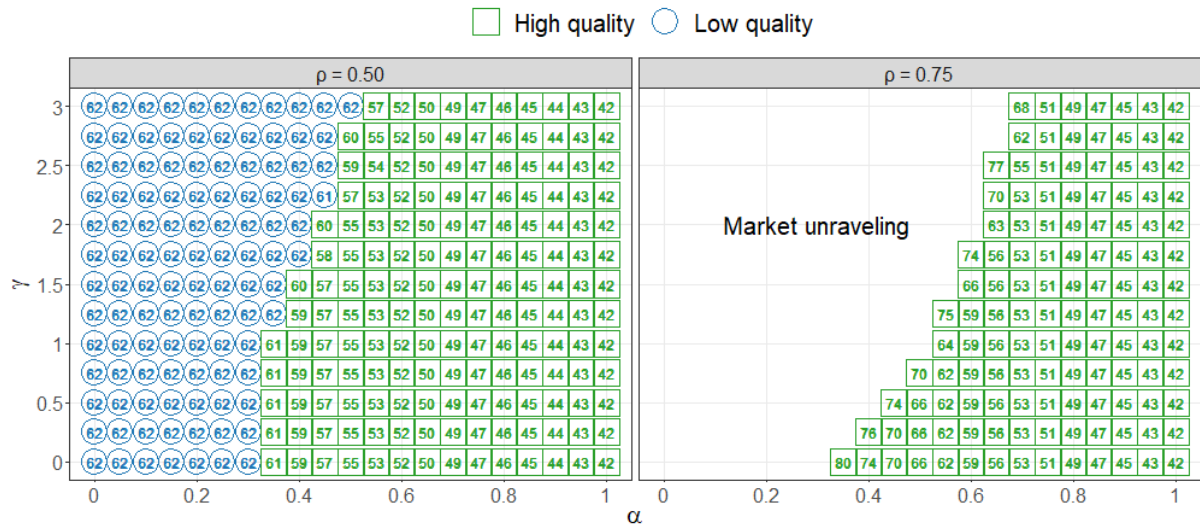
Notes: The figure displays equilibrium seller bidding strategies in (α, γ) space for $\rho = 0.5$ and $\rho = 0.75$. All computations are based on the following cost and valuation parameter values: $c^L = 0.30$, $c^H = 0.40$, $v^L = 0.35$ and $v^H = 0.80$. The numbers inside the shapes are the equilibrium bid amounts associated with the indicated quality level. The empty regions indicate market unraveling. Due to space constraints, we only consider $\gamma \leq 3$. Bids are discrete with minimum increment 0.01, which reflects the experimental implementation.

At this point, it is worth re-emphasizing that the experimental treatments are intended to approximate the market conditions (rather than specific parameters) observed anecdotally. That is, the $\rho = 0.5$ arrangement simulates a two-sided buyer and seller moral hazard problem. Nevertheless, it is interesting to extend the numerical analysis to incorporate a case of lower retainage ($\rho = 0.1$), which is closer to the retainage provisions observed in practice. Note that what constitutes a "low retainage" provision in our model is defined in relation to the cost and valuation parameters: from Proposition 2, a necessary condition for the implementability of high quality is that $c^L/c^H \geq (1 - \rho)$. Thus, to make this extension interesting, we lower the production cost differential, while holding constant the total surplus associated with low and high quality. The exact parameters and equilibrium outcomes are presented in the supplementary materials. The main insight is that a lower retainage provision requires a higher trust threshold for high quality to be implementable. The intuition here is that with less flexibility in the contract price, there is less scope for a trustworthy buyer to allocate profits in a way that rewards the seller for

delivering high quality. Thus, the probability of any such reward must be large enough to produce incentive compatibility in expected payoff terms.

Finally, we remark that if trustworthy buyers were also to reward *low* quality according to some – same or different – fairness norm γ, uncertainty about the buyer's type (i.e., $0 < \alpha < 1$) would become a necessary condition to incentivize sellers to deliver high quality. That is, a non-monotonicity would be introduced into the relationship between trust and the implementable retainage interval. Above a certain threshold, the high probability of encountering a trustworthy buyer would allow the seller to submit a low bid and still be insured against loss in case of producing a low quality unit. In this situation, the lower bound on the implementable retainage interval becomes $A/(D - C(1 - A))$, where $D = 1 - \alpha r_T^L$ is the expected retainage return lost if the seller produces a low quality unit. From Corollary 2, we have $D \geq C$, and so this new lower bound is non-decreasing in α. Therefore, the seller would have an incentive to undercut his competitor in the Bidding stage and produce low quality ($\underline{b}^L < \underline{b}^H$). This is always the case for $\alpha = 1$, i.e., when the seller knows that the buyer is trustworthy.

3. Experiment and Hypotheses

3.1 Experimental Design and Procedures

To isolate the effect of contractual retainage on trade outcomes in the absence of confounds typically observed in the field, we conduct an experiment. In our laboratory setting, we consider the case of two suppliers. We employed a between-subjects design with three treatments that varied the retainage provision, $\rho \in \{0, 0.50, 0.75\}$ (Table 2).

Table 2 – Experimental treatments and parameter values.

Treatment	Retainage Level	Buyer's valuation (q^L, q^H)	Seller's cost (q^L, q^H)
1	$\rho = 0.00$		
2	$\rho = 0.50$	$(35, 80)$	$(30, 40)$
3	$\rho = 0.75$		

The $\rho = 0$ treatment benchmarks previous experiments for which payment of the winning bid in full is binding on the buyer, i.e., a fixed-price contract (cf. *Auction* in Fugger, Katok and Wambach 2019). The $\rho = 0.50$ treatment simulates a setting in which suppliers can fully offset the risk of partial or non-receipt of retainage monies by increasing their bids at auction. By contrast, in the $\rho = 0.75$ treatment, there is no bid available at which a supplier can ensure to avoid a loss during trade. Thus, in this treatment, trade can proceed based on the buyer's intent to compensate the seller for costs incurred. We selected values $\rho = 0.50$ and $\rho = 0.75$ for the

non-zero retainage treatments because they are the easiest retainage provisions for subjects to comprehend within the appropriate intervals from Proposition 1. The valuation and cost parameters associated with high and low quality are also summarized in Table 2. These values were displayed on the computer screens of all subjects. We restrict bids to be integers.

Subjects were randomly allocated to one of the three treatments and no subject participated in more than one session. Each treatment included six independent cohorts. There were three cohorts of the same treatment in every session. Each cohort consisted of three buyers and six sellers, who were randomly matched across 30 procurement interactions.[19] A bespoke algorithm guaranteed that no participant in a cohort played with the same pair of individuals in any two consecutive interactions. A total of 162 human subjects participated in our experiment sessions, which were conducted at the laboratory of a large public university in the United States. Participants were students recruited using web-based recruitment software.

All sessions followed the same protocol. Upon arrival, participants were seated at computer terminals and handed a written copy of the instructions to read in private. Terminals had physical dividers to prevent subjects from seeing the screens of other participants. The instructions were played from an audio recording at the front to ensure the description of the game was common knowledge and delivery consistent across sessions. The task was explained to subjects using a cover story related to the application of interest and the instructions included concrete terms such as "Buyer" and "Seller". This was a deliberate choice to improve subject understanding (see Cooper and Kagel 2003) and increase external validity (see Katok 2017).[20] Participants completed a computerized test of understanding before being assigned to their role as a buyer or seller and matched into their first interaction group. Roles remained fixed throughout. Communication was prohibited and all interactions were anonymous. Own-group feedback was provided between periods. This information remained available in a history table to reinforce the game-theoretic assumption of "perfect recall". At the end of a session, participants answered a non-incentivized questionnaire to elicit demographic information, attitudes to trust and risk. The experimental interface was programmed using oTree software (Chen et al. 2016).

Subjects received monetary incentives for their participation. Each subject was paid his or her summed experiment earnings privately and in cash at the end of a session, in addition to a $5 show-up fee. Payment was made sequentially, with sufficient time intervals between any two subjects to mitigate against the possibility of side-payments. We used a symmetric exchange rate of 20 experiment currency units (ECU) to $1. Average subject earnings were $17.70 in the $\rho = 0$

---

[19] We did not inform subjects of the cohort size, to mitigate the possibility of tacit collusion in what might be considered a small cohort (see Katok 2011 for a discussion).
[20] The instructions were framed in neutral language (see Zizzo 2010). We avoided the term "retainage". Audio recordings are available on request.

treatment, \$22.70 in the $\rho = 0.50$ treatment and \$25.00 in the $\rho = 0.75$ treatment. Sessions lasted 60 to 75 minutes. Each subject received a non-refundable endowment of 7 ECU per period to cover potential losses. Subjects were informed that they would not leave the session with less than the show-up fee.[21] To reinforce the one-shot nature of interactions, we did not inform subjects about cumulative earnings until after the final period.

3.2   Hypotheses

Below we outline the hypotheses to be tested in our experiment, based on the standard theory (*ST*) and the behavioral model of fair payment norms (*BM*).

The first two hypotheses relate to the *ex-ante* competitive market.

Hypothesis 1. Participation. *ST* predicts that the market does not unravel for $\rho = 0$ or $\rho = 0.5$, but that the market strictly unravels for $\rho = 0.75$. *BM* predicts that the market does not unravel for $\rho = 0$ or $\rho = 0.5$ but may unravel under certain parameters for $\rho = 0.75$.

Hypothesis 1 captures the potential anti-competitive effect of a procurement arrangement in which the seller cannot guarantee to cover future production costs. Market unraveling is the unique equilibrium prediction under *ST*. Thus, observing substantial participation rates in $\rho = 0.75$ would offer support for the relevance of *BM*. No set of beliefs on α and γ can sustain non-participation in $\rho = 0$ or $\rho = 0.5$ (see Figure 2) and so market unraveling in these treatments is expected to be negligible.

Hypothesis 2. Prices. *ST* predicts that prices will be rank ordered $p_{\rho=0} < p_{\rho=0.5} < p_{\rho=0.75}$; *BM* predicts that bids will be lowest for $\rho = 0$.

Hypothesis 2 captures the potential inflationary effect of retainage, as reflected in tender prices. *ST* predicts that, conditional on trade, as the retainage proportion increases bidders will submit higher prices. *BM* offers no belief-independent comparative static between the non-zero retainage treatments: in both treatments, equilibrium bids are non-increasing in beliefs on α and non-decreasing in beliefs on γ; conditional on the same beliefs, bids are higher for $\rho = 0.75$ than for $\rho = 0.5$ (see Corollary 3).

The final two hypotheses related to the *ex-post* bilateral trade relationship between the buyer and seller. We define "trade efficiency" as the proportion of trade surplus realized out of

---

[21] In $\rho = 0.75$, limited liability was imposed for two sellers. All results reported below hold if we exclude these two subjects from our analysis. In a pilot experiment, we tested the most extreme seller trust arrangement of 100% retainage. Seller losses, however, became a problem. Summary statistics for this variant are available on request.

the total available given the sellers' bidding decisions. For our parameter values, low (high) quality corresponds to trade efficiency of 0.125 (1.00).

Hypothesis 3. Trade efficiency. *ST* predicts no difference in trade efficiency among treatments; *BM* predicts that trade efficiency is weakly higher for $\rho > 0$ than for $\rho = 0$.

Hypothesis 3 addresses the main research question: can retainage be used to mitigate the moral hazard problem and incentivize high quality? Our analysis suggests that in $\rho = 0.5$ and $\rho = 0.75$, high quality is implementable under *BM* but not under *ST*. In $\rho = 0$, neither model can rationalize high quality as an equilibrium outcome. In $\rho = 0.75$, if we observe trade then *BM* tells us that it is most likely to be efficient. The full set of equilibrium seller strategies under *BM* are provided in Figure 2. Our measure of trade efficiency does not consider the surplus-reducing effect of market unraveling. To this end, we construct an additional "global efficiency" measure defined as the proportion of surplus realized out of total attainable, i.e., if the market always attracts a bid. To what extent the higher likelihood of market unraveling in $\rho = 0.75$ undermines global efficiency, relative to the other retainage arrangements, is an empirical question and so we do not place a formal hypothesis on the trade-off between market unraveling and trade efficiency.

Hypothesis 4. Positive reciprocity. *ST* predicts that the retainage return will be independent of quality; *BM* predicts that the return will be positively correlated with the seller's quality choice.

Hypothesis 4 considers buyer trustworthiness. Prior economic experiments suggest that the existence of reciprocity is robust in the laboratory to the imposition of demanding market institutions (e.g., Fehr and Falk 1999). If a positive correlation is observed between quality and the retainage return in our experiment, this type of behavior would be consistent only with *BM*.

4. Experimental Results

We first outline the main aggregate results in relation to our four experimental hypotheses. We then analyze cohort outcomes over time, to gain greater insight into the market dynamics. We finish this section with an analysis of data at the individual-level, to better understand how retainage influences buyer and seller decision-making.

Table 3 summarizes average seller participation rates, prices and quality, along with buyer acceptance rates and retainage returns. The table also presents summary statistics on market unraveling, efficiency and profits.[22] Consistent with Hypothesis 1, sellers nearly always

---

[22] For each treatment, we have data from 540 matching groups. Since there is no interaction between subjects playing in different cohorts, each cohort is considered a statistically independent observation. We employ two-tailed Signed-Rank tests for one-sample comparisons and Wilcoxon-Mann-Whitney tests for two-sample comparisons, correcting for multiple testing using Holm's (1979) *p*-value adjustment method. We acknowledge the potential *caveat* of arbitrary static correlations within sessions (Fréchette, 2012).

bid in $\rho = 0$ and $\rho = 0.5$. By pairwise comparisons, the percentage of participating bidders is significantly lower (64%) in $\rho = 0.75$ (both p-values = 0.015). This variable seller participation is reflected in different rates of market unraveling. In $\rho = 0.75$, 22% of auctions fail to attract a single bidder. By contrast, no market unravels in the lower retainage arrangements.

Table 3 – Cohort means and standard deviations.

	$\rho = 0$	$\rho = 0.50$	$\rho = 0.75$
Panel A: Decision variables			
Seller participation	0.99	0.98	0.64
	(0.01)	(0.03)	(0.26)
Price	35.1	57.8	57.2
	(1.72)	(5.42)	(8.08)
Buyer acceptance	0.60	0.91	0.98
	(0.20)	(0.08)	(0.01)
High quality	0.06	0.30	0.50
	(0.02)	(0.20)	(0.19)
Retainage return (low)		0.06	0.16
		(0.07)	(0.08)
Retainage return (high)		0.32	0.40
		(0.16)	(0.16)
Panel B: Market outcomes			
Market unraveling	0.00	0.00	0.22
	(0.00)	(0.00)	(0.26)
Trade efficiency	0.11	0.36	0.56
	(0.03)	(0.17)	(0.17)
Global efficiency	0.11	0.36	0.46
	(0.03)	(0.17)	(0.25)
Buyer profit	2.83	15.8	31.4
	(1.07)	(7.65)	(2.49)
Seller profit	4.43	-0.25	-8.75
	(1.54)	(1.60)	(5.71)

Notes: Mean (SD) values for the key parameters in our experiment based on cohort averages. Profit data are per round and exclude the endowment.

Result 1. High retainage structures deter supplier participation in the contracting process.

That sellers in $\rho = 0.75$ still choose to submit a bid in most auctions, despite a high vulnerability to loss due to uncertain retainage return behavior of the buyer, suggests a role for trust in the decision-making process. Thus, *BM* provides a more satisfactory description of sellers' participation behaviors than *ST*. Further evidence that the retainage level shifts the burden of

trust in trade relationships away from the buyer can be inferred from buyers' acceptance behaviors. The acceptance rates are increasing in the retainage level, from 60% in $\rho = 0$ to 91% in $\rho = 0.5$ and 98% in $\rho = 0.75$.

Average contract prices are higher in $\rho = 0.5$ (57.8) and in $\rho = 0.75$ (57.2) than in $\rho = 0$ (35.1). These pairwise differences versus the zero-retainage treatment are highly significant (both p-values < 0.01). We also reject the point prediction of *ST* that prices equal 30 in $\rho = 0$ (p-value = 0.031); we fail to reject an average price of 60 in $\rho = 0.50$ (p-value = 0.563). There is no significant difference in prices between the two non-zero retainage treatments, although prices are more variable when $\rho = 0.75$. Thus, the aggregate price data supports the price differences of *BM* outlined in Hypothesis 2, but rejects the rank ordering of *ST*.

Result 2. Retainage provisions inflate tender prices.

Results 1 and 2 underscore the potential for retainage to have an anti-competitive effect on the procurement process. The flip side is that retainage significantly improves average quality levels. In $\rho = 0$, low quality is chosen by the seller in 93.9% of transactions. We fail to reject the null that trade efficiency attained its Nash equilibrium level of 12.5% (p-value = 0.31). In $\rho = 0.5$, the proportion of transactions associated with high quality is 30.2%, and in $\rho = 0.75$ it is 50.4%. The difference in quality levels between the non-zero retainage treatments is significant at the 10% level (p-value = 0.093) and is consistent with the prediction of *BM* that, conditional on attracting seller bids, trade in the high retainage arrangement is more likely to be of high quality. The quality choice frequencies in these treatments generate significant trade efficiency gains relative to $\rho = 0$ (both p-values < 0.01). As stated in Hypothesis 3, such an improvement can be rationalized only by *BM*.

Result 3a. Retainage mitigates the seller moral hazard problem and increases trade efficiency.

Due to the observed market unraveling, global efficiency in $\rho = 0.75$ is 10 percentage points lower than trade efficiency and is not significantly higher than in $\rho = 0.5$ (p-value = 0.589). To further investigate the trade-off between the seller's quality and participation decisions, in Figure 3 we plot a time series of the market unraveling complement and trade efficiency across the 30 periods in our experiment. The trend differences in trade efficiency are pronounced. In $\rho = 0$, trade efficiency fluctuates about its Nash equilibrium level within the 0-25% interval. In $\rho = 0.5$, trade efficiency fluctuates within the 25-50% interval. In both treatments, the variability of trade efficiency declines over time, while the market never unravels.

21

Figure 3. Decomposition of market unraveling and trade efficiency over time.
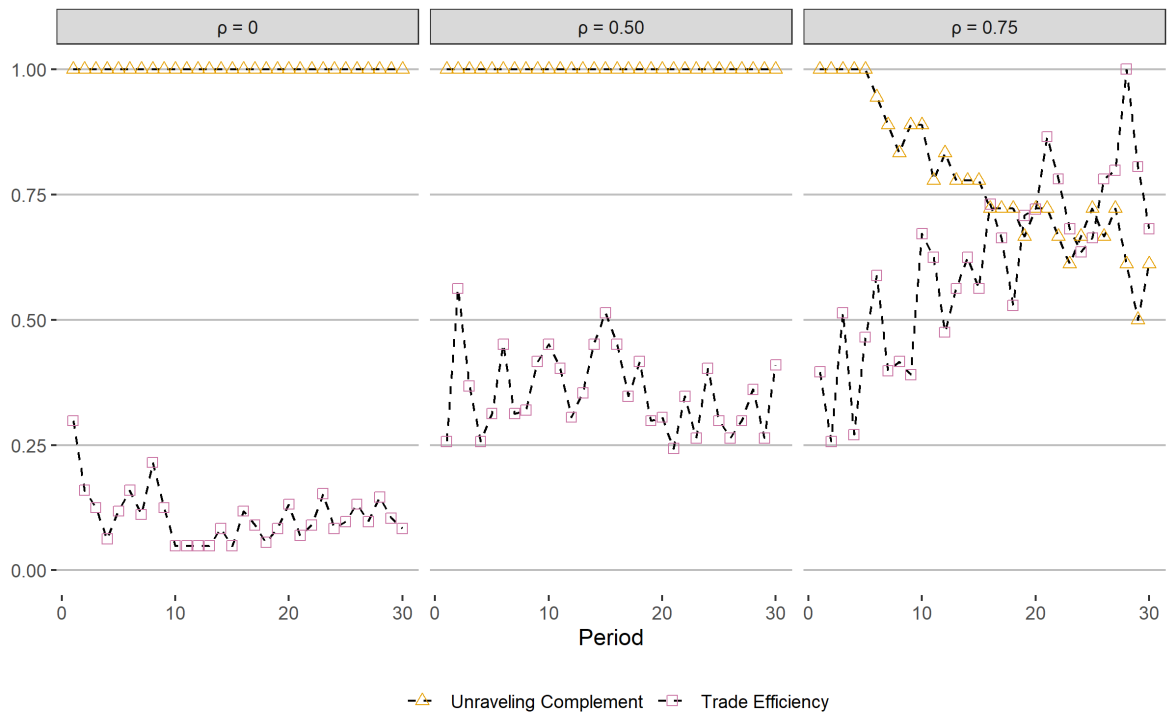
Notes: Based on 18 matching groups per treatment in a period. *Unraveling Complement* is one minus the proportion of auctions which failed to attract a single bidder. *Trade efficiency* is a measure of surplus realized divided by surplus made available.

Most interesting is $\rho = 0.75$, in which trade efficiency begins in the 25-50% interval then trends consistently upwards over time. By the end of the session and notwithstanding an end-game effect, nearly all transactions that take place are of high quality, which as we observed in Figure 2 (right panel) is a direct prediction of *BM*. Meanwhile, after period five, there is a marked fall in the number of auctions attracting at least one bidder. This trend continues into the final period, at which point around half of markets unraveled. Notably, in the second half of the experiment, the difference in global efficiency between $\rho = 0$ and $\rho = 0.75$ is not significant at conventional levels (*p*-value = 0.13). On the other hand, global efficiency remains significantly higher in $\rho = 0.5$ than in $\rho = 0$ after period 15 (*p*-value = 0.015).

**Result 3b.** High retainage structures undermine global efficiency.

In the supplementary materials, we present relative frequencies of trade by price and quality, conditional on at least one bid submitted and buyer acceptance. More than 85% of transactions in $\rho = 0$ are recorded in the 30-39 interval and the quality of these transactions is near-uniformly low. Most transactions in $\rho = 0.5$ fall into the interval 50-69, with low quality most commonly observed at prices above 60 and high quality more likely at prices below 60. In

$\rho = 0.75$, high quality is the majority choice for sellers at prices above 50. These patterns offer further indirect evidence to support *BM*.

Turning to Hypothesis 4, we reject the null of independence between the quality level and the buyer's retainage return predicted by *ST*, in favor of the positive relationship predicted by *BM*. In $\rho = 0.5$, the seller's probability of receiving a non-zero return is 31% after choosing low quality and 61.3% after choosing high quality. Buyers in this treatment return just 6.1% in exchange for low quality and 31.9% of retainage monies in exchange for high quality. In $\rho = 0.75$, where the retainage represents a larger share of the price, return rates are 15.6% and 40.4%, respectively. The premiums paid for high quality in each treatment are significant (both *p*-values = 0.031). Positive reciprocity is evident in the distributions of retainage returns by quality level (see Figure 4, which excludes periods 1 to 10 to mitigate learning effects – see discussion below). Consistent with our definition of an untrustworthy type offering zero return for low quality, and only a trustworthy type offering a positive return for high quality, we observe roughly twice as many zero returns for low quality as for high quality. There is also a greater mass on retainage return proportions above one-quarter for high quality.

Result 4. Trustworthy buyers reciprocate high quality with a more generous retainage return.

Figure 4. Distribution of retainage return proportions by quality level after period 10.
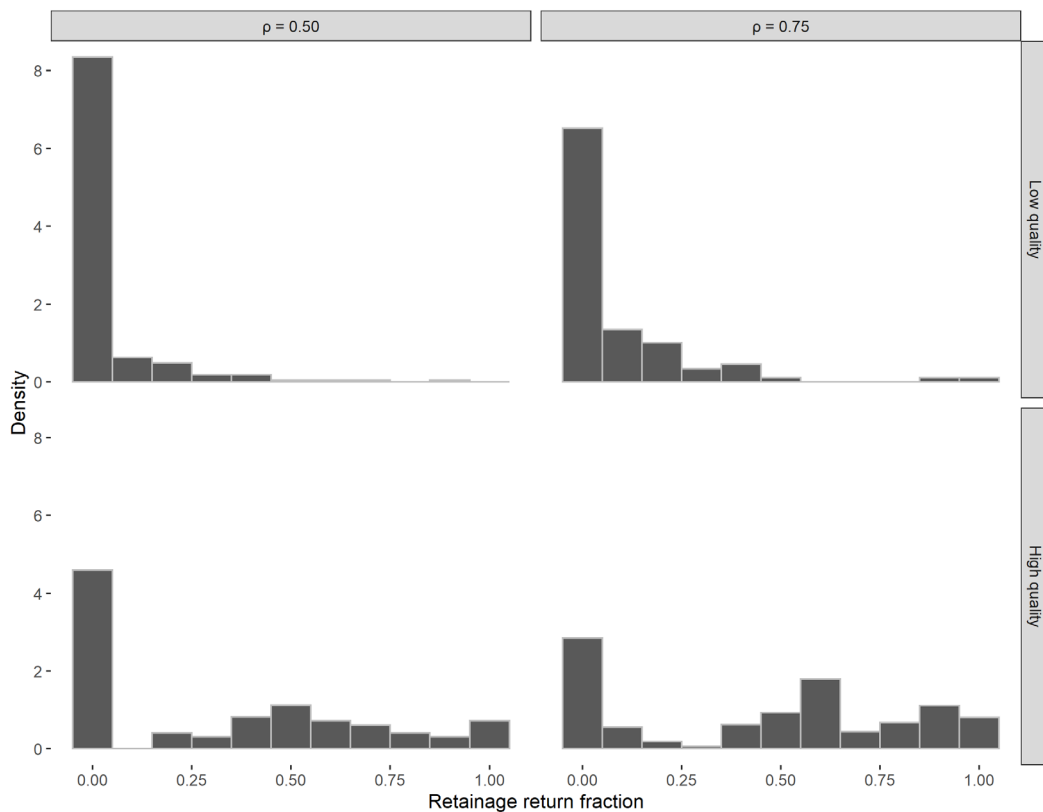
Notes: Histograms of retainage returns in the experimental treatments. The bin width is 0.1.

Retainage has implications for the distribution of profits. In $\rho = 0$, sellers can maintain a positive profit and even earn slightly more (4.43) than buyers (2.83) on average, although this difference is not significant (p-value = 0.156). Buyers gain substantially from the introduction of retainage. In $\rho = 0.5$, sellers earn approximately zero (which would be expected given the market imbalance) and buyers earn 15.8, a significant profit differential (p-value = 0.031). Sellers fare significantly worse in $\rho = 0.75$, incurring an average loss of 8.75, suggesting that their trust in the buyer's willingness to reciprocate is often misplaced.

Our theoretical analysis is equilibrium-based. To check whether subjects' learning in the experiment is an important behavioral factor, we split the dataset into three blocks of 10 periods and conduct a formal analysis of the differences in decision-making over time. The full results of this analysis are provided in the supplementary materials. It suffices to state here that there is some evidence of learning early on, but that behavior appears to converge after period 10, with no significant differences in buyer or seller decisions between the second and third blocks.

To obtain insight into the experiment dynamics, we consider each cohort separately and plot outcomes per group and period, in relation to the price (Figure 5). The top-left cohort exemplifies the disciplining power of competition in $\rho = 0$. After some early adjustments, sellers submit bids in the 30-35 interval and choose low quality. The top-right cohort reveals a different dynamic: sellers attempt to elicit acceptances at prices above the buyer's value of low quality. Buyers reject most such attempts. The middle two panels reveal a reverse temporal bidding trend in $\rho = 0.5$, as sellers learn to adjust their bids upwards to account for retainage monies lost. In the middle-left cohort, sellers produce low quality, marking up their bids proportional to the associated cost. In the middle-right cohort, sellers produce high quality and buyers reward this with a positive return, compensating sellers for the increased delivery cost. The bottom two panels in the figure highlight variable participation in $\rho = 0.75$. In the bottom-left cohort, sellers gradually choose not to participate in request for bids. Two-thirds of markets unravel in this cohort. Sellers in the bottom-right cohort are willing to participate and produce high quality. Where a buyer fails to reward high quality, the seller's downside is larger than in $\rho = 0.5$.

In the supplementary materials, we discuss individual-level regression analyses which were conducted to investigate the determinants of buyer and seller decisions in the *ex-ante* competitive market and the *ex-post* trade relationship. In all three treatments, bids exhibit a positive dependency on the most recently matched competitor's bid and subjects who traded in the prior period learn to submit lower bids. Retainage is most effective at mitigating seller-side moral hazard when the seller has recently received a positive retainage return. In $\rho = 0.75$, we also observe a strong positive time trend which appears to capture the withdrawal of less trusting sellers from the market and suggests that high retainage structures are most effective at inducing high quality once trust is established in the market.

Figure 5. Representative cohort outcomes over time.
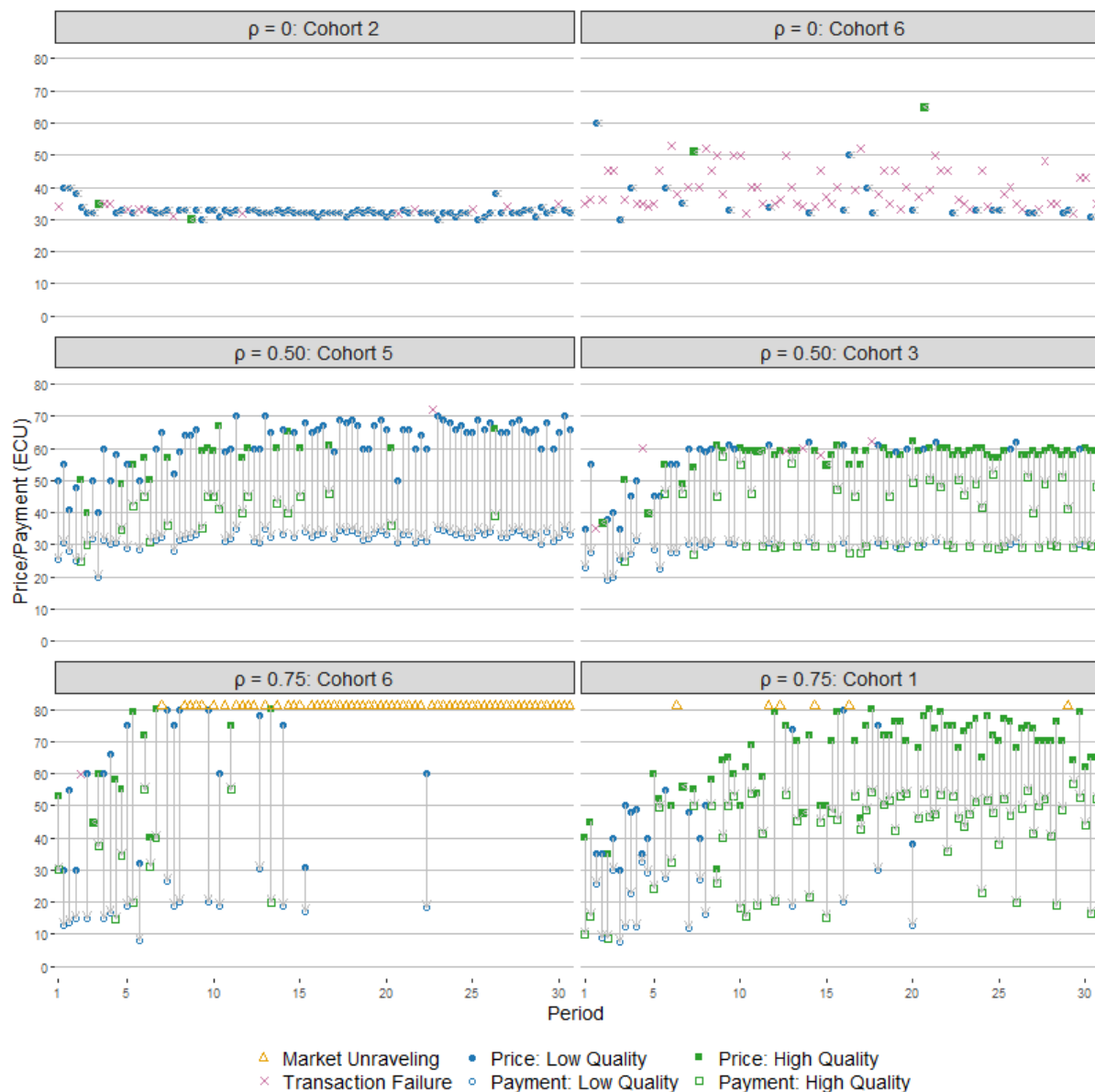
Notes: Panels display procurement outcomes over time in the specified treatment and cohort. An open triangle is an instance of market unraveling, in which neither seller submitted a bid. A cross is a transaction failure, in which a buyer rejected the winning bid. A solid circle is an accepted winning bid at which the seller produced low quality. A solid square is an accepted winning bid at which the seller produced high quality. An open circle (square) are the corresponding total payments in instances where these differ from the winning bid. In such instances, the vertical arrows represent the price-payment differential.

5. Behavioral Model Estimation

The findings in the previous section suggest that our model of fair payment norms can organize the data reasonably well. In the theoretical analysis, we defined the buyer's type on her preferred retainage return for high quality. A crude look at the individual-level return data after period 10 supports the existence of different buyer types in the experiment. Across the pooled retainage data, 18 out of 36 buyers either returned zero or made a partial retainage return in exchange for high quality which resulted in the seller incurring a loss. This increases to 28 out of 36 in exchange for low quality, which suggests that buyers exhibit positive reciprocity in the experiment and provides some empirical justification for our assumption that trustworthy buyers reward only high quality. Moreover, 15 buyers always returned some positive retainage return proportion to the seller in exchange for high quality, with mean returns among these subjects often exceeding 50% of the retainage amount.

5.1   Logit choice framework

We proceed to conduct a maximum likelihood estimation of our behavioral model's parameters in a logit choice framework. The probability that seller i chooses high quality in period t conditional on price p and retainage ρ is:

$$Pr_{it}(q^H) = \frac{e^{\lambda \cdot \Delta U}}{1 + e^{\lambda \cdot \Delta U}}, \tag{2}$$

where $\Delta U = \alpha r_T(p_{it}, q^H; \gamma)\rho p_{it} - (c^H - c^L)$, and λ is inversely related to the level of decision noise or randomness (McKelvey and Palfrey 1995). For the trust parameter, we specify:

$$\alpha = \alpha_0 + \alpha_1 I_{it, \rho=0.75} + \alpha_2 Period + \alpha_3 I_{it, \rho=0.75} \cdot Period \tag{3}$$

where $I_{it, \rho=0.75}$ is a dummy variable for random assignment to $\rho = 0.75$, which captures level and trend differences between our retainage treatments. We restrict $\gamma > 0$ by transforming $\tilde{\gamma} = \ln(\gamma)$ to keep the optimization problem unconstrained. The marginal likelihood of the choice sequence of seller i is given by:

$$L_i = \prod_{t=1}^{T} [Pr_{it}(q^H)]^{q_{it}} [1 - Pr_{it}(q^H)]^{1-q_{it}}, \tag{4}$$

The subject likelihood contributions are then logged and summed to obtain the sample log-likelihood function. We cluster robust standard errors at the subject level. Based on a statistical learning analysis presented in an appendix, we drop observations from the first block of 10

periods to mitigate issues of serial correlation (Davis 2015, p 334).[23] This leaves us with 573 transactions across the two treatments, of which 322 are from $\rho = 0.5$ and 251 are from $\rho = 0.75$. The results of this estimation procedure are presented in columns (1) to (3) of Table 4.

Table 4 – Results of the structural estimation.

	Baseline (1)	BM1 (2)	BM2 (3)	BM3 (4)	BM4 (5)
α					
Constant (α_0)		0.312***	0.392***	0.345***	0.380***
		(0.019)	(0.024)	(0.031)	(0.098)
$\rho = 0.75$ (α_1)			-0.170***	-0.156*	-0.177**
			(0.038)	(0.061)	(0.064)
Period (α_2)			-0.002	-0.004	-0.004
			(0.001)	(0.003)	(0.003)
Period * 0.75 (α_3)			0.005*	0.007+	0.007+
			(0.002)	(0.004)	(0.004)
Female					0.091*
					(0.036)
Age					-0.004+
					(0.002)
γ		1.297***	1.489***	1.222***	1.222***
		(0.291)	(0.195)	(0.000)	(0.000)
λ	0.019	0.452***	0.820***	0.662***	0.694***
	(0.020)	(0.076)	(0.204)	(0.174)	(0.176)
σ_u				0.094**	0.075**
				(0.030)	(0.025)
Trust random effect	No	No	No	Yes	Yes
Control variables[1]	No	No	No	No	Yes
AIC	791.44	661.58	638.42	548.25	509.02
BIC	795.79	674.63	664.53	578.70	581.40
Log Likelihood	-394.72	-327.79	-313.21	-267.12	-237.5
Observations	573	573	573	573	522
(Clusters)	(67)	(67)	(67)	(67)	(62)

Notes: Robust standard errors in parentheses, clustered at the subject level. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$. Models in columns (1) to (3) are estimated using maximum likelihood. Models in columns (4) and (5) are estimated using maximum simulated likelihood.
[1] Additional control variables (not reported in the table) are self-reported risk and trust indices, dummies for female and economics or business major, nationality, and income rank (coefficient estimates available on request). The number of observations in model (5) is lower due to missing responses in the questionnaire.

First, we consider the standard theory (*Baseline*) in which we restrict the behavioral model parameters to equal zero. The standard theory does not explain the data well, with an estimated value for λ not significantly different from zero (all noise). Then, we estimate the constant parameters of our behavioral model (*BM1*). The parameter estimates suggest that on

---

[23] The results are qualitatively unchanged by including the full dataset.

average, sellers believe that there is a one-in-three chance of encountering a trustworthy buyer and that the fair reference point yields the buyer approximately 1.3 times the seller's profit. The 95% confidence interval for our γ estimate includes equal profit-sharing. The α and γ parameter estimates are significantly different from zero (both p-values < 0.001) and the level of decision noise decreases. The behavioral model overwhelmingly outperforms the baseline when comparing the log-likelihood, AIC and BIC values and based on the results from a nested likelihood ratio test ($\chi^2 = 133.86$, p-value < 0.001).

Next, we re-estimate the behavioral model and allow α to vary as a function of the retainage level and over time (*BM2*).[24] In this specification, trust starts off higher in $\rho = 0.5$, with sellers assigning 39% probability of encountering a trustworthy buyer. The results reinforce our earlier observation that trust increases over time in $\rho = 0.75$, with a significant positive interaction between the treatment dummy and the time trend (p-value < 0.01). By contrast, there is some evidence of a fall in trust over time for $\rho = 0.5$, although this effect is statistically weak (p-value = 0.064). Thus, whereas trust is initially lower in the high retainage treatment, the trust differential disappears over time and ends up higher for $\rho = 0.75$. This provides a behavioral explanation for our earlier observations that trade efficiency builds over time in the $\rho = 0.75$ treatment (Figure 3) and average bids converge on similar levels in both non-zero retainage treatments, which requires higher beliefs on α or γ for higher values of ρ (Corollary 3). The fairness parameter γ is estimated at 1.3 to 1.5 times the seller's profit, although the confidence intervals between *BM1* and *BM2* overlap. Our λ estimate in this specification is not significantly below one. We prefer *BM2* based on all statistical comparisons ($\chi^2 = 29.16$, p-value < 0.001).

5.2 Endogenous selection effects

A potential confound in the above analysis are endogenous selection effects which may bias the estimation results. For example, sellers with higher trust, and by implication a greater expectation that high quality will be rewarded with a positive retainage return, may submit a lower bid in the auction. As a result, the estimated α using the quality choice data may be higher than the true α among all sellers. Using binomial and runs tests, we find statistical support for this conjecture (see the supplementary materials for details).

To mitigate the endogeneity issue, we augment equation (3) as follows:

$$\alpha = \alpha_0 + \alpha_1 I_{it,\rho=0.75} + \alpha_2 Period + \alpha_3 I_{it,\rho=0.75} \cdot Period + \beta X_i + u_i, \qquad (5)$$

---

[24] We also consider a specification in which the fair reference point γ is permitted to vary between treatments. We find no significant improvement in predictive power over the nested random effects model and so we maintain the assumption of a common reference point ($\chi^2 = 0.23$, p-value = 0.632).

where X_i is a vector of observed subject-level characteristics elicited in the post-experiment questionnaire, which includes self-reported risk and generalized trust attitudes, dummies for female, economics or business major, nationality, age, and income rank; and u_i is a subject-level random effect term which captures unobserved heterogeneity in trust levels, which we assume to be independent of the error term and of the observable characteristics, with a normal distribution with mean zero and variance σ_u^2. The revised subject likelihood function is given by:

$$L_i = \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} [Pr_{it}(q^H)]^{q_{it}} [1 - Pr_{it}(q^H)]^{1-q_{it}} \right] g(u_i) du_i, \tag{6}$$

where approximation of the integral with respect to the normally distributed variate is performed using maximum simulated likelihood.

The results of this estimation procedure are presented in columns (4) and (5) of Table 4. In model *BM3*, we include the subject-specific random effect but not the control variables in X_i. Accounting for unobserved seller heterogeneity in this way delivers a significant improvement in goodness-of-fit based on both AIC and BIC selection criteria. In Figure 6, we present the posterior random effect u_i for each subject in relation to the frequency of high-quality choices. This captures residual individual-level variability in trust levels. Those sellers who choose high quality more frequently are characterized by a significantly larger unobserved trust component, i.e., a larger positive deviation of the individual-level trust parameter from the population-level trust parameter.[25] In model *BM4*, we include the additional control variables in X_i.[26] The results suggest some further improvement in model fit based on the AIC but not the BIC measure, which reflects the higher penalty for over-fitting in the latter. We also find that trust is significantly higher among female subjects (*p*-value < 0.011) and (statistically weak) evidence that trust is lower among older subjects (*p*-value = 0.083) subjects.

In both models, our parameter estimates of α and γ are robust to – and qualitatively unchanged by – controlling for unobserved and observed individual variability. We caution that the estimates should be interpreted as upper bounds on the sellers' degree of trust and beliefs about fairness in the experiment because these estimates are inferred from the conditional (rather than unconditional) distribution of bids and quality. This is a limitation of our estimation approach. The conditional distribution is, however, the distribution typically observed

---

[25] We also find a small and significant negative correlation between u_i and bids.
[26] Only gender and age are found to have a statistically significant effect on trust and so the other control variables are not discussed further. It may seem surprising that trust and risk attitudes do not influence the α estimate. However, the generalized trust and risk measures elicited in the post-experiment questionnaire are non-incentivized. Moreover, Choi et al. (2020) suggest that *generalized* trust does not significantly predict *specific* trust behaviors. A correlation between attitudinal and behavioral measures of trust typically relies on them measuring the same target and context, which is not the case here.

empirically. It is also the distribution of most practical relevance for the performance of the retainage mechanism: retainage is only withheld from those suppliers who win the contract.

Figure 6. Deviation of seller-level trust from population-level trust as a function of quality.

$n = 67$ RMSE = 0.044
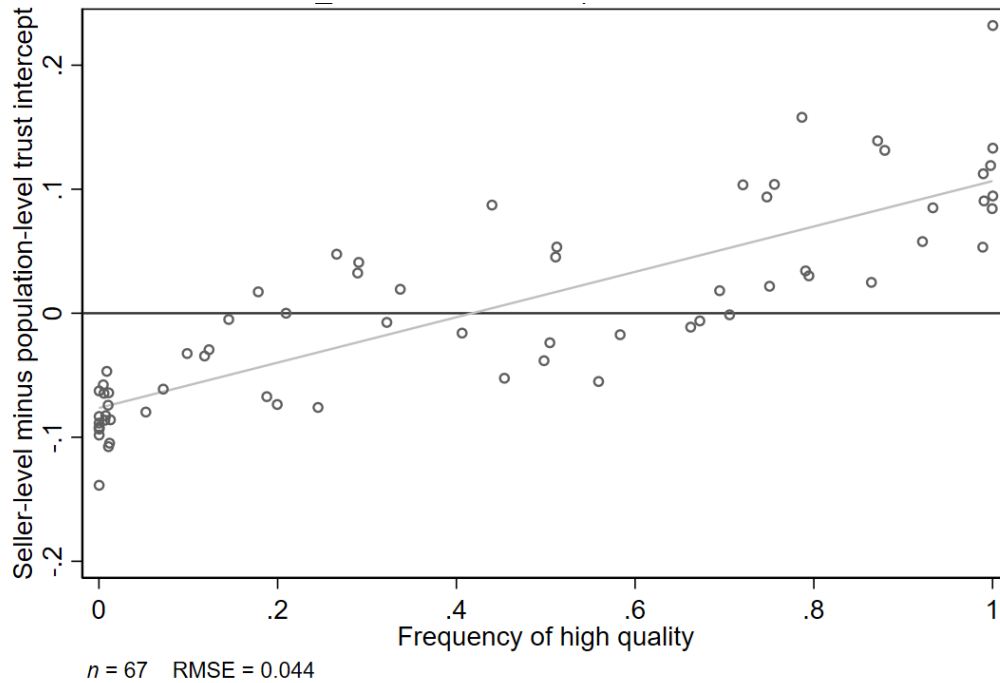
Notes: The scatterplot (with line of best fit) presents subject-specific estimates of the posterior trust random effect u_i against the relative frequency of high quality. Line of best fit: $u_i = -0.076 + 0.183\, q^H$, where the standard error on the constant term is 0.008, the standard error on q^H is 0.014, the model F-stat is 159.91 and the R^2 is 71.1%. Five subjects in a seller role did not win an auction in the last 20 periods and so are excluded.

5.3   Discussion

The behavioral model estimation implies that seller trust in our experiment was between $\alpha = 0.3$ and 0.4, and varied depending on the subject, period, and treatment; the fairness norm, common to both treatments, was estimated at between $\gamma = 1.2$ and 1.5. We benchmark these estimates against prior studies. Using a least absolute deviation approach, Bolton and Ockenfels (2000) estimate $\alpha = 0.50$ in the gift-exchange game of Fehr et al. (1993) and $\alpha = 0.42$ in the trust game of Berg et al. (1995). We obtain very similar estimates for α and γ when using a comparable least absolute deviation approach on the observable profit data.[27] The structure of our *ex-post* bilateral trade relationship is closer to the trust game and our estimate of α reflects this. Our estimate of γ implies that a fair outcome allocates the buyer more than half of the transaction profits. Thus, the market imbalance appears to work in the buyer's favor.

---

[27] See the supplementary materials for details.

It is interesting to infer what our parameter estimates imply for the optimal retainage level, as formulated in the equilibrium analysis. In Figure 7, we depict equilibrium outcomes in the (α, ρ) space, given a fairness norm of $\gamma = 1.35$, the midpoint across our model estimates. The retainage levels that correspond to our three experimental treatments are each marked on the figure with a red square. At $\rho = 0$, low quality is the only implementable outcome, and this was corroborated behaviorally in the experiment. At $\rho = 0.5$, our upper bound estimate of $\alpha = 0.4$ lies on the threshold over which high quality is an implementable seller strategy. The equilibrium bid at this coordinate (59) is very close to the mean price observed in the data (57.8).

Figure 7. Implementable retainage levels supported by the behavioral model estimation.

Notes: The figure displays equilibrium seller bidding strategies in (α, ρ) space for $\gamma = 1.35$, the midpoint estimate from Table 4. All computations are based on the following cost and valuation parameter values: $c^L = 0.30$, $c^H = 0.40$, $v^L = 0.35$ and $v^H = 0.80$. The numbers inside the shapes are the equilibrium bid amounts associated with the indicated quality level. The empty regions indicate market unraveling. The red squares indicate our experimental control on ρ. Bids are discrete with minimum increment 0.01, which reflects the experimental implementation.

To examine the predictive power of our model for bidding strategies in the $\rho = 0.5$ treatment, in Figure 8 we present actual and predicted price and quality based on the individual-level estimates of our behavioral model parameters. The estimated model fits the data closely. At

$\rho = 0.75$, the level of trust required to implement high quality is 0.55, greater than our estimate from the data. Thus, while our behavioral model can explain the increased market unraveling in this treatment, it cannot explain why the remaining sellers prefer to enter the market. There are two possible sources for the failure of our behavioral model to successfully predict price and quality outcomes in the high retainage treatment: (i) sellers' subjective beliefs are biased; or (ii) the model estimates of the sellers' beliefs are biased.

Figure 8. Actual and predicted price and quality levels in the $\rho = 0.5$ treatment.
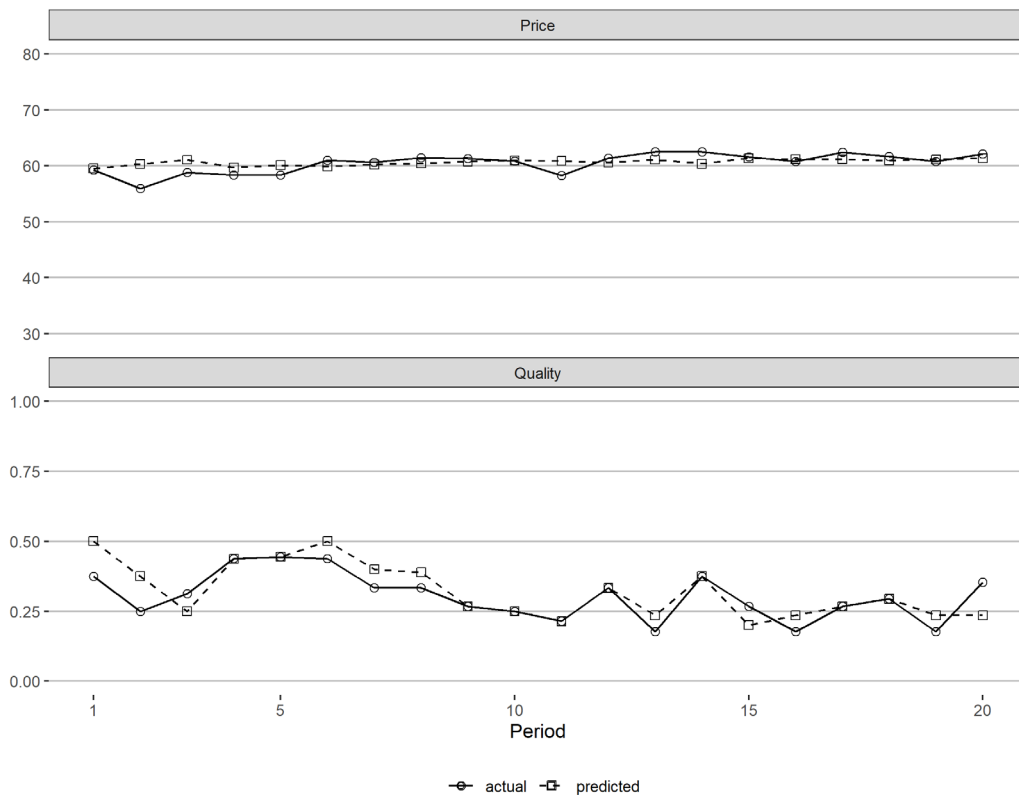
Notes: Actual and predicted seller price and quality decisions over the last twenty periods of the experiment, based on α and γ estimates from model *BM3*.

To provide insight into the source of predictive failure, in Figure 9 we compute predicted retainage return proportions under the assumption that sellers' beliefs about buyer types correspond to objective probabilities. Whereas in $\rho = 0.5$ (top panel), expected retainage returns closely track actual returns, there is a noticeable gap between expectations and actual returns in $\rho = 0.75$. On average, buyers return more than expected by sellers based on the behavioral model estimates, which may have encouraged sellers to bid for the contract more often than an objective assessment of the expected return would warrant. This provides some evidence to suggest that sellers' subjective beliefs are biased in this treatment. We cannot, however, rule out that this difference is due to model bias in our estimates of sellers' beliefs.

32

Figure 9. Actual and predicted retainage return proportions by treatment.
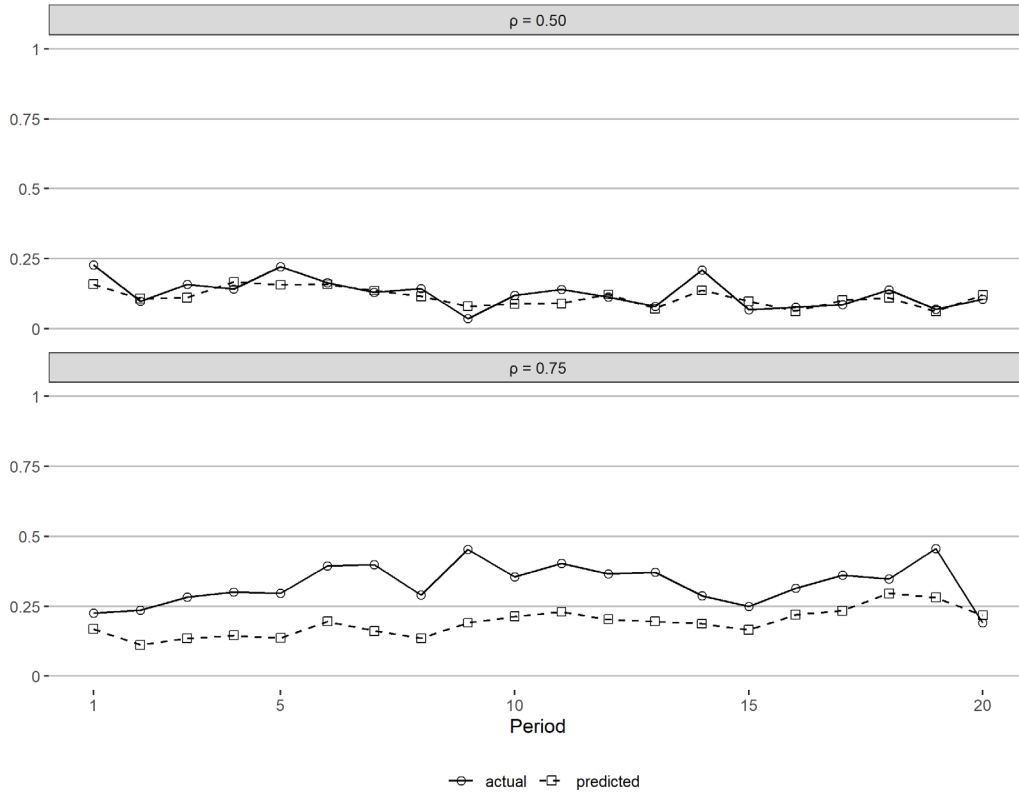
Notes: Actual and predicted buyer retainage return proportions over the last twenty periods of the experiment, based on α and γ estimates from model *BM3*.

6. Concluding Remarks

The managerial literature has devoted substantial attention to how auction institution design can improve procurement efficiency. In competitive procurement, standard principal-agent models show that high product quality equilibria do not exist due to moral hazard. Thus, we see preferences among procurers for repeat purchases from a smaller group of certified suppliers and the establishment of reputational mechanisms. The benefits of procurement mechanisms that permit buyers to incorporate aspects other than price in their decisions have been discussed in detail (see, e.g., Englebrecht-Wiggans et al. 2007). The identification of alternative incentive mechanisms that can reap the benefits of price competition while sustaining cooperation is of considerable managerial value.

We provide experimental evidence that retainage significantly increases the probability of high-quality project delivery relative to the benchmark zero retainage case. This is consistent with the observation of Fugger, Katok and Wambach (2019) that price-based auctions perform poorly with fixed-price contracts when there exists seller-side moral hazard. There is a trade-off

between trade efficiency and transaction fairness, due to the difficulty in sustaining trustworthy buyer payment behavior over time. When sellers are engaged without payment guarantee on production costs, we observe a marked unraveling in markets over time, which softens price competition. This reduces overall procurement efficiency, despite the higher quality of transactions remaining in the market.

We contribute two new managerial insights. First, retainage is an effective incentive mechanism to mitigate the seller-side moral hazard problem so long as suppliers are optimistic enough in their belief about interacting with a trustworthy buyer; in other words, if they have sufficient trust in the procurement relationship. An implication is that procurement managers may benefit from fostering trust with potential suppliers during bidding while letting the price mechanism allocate the contract. Such interventions may be low cost (e.g., increasing transparency of the tender process) and circumvent concerns about collusive behavior that can arise in buyer-determined selection processes (Fugger et al. 2016).

Second, confronted with the real business risk of encountering payment malpractice in the supply chain, retainage structures are liable to inflate tender prices and deter participation in the contracting process. The deductibility of retainage through the supply chain is likely to amplify these effects. Policymakers might consider regulations to increase confidence in the safeguarding and release of retainage monies withheld, for example via a protected deposit scheme. The introduction of an arbitration scheme to resolve payment disputes between buyers and suppliers has recently been debated in the UK parliament.[28] A related problem arises in the context of late payments, of which retainage monies are a case in point; regulatory penalties for reneging on agreed payment terms may have beneficial horizontal effects (see Walker and Hyndman 2021).

Further clarification as to the generalizability of our findings and conceptual simplifications is in order. First, our experimental setup excludes repeated interactions. This assumption is not necessarily violated in practice. We provide qualitative evidence in Section 2 that the UK construction supply chain is fragmented and that one-shot interactions are common. In this scenario, retainage may act as a substitute for reputational capital in the production of high quality. Moreover, there is an important methodological reason to make this assumption: it permits us to isolate the effect of retainage on issues of moral hazard without reputational confounds embodied in field data. Since the availability of reputational information is only likely to mitigate the seller-side moral hazard (see, e.g., Brosig-Koch and Heinrich 2014), our estimate as to the effectiveness of retainage is conservative. That is, reputational information may be used

---

[28] https://services.parliament.uk/bills/2017-19/constructionretentiondepositschemes.html.

to materially complement the retainage contract. Determining whether retainage can be used to dominate reputational capital is beyond the scope of our study.

Second, to ensure saliency of the trade-offs, our three experimental treatments map to somewhat "extreme" trust arrangements. That is, they span the buyer – seller trust space, rather than capturing exact contract parameters. This *caveat* should be kept in mind when interpreting our results. A numerical example demonstrated that similar inferences apply to low retainage provisions, in the range observed empirically in construction contracts. We do not claim that our high retainage treatment replicates a "letter of intent" type agreement as practiced; it does, however, capture the risk of payment at a level below sunk cost, due to legal vagaries of the agreement. Such incomplete contracting arrangements leave the supplier vulnerable to a "hold-up problem" which has broader application in supply chains characterized by relationship-specific investments (Haruvy et al. 2019).

Our behavioral model offers further insights into how to design retainage incentives. If high trust levels have been built up over time in the industry, then engaging suppliers based partially on intent may be an effective strategy to keep projects on schedule. If trust levels are low, agreeing a formal contract with retainage provision before initiating works is preferable. Factors determining trust attitudes and behaviors should be considered here. Exogenous market uncertainty, for example in the buyer's value, can affect beliefs about trustworthiness (Özer et al. 2011). Competition among buyers may reduce trustworthy behaviors (Spiliotopoulou et al. 2016). Özer and Zheng (2019) also emphasize that the target and context of trust matter. Perceptions of trust in a specific managerial network may substitute for generalized trust in supply chain interactions (Choi et al. 2020).

Practitioners should remain cognizant of prevailing industry norms. Trust and trustworthiness may be easier to facilitate in small pools of suppliers engaged in repeated interactions, in which case our estimates for the impact of retainage on transaction efficiency are conservative. Higher trust has been observed in collectivist societies (e.g., in China) when there is the prospect of long-term supply chain relationships, but it may be reduced in cross-border transactions if in-group bias emerges (Özer et al. 2014). Investigating cross-cultural differences in transaction norms would be an interesting avenue for further work. Consideration could also be given to the buyer's payment record in the procurement process. Sellers may trust more in those buyers who have previously proved themselves to be trustworthy payers.

A further implication of our analysis is that, with retainage, the first-order efficiency problem becomes a second-order distributional problem. Future research might therefore also examine the effects of incorporating dispute resolution into the setup. A limitation of our setup is that sellers are endowed with no bargaining power after the auction and cannot dispute the buyer's retainage return decision. In a consumer setting, Andreoni (2018) observes that some

legal enforcement is required to induce buyer trust when goods are sold with a "satisfaction guarantee". Nevertheless, formal litigation is often costly and alternative dispute mechanisms under-developed.

The type of incentives examined in this study are a feature of performance bonuses in principal-agent relationships more generally. Employment contracts often contain a fixed base payment, with the promise of discretionary rewards in the future (see also Fehr et al. 2007, Lee et al. 2018). In certain industries, such as finance or professional sports, these voluntary bonus payments may be the most lucrative part of the compensation package. A similar mechanism is used in the rental housing market by landlords, who withhold deposits as a percentage of the total rental price – typically amounting to one- or two-month's rent – to incentivize tenants to take good care of their property.[29] When there is an excess supply of agents, the motivating effects of performance bonuses are not obvious (MacLeod and Malcomson 1998). We provide evidence from a competitive bidding environment to suggest that implicit incentives can help mitigate the moral hazard problem, under such conditions of market imbalance.

Finally, an open question is the extent to which models of fairness, originally developed to explain personal exchanges, apply to the more impersonal setting of firm decision-making. We find that deriving common thresholds as to the likelihood of a buyer or seller in the market acting in good faith is a useful one. We acknowledge, however, that in repeated interactions the reference point may be path dependent. This possibility is not captured by our static equilibrium model and is a productive avenue for further theoretical and empirical work.

---

[29] Arrangements to pay "half now, half later" are also popular with criminal and gangster organizations, at least as portrayed in the movies.

References

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics, 97*(4), 543-569.

Anderhub, V., Engelmann, D., & Güth, W. (2002). An Experimental Study of the Repeated Trust Game with Incomplete Information. *Journal of Economic Behavior and Organization, 48*(2), 197-216.

Andreoni, J. (2018). Satisfaction Guaranteed: When Moral Hazard Meets Moral Preferences. *American Economic Journal: Microeconomics, 10*(4), 159-189.

Arcadis. (2020). Global Construction Disputes Report. Retrieved from https://www.arcadis.com/en/united-states/our-perspectives/global-construction-disputes-report-2020/.

ASA. (2018). Retainage Laws in the 50 States 2018. Alexandria, VA: Foundation of the American Subcontractors Association, Inc.

Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review, 104*(4), 1288-1319.

Bajari, P., & Tadelis, S. (2001). Incentives versus Transaction Costs: A Theory of Procurement Contracts. *RAND Journal of Economics, 32*(3), 387-407.

Bausman, D. C. (2004). *Retainage practice in the construction industry.* Alexandria, VA: Foundation of the American Subcontractors Association, Inc.

Beer, R., Ahn, H.-S., & Leider, S. (2018). Can trustworthiness in a supply chain be signaled? *Management Science, 64*(9), 3974-3994.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*(1), 166-193.

Brosig-Koch, J., & Heinrich, T. (2014). Reputation and mechanism choice in procurement auctions: An experiment. *Production and Operations Management, 23*(2), 210-220.

Cerić, A. (2016). *Trust in construction projects.* Oxon: Routledge.

Chakraborty, I., Khalil, F., & Lawarree, J. (2021). Competitive Procurement with Ex Post Moral Hazard. *CESifo Working Paper No. 8863, Available at SSRN: https://ssrn.com/abstract=3774670*.

Chakravarty, S., & MacLeod, W. B. (2009). Contracting in the shadow of the law. *RAND Journal of Economics, 40*(3), 533-557.

Chappell, D. (2021). *Construction Contracts: Questions and Answers* (4th ed.). Oxon: Routledge.

Chaturvedi, A. (2021). Excessive Competition and Supplier Non-Performance Risk: Trade-offs in

Reverse Auctions. *Production and Operations Management, Forthcoming*.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88-97.

Chen, L., & Lee, H. L. (2016). Sourcing Under Supplier Responsibility Risk: The Effects of Certification, Audit, and Contingency Payment. *Management Science, 63*(9), 2795-2812.

Choi, E. W., Özer, Ö., & Zheng, Y. (2020). Network Trust and Trust Behaviors Among Executives in Supply Chain Interactions. *Management Science, 66*(12), 5823-5849.

Cooper, D. J., & Kagel, J. H. (2003). The impact of meaningful context on strategic play in signaling games. *Journal of Economic Behavior & Organization, 50*(3), 311-337.

Corbin, A. L. (1919). Conditions in the Law of Contract. *Yale Law Journal, 28*(8), 739-768.

Cox, J. C., Isacc, R. M., Cech, P., & Conn, D. (1996). Moral Hazard and Adverse Selection in Procurement Contracting. *Games and Economic Behavior, 17*(2), 147-176.

Cox, S. C., H.; Udom, K.; Lupton, S. (2011). *Which contract? Choosing the appropriate building contract* (5th ed.). London: RIBA Publishing.

Cui, T. H., Raju, J. S., & Zhang, Z. J. (2007). Fairness and channel coordination. *Management Science, 53*(8), 1303-1314.

Davis, A. M. (2015). An Experimental Investigation of Pull Contracts in Supply Chains. *Production and Operations Management, 24*(2), 325-340.

Davis, A. M., & Hyndman, K. (2018). An Experimental Investigation of Managing Quality Through Monetary and Relational Incentives. *Management Science, 64*(5), 2345-2365.

Davis, A. M., & Leider, S. (2018). Contracts and Capacity Investment in Supply Chains. *Manufacturing & Service Operations Management, 20*(3), 403-421.

Djankov, S., Ghossein, T., Islam, A. M., & Saliola, F. (2017). Public procurement regulation and road quality. *The World Bank*.

Ellingsen, T., & Johannesson, M. (2004). Promises, threats and fairness. *Economic Journal, 114*(495), 397-420.

Emiliani, M., & Stec, D. (2005). Commentary on "Reverse auctions for relationship marketers" by Daly and Nath. *Industrial Marketing Management, 34*(2), 167-171.

Engelbrecht-Wiggans, R., Haruvy, E., & Katok, E. (2007). A comparison of buyer-determined and price-based multiattribute mechanisms. *Marketing Science, 26*(5), 629-641.

European Commission. (2009). International Accounting Standard 11: Construction Contracts. Retrieved from https://www.ucetni-portal.cz/stahnout/ias-11-en_842.pdf.

Fehr, E., & Falk, A. (1999). Wage rigidity in a competitive incomplete contract market. *Journal of Political Economy, 107*(1), 106-134.

Fehr, E., Gachter, S., & Kirchsteiger, G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica, 65*(4), 833-860.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics, 108*(2), 437-459.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and contract design. *Econometrica, 75*(1), 121-154.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817-868.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics, 15*(3), 485-498.

Fugger, N., Gretschko, V., & Pollrich, M. (2019). Sequential procurement with limited commitment. *ZEW - Centre for European Economic Research Discussion Paper No. 19-030*.

Fugger, N., Katok, E., & Wambach, A. (2016). Collusion in dynamic buyer-determined reverse auctions. *Management Science, 62*(2), 518-533.

Fugger, N., Katok, E., & Wambach, A. (2019). Trust in procurement interactions. *Management Science, 65*(11), 5110-5127.

Gretschko, V., & Pollrich, M. (2019). Incomplete contracts in dynamic procurement. *ZEW - Centre for European Economic Research Discussion Paper No. 19-040*.

Guo, R., Lee, H. L., & Swinney, R. (2015). Responsible Sourcing in Supply Chains. *Management Science, 62*(9), 2722-2744.

Hart, O. (1995). *Firms, contracts, and financial structure*. Oxford: Clarendon Press.

Haruvy, E., Katok, E., Ma, Z., & Sethi, S. (2019). Relationship-specific investment and hold-up problems in supply chains: Theory and experiments. *Business Research*, *12*(1), 45-74.

Herold, F. (2010). Contractual incompleteness as a signal of trust. *Games and Economic Behavior, 68*(1), 180-191.

Herweg, F., & Schmidt, K. M. (2017). Auctions versus Negotiations: The Effects of Inefficient Renegotiation. *RAND Journal of Economics, 48*(3), 647-672.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics, 6*(2), 65-70.

Hoppe, E., & Schmitz, P. (2011). Can contracts solve the hold-up problem? Experimental evidence. *Games and Economic Behavior, 73*(1), 186-199.

Hu, S., Wan, Z., Ye, Q., & Chi, W. (2017). Supplier behavior in capacity investment competition: An experimental study. *Production and Operations Management, 26*(2), 273-291.

Katok, E. (2017). Designing and Conducting Laboratory Experiments. In K. Donohue, E. Katok, & S. Leider (Eds.), *The Handbook of Behavioral Operations*. Hoboken, NJ: Wiley.

Katok, E., & Pavlov, V. (2013). Fairness in supply chain contracts: A laboratory study. *Journal of Operations Management, 31*(3), 129-137.

Lee, Y. S., Ribbink, D., & Eckerd, S. (2018). Effectiveness of bonus and penalty incentive contracts in supply chain exchanges: Does national culture matter? *Journal of Operations*

Management, 62, 59-74.

Li, S., Chen, K. Y., & Rong, Y. (2020). The behavioral promise and pitfalls in compensating store managers. *Management Science, 66*(10), 4899-4919.

Lo, W., Lin, C. L., & Yan, M. R. (2007). Contractor's Opportunistic Bidding Behavior and Equilibrium Price Level in the Construction Market. *Journal of Construction Engineering and Management, 133*(6), 409-416.

Loch, C. H., & Wu, Y. (2008). Social preferences and supply chain performance: An experimental study. *Management Science, 54*(11), 1835-1849.

MacLeod, W. B., & Malcomson, J. M. (1998). Motivation and Markets. *American Economic Review, 88*(3), 388-411.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior, 10*(1), 6-38.

Midler, P. (2007). 'Quality Fade': China's Great Business Challenge. *Knowledge@Wharton*, Retrieved from https://knowledge.wharton.upenn.edu/article/quality-fade-chinas-great-business-challenge/.

Nabi Mohamad, A., Ali, G., El-adaway Islam, H., Garza Louis, A., Tichy, S., & Girse, J. (2021). Contractual Guidelines for Substantial Completion under National Design-Build Standard Forms of Contract. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 13*(1), 04520043.

Özer, Ö., & Zheng, Y. (2019). Trust and Trustworthiness. In K. Donohue, E. Katok, & S. Leider (Eds.), *The Handbook of Behavioral Operations* (pp. 489–523). Hoboken, NJ: John Wiley & Sons Inc.

Özer, Ö., Zheng, Y., & Chen, K.-Y. (2011). Trust in forecast information sharing. *Management Science, 57*(6), 1111-1137.

Özer, Ö., Zheng, Y., & Ren, Y. (2014). Trust, Trustworthiness, and Information Sharing in Supply Chains Bridging China and the United States. *Management Science, 60*(10), 2435-2460.

Pye Tait. (2017). Retentions in the construction industry. BEIS Research Paper 17, Department for Business, Energy and Industrial Strategy. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/654399/Retention_Payments_Pye_Tait_report.pdf.

Raina, P., & Tookey, J. (2013). *The perceptions of retentions as held by clients, contractors and subcontractors*. Mimeo.

Rauwald, C. (2016). VW Said to Pay Compensation to Supplier to End Standoff. *Bloomberg* (August 22), https://www.bloomberg.com/news/articles/2016-2008-2023/vw-said-to-pay-compensation-to-supplier-to-end-six-day-standoff.

Shachat, J., & Tan, L. (2015). An experimental investigation of auctions and bargaining in

procurement. *Management Science, 61*(5), 1036-1051.

Spiliotopoulou, E., Donohue, K., & Gürbüz, M. Ç. (2016). Information Reliability in Supply Chains: The Case of Multiple Retailers. *Production and Operations Management, 25*(3), 548-567.

Thomas, H. R., Smith Gary, R., & Cummings Daniel, J. (1995). Have I Reached Substantial Completion? *Journal of Construction Engineering and Management, 121*(1), 121-129.

Verdeaux, J.-J. (2003). Public procurement in the European Union and in the United States: a comparative study. *Public Contract Law Journal, 32*(4), 713-738.

Waehrer, K. (1995). A Model of Auction Contracts with Liquidated Damages. *Journal of Economic Theory, 67*(2), 531-555.

Walker, M. J., & Hyndman, K. B. (2021). A Theoretical and Experimental Investigation into the Welfare Consequences of Late Payments. *Available at SSRN*: https://ssrn.com/abstract=3752360.

Wang, R. (2000). Bidding and Renegotiation in Procurement Auctions. *European Economic Review, 44*(8), 1577-1597.

Wevill, J. (2015). *Law in Practice: The RIBA Legal Handbook* (2nd ed.). London: RIBA.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*: Simon and Schuster.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75-98.