NET Institute*

www.NETinst.org

**An Empirical Study of Online Software Outsourcing:
Signals under Different Contract Regimes**

Mingfeng Lin, Siva Viswanathan and Ritu Agarwal
Robert H. Smith School of Business
University of Maryland

An Empirical Study of Online Software Outsourcing: Signals under Different Contract Regimes[1]

Mingfeng Lin, Siva Viswanathan and Ritu Agarwal

Department of Decision, Operations and Information Technologies
Robert H. Smith School of Business
University of Maryland, College Park

Abstract

We study whether and how contractual arrangements (fixed price vs. time-and-materials contracts) change the effect of reputation, certification, and language characteristics on the chances of winning outsourcing contracts. Using a comprehensive dataset from an online outsourcing marketplace, we model how buyers choose among bidding vendors, and how the effects of these variables change under different contract forms.   Our results show that online reputation is an important predictor of success only for fixed-price contracts, but not significant for times-and-materials contracts. In other words, contract forms can mitigate the typical Matthew Effect associated with online reputation systems. Contrary to popular belief, certifications do not increase the chances of winning regardless of the contract forms. Linguistic features of private communications from the vendor to the buyer also affect the chances of winning, and different dimensions have different effects when contract forms change. Our study is one of the first to study the interaction between contract formats and different signals that vendors can reveal to buyers in the competitive bidding process, and is also one of the first to investigate how texts of private communications affect buyers' contracting decisions.

This version: October 2010

Introduction

Recent developments in internet technologies have transformed many industries, and the market for labor, even offshore labor, is a vivid example (David H. Autor, 2001). While previously only large businesses could outsource (or offshore) software development activities, online markets now allow small buyers as well as small software developers to engage in transactions and even build long-term relationships. The lower barriers to enter the market now allow projects of much smaller sizes to be effectively outsourced online. Meanwhile, despite the rapid growth of these markets that bring together atomistic buyers and sellers, the "virtual" and "small stake" nature of these markets exacerbates issues of information asymmetry and the likelihood of opportunistic behaviors. Our goal in this paper is to better understand the process through which buyers and sellers (workers) are matched in this marketplace, how different variables affect buyers' choice, and how these variables function differently under different contractual regimes.

Asymmetric information in labor markets can be shown through the ex-ante adverse selection problems, worker signaling; as well as ex-post shirking or moral hazard problems. When a market like this becomes virtualized, one great advantage is that buyers and sellers of labor can engage in transactions that are much smaller in scale, which would not have been economically justifiable otherwise. There is a much larger potential for gains from trade in the electronic marketplace. Meanwhile, all the information asymmetry described in the last paragraph still applies. In fact, buyers and sellers are faced with many additional issues when transacting online:

(1) Because some jobs are much smaller, employers will have less incentive to conduct extensive screening prior to contracting; and if things don't work out, they will have less incentive to pursue recourse, seek arbitration, and enforce the contract. Opportunistic behavior could go unpunished and may even be encouraged.

(2) The "virtual" nature of the marketplace: employers usually do not know the real identity behind the worker's online ID. And, because this is a global marketplace, buyers are unlikely to find those outside the US.

To address these issues, almost all online labor markets (such as elance, freelancer or oDesk) carry certain functions to improve transaction efficiencies, such as identity verification, escrow, arbitration, skill certifications, and reputation systems such as rating. Effectiveness of these systems in online labor markets, however, are often assumed rather than tested empirically. Systematic studies in this regard will be highly valuable for practitioners.

From a research point of view, a significant number of academic studies have been done on reputation systems in online product markets such as eBay. Meanwhile, there have been many studies on contract formats, especially in the context of outsourcing. However, there has been little confluence between these two streams of literature. Our goal in this paper is to take a first step in that direction. We empirically examine how buyers' choice of vendors is associated with characteristics of the vendors at the time of their bids, including their reputation, certification and

textual communications; and more importantly, how different contract forms change the effect of these variables.

The problem addressed in the extant literature is mostly about how buyers decide which contract format to use, given that a buyer already knows which vendor to conduct business with. Much less is known about how buyers and sellers came into contact in the first place-- the process through which the buyer or the seller made the choice of a potential trading partner. One reason is that it is very difficult to obtain data on unsuccessful attempts from vendors. Our paper attempts to fulfill this gap in the literature by examining variables that affects buyers' choice of sellers, especially how their effects differ under different contractual mechanisms -- Fixed Price contracts, and Time-and-Materials contracts. These include: (1) The reputation of the seller reflecting his or her prior experience with other trading partners; (2) Third-party certification; and (3) Content and sentiment of how the seller communicates with the buyer. While the literature has separately documented the significance of variables (1) and (2) in product markets and outsourcing, to our knowledge there has been no study so far on how their effects differ under different contract mechanisms. More interestingly, not only are we able to capture the content of private communications between buyers and sellers, we also have the content of communications from other vendors who unsuccessfully attempted to win the contracts. Textual analyses of these messages reveal interesting semantic factors that have been largely missing in prior studies of contracts and the formation of transactional ties. Furthermore, we have access to virtually all information that the buyers have when they choose among the sellers, reducing the plausibility of alternative explanations related to variables unobservable to researchers.

To summarize, the main research questions that we address in the paper is:

(1) What variables affect a buyer's choice among potential sellers; and

(2) How do the effects of the following variables change under different contracting mechanisms?

　　(a) Reputation mechanism reflecting a vendors' past performance;

　　(b) Third-party certification that can potentially serve as a signal; and

　　(c) Content and sentiment of how vendors communicate with a potential buyer.

These variables (among others) are identified from our review of the literature on reputation systems, contracts, and outsourcing. Our working hypothesis is that, not only should these factors play a role in the buyer's decision process; their effects should differ under different contractual arrangements. While the choice of contract mechanisms is not new, its interaction effects with reputation systems, semantics and certifications are much less investigated.

The rest of the paper is organized as follows. Section 2 situates our study in the broader literature in labor, contracts, outsourcing, reputation and certification. Section 3 describes the context of our study, followed by a discussion of the data collection and the variables used. Section 5 provides the empirical model and results. Implications of the study are in Section 6, and Section 7 discusses limitations and future studies related to this paper. Section 8 concludes.

Literature Review

Our paper is situated at the intersection of reputation systems, certification, and contract formats, especially in the context of outsourcing. Here we make no attempt to be exhaustive in our review of such enormous literature, but rather, we focus on studies that most directly relates to our own.

Outsourcing has attracted significant interest in many disciplines, including economics and information systems (Christine Koh et al., 2004, Natalia Levina and Jeanne W. Ross, 2003, Hüseyin Tanriverdi et al., 2007). Two most popular forms of contracting are Fixed Price (FP) Contracts and Time-and-Materials (TM) Contracts. Fixed Price contracts specify a fixed price for an outsourcing project, and the vendor will be paid the agreed amount upon satisfactory delivery of projects. The risk is on the vendors: if they underestimate the cost of development, they cannot charge a higher price later. Time-and-Materials contract, on the other hand, is more flexible and shifts the risk to the buyer. Vendors are paid by time and the cost incurred, instead of the pre-specified amount.

The distinction between Fixed Price and Time-and-Materials contracts has been subject to many theoretical and empirical studies. Gopal, Sivaramakrishnan et al. (2003)investigate the determinants of contract choice, and further relate the choice to project outcomes using data from vendors located in India. Hasija, Pinker and Shumsky (2008) employ data from an outsourcing vendor to investigate the effect of different combinations of contract features. Through content analyses of actual contracts, Chen and Bharadwaj (2009) find that contract format has a moderating effect on the relationship between contract provisions and transactional characteristics. The stylized models of Dey, Fan and Zhang(2010) suggest that Fixed-Price

contracts are better for simple outsourcing projects, while Time-and-Materials contracts are better suited for complex ones. These results are echoed in Bajari and Tadelis (2001). These results indicate that in our model of vendor selection, project complexity should be accounted for.

A number of studies have also been conducted on the control and enforcement issues in outsourcing. For instance, Kirsch (1997) proposes that a portfolio of control modes can be adapted to outsourcing. Rustagi, King and Kirsch (2008) study variables that lead to the use of formal controls. Meanwhile, one of the classical issues in outsourcing is the hold-up problem where the party that makes buyer-specific investments will be at a disadvantage during negotiations. Susarla, Subramanyam and Karhade (2010) study IT outsourcing service contracts and finds that contract extensiveness could mitigate the hold-up problem, but this is moderated by the complexity of tasks.

Reputation is another important subject widely studied in outsourcing. Vendor reputation has been linked to contractual performance (Abhijit V. Banerjee and Esther Duflo, 2000, Tracy R. Lewis, 1986), since the concern for reputation can potentially "outweigh the temptation to renege on a given contract" (Tereza Tykvová, 2007). Jensen and Roy(2008) model the choice a trading partner as a two-stage process, in which reputation helps to decide among a bracket of alternatives.

While the issue of reputation systems in electronic commerce have been extensively studied in the context of product exchanges (such as those on eBay), there has been relatively little empirical study of reputation in the outsourcing context. One possible reason is that firms

rarely share their outsourcing experience with others, and there is also no central platform for them to do so even if they wish. By contrast, online outsourcing markets often extensively use such reputation systems to document a vendor's performance, therefore they provide an ideal context to study the use of reputation systems in the choice process of outsourcing clients.

One of our working hypotheses is that a vendor's reputation should have a much smaller effect on the choice of buyers under Time-and-Materials contracts. A typical issue in the traditional Time-and-Materials contract is that the effort level of vendors cannot be easily monitored or verified. This restriction has been significantly changed in the online outsourcing marketplace because of new technologies that allow buyers to effectively monitor the effort level of vendors. As buyers need to approve the billing hours submitted by the vendors, they can positively evaluate vendors' efforts if necessary. This makes it possible for them to cautiously take some risks and conduct business with vendors who have lesser experience on the marketplace, but are able to complete the task at lower costs. If they do turn out to be shirking, the buyers will be able to terminate the account instead of having to wait until the deadline under a FP contract, thereby reducing the loss.

Another topic that has received significant research is the role of certifications, especially those from third-parties. One such study in the context of outsourcing is Gopal and Gao (2009), who studied the effect of ISO certification on outsourcing vendors. Similar to ISO certifications, the online outsourcing marketplace provides links to a third-party certification website that exams the skill set of sellers on different subjects. When vendors pass these exams, an icon will be displayed next to their ID and prominently displayed to buyers when the vendor places a bid.

However, it should be noted that the threshold of such certification in this market is relatively low: Exams are available online, and vendors can take exams multiple times until they pass. Hence, the effect of such certification may not be practically significant, especially under pay-for-time contracts.

We further explore the role of textual communications in the buyers' decision process of choosing a vendor, focusing on their differing effects under different contract mechanisms. An analysis of language features used by vendors can not only illustrate the decision process of buyers, but also be very useful for vendors who attempt to expand their business online. Prior empirical studies often do not have information about alternative vendors who failed to secure the contract, yet the online market is able to archive all such communications. A study of the language used can be especially valuable to outsourcing vendors. Much like entrepreneurs pitching their ideas at venture capitalists, these vendors only have a limited opportunity to convince a potential buyer. Moreover, it is very likely that some language features are important for pay-for-time contracts, while others are more important for pay-for-deliverable contracts. We use linguistic analysis software to detect features of the vendors' communications, and incorporate them in our econometrical analyses.

Data

In this section we describe major features of the online labor outsourcing marketplace where we gathered our data, the "natural experiment" where a new contracting mechanism was

introduced, the construction of sample, and the derivation of various variables used in our main statistical models.

Context

Data for our study are provided by one of the leading online software outsourcing marketplaces. This marketplace is headquartered in the US, but buyers and sellers of the market come from all over the world.

(1) Overview

The primary dataset for this research is provided by a leading online software outsourcing market. This proprietary dataset contains extensive information about software buyers, vendors, their transactions, communications, mutual ratings and other information from October 2001 to October 2010 (Our analyses will not use the full sample, however, as we will discuss shortly). This is a marketplace of over 250,000 software developers and more than 120,000 software buyers from around the world. These are typically small software development projects, mostly ranging between $150 and $300. Prior to September 2009, all projects are "fixed-price contracts": coders are paid when they deliver satisfactory products according to buyer requirements.

(2) A Typical Process of Transaction

We will first briefly describe a typical process through which buyers and sellers are matched in this marketplace.

Buyers and sellers first sign up with an email address. Before they enter any contracts however, the website will verify their identity. For instance, US buyers are verified by phone, credit card, and driver's license information. Once the verification is complete, buyers post "request for bids" on the site. A typical request includes a title, a summary of requirements, and a rough budget (e.g. maximum $500).

Software vendors (sellers) can browse the requests, search for keywords, and they can be notified of new projects should they choose to receive such alert emails from the site. When they find a project of interest, they can post a bid, which is the amount they will charge for the delivered product. Along with the bid, they can (optionally) submit a text message trying to convince the buyer that they are a good candidate. This is much similar to the "pitch" that entrepreneurs make to venture capitalists when they seek funding. It should be noted that these are sealed bid auctions, in the sense that only the buyer can see the bids placed; peer sellers cannot observe each other's bids.

Buyers can communicate with the sellers, and then choose a seller to work with by accepting his or her bid. Buyers can choose any bid he or she wants, and lower price bids do not necessarily win. This is an important feature of the online labor market that distinguishes itself from websites such as eBay.

Once the bid is accepted, the buyer will first pay the amount of the bid by credit card or electronic check into an escrow account of the site. Then the site sends a notification to the seller that they can start working on the project. A contract is thus created.

The buyer and seller communicate with each other through the website to clarify requirements and other details. When the seller finishes the project, he or she uploads programs to the site, and the buyer can download it to test whether the requirements are met. If so, buyers accept the project as 100% complete, and the funds are released from the escrow account to the seller. Buyers and sellers can then rate each other on a scale of 1 to 10, and also provide comments on each other. These ratings become public information on their profiles, and form the "reputation" system on the marketplace.

The website deducts a percentage of fees from the escrowed amount when it is released to the vendor. These fees are not only for the provision of an infrastructure and possible arbitrations (see next paragraph), but also taking care of paperwork related to taxes and other issues involved in paying another person, especially those in a foreign country.

If the project is not completed due to any reason, it typically enters arbitration. The arbitrator is a staff member of the site and the arbitration process can be initiated by either the buyer or the seller. Arbitrator will review all communications on site, including the original requirements, and will contact both parties. Offline communications, if any, are not considered in the process. If either party fails to respond, he or she receives a low rating, and loses.

(3) A "Natural Experiment": Emergence of Pay-for-Time Micro Contracts

Up until September of 2009, all projects in this online software outsourcing marketplace are "pay-by-deliverables"; that is, the buyer and seller agree on the requirements of the project at the beginning, and the cost of the project is fixed. This is comparable to the "fixed price" contracts in outsourcing, where the buyer's obligation is limited a priori and the burden of risks falls on the vendors. Since September 2009, the website started to allow buyers and sellers to enter "pay-for-time" contracts. In this arrangement, sellers are merely paid by the number of hours they work on the project without guaranteeing the outcome, and the buyer can terminate the contract at any time. If mutually agreed, the contract can extend at the agreed hourly rate. This is made possible only because the website created an application to allow the buyers to closely monitor the efforts of the sellers. When the seller starts working on a pay-for-time project, they will log into the monitoring software, which will take screenshots, random keystrokes as well as webcam pictures at given intervals. The buyer can also manually take additional pictures or keystroke recordings as required by the contract. These records are kept for the purpose of arbitration: if the buyer believes that the coder has inflated the number of hours, the arbitrator can use these recordings as evidence.

This change provides an exogenous "shock" to the marketplace that we exploit to study whether and how changing contract formats will change the effect of vendor characteristics on their chances of winning.

(4) Reputation Systems

Much like eBay, the website has developed extensive reputation systems for the vendors so as to assist buyer's choice among candidates. When buyers and sellers first sign up, they have no ratings. When a project is completely successfully, buyers and sellers can both rate the other party. The rating has a numeric part that ranges from 1 to 10 stars, as well as a textual part that they can comment on the rating. Information about number of ratings that the seller has received up to the time of the bids, as well as the average of those rating, are displayed prominently to the buyer when they look at the list of bids received.

(5) Certifications

The website also works with a third party provider who allows vendors to take exams online on different subjects. Vendors need to pay about $50 US Dollars to take an exam. If they fail an exam, they can wait a few days to re-take it. After they have passed an exam, the website will display an "Expert" icon next to their bids. It should be noted that the icon itself does not describe the subject of the exams taken, just merely the fact that the vendor has taken at least one exam. Buyers need to click through the icon and find out the subject of the exams.

Sample Construction

To understand the factors affecting buyers' choice of vendors, we need to rule out alternative explanations. We therefore construct two samples of bid requests that consummated

in actual contracts (i.e. buyers matched to a vendor), all bids placed in those auctions, and information about buyers, sellers, project descriptions and communications. The first sample includes all pay-for-time (c.f. Time and Materials) contracts that were reached between buyers and vendors between September 2009 and September 2010. The second sample is a random sample of pay-for-deliverables (c.f. Fixed Price) contracts. By conducting analyses on these two samples -- both jointly and separately -- we will be able to test whether contract format makes a different in the effect of various variables on the probability that a bid is chosen. This however, also depends on careful construction of the two samples themselves.

Pay-for-deliverable contracts are the original format used on the website, and it is also what most users are accustomed with. Hence, even after pay-for-time contracts are made available; many users (buyers and vendors) continue to use pay-for-deliverable contracts. This is especially true among buyer-vendor pairs that already have repeated transactions. In our sample, almost all pay-for-time contracts in our sample (one year after the implementation of the new method) are between dyads that do not previously have engaged in transactions. Hence, for the first sub-sample (PFT contracts), we remove the contracts between parties with prior experience between them, so that all buyers face comparable degrees of uncertainty when they choose among the vendors.

While we have data on bid requests (auctions) that do not have any winning bids at all, we decide not to incorporate them in the current analyses. This is because a contract failing to consummate may be due to unrealistic requirements of the buyers, instead of any characteristics of the vendors. Hence, by focusing on contracts that are actually agreed on, we eliminate the

confounding factors from the buyer's side. And as mentioned earlier, we removed contracts between buyers and vendors who had prior experience with each other because the degree of uncertainty between them is significantly lower.

We now turn to the construction of the second sample - auctions and bids related to Pay-for-Deliverables (PFD) contracts.

To match the construction of the PFT sample, we first removed PFD auctions that did not result in an actual contract for the same reason. We then focus on only the PFD contracts in the three months prior to the introduction of PFT contracts (June - August 2009). This is to ensure that vendors do not face resource constraints and have to choose between PFD and PFT auctions that are posted at the same period of time. For consistency, we also eliminated from this sample dyads that had prior experience working with each other. From this reduced dataset, we take a random sample of the same number of auctions as the PFT sub-sample described above.

Then, for both PFT and PFD sub-samples, we extracted all bids related to those auctions, all comments posted by the vendors in their "first contact" with the potential buyer, information about the vendors at the time of their bids, and requirement documentations related to these projects. The textual comments were then submitted to two programs for analyses. The first one is GNU Aspell[2] which contains a dictionary of English words. We compare each word contained in the vendors' emails, and count the number of typos contained therein: it is possible that more typos may reflect difficulty of communication, which could affect the chances that the bids are accepted. The second one is LIWC (Linguistic Inquiry and Word Count), a specialized program

---

[2] http://aspell.net/

for computational linguistic analyses that generate variables characterizing the texts on a large number of dimensions. We go through all these categories and identify the ones that have justifiable possible effects on the bids being chosen. These and other variables are explained in greater details in the next section.

Derivation of Variables

The level of analysis in our model is each bid: we study how buyer's characteristics and their comment's characteristics are associated with the outcome of their bids (winning or losing). In this section we describe the construction of the major variables used in our empirical models.

Dependent variable: a dichotomous variable that takes on a value of 1 when a bid is accepted and 0 otherwise.

Independent variables:

(1) *ExpertCertification*: Whether or not there is an "Expert Certification" icon next to the vendor's bids. As we described earlier, these icons do not directly indicate whether the exams that the vendor has taken were on a subject relevant to the current project. However, the icon does increase the visibility of a seller, and could potentially create an advantage for them. On the other hand, the cost of obtaining the signal is relatively low. Whether or not this is a useful signal is an empirical question. In a robustness test, we also included the number of tests that the vendor has taken up to the time that the bid was placed.

(2) *noRating*: An indicator variable that the vendor has no ratings at all.

(3) *AvgRating*: The mean of ratings that the vendor has received up to the time that the bid was placed.

(4) *logRatingsCount*: Logarithm of the number of ratings that the vendor received up to the time of the bid.

(5) *BuyerCoderSameCountry*: An indicator variable that the buyer and seller are residents of a same country.

(6) *BuyerCoderBothUS*: An indicator that both parties are from US.

(7) *logCoderMonth*: Logarithm of the number of months that the coder has signed up on this market.

(8) *logExpertiseLength*: Each vendor has a "resume" page where they can post their resumes or further describe their experiences and expertise. This variable captures the length of the document.

(9) *logBidAmount*: Logarithm of the dollar amount of the bid.

(10) *logBidOrder*: Logarithm of the order in which the bid was placed. A larger number suggests that the bid was placed later. Since the bids are displayed in the order they are received by default, earlier bids are more likely to be noticed and accepted.

(11) *noCommentBid*: An indicator variable that the bid does not come with a message.

(12) *noTypo*: An indicator variable that there is no typo in the comment posted by the vendor.

(13) *ProjectAmtRange*: to control for the size of projects, we first calculated the total amount that the buyer accepted for each project when he or she chooses a vendor. For pay-for-time contracts, this is the hourly rating that the vendor bid, and the estimated number of hours. We then "bin" this amount into different intervals: 1 if it's lower than $100, 2 if it's between $100 and $200, 3 if

between $200 and $300, 4 if between $300 and $400, and 5 for $400 and above. These are then included in the estimation as a series of dummies in a saturated model specification.

(14) *Variables of linguistic features*. We use LIWC 2007 to parse bid comments that accompany the bids placed by vendors. LIWC is a computational text analysis program developed by psychologists James W. Pennebaker and his colleagues. LIWC takes the text files as input, and produce numerical characterization of the file by categorizing words and phrases into approximately 80 output variables. These include "4 general descriptor categories" such as word count, and number of words longer than six letters; "22 standard linguistic dimensions", which include percentage of words that are nouns, auxiliary verbs, adverbs and so on; "32 word categories tapping psychological constructs" such as affect and cognition; "7 personal concern categories" such as home, leisure, work; and 3 "paralinguistic dimensions" such as fillers ("um", "you know"); as well as 12 punctuation categories.

Our study is one of the first attempts to apply such analysis to vendors' private communications with the buyers in an outsourcing process. We focus on the following variables because the other variables either cannot be theoretically justified in our context, or they occur very rarely in the sample (such as those representing biological processes). In addition to the variables indicate the count of each category of words, we also tested the percentage of these words, and obtained similar results.

(a) *"We" words* including "we", "our", "us" and so on. "We" could refer to the vendor themselves - in which case this variable can indicate that the vendor is representing a collective of developers; or it can be used to address the buyer as well ("we can start working soon"). Our

examination of a sample of bid comments indicates that the previous use is more frequent. But in either case, the use of "we" can indicate either the resources of the vendor, or as a subjective way of reducing the psychological distance from the buyer. Either rationally or emotionally, this category could potentially affect buyers' choice.

(b) *Words with more than 6 letters*. Long words (which are often jargons as well) can increase the difficulty of communication, hence likely to reduce the chances of winning.

(c) *Auxiliary verbs* such as "will, am, have".

(d) *Adverbs* such as "quickly" "shortly" and "satisfactorily".

(e) *"Time" words* such as "time", "end", "until". These words should be a significant predictor in Pay-for-Time contracts, although the perception of these words are still subject to the buyer, therefore their direction is not immediately clear.

(f) *"Money" words* such as "owe" "cash". Note that the bid amount itself is not included in this analysis. Hence, "money" words refer to the discussion of money-related issues in the body of the text. Discussion of the dollar amount could represent a tradeoff with discussions on the requirements of the product, and could negatively affect the chances of winning the bid.

(15) One other linguistic variable that we are interested in studying is the prevalence of *typos* in the comment submitted by vendors. While it is possible that too many typos make it difficult to communicate and therefore should make a bid less attractive, it is equally likely that buyers can be tolerant of these typos in search of a good deal. We use open-source software GNU Aspell to achieve this by submitting these text files (via Perl scripts) to an English dictionary[3] associated with GNU Aspell, and compare each word against the dictionary. The number of typos is

---

[3] Copyright Kevin Atkinson, http://wordlist.sourceforge.net/

recorded for each comment associated with bids. In the estimation, we use logarithm of this number. We also calculate the percentage of typo (i.e. number of typos versus total number of words in the text) as a candidate variable.

Model

The main goal of our paper is to understand how buyers' choice criteria of potential vendors change under different contracting mechanisms. For this purpose, our dataset have some very ideal features that allow us to model the buyers' choice. First, since we focus on buyer and vendors who do not have prior exchange experiences, hence, all information that leads to a buyer's decision is captured in our data. Unobserved factors that may have contributed to these choices are minimized, and can be considered orthogonal to our variables of interest. This is consistent with the identification strategy in Angrist (1998). Second, unlike prior studies that only have information about chosen vendors, we have information about all others who were rejected by the buyer. Information about these unsuccessful bidders sheds light on how buyers made their decisions. Third, this website uses a sealed auction format; only the buyers see who the bidders are, and how much was bid. This ensures that the bids among vendors are independent of each other, allowing for proper statistical modeling.

Given these features, we use maximum likelihood estimation of logistic models, where the dependent variable is whether or not a bid is placed. Independent variables are virtually all the information that buyers were presented with when they decide whom to contract with, as described in the previous version. As mentioned earlier, all bids placed in auction are

independent of each other because they are sealed. Out of an abundance of caution, we estimated

the standard errors using clustered sandwich estimators to allow for intragroup correlation, where

we specify a cluster to be an auction (a request for bid).

Our main hypothesis is that the effect of reputation, certification, and linguistic variables

change when contract forms change. To test it, we first estimate a logistic model where a dummy

variable that indicates a Pay-for-Time contract is interacted with the variables of our interest. For

the overall sample - which includes bids from bidders who have no ratings, no expert

certifications, or bids that were placed without textual comments - we estimate the following

model:

$$
\begin{aligned}
\text{Prob}(BidWins = 1 \mid x) = {} & \beta_0 + \beta_1 noRating + \beta_2 PFT * noRating \\
& + \beta_3 ExpertCertified + \beta_4 PFT * ExpertCertified \\
& + \beta_5 noBidComment + \beta_6 BuyerCoderSameCountry \\
& + \beta_7 \log BidAmout + \beta_8 \log BidOrder \\
& + \beta_9 \log CoderMonths + \beta_{10} \log ExpertiseLength \\
& + \beta_{11} PFT + \varepsilon
\end{aligned}
$$

In other words, we multiplied the PFT dummy with dichotomous variables that indicate

no rating bids, no comment bids, and no certification bids - respectively. Results of this model

are shown in Table 1. We can see that PFT itself is statistically significant, and so is its

interaction with the no-rating dummy. The other two interactions are however not statistically

significant. It appears that ratings play a different role under different contract forms, but

certification and comment dummies are less clear.

For the next step, we excluded interaction terms and estimated them separately on the PFT contract subsample, and the PFD contract subsample. Results (not reported) are consistent with Table 1: dummy variables indicated whether or not there is no comment, and whether or not there is no certification, are statistically insignificant for both subsamples. The dummy variable for no-Rating, however, show a pattern consistent with our hypothesis: having no ratings negatively affects the chances of being selected; this effect is statistically significant at the 1% level for PFD or Fixed Price contracts, but only marginally significant for PFT contracts. In addition, the magnitude of the effect is also stronger for PFD contracts: placing a bid without a rating reduces the odds of winning by over 58% in PFD (cf. Fixed Price) contracts, but by only about 30% in PFT (cf. Time and Materials) contracts.

Some auxiliary results are also worth mentioning. We find evidence that on average, buyers actually prefer vendors who are from their same country. The tendency to offshore is actually less than what mass media would have us believe: The odds ratio of a same-country vendor being chosen is actually over 250%. This pattern persists in many more specification that we shall discuss, and is robust to the inclusion of variables such as the time zone difference and whether English is the official language. In addition, when we replace this variable with a dummy that takes the value of 1 when both buyer and vendor are from US, the same result obtains. In other words, under comparable degrees of uncertainty (first time interactions), US buyers also prefer domestic vendors rather than foreign.

Some auction variables are also statistically significant predictors of bidding outcomes: bids placed earlier are more likely to be successful, and higher amount of bids are less likely to be chosen.

In unreported analyses, we also investigated *interaction* between some variables. One example is between the indicator variable for "*no rating*" and "*number of months since vendor signed up*". While "no rating" is shown to be negatively associated with the chances of winning, it is significantly worse for vendors who are on the market longer. In other words, between two vendors who are not rated, the one who joined the site earlier are even less likely to win a contract. This is because no-rating suggests that the vendor has not been chosen by any other buyer so far. The longer they stay in that situation, the less attractive they become.

The above analyses however, only use dichotomous variables for comment and rating. This may be sufficient for Expert Certification (bids either have an "expert" icon next to it, or it does not), but it is certainly worth exploring the actual level of rating and the number of ratings. In addition, among the bids that do include textual messages, we will investigate how linguistic features of these messages affect the choice of buyers.

We first analyze the number of ratings as well as the average rating of vendors when they place bids. These variables are displayed prominently to buyers when bids are placed. Replacing the dummy variable of "no rating', we include two new variables: (1) logarithm of the number of ratings that the vendor has at the time of the bid; and (2) the average of ratings that the vendor received at the time of the bid. We report results using subsamples, instead of the overall sample

with interaction terms. Results are consistent, although subsample analyses are easier to interpret and less dependent on statistical assumptions. In addition, we exclude auctions that choose bidders who do not have ratings at the time of the request in this estimation.

Results of the estimation on these two subsamples are reported in Table 2. Variables that we have examined previously display very consistent results as before: certification is insignificant, while bidding order and bid amount matters. On the other hand, ratings variables show some interesting patterns. For instance, while having a large number of ratings in PFD (cf. Fixed Price) contracts significantly increases the chances of securing the contract, the effect is statistically insignificant in PFT (cf. Time and Materials) contracts. Meanwhile, while the average rating has a positive and statistically effect on the chances of winning a contract under PFD arrangements, the effect is insignificant for PFT contracts. In fact, the magnitudes of these coefficients are also smaller in PFT contracts. All these findings are consistent with our working hypotheses.

We now turn to the textual content of comments that accompany bids, and identify characteristics that are associated with better chances of winning under different contract formats. To this end, we also first limit the estimation sample to auctions that do not choose a bid without comments.

The first variable of interest is the number of typos. We studied two alternative metrics for typos: total number of typos in the vendor's message, and the ratio of this number to the total number of words. While the odds ratio associated with these variables are indeed smaller than 1,

they are not statistically significant in all specifications that we described earlier. It thus appears that buyers in this market seem to be tolerant of typos, and do not consider them a signal of communication difficulties.

The second sets of variables are generated from LIWC. The two main variables that we are interested in are "Time" and "Money" variables, as they represent two of the main dimensions that buyers consider when they choose a vendor. We find that "time" words are a statistically significant predictor of bid success only for pay-for-time contracts, and insignificant in pay-for-deliverable (fixed price) contracts. For PFT contracts, a larger number of "time" words are associated with higher chances of winning. "Money" words, on the other hand, are also significant only for PFT contracts. By contrast and as predicted, a larger number of "Money" words are associated with lower chances of winning.

A few other variables also show some interesting patterns. We find that the number of words with more than 6 letters is negatively associated with chances of winning, although the odds ratio is relatively small in scale. "We" words are not significant predictors in either PFT or PFD contracts. Adverbs such as "quickly" increases the chance of winning for PFT contracts only, but not for PFD contracts. Auxiliary verbs decrease the chance of winning, although only for PFD contracts. Many of these results are robust to specifications. For instance, removing Auxiliary verbs do not change the results, especially the other categories of words. Their effects seem to be orthogonal to each other.

Discussions and Implications

One implication of our study is that Matthew Effects (Robert K. Merton, 1968) can be mitigated by changes in contractual forms or other incentive design. Matthew Effect refers to the phenomenon that in a competitive environment, individuals, organizations and entities that were previous in an advantageous position are very likely to continue their advantage. This is similar to the idea of "preferential attachment" (Thomas T. Hills et al., 2009) and other concepts; basically, "the rich grows richer, while the poor grows poorer". For electronic commerce websites, such tendency may not be ideal as it is likely to drive away new vendors, yet it is indeed happening: Consumers are more likely to buy from sellers with more ratings and higher ratings. Our results on the ratings suggest that this effect can be at least mitigated by changes in the contract format, at least in the labor and outsourcing markets. Buyers have an incentive to hire less known, less experience vendors when the design of the contract allows them to try out these vendors, and at the same time allow them to stop paying them if they turn out to be of low-ability type. This not only has theoretical implications, but also practical: electronic commerce websites concerned about expanding customer base can design mechanisms to redistribute the risks among buyers and sellers.

The second implication from our findings is that certifications may not always be effective. It is possible that this is unique only to this website, and only to the particular types of certifications. However, given the popularity of all kinds of certifications and third-parties in peer-to-peer markets such as eBay, these results suggest that we should not take certifications' effectiveness for granted. While vendors often seek such certifications (at a financial cost), and

they do obtain and extra "icon", they unfortunately do not appear to increase vendors' chances of obtaining contracts in this market.

Last but not least, our analyses show some potential for automated text analysis. Textual analysis is still an emerging field. Our analyses show that buyers do appear to take into account what was written by the vendors, and their consideration changes under different contract regimes. For platforms such as online outsourcing markets, implementing automatic text analyses programs can potentially increase the efficiency of screening vendors, especially as when we are able to link textual cues to project outcomes, which we are investigating in a separate paper.

Limitations and Future Research

Our goal in this paper is not about the choice of contract formats. By constructing two non-overlapping subsamples of different contractual forms, we seek to understand how contract forms moderate the relationship between a vendor's reputation and their chance of winning a contract. A natural extension of this analysis is certainly to go beyond dyads of first-time interactions, and understand better the endogenous choice of contractual forms in this context, especially between buyers and sellers who have repeated interactions. As described in the paper, the proportion of buyer-seller pairs that switched from pay-for-deliverables to pay-for-time contracts is very small. It is possible, however, that as users become more familiar with the context, we will be able to observe more "switching" of contract forms. At that time, we will be able to extend the analyses in this paper to model the contract choice endogenously.

A second extension of our study will be incorporating some metrics for the outcome of projects. We address the outcome metrics of projects in a separate paper.

Our analysis of the textual comments is one of the first efforts to study the effect of written language on buyer choice of vendors. We recognize that there are much more advanced text mining techniques available. However, LIWC has been quite broadly used in psychology and management, and it is very similar to packages used in finance (P. C. Tetlock et al., 2008), *General Inquirer*. Another valid critique of this analysis is analogous to "Lucas Critiques" in economics, in the sense that when vendors realize how buyers respond to their languages, it will affect how they write in the future, which in turn will change how buyers screen vendors. These are certainly interesting dynamic interactions that can be explored in future research. However it does not affect the validity of our current research: these communications are private between buyers and vendors. The website has not done analyses like this before, and at least in the time frame that we study, no such results were revealed to vendors.

Conclusions

The advancement of information technologies, especially internet technologies, promises to change the landscape of labor markets forever (David H. Autor, 2001). How buyers and sellers (workers) are matched, how services are delivered, and how efforts are monitored, all these dimensions that economists have long studied will be dramatically different in the online market. The monitoring technology that we describe in the paper is but one such development. All these

changes create an abundance of new research arenas. For instance, while researchers have extensively investigated contract forms as well as online reputation systems, these two literatures rarely cross paths. Our unique dataset allows us to investigate whether and how the effect of reputation systems, certifications, and text features change under different contract regimes. Given the popularity of reputation, certification and abundance of text in electronic markets, our results can have much wider implications beyond this website itself.

Tables

Table 1: Full model with interactions

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; standard errors in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Dummies for project amount range suppressed for brevity. * $p<0.1$; ** $p<0.05$; *** $p<0.001$

Variable	Odds Ratio
ExpertCertified	1.195
	(0.768)
noRating	0.339***
	(0.093)
noCommentBid	0.899
	(0.250)
BuyerCoderSameCountry	2.036***
	(0.391)
logBidAmount	0.636***
	(0.038)
logBidOrder	0.628***
	(0.029)
logCoderMonths	1.592***
	(0.132)
logExpertiseLength	1.038
	(0.043)
PFT	2.846***
	(0.363)
PFT*Expert	1.025
	(0.683)
PFT*Rating	2.380***
	(0.793)
Intercept	0.116***
	(0.049)
N (number of bids)	**5670**

Table 2: Effect of Ratings on PFD and PFT subsamples

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; standard errors in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Dummies for project amount range suppressed for brevity. * $p<0.1$; ** $p<0.05$; *** $p<0.001$

Variable	Odds Ratio under Pay-for-Delivery Contracts	Odds Ratio under Pay-for-Time Contracts
ExpertCertified	1.017	1.170
	(0.651)	(0.256)
logRatingsCount	1.229***	1.094
	(0.095)	(0.097)
AvgRating	1.255*	1.166
	(0.154)	(0.117)
BuyerCoderSameCountry	1.395	3.378***
	(0.470)	(1.103)
logBidAmount	0.665***	0.553***
	(0.061)	(0.058)
logBidOrder	0.725***	0.553***
	(0.050)	(0.052)
logCoderMonths	1.616**	1.354**
	(0.341)	(0.184)
logExpertiseLength	1.093	0.975
	(0.068)	(0.074)
Intercept	0.004***	0.400
	(0.005)	(0.449)
N (number of bids)	**2607**	**974**

Table 3: Effect of Textual Communications under Pay-for-Deliverable vs. Pay-for-Time Contracts

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; standard errors in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Dummies for project amount range suppressed for brevity. * $p<0.1$; ** $p<0.05$; *** $p<0.001$

Variables	Odds Ratio Under Pay-for-Deliverable Contracts	Odds Ratio Under Pay-for-Time Contracts
ExpertCertified	0.988	1.300
	(0.658)	(0.260)
noRating	0.450***	0.708
	(0.126)	(0.161)
BuyerCoderSameCountry	1.861**	2.538***
	(0.506)	(0.811)
logBidAmount	0.632***	0.591***
	(0.063)	(0.062)
logBidOrder	0.707***	0.545***
	(0.047)	(0.042)
logCoderMonths	2.152***	1.340***
	(0.350)	(0.138)
logExpertiseLength	1.099*	1.033
	(0.060)	(0.072)
Sixltr	0.969***	0.987
	(0.011)	(0.012)
we	0.972	0.994
	(0.027)	(0.039)
auxverb	0.959***	0.971
	(0.014)	(0.021)
adverb	1.022	1.063**
	(0.027)	(0.032)
time	0.981	1.049**
	(0.017)	(0.020)
money	1.005	0.897**
	(0.030)	(0.049)
noTypo	0.745	0.843
	(0.190)	(0.230)
Intercept	0.080***	1.348
	(0.062)	(0.965)
N (number of bids)	**3653**	**1490**

References

Angrist, JD. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica*, 66(2), pp. 249-88.

Autor, David H. 2001. "Wiring the Labor Market." *Journal of Economic Perspectives*, 15(1), pp. 25-40.

Bajari, Patrick and Steven Tadelis. 2001. "Incentives Versus Transaction Costs: A Theory of Procurement Contracts." *RAND Journal of Economics (RAND Journal of Economics)*, 32(3), pp. 387-407.

Banerjee, Abhijit V. and Esther Duflo. 2000. "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry." *Quarterly Journal of Economics*, 115(3), pp. 989-1017.

Chen, Yuanyuan and Anandhi Bharadwaj. 2009. "An Empirical Analysis of Contract Structures in It Outsourcing." *Information Systems Research*, 20(4), pp. 484-506.

Dey, Debabrata; Fan Ming and Zhang Conglei. 2010. "Design and Analysis of Contracts for Software Outsourcing." *Information Systems Research*, 21(1), pp. 93-114.

Gopal, Anandasivam and Guodong Gao. 2009. "Certification in the Indian Offshore It Services Industry." *Manufacturing & Service Operations Management*, 11(3), pp. 471-92.

Gopal, Anandasivam; Konduru Sivaramakrishnan; M. S. Krishnan and Tridas Mukhopadhyay. 2003. "Contracts in Offshore Software Development: An Empirical Analysis." *Management Science*, 49(12), pp. 1671-83.

Hasija, Sameer; Edieal J. Pinker and Robert A. Shumsky. 2008. "Call Center Outsourcing Contracts under Information Asymmetry." *Management Science*, 54(4), pp. 793-807.

Hills, Thomas T.; Mounir Maouene; Josita Maouene; Adam Sheya and Linda Smith. 2009. "Longitudinal Analysis of Early Semantic Networks: Preferential Attachment or Preferential Acquisition?" *Psychological Science*, 20(6), pp. 729-39.

Jensen, Michael and Aradhana Roy. 2008. "Staging Exchange Partner Choices: When Do Status and Reputation Matter?" *Academy of Management Journal*, 51(3), pp. 495-516.

Kirsch, Laurie J. 1997. "Portfolios of Control Modes and Is Project Management." *Information Systems Research*, 8(3), pp. 215.

Koh, Christine; Ang Soon and Detmar W. Straub. 2004. "It Outsourcing Success: A Psychological Contract Perspective." *Information Systems Research*, 15(4), pp. 356-73.

Levina, Natalia and Jeanne W. Ross. 2003. "From the Vendor's Perspective: Exploring the Value Proposition in Information Technology Outsourcing." *MIS Quarterly*, 27(3), pp. 331-64.

Lewis, Tracy R. 1986. "Reputation and Contractual Performance in Long-Term Projects." *RAND Journal of Economics (RAND Journal of Economics)*, 17(2), pp. 141-57.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science*, 159(3810), pp. 56-63.

Rustagi, Sandeep; William R. King and Laurie J. Kirsch. 2008. "Predictors of Formal Control Usage in It Outsourcing Partnerships." *Information Systems Research*, 19(2), pp. 126-43.

Susarla, Anjana; Ramanath Subramanyam and Prasanna Karhade. 2010. "Contractual Provisions to Mitigate Holdup: Evidence from Information Technology Outsourcing." *Information Systems Research*, 21(1), pp. 37-55.

Tanriverdi, Hüseyin; Prabhudev Konana and Ge Ling. 2007. "The Choice of Sourcing Mechanisms for Business Processes." *Information Systems Research*, 18(3), pp. 280-99.

Tetlock, P. C.; M. Saar-Tsechansky and S. Macskassy. 2008. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *The Journal of Finance*.

Tykvová, Tereza. 2007. "Who Chooses Whom? Syndication, Skills and Reputation." *Review of Financial Economics*, 16(1), pp. 5-28.