Contract Choice, Moral Hazard, and Performance Evaluation: Evidence from Online Labor Markets

Abstract

Due to the spatial and temporal separations between clients and freelancers, online labor markets (OLMs) are particularly susceptible to issues related to information asymmetry. Based on the economics of information, we hypothesize that the choice of contract type—i.e., between the fixed-priced (FP) contract and the time-and-materials (TM) contract—has important implications for curbing moral hazard during contract execution, and therefore will influence the client's perceived contractual performance upon project completion. We test the predictions by assembling a dataset of data analytics projects completed by freelancers on *Upwork*, the largest online freelancing platform. We find that, consistent with our hypothesis, freelancers under a TM contract receive significantly lower performance ratings by their clients on average compared to those under an FP contract. Interestingly, we also find that the level of expertise required for a project moderates the effect of contract choice on client satisfaction; the negative impact of a TM contract is smaller (i.e., less negative) when a project requires intermediate-level or expert-level skills. Our study offers useful insights into an important institutional determinant of contractual performance evaluation, which has profound implications for freelancers' reputations in OLMs.

Keywords: *online labor market, contract choice, contractual performance evaluation, information asymmetry, moral hazard*

I. Introduction

Online labor markets (OLMs) are marketplaces that connect workers with short-term job opportunities in the context of knowledge work. These platforms, such as *Upwork* and *Freelancers.com*, help independent workers, many of whom could not have a traditional job due to personal circumstances, find employments that offer flexible work schedules. Furthermore, OLMs help clients recruit contingent service providers, often with rare skills that are difficult to source in traditional labor markets. In recent years, OLMs brokered labor relationships in a wide range of occupations, giving rise to the gig economy. Statistics reveal that over a third of the US workforce involves in gig work to some extent: more than 57 million workers in the US economy participated in freelancing in 2019, accounting for 35% of all US workers. Moreover, the direct contribution of freelancing to the economy is over $1 trillion, nearly 5% of the U.S. GDP.[1] Recent studies also show that many workers consider OLMs as a substitute source of income for full-time employment because online work shortens the search time between jobs, reducing the duration of unemployment (Borchert et al. 2018, Cantarella and Strozzi 2021).

Despite the merits of these platforms, more than 60% of OLM projects fail to reach a contract, with the client unable to hire any freelancer (Zheng et al. 2015). Notably, one of the most significant challenges faced by OLMs is the issue of information asymmetry due to the spatial and temporal separation of the client and the worker (Benson et al. 2020, Pelletier and Thomas 2018). For example, a client on an OLM may be unable to differentiate high-quality freelancers from low-quality ones, and she cannot effectively screen those who bid on her projects, leading to adverse selection. Earlier research has examined several ways of addressing these pre-contractual information asymmetries in the context of OLMs. For example, studies have shown that signaling mechanisms such as reputation (Lin et al. 2018, Moreno and Terwiesch 2014) and experience in related fields (Agrawal et al. 2015) help mitigate this type of information asymmetry. Despite the progress, this line of literature has mainly focused on the issue of adverse selection, such as the client's decision on which contract type to use (Chen and Bharadwaj 2009, Yao et al. 2010) or which vendor to hire (Lin et al. 2018). In contrast, much less is known about the role of post-contractual information asymmetry—i.e., moral hazard—in determining the outcomes of a contractual relationship.

In this work, we aim to bridge this gap and investigate how the contract type used in an OLM project—through its role in curbing moral hazard during contract execution—will influence the client's perceived contractual performance upon project completion. We examine the difference between two contractual formats commonly used in OLMs: the fixed-priced (FP) contract and the time-and-materials (TM) contract. Particularly, building on agency theory and the economics of information (Stiglitz 2000), we propose that

[1] See https://www.upwork.com/press/releases/freelancing-and-the-economy-in-2019.

the use of different contract types in an OLM project will influence the likelihood of moral hazard taking place and the cost of monitoring, and therefore will result in differences in the client's perceived contractual performance. We further hypothesize that the relationship is moderated by the level of expertise required for the project, because the degree of contract incompleteness and the difficulty in outcome verification are both increasing with project complexity (Al-Najjar 1995, Bapna et al. 2010), therefore making moral hazard more difficult to prevent for expert-level projects, regardless of the contract choice.

To test these predictions, we conduct empirical investigations by examining a sample of data analytics projects collected from *Upwork*, each with detailed information on the project characteristics, the freelancer characteristics, the type of contract adopted, the level of expertise required for the project, and the client's evaluation of the contractual performance. To address the potential endogeneity of the choice of contract type, we employ an endogenous treatment regression model in which we use instruments that exogenously shift the contract type choice. As an alternative identification strategy, we further present analyses based on a matched sample in which each observation under a TM contract is matched to one under an FP contract using a propensity score matching method. Consistent with our theorizing, we find that, with everything else being equal, the perceived contractual performance—as measured by client satisfaction—is significantly lower under a TM contract than under an FP contract. Interestingly, the results also show that the negative impact of a TM contract on the client's satisfaction is weaker (i.e., less negative) for intermediate-level or expert-level projects than for entry-level ones. Taken together, these findings deepen our understanding of the relationships between contract choice, moral hazard, and contractual performance evaluation in OLMs, and lead to some important managerial implications.

The findings regarding the relationships between contract choice, moral hazard, and performance evaluation are important because they have far-reaching implications for OLMs. For example, it is well known that clients increasingly rely on OLM reputation systems to screen potential vendors and they are willing to pay a premium for more reputable workers (Moreno and Terwiesch 2014). However, if the client's rating of contractual performance—which forms the foundation of the freelancers' reputation in OLMs—is determined by institutional factors in addition to freelancer characteristics, clients need to be cognizant of potential biases in the generating process of vendor reputation and use these reputation scores with caution. Furthermore, freelancers who accept jobs under a time-and-material contract also need to anticipate the potential negative impact of the contract choice on their performance ratings, and therefore may preemptively take actions to reduce the client's concerns over moral hazard—such as initiating more frequent and transparent communications with the client to report their work progresses—and be mindful of the client's expectations regarding project timeline and cost.

II. Literature Review

Contracts in IT Outsourcing

A contract is defined as an agreement between two contractual parties before the start of a transaction that provides a legally binding institutional framework enunciating the client and the vendor's rights, duties, and responsibilities (Chen and Bharadwaj 2009). In the context of IT outsourcing, prior literature has studied how various contractual features and structures can be employed to ensure desirable behavior, often through the lens of agency theory. For example, Osei-Bryson and Ngwenyama (2006) propose an approach to analyzing risks in IS outsourcing and structuring incentive schemes that can be used to improve vendor performance. Susarla et al. (2010) identified three features of contracts—extensiveness, duration, and extension clause—that serve as a remedy to holdup problems and encourage relationship-specific investments.

Most relevant to our study is the stream of literature that investigates the choice between time-and-material (TM) contracts and fixed-price (FP) contracts. For example, Gopal et al. (2003) identify a set of factors that contribute to the choice of the contract type in the context of software offshoring; they also find that a vendor's profit is higher under a TM contract. Kalnins and Mayer (2004) argue that uncertainties and measurement issues affect the choice of contract. They find that TM contracts are preferred by the employer when it is difficult to estimate the contract cost ex-ante or when the quality is difficult to measure ex-post. Corts and Singh (2004) find that repeated interaction and high-powered formal contracts are substitutes in the context of offshore drilling, and clients are less likely to choose an FP contract as the frequency of client-vendor interaction increases. Dey et al. (2010) suggested that FP contracts are often more appropriate for simple software projects that require shorter development time, while TM contracts work better for more complex projects and when the auditing process is effective.

Contractual Performance

IS outsourcing research has mainly examined contractual performance at the organizational level, such as lower labor and production costs (Larsen et al. 2013), improvement in innovation capabilities (Baier et al. 2015, Nieto and Rodríguez 2011), and opportunities to learn (Jensen 2009). Since direct measurements of IT outsourcing success are difficult to obtain, IS researchers have identified several ways to measure outsourcing performance indirectly. For example, Doll and Torkzadeh (1988) developed a widely used client satisfaction instrument to measure project outcomes. Some researchers have also investigated the determinants of contractual performance in IT outsourcing. Notably, Grover et al. (1996) find that the quality of the vendor and characteristics of client-vendor relationships such as trust, cooperation, and communication are important predictors of outsourcing success. Levina and Ross (2003), through a detailed case study, draw the conclusion that the value proposition of an outsourcing vendor and its client's satisfaction depends critically on its core competencies and complementary organizational design, which cannot be easily replicated by the client.

Prior research has also identified service quality, or the degree and direction of the discrepancy between a service receiver's expectations and her perceptions (Parasuraman et al. 1988), as an important predictor of the vendor's contractual performance (Grover et al. 1996). Particularly, Kim et al. (2005) examine how the quality of service provided by information systems outsourcing vendors affects their customers' perceptions. Their results show that the image projected by the vendor is more important than reliability, responsiveness, and customer empathy in determining client satisfaction. Yoon and Suh (2004) adapt *SERVQUAL*[2] to the IT consulting context as a measure of IS customers' perceived quality of the consulting services and find that IT consulting *SERVQUAL* is significantly related to the level of customer satisfaction.

Information Asymmetry in Online Labor Markets

Online labor markets differ from offline, physical labor markets in several aspects. First, OLMs facilitate labor relationships, i.e., matching client tasks with freelancers' capabilities, through a bidding process enabled by technology platforms (Horton 2010), therefore greatly relaxing the geographical constraints of traditional labor markets and allowing a client to outsource jobs to freelancers around the globe (Agrawal et al. 2015). In addition, OLMs are often used for projects with limited scope over a relatively short period and therefore are not conducive to developing long-term, stable vendor-client relationships typically observed in traditional IT outsourcing contexts.

Due to these differences, some market inefficiencies in traditional labor markets—such as information asymmetry—are exacerbated in OLMs. For example, the screening and selection of freelancers are typically based on very limited information about the bidder, and important freelancer characteristics such as perseverance, trustworthiness, or motivation are difficult to observe through online interactions alone (David 2001). As a result, adverse selection is particularly prevalent in the context of OLMs (Agrawal et al. 2015, Hong and Pavlou 2017, Horton 2010). A robust stream of research has investigated the role of reputation systems as a remedy to the adverse selection issue in OLMs, which concludes that freelancers with high reputation ratings are more likely to win contracts (Lin et al. 2018), that more reputable bidders

[2] *SERVQUAL* is a multidimensional research instrument designed to measure service quality by capturing respondents' expectations and perceptions along five dimensions of service quality: reliability, responsiveness, assurance, empathy, and tangibles.

can earn a price premium (Moreno and Terwiesch 2014), and that third-party certification systems serve as a surrogate signaling mechanism for platform-specific reputation (Goes and Lin 2012)

Once a contract is signed, moral hazard problems may arise when the freelancer acts opportunistically, and the client cannot perfectly monitor the freelancer's effort. Related work has suggested a few solutions to this information asymmetry problem. For example, Liang et al. (2019) show that the implementation of IT-enabled monitoring systems effectively mitigates moral hazard such that clients reduce their reliance on freelancer reputation in hiring decisions. Moreover, some researchers have examined the effectiveness of financial incentives in curbing moral hazard issues, but the findings have been mixed. On the one hand, Yin et al. (2013) study the effect of performance-contingent financial rewards on work quality and worker effort in the context of *Amazon Mechanical Turk* and find that the financial rewards alone affect neither quality nor effort. Mason and Watts (2009), on the other hand, find that financial incentives increase the quantity (i.e., output volume) but not the quality (as measured by accuracy) of work by the participants of their experiment.

III. Theory and Hypotheses

Contract Types and Client Satisfaction in OLMs

We start by drawing on the economics of information (Macho-Stadler and Pérez-Castrillo 2001) as a unifying theoretical framework to understand how the choice between TM and FP contracts can lead to different implications for moral hazard during project execution, which will, in turn, affect client satisfaction upon the conclusion of the project. Because an OLM is mediated through an online platform, monitoring is particularly difficult compared to employment relationships in an offline, physical environment due to the spatial and temporal separations between clients and freelancers (Liang et al. 2019). Although most OLMs have some monitoring systems that allow a client to keep track of the work progress remotely,[3] the client and the worker typically do not have direct, face-to-face interaction that offers richer nonverbal cues. Under a TM contract, the client bears significant risks because work time is typically self-reported, leaving room for opportunistic behaviors on the part of the freelancer such as inflating the reported work hours (Corts and Singh 2004, Liang et al. 2019). Facing this issue, the client may have to incur greater monitoring costs to reduce the likelihood of moral hazard. In contrast, under an FP contract, the worker bears a significant part of the project risk because project cost or time overruns could affect the worker's project profitability (Gopal and Koka 2012, Gopal et al. 2003). Therefore, under an FP contract, the worker

[3] For example, *Upwork* has implemented a mechanism of "Work Diary" that allows freelancers to record their hours and show their clients work-in-progress screenshots through a desktop app.

has strong incentives to execute the project and manage her progress efficiently, reducing the likelihood of moral hazard. This, in turn, will relieve the burden of monitoring on the part of the client.

Furthermore, due to the complex nature of IT projects, it is often difficult for a client to estimate the amount of effort involved in a project accurately ex-ante (Larsen et al. 2013). Therefore, before contracting, the client often underestimates the complexity of a project, its scope, and its true cost (Conrow and Shishido 1997). Under an FP contract, the client and the worker may engage in negotiation to adjust the client's expectations if the two parties' estimates over the project budget diverge significantly. Under a TM contract, however, the discrepancy in expectations is less likely to be discovered and corrected ex-ante because the payment terms are based on an hourly rate rather than a lump-sum payment. As a result, longer-than-expected project duration or budget overrun may come as a surprise if the client underestimates the cost initially. Because making adjustments to contract terms in the middle of a project is particularly complex, significant adaptation costs will occur if the client and the worker engage in renegotiation (Bajari and Tadelis 2001).

Given these differences, we expect that client satisfaction in OLM projects upon project completion will vary between the two contract types. Particularly, because a TM contract is associated with higher monitor costs during project execution and/or higher adaptation costs when a budget overrun occurs, we hypothesize that:

H1. Upon the completion of an OLM project, the client's satisfaction is lower under a TM contract than under an FP contract, with everything else being equal.

The Moderating Role of Project Expertise Requirement

In OLMs, jobs are associated with varying degrees of expertise requirement. For example, an entry-level project such as data entry typically requires minimum domain knowledge. The task is usually repetitive and straightforward, and the deliverables are easy to verify. In contrast, an expert-level project may require a complex set of skills and years of professional experience in some specific knowledge domains. We propose that the level of expertise requirement associated with a project moderates the relationship between the contract type and perceived contractual performance. Particularly, we expect that the difference in client satisfaction between an FP contract and a TM contract will be smaller when a project is more complex and requires high-level expertise.

As we argued earlier, in OLMs the client cannot perfectly observe the effort of a worker and moral hazard may arise due to opportunistic freelancer behaviors. Under such conditions, an FP contract is preferred by a client because it has the merit of preventing moral hazard and reducing the need for monitoring. However, the effectiveness of an FP contract in curbing moral hazard during contract execution also depends on the ease with which the project output can be verified against the contract terms (Bapna et al. 2010, Eisenhardt 1989). When the goal of a project is clearly defined and its outcome is easy to measure, such as for an entry-level task, the use of an FP contract (an outcome-based contract) has a clear advantage over a TM contract (a behavior-based contract) in reducing the likelihood of moral hazard (Baron and Besanko 1987). Therefore, the use of an FP contract likely results in significantly higher client satisfaction than a TM contract.

However, expert-level projects have two distinct features: 1) it is difficult for the client to completely specify in the contract all the project requirements and all the contingencies that may arise during project execution (Al-Najjar 1995, Susarla et al. 2010), and 2) the outcome of the project is usually difficult to verify for clients unfamiliar with the knowledge domain (Aubert et al. 2002, Bapna et al. 2010). Under such conditions, moral hazard can still arise even when an FP contract is used, because a freelancer's opportunistic behavior cannot be easily detected by the client. For example, in a project that involves the development of a data processing application, the freelancer may produce a program that meets all the functional requirements but does not scale well for large data sets or breaks down when the number of users increases. The detection of these quality issues often requires sophisticated knowledge and rigorous testing beyond the client's capabilities. Therefore, for an expert-level project, the use of an FP contract is not as effective in dispelling the client's concerns over moral hazard as it is for a project that requires entry-level skills. In other words, both the incompleteness of an FP contract and the difficulty in verifying the project deliverables are increasing in the complexity of the underlying project, therefore creating room for opportunistic behaviors by the freelancer and reducing the effectiveness of an FP contract in curbing moral hazard. As a result, we hypothesize that:

H2. The negative impact of a TM contract on the client's satisfaction will be weaker (i.e., less negative) for expert-level projects than for entry-level projects.

IV. Method

Research Context

We conduct empirical investigations using data collected from *Upwork*, a freelancing platform formerly known as *Elance-oDesk* that resulted from a merger between two companies, *oDesk* and *Elance*, in December 2013. It is currently the largest freelancer marketplace with $1 billion worth of jobs posted annually.[4] It connects businesses with freelancers around the globe in more than 70 job categories, ranging from video editing, graphic design, software development, social media solutions, financial planning to

[4] See https://en.wikipedia.org/wiki/Upwork.

administrative support. A freelancer registers an account and builds a profile by furnishing basic contact information. Many freelancers also populate their profiles with information regarding their skills, educational attainments, certifications, professional experiences, and exhibits of sample projects. To post jobs on *Upwork*, a client registers an account by providing information such as the company name, website URL, and verification of a payment method. The client can then post a job by either creating her post from scratch or using a template in which many fields are pre-populated with suggestions that *Upwork* has adapted from similar projects. A typical job post includes a job post title, the job category, a job description, screening questions, relevant skills, and the level of expertise required.

An important part of a job post is the way that the client budgets for the project, in which she chooses to pay the freelancer either on an hourly basis or a fixed price. With an hourly project, the professional tracks the time he spends working, and the client is billed weekly. With a fixed-price project, pricing is predetermined and the client either pays all at once or by milestones—i.e., predetermined deadlines that break the project into smaller pieces of work.[5] The funds are deposited into escrow at the beginning of the project and/or each milestone, and then released as the client approves the work by the freelancer.

Once a job is posted, freelancers bid for the project by submitting their cover letters and proposals. For fixed-price projects, freelancers can also propose milestones that divide the payment for a project into predefined pieces with specific goals. For hourly contracts, freelancers may include their hourly rate when submitting a contract proposal. Clients then interview and negotiate with applicants before hiring. During the negotiation, the freelance may choose to update the proposal terms such as the bid or hourly rate before creating a final contract. Once the proposal is accepted and a contract is signed, the freelancer starts working on the project and logs his work time using a virtual monitoring system called 'Work Diary,' which tracks time and records the progress made by the freelancer through a desktop app.[6]

Payments are made through *Upwork*'s online automated billing and payment systems. Hourly contracts are paid weekly based on the billable hours logged in the freelancer's Work Diary, and fixed-price contracts are paid by predefined milestones upon the client's approval of the deliverables, at which point funds are released from escrow. Upon project completion, both the client and the freelancer can provide feedback and evaluate the other party on several processes- and outcome-related criteria.

[5] For details, see https://www.upwork.com/infographics/fixed-price-vs-hourly-project.

[6] See https://support.upwork.com/hc/en-us/articles/211068518-Use-Your-Work-Diary.

Data and Variables

We assemble a dataset of projects completed by freelancers with data analytics skills from *Upwork*. As a first step, we identify all independent freelancers[7] who identify themselves as professionals in the domain of data analytics and who reside in the United States, resulting in 1,075 freelancers. We then collect their complete job histories on *Upwork* during the period between January 2014 (which is when the company first started operation as *Upwork* after the merger of *oDesk* and *Elance*) and August 2021. To limit the impact of unobservable, confounding factors of the clients, we restrict the sample to projects posted by independent, non-enterprise clients. We choose data analytics projects as the sample because there are significant variations in project size and the level of expertise involved, and there is a balanced use of the two contract types. The final sample includes 12,388 projects completed by 1,075 freelancers.

Dependent variable. The dependent variable, *client satisfaction,* is measured by the client's overall rating of the project performance on a scale of 1-5 upon project completion. The perceived project performance is calculated as the average of six components: *skills* (i.e., how skillful the worker is), *availability* (i.e., how flexible the freelancer is regarding her availability), *communication* (i.e., the degree of effectiveness of the freelancer's communication), *quality* (i.e., the quality of the deliverables), *deadlines* (i.e., how well the freelancer meets deadlines), and *cooperation* (i.e., how easy it is to cooperate with the freelancer). Each of the six components is rated by the client separately on a 1-5 scale.

Independent variable and moderator variable. The main independent variable of interest is the *contract type* associated with a project. The contract type is selected by the client when a job is posted on *Upwork*, which takes the form of either a TM or an FP contract. Furthermore, we use the *expertise* level of the project (which can be *entry*, *intermediate*, or *expert*) as a moderator variable. The level of *expertise* required for the project is specified by the client in the job posting.

Control variables. We control for various individual-level and project-level characteristics. At the individual level, we measure an individual's *platform experience* at the beginning of the project by calculating the difference (in days) between the project start date and the user's registration date. At the project level, we control for the total amount of *earnings* the worker made from completing the project. Using the information on the starting and ending dates of a project, we calculated the variable *project duration* to account for the length of the project. To capture the degree of skill match between the freelancer's skill set and the project's skill requirement, we calculated the similarity score of the two using well-established text mining techniques. Particularly, in our research context, both the skill requirement of

[7] There are two types of freelancer accounts: an individual freelancer and an agency. We exclude agencies from the sample to obtain a more homogenous set of projects.

an *Upwork* project and a freelancer's skillset as described in his profile are specified by choosing from a large collection of predefined hashtags (e.g., #DataVisualization, #MachineLearning, etc.), and we compute the *skill match* score between the two sets of hashtags using the Jaccard similarity coefficient (Burtch et al. 2021, Hass 2017). Because our sample consists of projects in the data analytics domain, most of them involve some computer programming tasks. Therefore, we also control for the primary programming language by extracting the first programming languages specified on the list of skill hashtags associated with each project. The variable *programming language* is coded as a categorical variable that consists of seven different languages. Finally, to control for the client-vendor trust that may have been developed through prior interactions (Corts and Singh 2004), we create a binary control variable *first-time interaction* that is set to 1 if the freelancer and the client transact for the first time.

Tables 1 and 2 show the summary statistics of the key variables and the correlation between them. In our sample, TM contract type was used in approximately 49.7% of the projects. Our data indicate that on average there is a good match between the project skill requirement and the freelancer's self-reported skillset, with a mean Jaccard similarity coefficient of 84%. In addition, 52% of the projects require entry-level expertise, while intermediate-level and expert-level projects make up 19% and 29% of the sample, respectively. Among these data science and analytics projects, *python* and *R* appear to be the most popular programming languages, accounting for the primary language of 37% and 17% of the sample, respectively.

*** Insert Table 1 and 2 about here***

Empirical Specification

We start with a two-way (freelancer and year) fixed effects panel data approach to evaluate the hypotheses. Specifically, let Y_{ijt} be the measure of client satisfaction with freelancer i for project j in year t, X_j denote the contract type (with 1 being a TM contract), and E_j denote project j's required expertise level. Let $\boldsymbol{Z}_{ij}$ represent a vector of the time-varying individual- and project-level control variables (such as platform experience, project duration, project earnings, skill match between the freelancer and the project, etc.). The baseline model is specified in the form of Equation 1:

$$Y_{ijt} = \beta X_j + \gamma E_j + \boldsymbol{\rho Z}_{ij} + \alpha_i + \mu_t + \varepsilon_{ijt} \quad (1)$$

where α_i and μ_t represent a set of freelancer and year fixed effects, respectively, and ε_{ijt} captures the idiosyncratic error. Because we hypothesize that client satisfaction is lower under a TM contract, we expect β to be negative. To examine the moderating role of the project's required expertise level, we add the interaction term $X_j * E_j$ to the model, leading to Equation 2:

$$Y_{ijt} = \beta X_j + \gamma E_j + \delta(X_j * E_j) + \boldsymbol{\rho Z}_{ij} + \alpha_i + \mu_t + \varepsilon_{ijt} \quad (2)$$

where coefficient δ captures the moderating effect.

Addressing Endogenous Contract Type Choice

Despite the use of freelancer fixed effects models and including relevant controls to rule out the effect of unobserved heterogeneities, some unobserved factors that influence the contract type choice may be also correlated with the client's perceived project performance, leading to potential biases in our estimation. We address this concern by treating the choice of contract type as endogenous and identifying a couple of variables that serve as instruments for the nonrandom assignment of contract type. Using these instruments, we employ a linear regression model with endogenous treatment effects to account for the endogeneity of the contract type (using the "*etregress*" command in Stata, StataCorp LP 2015). The model consists of a linear equation for the outcome (second stage) and a probit equation for the assignment of treatment (first stage), while the error terms of the two equations follow a bivariate normal distribution.

As an alternative identification strategy, we also employ a matching sample approach in which each treatment observation (a project under a TM contract) is matched to a control observation (a project under an FP contract) using a propensity score matching algorithm. Particularly, for each observation associated with a TM contract, we apply a one-to-one nearest neighbor matching without replacement to identify a matched control observation under an FP contract that is comparable in its probability of treatment assignment based on observed freelancer and project characteristics. We then test the regression models using the resulting matching sample. The use of this matching sample method therefore helps overcome issues of selection bias under our non-experimental setting.

V. Results

Baseline Results

We show the regression results from the fixed-effects models as specified in equations 1 and 2 in Table 3. We take the log transformation of *project duration, platform experience*, and *earnings* to account for the right skewness of the variables. We first run the baseline model with the main independent variables of interest, *contract type*, along with other control variables as predictors (Column I). We then add the interaction between *contract type* and *expertise* to examine the moderating effect of expertise level (Column II).

*** Insert Table 3 about here ***

In column I where we show the direct effect of contract type on client satisfaction, we find that compared to an FP contract, a freelancer's performance rating under a TM contract is lower by approximately 0.18 points on a 1-5 scale ($\beta = -0.182$, $p < 0.01$). Given the sample mean client satisfaction of 4.54, this translates to a 3.9% decrease in performance rating. The result is consistent with our argument that under a TM contract a freelancer has less incentive to execute the project and manage her progress efficiently since she is not responsible for time and material overruns, and thus moral hazard is more likely to occur. Unable to perfectly monitor the freelancer's effort, the client is uncertain about the degree of moral hazard and bears considerable risk under an hourly payment scheme, and therefore is less satisfied with the project outcome than under an FP contract. We further note that intermediate-level and expert-level projects on average receive a higher rating than entry-level projects ($\beta = 0.075$, $p < 0.01$ and $\beta = 0.241$, $p < 0.01$, respectively), and clients on average give higher ratings to freelancers with higher match with the project skillset requirement ($\beta = 0.169$, $p < 0.01$).

Column II shows the interaction effects of expertise levels and the TM contract type. We find that when a project requires intermediate-level or expert-level skills, the negative impact of a TM contract on the client's satisfaction is not as severe as a project that requires entry-level skills ($\beta = 0.186$, $p < 0.01$ and $\beta = 0.124$, $p < 0.01$, respectively). To evaluate Hypothesis 2, we conduct a likelihood ratio (LR) test comparing column I—which excludes the moderating effect of expertise—and column II—which includes the moderating effect—and the result provides strong support to the hypothesis ($\chi^2(2) = 47.34$, $p < 0.01$).

Based on results from column II, Figure 1 presents a plot showing the difference in the marginal effects of contract type on perceived performance under different levels of expertise. For an average entry-level project, the value of predicted client satisfaction under a TM contract is significantly lower than under an FP contract (*diff* = -0.305, on a scale of 1 to 5, $p < 0.01$). In contrast, the difference in predictive margins between the two contract types is not salient when the project requires intermediate-level skills (*diff* = -0.008, not significant), nor is it significant when the project requires expert-level of skills (*diff* = -0.039, not significant).

*** Insert Figure 1 about here ***

Models with Endogenous Treatment Effects

We further examine the degree to which our findings may suffer from estimation biases due to endogenous contract type and test the endogenous treatment effects model as described earlier. For this exercise, we identify a couple of instruments for the endogenous choice of contract type. First, because monitoring effort is greater if the client and the worker are geographically distant (McElheran 2014), a client is more inclined to use a TM contract if she is physically close to the freelancer she hires. OLMs make it possible for workers and clients from across the globe to connect, overcoming traditional geographical boundaries. As a result, workers and clients may have to work across different cultures and time zones, potentially generating more risk in the coordination process and increasing monitoring costs (Handley and Benton Jr 2013). We determine whether both contract parties are in the same country and use the binary variable *same country* as an instrument. Since the freelancers in our sample are all from the U.S., the *same country* variable is set to 1 if the client is also located in the U.S., and to 0 otherwise.

Second, we consider a significant change in the monitoring mechanism provided by *Upwork* during our sample period which involves the debut of a real-time chat service on the platform. This new service, introduced on May 5th, 2015, features *Slack*-like, real-time instant messaging capabilities, allowing clients to see if workers are online and start a conversation right away to discuss the project's progress.[8] We expect that the debut of the real-time monitoring feature would exogenously shift the likelihood of using a TM contract in a project. On the one hand, the introduction of such a mechanism reduces the cost of monitoring, leading to an increased propensity of clients employing a TM contract (Liang et al. 2019). On the other hand, earlier research shows that freelancers on gig platforms often resent *Slack*-like monitoring tools to protect their privacy, so much so that they may avoid bidding on hourly contracts altogether (Sutherland et al. 2020). Many freelancers are drawn to gig platforms by the promise of flexible work schedules, and real-time monitoring systems take the flexibility and autonomy away from the workers. Therefore, the effect of the debut of the new feature on the likelihood of a TM contract will likely depend on the interplay of the two countervailing forces. We create a dummy variable, *monitoring system*, as the second instrument, with its value set to 1 if a project has a start date later than May 15th, 2015, and to 0 otherwise.

The results of the endogenous treatment effects regressions, which make use of the two instruments, are reported in Table 4. In keeping with the prior literature (e.g., Liang et al. 2016), we also include project expertise levels as predictors in the first-stage contract choice equation. The first stage results show how the instrumental variables affect the choice of contract type. As expected, we find that the variable *same country* is positively associated with the use of a TM contract ($p < 0.01$), confirming that physical proximity reduces monitoring costs. In contrast, the variable *monitoring system* negatively predicts the selection of a TM contract ($p < 0.01$), suggesting that freelancers eschew TM contracts after the introduction of the real-time chat feature due to privacy concerns. The second stage regressions model the outcome equations with perceived contractual performance as the dependent variable. In the baseline model (Column I), we again find results in support of Hypothesis 1 that the use of TM contract negatively affects the perceived performance ($\beta = -0.198$, $p < 0.01$). In the full model (Column II), we find supportive evidence that the negative impact of a TM contract on client satisfaction is moderated by a project's expertise requirement: the difference in performance ratings between the two contract types are reduced for a project with

[8] For details, see https://techcrunch.com/2015/05/05/elance-odesk-rebrands-as-upwork-debuts-slack-like-chat-platform.

intermediate-level or expert-level skill requirement relative to an entry-level project ($\beta = 0.148$, $p < 0.01$ and $\beta = 0.087$, $p < 0.01$). Again, an LR test comparing the models of column I and column II provides support to Hypothesis 2 ($\chi^2(2) = 25.35$, $p < 0.01$). In summary, both hypotheses are supported under the endogenous treatment effects model, suggesting that our findings are robust to endogenous contract type choice.

*** Insert Table 4 about here ***

Other Robustness Tests

To further evaluate the robustness of our results, we test several additional model specifications. First, although the endogenous treatment regression model addresses the issue of nonrandom assignment of the contract type, it is less effective in controlling for unobserved freelancer heterogeneities because the model cannot accommodate freelancer fixed effects due to the well-known issue of incidental parameters problem (Lancaster 2000). We address this limitation by employing an Extended Regression Model (ERM) that accounts for both the correlation of observations within panels and the non-random treatment assignment.[9] Similar to the endogenous treatment regression model, we use both *same country* and *monitoring system* as excluded instruments in the first stage equation of the ERM that models the contract type choice. Different from the endogenous treatment model, ERM incorporates freelancer random effects into the estimation process to account for unobserved freelancer heterogeneities (StataCorp LP 2021).

We show the results from ERM in Table 5. The regression output consists of separate parameter estimates for two potential contract types, presented in Columns I (under a TM contract) and II (under an FP contract), respectively. The estimated correlation between the error terms of the contract choice equation and the performance equation is 0.17, and its associated z statistic rejects the null that there is no endogenous treatment ($p < 0.01$). In addition, post estimation calculations suggest that the average treatment effect (ATE) of the TM contract is -0.297 ($p < 0.01$), which lends support to Hypothesis 1.

*** Insert Table 5 about here ***

The moderating effect of project expertise level can be evaluated by testing the contrasts of the coefficients of the expertise variables under two different contract types. We calculate the contrasts associated with the covariates under TM and FP contract types. Consistent with our earlier findings, the difference in performance ratings between TM and FP contract types is smaller (less negative) for a project with intermediate-level skill requirement relative to an entry-level project ($\beta = 0.298$, $p < 0.01$), and it is

[9] The model is estimated using the *xteregress* command with the *entreat* option.

also smaller for a project with expert-level skill requirement relative to an entry-level project ($\beta = 0.250$, $p < 0.01$), supporting Hypothesis 2.

Second, in the regression models that we have tested thus far, the project expertise variable has been coded as a set of dummy variables, with one for each expertise level. As a result of multiple levels of expertise, the test of Hypothesis 2 has relied on LR tests comparing the restricted model without the moderating effect of expertise and the full model with the moderating effect. In Table 6 we present results from a set of models in which we treat the expertise variable as continuous (with values 1, 2, and 3 representing *entry*, *intermediate* and *expert* levels, respectively). This transformation results in a more intuitive way of testing the hypotheses, especially for Hypothesis 2 regarding the moderating effect of expertise. In both the fixed effects model and the endogenous treatment regression model, we find that the use of TM contract results in lower performance rating ($\beta = -0.183$, $p < 0.01$ in Columns I and $\beta = -0.184$, $p < 0.05$ in Column III), and that the negative effect of TM contract is weaker (less negative) when the project requires greater expertise ($\beta = 0.142$, $p < 0.01$ in Columns II and $\beta = 0.134$, $p < 0.01$ in Column IV), supporting both hypotheses.

*** Insert Table 6 about here ***

Matched Sample Analyses

To further address the non-random assignment of contract type, we construct a sample composed of a treatment group and a control group that is comparable on the probability of treatment assignment using propensity score matching (PSM), a method that has been employed in various non-experimental settings when the assignment of treatment is not controlled by the researcher (Dehejia and Wahba 2002). We first predict the propensity score of a project choosing the FP contract type using a logistic regression in which project- and individual-level covariates (such as skill similarity score, project expertise requirement, the programming language, project length, and the worker's platform experience, etc.) are used as explanatory variables. To minimize the bias in the estimated contract type effect, for every observation of an FP contract we apply a one-to-one nearest neighbor matching without replacement to identify a matched control observation under a TM contract (Austin et al. 2010). The PSM process results in 3,847 projects under an FP contract and a matched sample of projects under a TM contract with the same sample size. A balance check of the covariates is shown in Table 7, which reveals that the control sample and the treatment sample are not significantly different in observed freelancer and project characteristics after matching. Figure 2 shows the distributions of the propensity scores for the two subsamples before and after matching, again confirming that the control sample and treatment sample are similar after PSM.

*** Insert Table 7 and Figure 2 about here ***

We replicate the fixed effect models using the matched sample and report the results in columns I and II of Table 8. We find a consistent result that projects under a TM contract receive a lower performance rating compared to those under an FP contract on average (β = -0.195, $p < 0.01$), supporting Hypothesis 1. In addition, the effect of contract type is again shown to be moderated by project expertise requirement: the negative effect of a TM contract is weaker when a project requires intermediate-level skills (β = 0.167, $p < 0.01$) or expert level skills (β = 0.098, $p < 0.01$). An LR test comparing the models of column I and column II lends support to Hypothesis 2 as well ($\chi^2(2) = 24.73$, $p < 0.01$).

*** insert Table 8 about here ***

VI. Conclusions and Discussion

Due to their unique characteristics, OLMs are particularly susceptible to information asymmetry both before and after contracting (Kanat et al. 2018). Based on the economics of information, we advance arguments that contract type choice—i.e., between the fixed-price contract and the time-and-material contract—has important implications for preventing moral hazard during contract execution, and therefore will influence the client's perceived contractual performance upon project completion. We assemble a dataset of data analytics projects completed by freelancers on *Upwork* and empirically evaluate the propositions. We find that freelancers under a TM contract receive significantly lower ratings by their clients on average compared to those under an FP contract, consistent with our theorizing that the use of a TM contract leads to increased monitoring costs and the client's greater concerns over moral hazard (Liang et al. 2019). Notably, we also find that the expertise required for a project moderates the effect of contract choice on client satisfaction: particularly, the negative impact of TM contract is weaker when a project requires intermediate-level or expert-level skills. Our interpretation is that the degree of contract incompleteness and the difficulty in outcome verification are both increasing with project complexity (Al-Najjar 1995, Bapna et al. 2010). Therefore, for an expert-level project, the advantage of an FP contract over a TM contract in curbing moral hazard is greatly reduced.

Our study makes several contributions to the existing IS research. First, although earlier studies in the IT outsourcing literature have examined the various factors that contribute to the choice of contract type (e.g., Gopal et al. 2003, Kalnins and Mayer 2004), very few of them examine the implications of such choice on the outcome of a contractual relationship. By studying the relationship between the contract choice and the client's perceived contractual performance, we reveal how the incentive problems unfold under the different contract terms, which in turn influences vendor performance evaluation. Second, our analyses also add to the current understanding of the issue of information asymmetry in OLMs. Compared to the traditional labor market, OLMs have a lower entry barrier and lack effective screening and monitoring mechanisms, leading to more severe information asymmetry problems. Whereas much of this line of literature has been focusing on solutions to the adverse selection issues and the role of reputation in contractor selection in particular (e.g., Hong and Pavlou 2017, Lin et al. 2018, Moreno and Terwiesch 2014), our study approaches the research topic from a different angle and focuses on the issue of moral hazard that, with a few exceptions (e.g., Liang et al. 2016), has not been thoroughly investigated. Finally, earlier studies of OLMs reveal that clients frequently use the reputation systems in OLMs as a basis for vendor screening (Lin et al. 2018, Moreno and Terwiesch 2014), but the effectiveness of these reputation systems is predicated on the assumption that a freelancer's online reputation genuinely reflects her innate ability and/or work ethics. Our study challenges this assumption and points to the possibility that institutional factors—such as the contract type used for the project—may contribute to potential bias in the reputation generation process, and therefore clients should take the reputation ratings with a grain of salt.

Our research also reveals some important managerial implications for practitioners in OLMs. For example, freelancers' historical performance ratings are prominently displayed in their OLM profiles and form the basis of their OLM reputations (Lin et al. 2018, Moreno and Terwiesch 2014). Our findings help freelancers better understand how different contract types may impact their performance ratings, and these insights can be used to guide their bidding behavior and maintain their reputation. They inform freelancers to anticipate a lower performance evaluation when a TM contract is selected by the client, and they should take preemptive actions to alleviate the client's concerns over moral hazard under such conditions. In addition, to the extent that clients frequently make vendor selection decisions based on the bidders' reputation, they need to be cognizant of the potential biases caused by the contract type under which the freelancer had worked in the past and correct for such biases in their decision process if necessary. Our finding regarding the moderating effect of project expertise requirement suggests that the bias induced by contract type is most significant when the freelancer frequently works entry-level jobs under a TM contract, and clients need to be particularly mindful in screening bidders of this type. Finally, for OLM platforms, our study suggests that it might be helpful to present freelancers' performance ratings under different contract types separately in their profile and provide guidelines to assist clients in interpreting this information, which will increase transparency and aid the vendor screening process.

Several limitations of our study lead to avenues for future research. First, due to data availability, we cannot determine whether the lower performance evaluation under a TM contract is attributed to real moral hazard or the clients' perception of moral hazard (Bellavitis et al. 2019), whereas the latter can also be affected by other factors such as the variance in freelancers' innate abilities or the clients' unrealistic expectations. Second, because our sample consists of contractual relationships between non-enterprise clients and individual freelancers, one should exercise caution when generalizing the findings of this study to other IT outsourcing contexts where the contracting parties are commercial enterprises that often forge long-term contractual relationships with repeated interactions (Corts and Singh 2004, Gulati 1995). We hope our work will ignite sparks of interest in pursuing these potentially fruitful research directions.

References

Agrawal, A., J. Horton, N. Lacetera, E. Lyons. 2015. Digitization and the contract labor market: A research agenda. *Economic analysis of the digital economy*, University of Chicago Press, 219-250.

Al-Najjar, N.I. 1995. Incomplete contracts and the governance of complex contractual relationships. *The American Economic Review*. **85**(2) 432-436.

Aubert, B.A., M. Patry, S. Rivard. 2002. Managing IT outsourcing risk: lessons learned. *Information Systems Outsourcing*, Springer, 155-176.

Austin, P.C., M. Chiu, D.T. Ko, R. Goeree, J.V. Tu. 2010. Propensity score matching for estimating treatment effects. *Analysis of observational health care data using SAS*. **3** 51-84.

Baier, E., C. Rammer, T. Schubert. 2015. The impact of captive innovation offshoring on the effectiveness of organizational adaptation. *Journal of International Management*. **21**(2) 150-165.

Bajari, P., S. Tadelis. 2001. Incentives versus transaction costs: A theory of procurement contracts. *Rand Journal of Economics*. **32**(3) 387-407.

Bapna, R., A. Barua, D. Mani, A. Mehra. 2010. Research commentary—Cooperation, coordination, and governance in multisourcing: An agenda for analytical and empirical research. *Information Systems Research*. **21**(4) 785-795.

Baron, D.P., D. Besanko. 1987. Monitoring, moral hazard, asymmetric information, and risk sharing in procurement contracting. *The RAND Journal of Economics*. **18**(4) 509-532.

Bellavitis, C., D.S. Kamuriwo, U. Hommel. 2019. Mitigation of Moral Hazard and Adverse Selection in Venture Capital Financing: The Influence of the Country's Institutional Setting. *Journal of Small Business Management*. **57**(4) 1328-1349.

Benson, A., A. Sojourner, A. Umyarov. 2020. Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Science*. **66**(5) 1802-1825.

Borchert, K., M. Hirth, M.E. Kummer, U. Laitenberger, O. Slivko, S. Viete. 2018. Unemployment and online labor. *ZEW-Centre for European Economic Research Discussion Paper, No. 18-023*.

Burtch, G., Q. He, Y. Hong, D. Lee. 2021. How Do Peer Awards Motivate Creative Content? Experimental Evidence from Reddit. *Management Science*. **Articles In Advance**.

Cantarella, M., C. Strozzi. 2021. Workers in the crowd: the labor market impact of the online platform economy. *Industrial and Corporate Change*. **30**(6) 1429-1458.

Chen, Y., A. Bharadwaj. 2009. An empirical analysis of contract structures in IT outsourcing. *Information Systems Research*. **20**(4) 484-506.

Conrow, E.H., P.S. Shishido. 1997. Implementing risk management on software intensive projects. *IEEE Software*. **14**(3) 83-89.

Corts, K.S., J. Singh. 2004. The effect of repeated interaction on contract choice: Evidence from offshore drilling. *Journal of Law, Economics, and Organization*. **20**(1) 230-260.

David, H. 2001. Wiring the labor market. *Journal of Economic Perspectives*. **15**(1) 25-40.

Dehejia, R.H., S. Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*. **84**(1) 151-161.

Dey, D., M. Fan, C. Zhang. 2010. Design and analysis of contracts for software outsourcing. *Information Systems Research*. **21**(1) 93-114.

Doll, W.J., G. Torkzadeh. 1988. The measurement of end-user computing satisfaction. *MIS Quarterly*. **12**(2) 259-274.

Eisenhardt, K.M. 1989. Agency theory: An assessment and review. *Academy of management review*. **14**(1) 57-74.

Goes, P., M. Lin. 2012. Does information really "unravel"? Understanding factors that motivate sellers to seek third-party certifications in an online labor market. *NET Institute Working Paper No. 12-02*.

Gopal, A., B.R. Koka. 2012. The asymmetric benefits of relational flexibility: Evidence from software development outsourcing. *MIS Quarterly*. **36**(2) 553-576.

Gopal, A., K. Sivaramakrishnan, M.S. Krishnan, T. Mukhopadhyay. 2003. Contracts in offshore software development: An empirical analysis. *Management Science*. **49**(12) 1671-1683.

Grover, V., M.J. Cheon, J.T. Teng. 1996. The effect of service quality and partnership on the outsourcing of information systems functions. *Journal of Management Information Systems*. **12**(4) 89-116.

Gulati, R. 1995. Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*. **38**(1) 85-112.

Handley, S.M., W. Benton Jr. 2013. The influence of task-and location-specific complexity on the control and coordination costs in global outsourcing relationships. *Journal of Operations Management*. **31**(3) 109-128.

Hass, R.W. 2017. Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications. *Memory & Cognition*. **45**(2) 233-244.

Hong, Y., P.A. Pavlou. 2017. On buyer selection of service providers in online outsourcing platforms for IT services. *Information Systems Research*. **28**(3) 547-562.

Horton, J.J. 2010. Online labor markets. *The 6th Workshop on Internet and Network Economics*.

Jensen, P.D.Ø. 2009. A learning perspective on the offshoring of advanced services. *Journal of International Management*. **15**(2) 181-193.

Kalnins, A., K.J. Mayer. 2004. Relationships and hybrid contracts: An analysis of contract choice in information technology. *Journal of Law, Economics, and Organization*. **20**(1) 207-229.

Kanat, I., Y. Hong, T. Raghu. 2018. Surviving in global online labor markets for IT services: A geo-economic analysis. *Information Systems Research*. **29**(4) 893-909.

Kim, D.R., M.J. Chen, M. Aiken. 2005. Towards an understanding of the relationship between IS outsourcing vendors' service quality and outsourcing effects. *International Journal of Information Technology and Management*. **4**(1) 12-24.

Lancaster, T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics*. **95**(2) 391-413.

Larsen, M.M., S. Manning, T. Pedersen. 2013. Uncovering the hidden costs of offshoring: The interplay of complexity, organizational design, and experience. *Strategic Management Journal*. **34**(5) 533-552.

Levina, N., J.W. Ross. 2003. From the vendor's perspective: Exploring the value proposition in information technology outsourcing. *MIS Quarterly*. **27**(3) 331-364.

Liang, C., Y. Hong, B. Gu. 2016. Effects of IT-enabled monitoring systems in online labor markets. *Proceedings of International Conference on Information Systems*, Dublin, Ireland.

Liang, C., J. Peng, Y. Hong, B. Gu. 2019. Monitoring Policies and Gig Workers' Job Preferences. *SSRN Working Paper No. 3772671*.

Lin, M., Y. Liu, S. Viswanathan. 2018. Effectiveness of reputation in contracting for customized production: Evidence from online labor markets. *Management Science*. **64**(1) 345-359.

Macho-Stadler, I., J.D. Pérez-Castrillo. 2001. *An introduction to the economics of information: incentives and contracts*. Oxford University Press on Demand.

Mason, W., D.J. Watts. 2009. Financial incentives and the" performance of crowds". *Proceedings of the ACM SIGKDD workshop on human computation*.

McElheran, K. 2014. Delegation in Multi-Establishment Firms: Evidence from I.T. Purchasing. *Journal of Economics & Management Strategy*. **23**(2) 225-258.

Moreno, A., C. Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research*. **25**(4) 865-886.

Nieto, M.J., A. Rodríguez. 2011. Offshoring of R&D: Looking abroad to improve innovation performance. *Journal of International Business Studies*. **42**(3) 345-361.

Osei-Bryson, K.-M., O.K. Ngwenyama. 2006. Managing risks in information systems outsourcing: An approach to analyzing outsourcing risks and structuring incentive contracts. *European Journal of Operational Research*. **174**(1) 245-264.

Parasuraman, A., V.A. Zeithaml, L. Berry. 1988. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*. **64**(1) 12-40.

Pelletier, A., C. Thomas. 2018. Information in online labour markets. *Oxford Review of Economic Policy*. **34**(3) 376-392.

StataCorp LP. 2015. *Stata treatment-effects reference manual*. Stata Press, College Station, TX.

StataCorp LP. 2021. *Stata longitudinal-data/panel-data reference manual: release 17*. Stata Press, College Station, TX.

Stiglitz, J.E. 2000. The contributions of the economics of information to twentieth century economics. *The Quarterly Journal of Economics*. **115**(4) 1441-1478.

Susarla, A., R. Subramanyam, P. Karhade. 2010. Contractual provisions to mitigate holdup: Evidence from information technology outsourcing. *Information Systems Research*. **21**(1) 37-55.

Sutherland, W., M.H. Jarrahi, M. Dunn, S.B. Nelson. 2020. Work precarity and gig literacies in online freelancing. *Work, Employment and Society*. **34**(3) 457-475.

Yao, T., B. Jiang, S.T. Young, S. Talluri. 2010. Outsourcing timing, contract selection, and negotiation. *International Journal of Production Research*. **48**(2) 305-326.

Yin, M., Y. Chen, Y.-A. Sun. 2013. The effects of performance-contingent financial incentives in online labor markets. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Yoon, S., H. Suh. 2004. Ensuring IT consulting SERVQUAL and user satisfaction: a modified measurement tool. *Information Systems Frontiers*. **6**(4) 341-351.

Zheng, A., Y. Hong, P. Pavlou. 2015. Value uncertainty and buyer contracting: Evidence from online labor markets. *Proceedings of International Conference on Information Systems*, Fort Worth, TX.

Tables and Figures

Table 1. Summary Statistics

Variable	Unit or Range	Mean	Std. Dev.	Min	Max
client satisfaction	1-5	4.541	0.772	1.000	5.000
TM contract	binary	0.497	0.500	0.000	1.000
expertise requirement					
entry	binary	0.521	0.500	0.000	1.000
intermediate	binary	0.188	0.391	0.000	1.000
expert	binary	0.291	0.454	0.000	1.000
skill match	0-1	0.843	0.190	0.000	1.000
project duration	log of days	2.905	1.703	0.000	7.679
platform experience	log of days	6.909	1.008	0.693	8.303
earnings	log of $	6.465	1.845	0.000	12.091
first-time interaction	binary	0.882	0.322	0.000	1.000
programming language					
Python	binary	0.367	0.482	0.000	1.000
R	binary	0.166	0.372	0.000	1.000
JavaScript	binary	0.095	0.293	0.000	1.000
Google analytics	binary	0.058	0.234	0.000	1.000
VBA	binary	0.086	0.281	0.000	1.000
Tableau	binary	0.041	0.199	0.000	1.000
SQL	binary	0.134	0.341	0.000	1.000
AWS	binary	0.052	0.222	0.000	1.000
same country	binary	0.918	0.274	0.000	1.000
monitoring system	binary	0.895	0.306	0.000	1.000

Note: The summary statistics are based on 12,388 projects completed by 1,075 freelancers.

Table 2. Correlation Table

		1	2	3	4	5	6	7	8	9	10	11
1	client satisfaction	1.000										
2	TM contract	-0.103*	1.000									
3	expertise = entry	-0.105*	0.184*	1.000								
4	expertise = intermediate	-0.002	-0.048*	-0.503*	1.000							
5	expertise = expert	0.117*	-0.161*	-0.666*	-0.310*	1.000						
6	skill match	-0.002	0.050*	0.030*	-0.011	-0.024*	1.000					
7	project duration	-0.079*	0.164*	0.840*	-0.186*	-0.763*	0.019*	1.000				
8	platform experience	-0.011	0.018*	-0.004	0.003	0.002	0.017	0.061*	1.000			
9	earnings	-0.093*	0.259*	0.522*	-0.211*	-0.392*	0.067*	0.426*	0.070*	1.000		
10	first-time interaction	-0.063*	0.073*	0.391*	0.182*	-0.588*	-0.005	0.412*	-0.009	0.143*	1.000	
11	same country	-0.053*	0.053*	0.320*	0.149*	-0.481*	-0.008	0.339*	-0.002	0.109*	0.716*	1.000
12	monitoring system	-0.036*	0.010	-0.024*	0.003	0.024*	0.019*	-0.029*	0.007	-0.022*	0.009	0.015

Note: * $p < 0.05$.

Table 3. Fixed Effects Model

	DV = Client Satisfaction	
	I	II
TM contract	-0.182***	-0.230***
	(0.017)	(0.019)
expertise = intermediate	0.075***	-0.016
	(0.025)	(0.029)
expertise = expert	0.241***	0.184***
	(0.029)	(0.030)
intermediate X TM contract		0.186***
		(0.032)
expert X TM contract		0.124***
		(0.028)
skill match	0.169***	0.162***
	(0.044)	(0.045)
earnings	0.003	0.003
	(0.005)	(0.005)
project duration	0.010	0.011*
	(0.006)	(0.006)
platform experience	0.010	0.010
	(0.008)	(0.008)
first-time interaction	-0.003	-0.004
	(0.023)	(0.023)
constant	4.303***	4.346***
	(0.117)	(0.117)
Observations	12,388	12,388
R-squared	0.232	0.235
Freelancer FE	YES	YES
Programming language dummies	YES	YES
Year FE	YES	YES
Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Point estimates of programming language dummies are suppressed.		

Table 4. Endogenous Treatment Effects Model

Second Stage	**DV = Client Satisfaction**	
	I	II
TM contract	-0.198***	-0.227***
	(0.067)	(0.077)
expertise = intermediate	0.052***	-0.007
	(0.020)	(0.023)
expertise = expert	0.162***	0.137***
	(0.023)	(0.024)
intermediate X TM contract		0.148***
		(0.029)
expert X TM contract		0.087***
		(0.025)
skill match	0.023	0.021
	(0.033)	(0.033)
earnings	-0.017***	-0.016***
	(0.004)	(0.004)
project duration	0.009	0.012**
	(0.006)	(0.006)
platform experience	-0.006	-0.007
	(0.007)	(0.007)
first-time interaction	-0.031	-0.030
	(0.021)	(0.021)
constant	4.779***	4.791***
	(0.105)	(0.108)
First Stage	**DV = TM contract**	
expertise = intermediate	-0.090***	
	(0.030)	
expertise = expert	-0.203***	
	(0.028)	
same country	0.126***	
	(0.045)	
monitoring system	-0.494***	
	(0.139)	
Observations	12,388	
Programming language dummies	YES	
Year FE	YES	
Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Point estimates of programming language dummies are suppressed.		

Table 5. Panel-data Extended Regression Model

Second Stage	DV = Client Satisfaction	
	I	II
	TM Contract	FP Contract
expertise = intermediate	0.209***	-0.089**
	(0.051)	(0.038)
expertise = expert	0.336***	0.086**
	(0.055)	(0.043)
skill match	0.027	0.140**
	(0.061)	(0.056)
project duration	0.012	0.008
	(0.009)	(0.007)
platform experience	0.006	0.001
	(0.011)	(0.009)
earnings	-0.002	0.006
	(0.008)	(0.007)
first-time interaction	-0.045	-0.005
	(0.038)	(0.025)
First Stage	**DV= TM contract**	
expertise = intermediate	-0.265***	
	(0.052)	
expertise = expert	-0.419***	
	(0.047)	
same country	0.053	
	(0.068)	
monitoring system	-0.166	
	(0.134)	
constant	0.228	
	(0.153)	
Observations	12,388	
Freelancer RE	YES	
Programming language dummies	YES	
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Point estimates of programming language dummies are suppressed.		

Table 6. Continuous Measure of Expertise

	Fixed effects models		Endogenous treatment regressions Second stage	
	I	II	III	IV
TM contract	-0.183***	-0.438***	-0.184**	-0.787***
	(0.017)	(0.037)	(0.088)	(0.228)
expertise	0.119***	0.058***	0.078***	0.022*
	(0.015)	(0.015)	(0.011)	(0.013)
expertise X TM contract		0.142***		0.134***
		(0.017)		(0.015)
skill match	0.168***	0.160***	0.023	0.018
	(0.045)	(0.045)	(0.033)	(0.033)
project duration	0.004	0.008*	-0.016***	-0.011**
	(0.005)	(0.005)	(0.004)	(0.004)
platform experience	0.011*	0.013**	0.010*	0.011**
	(0.006)	(0.006)	(0.006)	(0.006)
earnings	0.010	0.011	-0.007	-0.003
	(0.008)	(0.008)	(0.007)	(0.007)
first-time interaction	-0.017	-0.018	-0.040**	-0.048**
	(0.022)	(0.022)	(0.020)	(0.020)
constant	4.176***	4.266***	4.706***	4.738***
	(0.122)	(0.122)	(0.108)	(0.127)
Observations	12,388	12,388	12,388	12,388
R-squared	0.229	0.234	-	-
Freelancer FE	YES	YES	NO	NO
Year FE	YES	YES	YES	YES
Programming language dummies	YES	YES	YES	YES
Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Point estimates of programming language dummies are suppressed.				

Table 7. PSM Balance Check

	Contract Type = TM		Contract Type = FP		t-test	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	t	p>t
expertise = intermediate	0.219	0.413	0.205	0.403	1.51	0.132
expertise = expert	0.317	0.466	0.301	0.459	1.53	0.126
skill match	0.839	0.196	0.845	0.185	-1.41	0.159
project duration	2.812	1.650	2.823	1.629	-0.31	0.755
platform experience	6.908	1.016	6.914	1.019	-0.25	0.801
earnings	6.277	1.686	6.333	1.534	-1.52	0.128
programming language						
Python	0.381	0.486	0.376	0.484	0.44	0.659
R	0.152	0.359	0.159	0.366	-0.84	0.401
JavaScript	0.102	0.302	0.096	0.295	0.72	0.474
Google analytics	0.061	0.240	0.059	0.236	0.39	0.698
VBA	0.080	0.272	0.086	0.280	-0.76	0.448
Tableau	0.041	0.198	0.039	0.193	0.44	0.661
SQL	0.132	0.338	0.134	0.341	-0.30	0.765
AWS	0.051	0.220	0.051	0.220	0.04	0.965
first-time interaction	0.892	0.310	0.881	0.323	1.48	0.140
Observation	3,847		3,847		7,694	

Table 8. Matched Sample Analyses

	Fixed effects models	
	I	II
TM contract	-0.195***	-0.237***
	(0.022)	(0.024)
expertise = intermediate	0.061*	-0.024
	(0.033)	(0.036)
expertise = expert	0.231***	0.177***
	(0.037)	(0.039)
intermediate X TM contract		0.167***
		(0.040)
expert X TM contract		0.098***
		(0.034)
skill match	0.174***	0.162***
	(0.060)	(0.060)
earnings	0.007	0.007
	(0.007)	(0.007)
project duration	-0.000	-0.002
	(0.008)	(0.008)
platform experience	0.008	0.008
	(0.010)	(0.010)
first-time interaction	0.002	0.001
	(0.030)	(0.029)
constant	4.247***	4.299***
	(0.156)	(0.156)
Observations	7,694	7,694
R-squared	0.271	0.273
Freelancer FE	YES	YES
Programming language dummies	YES	YES
Year FE	YES	YES
Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Point estimates of programming language dummies are suppressed.		

Figure 1. Predictive Margins of Expertise Level

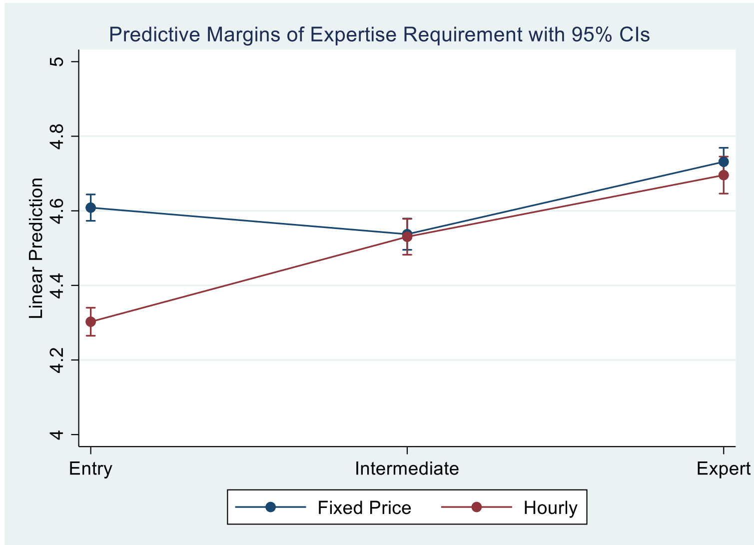

Figure 2. Kernel Density Plots Before and After Matching