




Monitoring and the Cold Start Problem in Digital Platforms: Theory and Evidence from Online Labor Markets

Chen Liang,^a Yili Hong,^b Bin Gu^{c,*}
^aSchool of Business, University of Connecticut, Storrs, Connecticut 06269; ^bMiami Herbert Business School, University of Miami, Coral Gables, Florida 33146; ^cQuestrom School of Business, Boston University, Boston, Massachusetts 02215

*Corresponding author

Contact: chenliang@uconn.edu,  <https://orcid.org/0000-0002-8444-8050> (CL); khong@miami.edu,  <https://orcid.org/0000-0002-0577-7877> (YH); bgu@bu.edu,  <https://orcid.org/0000-0002-0396-8899> (BG)

Received: March 12, 2021

Revised: December 29, 2022;
November 20, 2023

Accepted: January 1, 2024

Published Online in Articles in Advance:
March 6, 2024

<https://doi.org/10.1287/isre.2021.0146>

Copyright: © 2024 INFORMS

Abstract. Many online labor platforms employ reputation systems and monitoring systems to mitigate moral hazard. Whereas reputation systems have the potential to reduce moral hazard, they suffer from the cold-start problem, in which new entrants without an established reputation face a high entry barrier as employers predominantly select workers based on their existing reputation. Monitoring systems, providing employers with direct oversight of workers' actions, offer a different approach. By tracking and reporting workers' effort levels, monitoring systems reduce ex post information asymmetry and, thus, lower employers' expected moral hazard risk from workers. However, unlike reputation systems, monitoring systems do not directly address ex ante information asymmetry, failing to assist employers in identifying the right workers. This inherent limitation raises questions about their effectiveness in resolving the cold-start problem. In this paper, we first propose a stylized theoretical model that characterizes worker entry in the presence of reputation and monitoring systems. Based on a unique data set from a leading online labor platform, we then empirically investigate the effect of monitoring systems on the entry barriers by examining the change in workers' entry behaviors after the introduction of the monitoring system along with associated project outcomes, which include employers' hiring preferences, hiring prices, and project performance. We exploit the differential availability of the monitoring system across two project types: time-based projects, for which the monitoring system is accessible, and fixed-price projects, for which it is not. Employing a difference-in-differences estimation with a sample including 9,344 fixed-price projects and 3,118 time-based projects, we report that the introduction of the monitoring system increases the number of bids on time-based projects by 27.8%, and the incremental bids predominantly originate from inexperienced workers who lack platform reputation. We further find that, following the introduction of the monitoring system, employers' preference for experienced workers diminishes, accompanied by an average reduction of 19.5% in labor costs, whereas we observe no significant decrease in project completion and review rating. Our results collectively suggest that monitoring systems alleviate the cold-start problem in online platforms.

History: Xiaoquan (Michael) Zhang, Senior Editor; Beibei Li Associate Editor.

Funding: This work was supported by the NET Institute Grant and the Robert Wood Johnson Foundation (RWJF) [Grant 74503].

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/isre.2021.0146>.

Keywords: cold-start problem • online platforms • monitoring systems • entry barrier • reputation systems

1. Introduction

Moral hazard is a long-standing problem with a significant economic cost to market participants. According to a recent study, workers' idle time due to shirking costs U.S. employers more than one billion dollars per year (Brodsky and Amabile 2018). The shirking problem is especially salient for remote work in online platforms in which employers and workers are strangers and employers have limited control and high information asymmetry over workers' effort (Moreno and Terwiesch 2014).¹

A common solution for the moral hazard problem in online platforms is the reputation system. Specifically, reputation systems share workers' past performance information with prospective employers. They reduce employers' uncertainty about workers' potential shirking behavior (Banker and Hwang 2008, Tadelis 2016) by serving as a sanction mechanism that curbs moral hazard behaviors (e.g., Dellarocas 2006, Hui et al. 2016). However, recent research identifies a serious unintended consequence of reputation systems: that is, they create an

entry barrier for qualified entrants who have not yet established their platform reputation, also known as the cold-start problem (Pallais 2014, Butler et al. 2020).

The cold-start problem in online platforms relates to the literature on entry deterrence, which suggests that reputation is a type of intangible resource that can bring a sustainable competitive advantage for incumbents (Bunch and Smiley 1992, Gruca and Sudharshan 1995, Gao et al. 2017). This issue stems from *ex ante* information asymmetry, with which workers have private information about their abilities and work ethic, etc., which remains unknown to employers during hiring (Shapiro 1982, Kokkodis and Ipeirotis 2016). Consequently, employers often rely on reputation to aid in identifying good workers, leading to their reluctance to engage with new entrants and creating significant entry barriers (Farrell 1986, Dellarocas et al. 2006).

The cold-start problem is particularly significant for online labor platforms on which employers constantly look for workers. Screening workers based on reputation causes inefficient hiring because newly registered workers lack the opportunities to demonstrate their abilities and diligence (Pallais 2014). The inability for workers to transfer their reputation between different online platforms further exacerbates the problem. This is especially true in online labor platforms in which many workers are moonlighters² and multihoming (Wood et al. 2019, International Labour Organization 2021). From the platform's perspective, the cold-start problem can significantly slow down platform growth and negatively affect competition. First, given that all workers have capacity limits (Horton 2019), the cold-start problem prevents online labor platforms from continuously attracting new workers to fulfill the labor demand of an increasing number of employers.³ This suggests that it is important for online labor platforms to attract and onboard more new workers by alleviating the cold-start problem. Second, given that employers and workers are commonly multihoming (Wood et al. 2019, International Labour Organization 2021), online labor platforms all attempt to prevent their existing workers from multihoming on or migrating to rival platforms, attracting workers on rival platforms to enter their own platforms (Li and Zhu 2021). As such, platforms with a more serious cold-start problem are in a more disadvantaged position in competing with other platforms. To the best of our knowledge, there is limited attention in the literature (Hui et al. 2020) on mechanisms that can alleviate the cold-start problem caused by reputation systems.

Unlike reputation systems that rely on the codification of workers' history-dependent performance, which creates biases against new entrants, monitoring systems facilitate the direct observation of workers' actions in current projects and do not differentiate new entrants (referred to as inexperienced workers) from established workers. In the traditional transaction cost economics literature (e.g., Williamson 1981, Bajari and Tadelis 2001),

monitoring is often assumed to be manual and costly. However, with the advance of information technology, automated monitoring systems are nowadays increasingly prevalent and less costly. For example, with the increasing adoption of remote work arrangements since the COVID-19 pandemic, 60% of U.S. firms with remote workers use work-monitoring systems.⁴ Such monitoring systems typically provide employers with firsthand, real-time information on workers' progress by automatically recording screenshots, webcam images, and even keystrokes from workers' devices.⁵ Owing to the transparency of workers' actions to employers, workers refrain from moral hazard behavior because monitoring evidence of shirking may lead to payment disputes or employment termination (Liang et al. 2023).

Existing literature on monitoring predominantly examines its impact on workers' performance in traditional labor market scenarios in which monitoring is implemented after the employment arrangement (Hubbard 2000, Duflo et al. 2012, Pierce et al. 2015, Staats et al. 2017). In such contexts, the hired workforce is predetermined, and employers only encounter *ex post* information asymmetry. This setup precludes the possibility for employers to strategically hire different workers or for workers to strategically bid for different projects. However, unlike traditional labor markets, online labor platforms typically involve short-term transactions between strangers with both employers and workers engaged in hiring and bidding decisions on a project-to-project basis. The introduction of monitoring systems in these platforms has the potential to markedly shape both employers' strategic hiring behavior and workers' strategic bidding behavior, which has received little scholarly attention to date.

In contrast to reputation systems, monitoring systems do not address *ex ante* information asymmetry, failing to assist employers in identifying good workers. This raises an important question: can monitoring systems effectively mitigate the cold-start problem in online labor markets? Furthermore, it's essential to understand the impact of these systems on employment contracting outcomes: do they enhance project performance, reduce hiring costs, or both? Such critical questions bear profound implications for both the platform and its stakeholders, including employers and workers.

In this paper, we take a first step toward investigating whether and the degree to which monitoring systems alleviate the cold-start problem in online platforms by examining the associated changes in worker entry. We employ a simple conceptual framework in which the quality of labor service remains only partially observable (Shapiro 1982, Laouénan and Rathelot 2022). Notably, the unobservable quality component is predominantly related to the worker's effort, thereby moral hazard. By explicitly modeling the labor service supply and demand, differentiating between inexperienced and experienced workers, we elucidate how monitoring systems can

potentially change market dynamics and, thus, affect workers' bid entry decisions by lowering the moral hazard risk. In particular, with regard to hiring experienced workers, because reputation systems have already helped reduce the moral hazard risk, the incremental elevation in employers' expectations concerning their labor service quality, owing to monitoring, is inherently limited. By comparison, hiring inexperienced workers used to be considered as high risk because of the high uncertainty in their effort before monitoring systems are available. Once monitoring systems are introduced, employers are less concerned about the quality of inexperienced workers' labor service than before because they are now able to keep better track of worker effort. As such, monitoring systems are expected to have a stronger positive effect on employers' hiring of inexperienced workers and make the platform a more level playing field, thereby attracting more bids from inexperienced workers.

To empirically test our theoretical conjectures, we leverage a quasi-experiment when a leading online labor platform officially introduced a monitoring system on February 5, 2014. We leverage two types of projects (i.e., time-based projects in which workers are paid for the number of working hours and fixed-price projects in which workers are paid at a fixed-price) on the platform that were differentially impacted by the introduction of the monitoring system. Our econometric identification hinges on the fact that monitoring was only available for time-based projects and not for fixed-price projects, which allows us to employ the difference-in-differences (DID) estimation. Leveraging a data set including 12,462 projects posted on this platform, we first used the covariate balancing inverse probability of treatment weighting (IPTW) and coarsened exact matching (CEM) methods to match and weight fixed-price and time-based projects. In the resulting matched and reweighted sample, two groups of projects are similar in terms of all observable characteristics (e.g., category, required skills, description keywords, length of description, and employer tenure). We then use DID models to identify the effect of the introduction of the monitoring system on workers' entries. We find that, using comparable fixed-price projects as the control, the introduction of the monitoring system increases the number of bids (entries) in time-based projects by an average of 27.8%. Our further analysis shows that the number of bids (for time-based projects) from inexperienced workers increases by 44.5%, whereas the increase in the number of bids from experienced workers is much smaller and only sporadically significant. We further conduct a series of comprehensive robustness checks, including endogenous treatment effect model, instrumental variable (IV) analysis, doubly robust (DR) estimator, interrupted time series (ITS) analysis, and sensitivity analysis. And we observe highly robust and consistent results.

Furthermore, our additional analyses on employment contracting show that the introduction of the monitoring

system reduces employers' preference for experienced workers and, thus, lowers hiring prices by 19.5% on average. Interestingly, the overall project performance does not suffer as the hiring prices decrease, suggesting an improvement in cost efficiency. Overall, our results suggest that monitoring systems reduce the entry barrier for inexperienced workers and, thus, help alleviate the cold-start problem.

This paper contributes to two streams of literature. First, we contribute to the literature on online platforms by considering and demonstrating that monitoring systems lower the entry barrier for inexperienced workers. Although the cold-start problem is recognized as an important impediment to the future development of online platforms (Pallais 2014), little is known on how platforms can use market design to alleviate this problem. Our study expands on the extant literature by providing an actionable solution for platforms, that is, monitoring. Second, our study advances the prior literature on monitoring that primarily focuses on the effect of monitoring on workers' performance after the employment arrangement is made with a workforce that is predetermined (Pierce et al. 2015, Staats et al. 2017). We contribute to the increasingly important monitoring literature (Liang et al. 2023) by rigorously examining the impact of monitoring on workers' strategic bidding decisions (increased bids for projects with monitoring among inexperienced workers) and employers' hiring decisions (greater willingness to hire inexperienced workers and reduced inclination to pay reputation premiums) as well as labor costs and project delivery. Monitoring in online settings is increasingly ubiquitous and important today⁶ as remote work has emerged as a prevalent work arrangement in modern society. This study deepens our understanding of how monitoring can affect the competition and dynamics in online labor markets.

2. Theoretical Background

Because prospective employers cannot perfectly observe the quality of each worker's labor service, the moral hazard (hidden actions) problem arises. Prospective employers typically infer the labor service quality based on some observable worker characteristics (e.g., various types of verification, preferred worker badge) and existing reviews. In this section, we first characterize how prospective employers' learning the quality of the worker's labor service from existing reviews can lead to the cold-start problem with a stylized theoretical model when the worker has outside options. Following this model, we explain how the introduction of the monitoring system can alleviate the cold-start problem.

2.1. Moral Hazard and the Cold-Start Problem

2.1.1. Workers' Labor Allocation. Given that most workers in an online labor platform have their outside options (e.g., regular full-time/part-time jobs), a worker

needs to allocate labor (working time) between two activities: launching the worker's career in the given online labor platform (looking for projects, submitting bids, working on the hired projects, etc.) and working for outside options (e.g., regular jobs). Here, L is the amount of labor (time) allocated to the given online labor platform, and $1 - L$ is the amount dedicated to the outside option with the wage W . Similar to most common input elements for production (e.g., labor, capital) (Horton 2019, Laouénan and Rathelot 2022), we assume the worker's labor input in the given online labor platform has decreasing returns to scale and labor service output is equal to L^α ($0 < \alpha < 1$). Therefore, the worker's revenue at each period (e.g., one month) is

$$PL^\alpha + W(1 - L) \text{ with } \alpha \in (0, 1).$$

From the employers' perspective, they choose workers mainly based on two attributes: price P and quality Q (Che 1993, Adomavicius et al. 2012). The demand D for a worker's labor service with the quality Q and price P in the online labor platform is as follows:

$$D = \frac{Q^m}{P^n} \text{ with } m > 0 \text{ and } n > 0.$$

In the given platform, the worker of service quality Q set a service price P and the labor (time) allocated to the platform L to maximize revenue, subject to the demand constraint. Thus, in a scenario in which L holds positive values,⁷ the optimal values for log service price p and log demand d (which corresponds to the log labor service output $\alpha \log L$ because of market clearing) can be derived as follows:

$$p = \beta p_0 + \beta w + \beta \gamma q, \quad (1)$$

$$d = -n\beta p_0 - n\beta w + \tau q, \quad (2)$$

where $p = \log P$, $w = \log W$, $q = \log Q$, $\beta = \frac{1}{1+\frac{n}{\alpha}-n}$, $p_0 = \log\left(\frac{n}{\alpha(n-1)}\right)$, $\gamma = m\left(\frac{1}{\alpha} - 1\right)$, and $\tau = \frac{m}{1+n\left(\frac{1}{\alpha} - 1\right)}$. Given that $\alpha \in (0, 1)$, $m > 0$, and $n > 0$, this indicates that $\beta > 0$, $p_0 > 0$, $\gamma > 0$, and $\tau > 0$. A detailed proof of Equations (1) and (2) can be found in the online theoretical supplementary appendix.

2.1.2. Moral Hazard and Employers' Expectation of Quality. We assume that the quality of the worker labor service q is the sum of the two orthogonal components: $q = \delta + u$, where δ is directly observable by prospective employers and u is not. The unobserved quality component is more related to workers' effort and thereby moral hazard. According to the prior literature, most of the worker's labor service quality is unobservable (Shapiro 1982, Kokkodis and Ipeiritos 2016), suggesting that prospective employers' belief about u plays a primary role in online labor employment. To help employers learn about u , most online labor platforms provide reputation systems.

Following Laouénan and Rathelot (2022), we assume that prospective employers' prior belief about the distribution of the labor service quality is $N(\bar{u}_0, \sigma_{u_0}^2)$ before observing any reputation signals of workers.⁸ In particular, because inexperienced workers have not accumulated any reputation on the platform yet, the distribution of their labor service quality is $N(\bar{u}_0, \sigma_{u_0}^2)$. Further, we model how employers update their beliefs when observing the reputation signals of experienced workers (Laouénan and Rathelot 2022). We assume that each project review transmits a signal, which is a random draw around u in a normal distribution with the variance σ^2 . For a worker with K existing reviews, prospective employers observe the average signal transmitted by all the worker's existing reviews as g with the variance σ^2/K . Let $\rho = \sigma^2/\sigma_{u_0}^2$, and prospective employers' expectation about u is the weighted average between the prior \bar{u}_0 , and the average reputation signal g becomes

$$E(u|g, K) = \frac{Kg + \rho\bar{u}_0}{K + \rho}.$$

We further write prospective employers' expectation about a worker's service quality as

$$E(q|g, K) = \delta + \frac{Kg + \rho\bar{u}_0}{K + \rho}. \quad (3)$$

2.1.3. The Cold-Start Problem. When the labor service quality is not perfectly observed, a worker sets a price based on Equation (1) by considering the worker's wage from outside options and the expected quality inferred from employers' information set (Equation (3)). The log price p and log demand d are

$$p = \beta p_0 + \beta w + \beta \gamma \delta + \beta \gamma \frac{Kg + \rho\bar{u}_0}{K + \rho},$$

$$d = -n\beta p_0 - n\beta w + \tau \delta + \tau \frac{Kg + \rho\bar{u}_0}{K + \rho},$$

where $\beta > 0$, $p_0 > 0$, $\tau > 0$, $n > 0$, $\gamma > 0$, $K > 0$, $\rho > 0$, $\bar{u}_0 > 0$, and $g > 0$. In reality, because most employers view the reputation sign very positively (e.g., Banerjee and Duflo 2000, Moreno and Terwiesch 2014), g is expected to be much greater than \bar{u}_0 . Moreover, to illustrate the cold-start problem, we focus on the comparison between an inexperienced and an experienced worker who have the same observable quality δ_0 . Given this, prospective employers' expectation about the unobserved quality of an experienced worker's labor service is greater than that of an inexperienced worker's labor service ($\Delta_u = E(u|g, K) - \bar{u}_0 = \frac{Kg + \rho\bar{u}_0}{K + \rho} - \bar{u}_0 = \frac{K(g - \bar{u}_0)}{K + \rho} > 0$). Given that $\frac{\partial p}{\partial u} = \beta \gamma > 0$ and $\frac{\partial d}{\partial u} = \tau > 0$, compared with experienced workers, inexperienced workers need to set a lower price and face a lower demand. In this case, inexperienced workers only participate in the given online labor platform if $d \geq 0$, which requires $\bar{u}_0 \geq \frac{n\beta(p_0 + w)}{\tau} - \delta_0$.

When employers' prior belief about the average labor service quality in the absence of reputation signals \bar{u}_0 is too low, many inexperienced workers will not participate in the online labor platform.

2.2. Impact of the Introduction of the Monitoring System

In the preceding discussion, reputation systems can only help employers positively update their beliefs about the service quality of experienced workers (Laouénan and Rathelot 2022). This leads to the cold-start problem; that is, inexperienced workers face a high entry barrier because the expected quality of their labor service inferred from employers' information set is too low. By comparison, a monitoring system can effectively mitigate moral hazard regardless of whether experienced or inexperienced workers are hired, suggesting that it could be a potential equalizer and alleviate the cold-start problem brought by reputation systems. Specifically, monitoring systems are found to effectively enhance workers' performance in multiple off-line employment contexts, such as the trucking industry (Hubbard 2000), schools (Duflo et al. 2012), restaurants (Pierce et al. 2015), and hospitals (Staats et al. 2017). In online labor markets, monitoring systems allow employers to track workers' effort more precisely and efficiently, which significantly increases the probability of shirkers being caught and, thus, decreases workers' expected payoff from shirking. Monitoring systems discourage both experienced and inexperienced workers from shirking and compel them to perform at the customary level of effort (Shapiro and Stiglitz 1984).

Therefore, after the introduction of monitoring systems, prospective employers' prior belief about the unobserved quality \bar{u}_0 increases regardless of whether the worker has reputation or not. Because $\frac{\partial p}{\partial \bar{u}_0} = \beta\gamma > 0$ and $\frac{\partial d}{\partial \bar{u}_0} = \tau > 0$, the log price p of the labor service of inexperienced workers increases as well as the log

demand d . As \bar{u}_0 increases, the entry constraint for inexperienced workers ($\bar{u}_0 \geq \frac{n\beta(p_0+w)}{\tau} - \delta_0$) is relaxed. This implies that some inexperienced workers who previously would not allocate labor to the given online labor platform would now be willing to allocate more labor to the platform when monitoring systems are in place. The cold-start problem for inexperienced workers is alleviated. Bearing this in mind, we expect that monitoring systems disproportionately attract more bids from inexperienced workers.

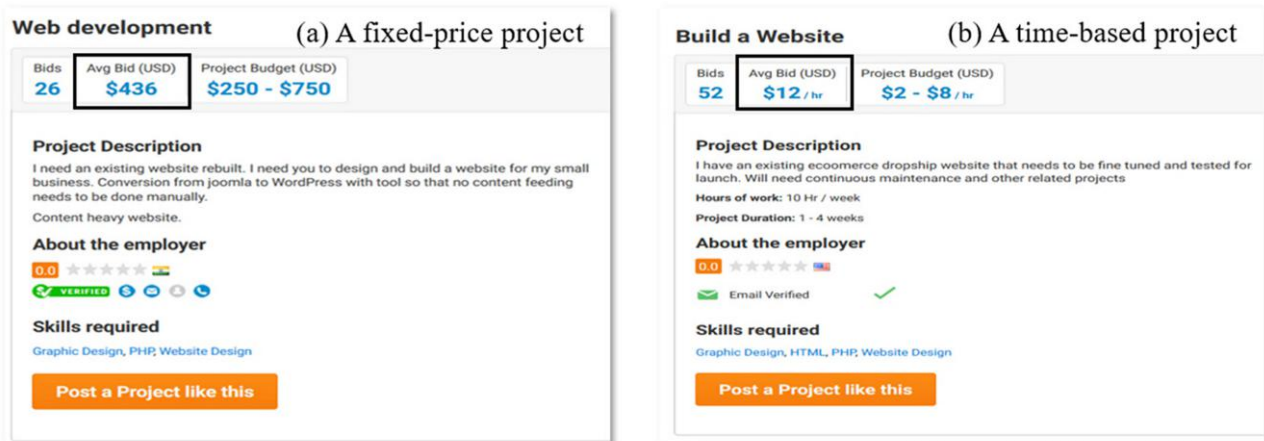
3. Empirical Research Design

3.1. Research Context

We obtained our data from a leading online labor market platform. On this platform, an employer can post a project with a description, estimated budget, and required skills. There are two types of projects on this platform: fixed-price projects (Figure 1(a)), for which the employer pays a fixed-price for completing the entire project and each worker bids on the fixed-price that the worker is willing to accept, or time-based projects (Figure 1(b)) for which the employer pays an hourly rate (in dollars per hour) for completing the project and each worker bids on the hourly rate that the worker is willing to accept. As shown in Online Appendix A, for tasks of similar requirements (e.g., website design, writing), the project can be set as either fixed-price or time-based. Typically, a project is open for bidding for one week, and any worker who is interested in the project can bid on the project. At the end of the bidding period, the employer reviews workers' information (e.g., bid amount and reputation) and awards the project to one worker who best satisfies the requirements.

On February 5, 2014, the platform officially introduced a monitoring system.⁹ The system takes the form of a desktop app that tracks hours and takes screenshots on the workers' end. The desktop app works with time-based projects but not with fixed-price projects. The use

Figure 1. (Color online) Screenshots of Web Pages for a Fixed-Price vs. a Time-Based Project



of monitoring software is required for workers of time-based projects with a few exceptions. If time-based workers do not use the monitoring software, they won't receive automatic payment and need to present evidence for payments. Further, without monitoring, when workers are engaged in disputes with employers, they have the burden of proof, which suggests higher non-payment risks. Figure 2 illustrates the observation window timeline.

The monitoring system allows employers to effortlessly monitor workers. It randomly takes several screenshots roughly every 10 minutes and continuously tracks the time the worker has spent on each time-based project.¹⁰ Specifically, it automatically tracks when and for how long the worker has worked, the accrued compensation the worker has earned, and corresponding screenshots with precise time stamps. Therefore, it effectively keeps detailed records of the worker's actions and provides the employer with up-to-date information on the project progress. Detailed monitoring records serve as evidence of the worker's effort (or the lack thereof) when a dispute is filed, which protects workers (employers) from unjustified rejections (payments) (Moore and Hayes 2018). Figure 3 provides screenshots of the platform's monitoring application.

3.2. Empirical Model

Following prior studies that leverage policy changes to estimate causal effects (e.g., Chen et al. 2011, 2017; Dewan et al. 2017), we leverage the differential availability of the monitoring system for different types of projects (time-based versus fixed-price) to conduct a DID analysis. Specifically, the official monitoring app introduction is an exogenous event that only affects time-based projects. With this external shock based on platform policy change, we perform a DID estimation, which is used extensively in information systems research when exogenous changes are introduced (Chen et al. 2011, 2017; Dewan et al. 2017; Powell and Seabury 2018; Wang et al. 2018). In line with the common practices (Bertrand et al. 2004, Angrist and Pischke 2008), our main DID models are specified in Equation (4):

$$Y_{ij} = \beta_0 + \beta_1 \text{Time_based}_j + \beta_2 \text{Time_based}_j \times \text{After}_j + \delta P_j + \gamma_i + \tau_{t(j)} + \varepsilon_{ij}, \quad (4)$$

where i and j index employer and project, respectively. We consider the following dependent variables (Y_{ij}): (1)

the log-transformed number of bids for project j posted by employer i ($\text{Log_bid_count}_{ij}$) and (2) the percentage of inexperienced workers (i.e., workers with no existing platform reputation) in project j posted by employer i ($\text{Pct_inexperienced}_{ij}$). We further decompose the bids for project j into those from inexperienced workers ($\text{Log_inexperienced}_{ij}$) and experienced workers ($\text{Log_experienced}_{ij}$), respectively. The project type is indicated by Time_based_j , which equals one if project j is time-based and zero if it is fixed-price. Here, After_j is the dummy variable equal to one if project j is awarded after the month when the monitoring system was introduced (zero otherwise). The coefficient of the interaction term $\text{Time_based}_j \times \text{After}_j$ (β_2), thus, identifies the effect of introducing the monitoring system on time-based projects relative to fixed-price projects. To further control for project heterogeneity, we add project characteristic controls (P_j), a vector of employer fixed effects (γ_i), and a vector of month dummies ($\tau_{t(j)}$) into the DID models, and ε_{ij} denotes the robust standard errors clustered on employers. It is notable that the main effect of After_j is subsumed by month dummies.

Note that the identification of the DID estimation relies on the parallel trend assumption instead of strict exogeneity (Abadie 2005). As explained by Angrist and Pischke (2008), although the treatment and control groups can differ, this difference is captured by the group dummy. In our case, even if time-based projects are different from fixed-price projects in terms of their attractiveness to inexperienced workers, such difference is meant to be captured by the group dummy (i.e., Time_based) as long as the parallel trend assumption holds. In the robustness check section, we provide supporting evidence for the parallel trend assumption and further leverage the exogenous launch date of monitoring systems to reexamine our findings with an alternative design wherein both the treatment and control groups are time-based. Furthermore, we find highly consistent results when we match fixed-price and time-based projects with covariate balancing IPTW or CEM.

3.3. Alternative Causal Inference Approaches Dealing with Selection Issues

Besides the DID estimations that leverage the shock induced by platform policy change, we also explore multiple alternative empirical approaches that explicitly deal with selection issues related to contract type, which do not rely on the parallel trend assumption, therefore

Figure 2. A Timeline of Our Observation Window

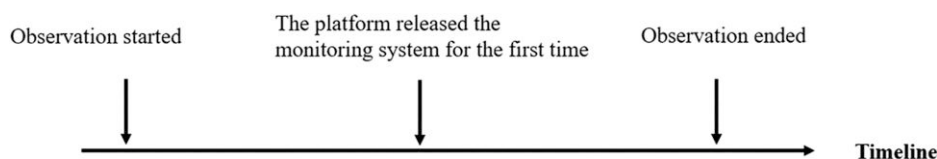
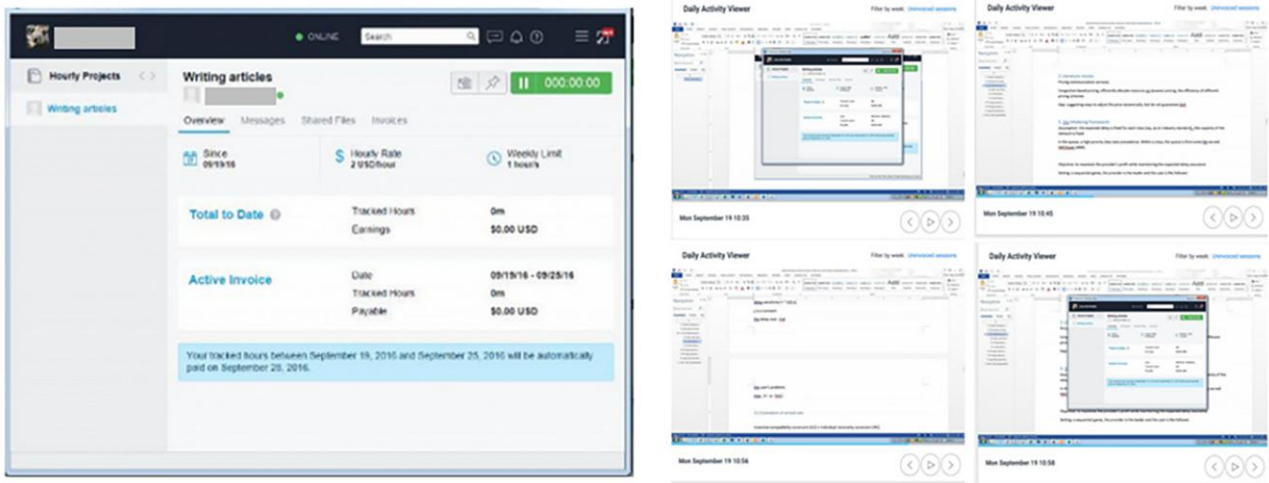


Figure 3. (Color online) Screenshots of the Monitoring System



corroborating the findings from the DID estimations. There are two types of selection issues, namely, selection on observables and selection on unobservables. In our context, we use fixed-price projects as the counterfactual to estimate the change in workers’ bids for time-based projects in the absence of the monitoring system. Here, selection on observables refers to the case wherein workers’ bids for these two types of projects are merely related to the observable project characteristics, whereas selection on unobservables represents the scenario wherein workers’ bids for these two types of projects are associated with some unobserved aspects.¹¹ In particular, during our observational period, the employers are anonymous when they post the projects and the workers cannot initiate conversations until after they are hired, workers make their bid decisions solely based on those observable project characteristics. In other words, employers’ private information regarding the projects is not in the workers’ information set and workers’ bid entries can only be affected by observable project information, which suggests that selection on unobservables is not most concerning. Nevertheless, in an attempt to account for both types of selection concerns, as laid out in Table 1, we present a road map for our alternative empirical approaches.

To address selection on observables, we employ various text-mining techniques to generate textual features from project descriptions and further adopt covariate balancing IPTW and CEM approaches to match and weight two groups of projects. In addition, in order to account for the potential composition change across or within groups over time, we perform matching in causal inference by ensuring the comparability of both groups both before and after the introduction of monitoring systems.

Furthermore, we conduct a series of robustness checks to deal with selection on unobservables. We use both a model-based identification approach (i.e., the

endogenous treatment effect model) and an instrument-based identification approach to account for the potential endogeneity of contract type. Moreover, we employ the DR estimator (Funk et al. 2011, Chernozhukov et al. 2017, Sant’Anna and Zhao 2020), which can recover the causal treatment effect of monitoring by combining weighting with the imputation of missing outcomes. Further, following the recent literature (Altonji et al. 2005, Oster 2019), we assess the sensitivity of our findings to the selection on unobservables and to what extent the selection on unobservables could affect the significance level of the observed treatment effect. In addition, we adopt the ITS analysis approach (Zhang and Zhu 2011, Pu et al. 2020) to quantify the change in workers’ bids for time-based projects posted after the system launch date by using time-based projects posted before the launch as the counterfactual. We also rule out the possibility of spurious causality with two placebo tests based on pseudo treatment time and treatment assignment, respectively.

4. Data

4.1. Variables

We obtained a unique archival data set from a leading online labor platform that includes detailed project information and worker information from September 1, 2013, to August 31, 2014. Variable definitions and descriptive statistics of the full sample are provided in Table 2.

To show the prevalence of inexperienced workers being plagued by the cold-start problem, we plot the distribution of the number of ratings at the worker level in Figure 4. As shown, around 80% of workers have no established reputation (e.g., ratings submitted by prior employers) during our observation window, suggesting that inexperienced workers are a very important portion

Table 1. Overview of Empirical Analyses for Causal Inference

Concern	Analysis	Objective	Section
Selection on observables	Covariate balancing IPTW	Controlling for the differences in observable project characteristics with weighting	Section 5.1
	CEM	Accounting for the differences in observable project characteristics with matching	Online Appendix D
	Matching in causal inference	Controlling for the differences in observable project characteristics and the potential composition change	Online Appendix F
Selection on unobservables	Endogenous treatment effect model	Using a model-based identification approach to account for the potential endogeneity of contract type with a two-stage endogenous treatment effect model	Section 5.2
	IV	Using an instrument-based identification approach to account for the potential endogeneity of contract type with the interaction (between a leave-out-mean residual of employers' other contract type choices and the approximate project duration measure) as the instrument for contract type	Online Appendix G
	DR estimator	Using the DR estimator that can provide an unbiased estimate of the causal treatment effect if either the treatment assignment or outcome regression model is correctly specified	Section 5.3
	Sensitivity analysis	Assessing the overall sensitivity of our findings to any omitted variable bias	Online Appendix I
	ITS analysis	Using an ITS design to estimate the treatment effect based on the shift of bidder entries in time-based projects pre and post system introduction	Section 5.4
Spurious causality	Placebo test (pseudo treatment time or treatment assignment)	Examining the possibility of spurious causality with a pseudo treatment time/assignment	Online Appendix H

of the workforce, and the cold-start problem could be a prominent impediment in the online labor market. Even among the workers with at least one rating, about 50% of them have no more than five ratings.

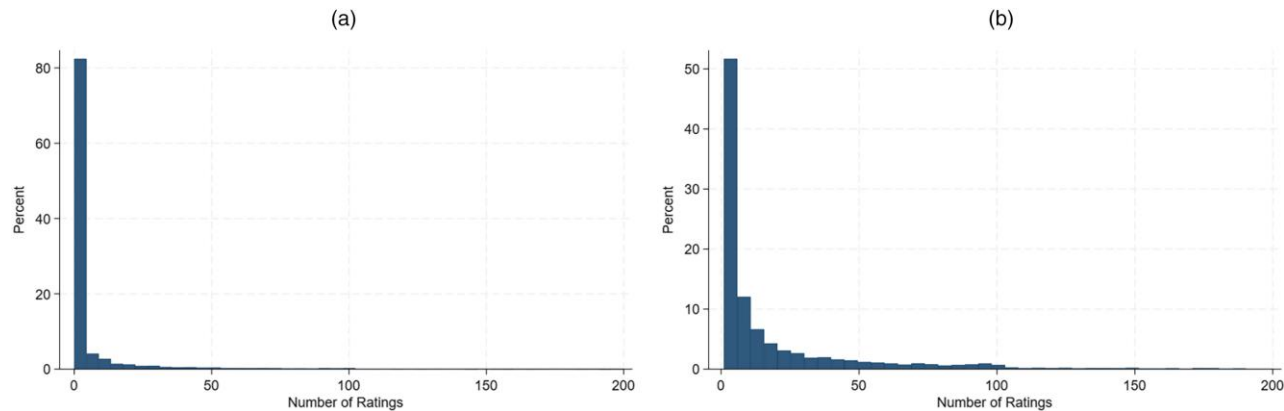
4.2. Model-Free Evidence
Based on the matched sample, we first present model-free evidence of the change in the key dependent variables, that is, the number of total bids and number of

Table 2. Definitions and Summary Statistics of Project-Level Variables

Variable	Variable definition	Observations	Mean	Standard deviation	Minimum	Maximum
<i>Budget_max</i>	The maximum project budget set by the employer	12,462	149.45	270.77	2.00	10,000.00
<i>Time_based</i>	A dummy variable = 1 if the project is a time-based project, = 0 if the project is a fixed-price project	12,462	0.25	0.43	0.00	1.00
<i>Bid_count</i>	Total number of bids received by the project	12,462	13.61	14.67	1.00	139.00
<i>Bid_inexperienced</i>	Total number of bids for the project submitted by inexperienced workers	12,462	2.77	4.36	0.00	84.00
<i>Bid_experienced</i>	Total number of bids for the project submitted by experienced workers	12,462	10.84	11.99	0.00	122.00
<i>Pct_inexperienced</i>	Percentage of the bids submitted by inexperienced workers	12,462	0.16	0.18	0.00	1.00
<i>Project_title_length</i>	Number of words in the project title	12,462	5.60	3.16	1.00	40.00
<i>Project_desc_length</i>	Number of words in the project description	12,462	72.49	81.23	0.00	1,350.00

Notes. There are a few projects whose descriptions are meaningless and do not provide any information (e.g., “oooooooooooo”). We recode the description length of these projects as zero.

Figure 4. (Color online) Relative Frequency Histograms of Workers’ Number of Ratings



Notes. (a) Histogram among all workers. (b) Histogram among all workers with ratings. We have ignored workers with more than 200 reviews to make the plots readable.

bids from inexperienced bidders. As Figure 5 shows, there is a disproportionate increase in two dependent variables regarding the aggregate worker entry among time-based projects (i.e., the treatment group) than among fixed-price projects (i.e., the control group) without controlling for the effect of project characteristics and employer characteristics.

4.3. Covariates for Matching and Weighting

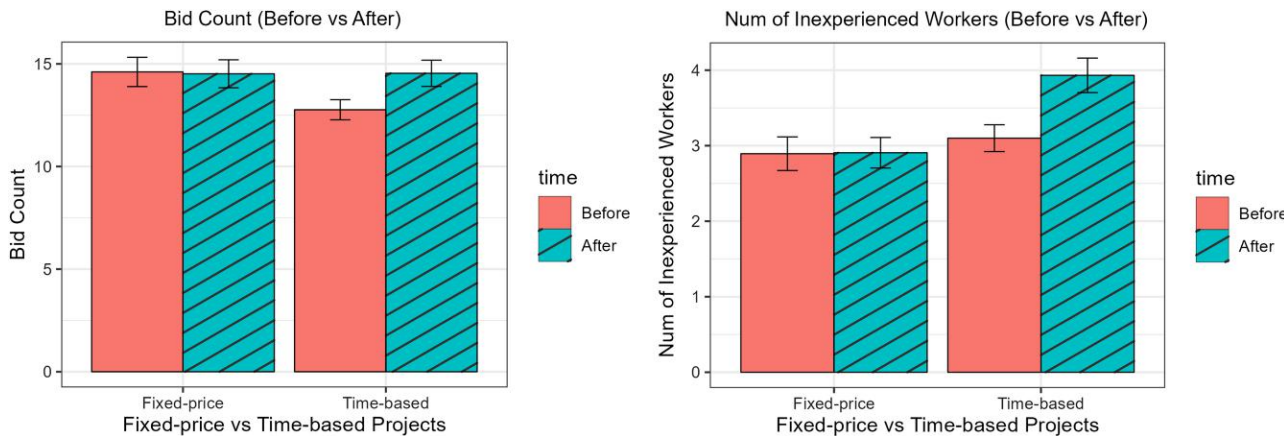
To ensure comparability between time-based and fixed-price projects from the perspective of workers (the decision makers for our dependent variables), we employ CEM (Iacus et al. 2012, Wang et al. 2018) and covariate balancing IPTW (Rosenbaum and Rubin 1983, Imai and Ratkovic 2014) jointly with text-mining techniques to match time-based and fixed-price projects based on various characteristics (e.g., category, employer experience, and reputation). Among these, the project description is

likely to be the most informative one as it can capture the communication between employers and prospective workers. As such, we have extracted a series of important textual features by performing text mining on project descriptions (see Table 3).

In the context of buyer–seller communication, the extant literature highlights two fundamental elements: content and style (Sheth 1976, Williams and Spiro 1985). Content pertains to the information conveyed in the messages (e.g., topics, instructions), whereas style refers to the way individuals interact verbally (e.g., politeness, informal tone).

Regarding communication content, we mainly focus on two aspects: the topics of project descriptions and whether these descriptions contain information aimed at mitigating uncertainty and reducing information asymmetry between employers and workers. First, to better match projects based on the semantic similarity of

Figure 5. (Color online) Model-Free Evidence of the Change in Worker Entry



Notes. The full sample of all the projects is used and weighted by the covariate balancing IPTW. The bars represent the average number of total bids (*Bid_count*) and the number of bids from inexperienced workers (*Bid_inexperienced*) in fixed-price and time-based projects before and after the introduction of the monitoring system. Error bars represent the 95% confidence intervals.

Table 3. Covariates Used for Matching and Weighting

Dimension	Variable	Variable description
Task complexity, risk of the project	<i>Project category dummies</i>	Dummy variables for various project categories, including software, design, marketing, administrative, etc.
Project title length	<i>Project_title_length</i>	Number of words in the project title
Project description length	<i>Project_desc_length</i>	Number of words in the project description shown on the project page
Experience and reputation	<i>Employer_tenure_month;</i> <i>Employer_overall_rating</i>	Employer's tenure measured in months, which is also a proxy of employers' experience and relevant knowledge; employers' overall rating indicating employers' reputations
Communication content		
Description semantics	<i>Topic_cluster</i>	A categorical variable denoting to which topic cluster the project has been assigned based on the topic analysis with the Top2Vec approach
Expressions of uncertainty	<i>Hedges</i>	A linguistic feature score indicating the uncertainty and ambiguity
Clear suggestions for expected actions	<i>Give_agency</i>	A linguistic feature score indicating to what extent employers clearly suggest an action for others
Communication style		
Text (in)formality	<i>Informal_title,</i> <i>Impersonal_pronoun</i>	Linguistic feature scores indicating to what extent employers tend to use informal titles and nonperson referents, respectively
Politeness and gratitude	<i>Please, Gratitude</i>	Linguistic feature scores indicating to what extent employers express please and gratitude, respectively

descriptions, we jointly embed documents (project descriptions) and words in the same semantic space to discover task topics and classify project clusters based on their task topics with the Top2Vec topic modeling approach (Angelov 2020, Liang et al. 2022). Owing to its joint semantic embedding of document and word vectors, the Top2Vec approach effectively accounts for the semantic association between documents and words and, thus, discovers more representative and informative topics than bag-of-word topic modeling approaches, such as latent dirichlet allocation (Angelov 2020). Second, drawing from the extant literature, we recognize that project uncertainty (Pich et al. 2002) and the concreteness of task information (Ludwig et al. 2022) have the potential to influence workers' uncertainty regarding task requirements and the project's appeal. To this end, we assimilate two measures: one capturing employers' uncertainty about the project (*Hedges*) and the other one assessing whether employers tend to provide clear instructions in project descriptions (*Give_agency*).

Furthermore, the (in)formal tone and politeness, as delineated in existing research, emerge as two salient aspects of communication style that can either augment positive outcomes (Jessmer and Anderson 2001, Lee et al. 2019) or introduce downsides (Jeong et al. 2019). In light of this, we also incorporate the following linguistic features into the matching process, including whether employers tend to write project descriptions in an (in)formal way (*Informal_title* and *Impersonal_pronoun*) and whether employers show their politeness and gratitude to workers in project descriptions (*Please, Gratitude*). All

these linguistic features are extracted with the Politeness package in R (Yeomans et al. 2018).

With the aforementioned covariates, we use both covariate balancing IPTW and CEM to improve our DID estimation. Both approaches can help us focus on those fixed-price projects that are more similar to time-based projects based on the distributions of covariates and propensity score. As shown in Online Appendices B and C, these covariates are highly comparable and have similar means across project types in both the IPTW and CEM samples. Moreover, it is worth noting that projects in the two groups are required to exactly match on the long list of binary covariates in CEM, which leads to a relatively small size of the CEM sample. In contrast, IPTW preserves the sample size by weighting the importance of observations in the DID regressions. As such, we mainly report the IPTW-DID estimation in Section 5.1 and provide the related results based on the full sample and the CEM sample in Online Appendix D.

5. Results

5.1. DID Estimation

Table 4 presents the result of our DID model regarding workers' entries. We find that the coefficient of the interaction term *Time_based* × *After* in the model for *Log_bid_count* (column (1)) is significantly positive, which suggests that introducing the monitoring system significantly increases the number of bids for time-based projects. The coefficient of the interaction term is 0.245, which translates to a 27.8% increase in the number of bids.¹²

Table 4. Results of the IPTW-DID Estimation

Model	(1)	(2)	(3)	(4)
Dependent variable	<i>Log_bid_count</i>	<i>Pct_inexperienced</i>	<i>Log_inexperienced</i>	<i>Log_experienced</i>
<i>Time_based</i>	0.215*** (0.079)	−0.002 (0.016)	0.027 (0.077)	0.235*** (0.073)
<i>Time_based</i> × <i>After</i>	0.245*** (0.072)	0.071*** (0.016)	0.368*** (0.074)	0.170** (0.068)
<i>Log_budget_max</i>	0.152*** (0.020)	−0.005 (0.004)	0.040** (0.018)	0.165*** (0.019)
<i>Log_title_length</i>	−0.062 (0.041)	−0.003 (0.010)	−0.067 (0.045)	−0.063 (0.040)
<i>Log_desc_length</i>	0.180*** (0.023)	0.012** (0.005)	0.123*** (0.021)	0.167*** (0.022)
Category dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
Employer dummies	Yes	Yes	Yes	Yes
Clusters (employers)	1,941	1,941	1,941	1,941
Observations	5,413	5,413	5,413	5,413
R ²	0.636	0.596	0.580	0.635

Notes. When we limit to those projects posted by employers with more than one project, the sample size decreases from 12,462 to 5,413. Results are highly consistent if we do not include employer fixed effects. We calculate the dependent variables of the last two columns according to the following equations: $\text{Log_inexperienced} = \ln(\text{Bid_inexperienced} + 1)$ and $\text{Log_experienced} = \ln(\text{Bid_experienced} + 1)$. The results are highly consistent when we control for the week dummies instead of month dummies. Robust standard errors clustered on employers are reported in parentheses. The results are consistent when we use the top 1, 5, or 10 quantiles of the Mahalanobis distance as a threshold to exclude outliers.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

We further assess whether the introduction of monitoring systems reduces the entry barrier for new entrants and disproportionately attracts more inexperienced workers. As reported in column (2) of Table 4, the marginal effect of the *Time_based* dummy is insignificant, indicating that the percentage of inexperienced workers for time-based projects is roughly the same as the percentage of inexperienced workers for fixed-price projects before the introduction of the monitoring system. However, after the introduction of the monitoring system, the coefficient of *Time_based* increases significantly. This increase suggests that the percentage of inexperienced workers increases more in time-based projects relative to fixed-price projects. Specifically, the marginal effect estimate based on the delta method indicates that the percentage increases by 7.10%. In line with this, columns (3) and (4) of Table 4 also reveal that the increase in number of bids from inexperienced workers (*Log_inexperienced*)¹³ is more pronounced and robust than that from experienced workers (*Log_experienced*).¹⁴ Notably, the introduction of monitoring systems is associated with a 44.5% increase in the number of bids from inexperienced workers. As shown in Online Appendix D, the DID estimations based on the full sample and the CEM sample are highly consistent, suggesting the increase in bid entries of inexperienced workers following the introduction of monitoring systems.

In order to further test the parallel trend assumption of the DID model (Autor 2003, Angrist and Pischke 2008), we employ the relative time model to assess whether time-based projects and fixed-price projects have a common trend during the pretreatment period. Because of the page limit, we present detailed results in Online Appendix E and visualize the results in Figure 6.

The plot shows that all the relative time parameters are insignificant before the introduction of the monitoring system, whereas most of the relative time parameters are significantly positive after February 2014 wherein the platform officially introduced the monitoring system, suggesting the validity of the parallel trend assumption and the DID design.

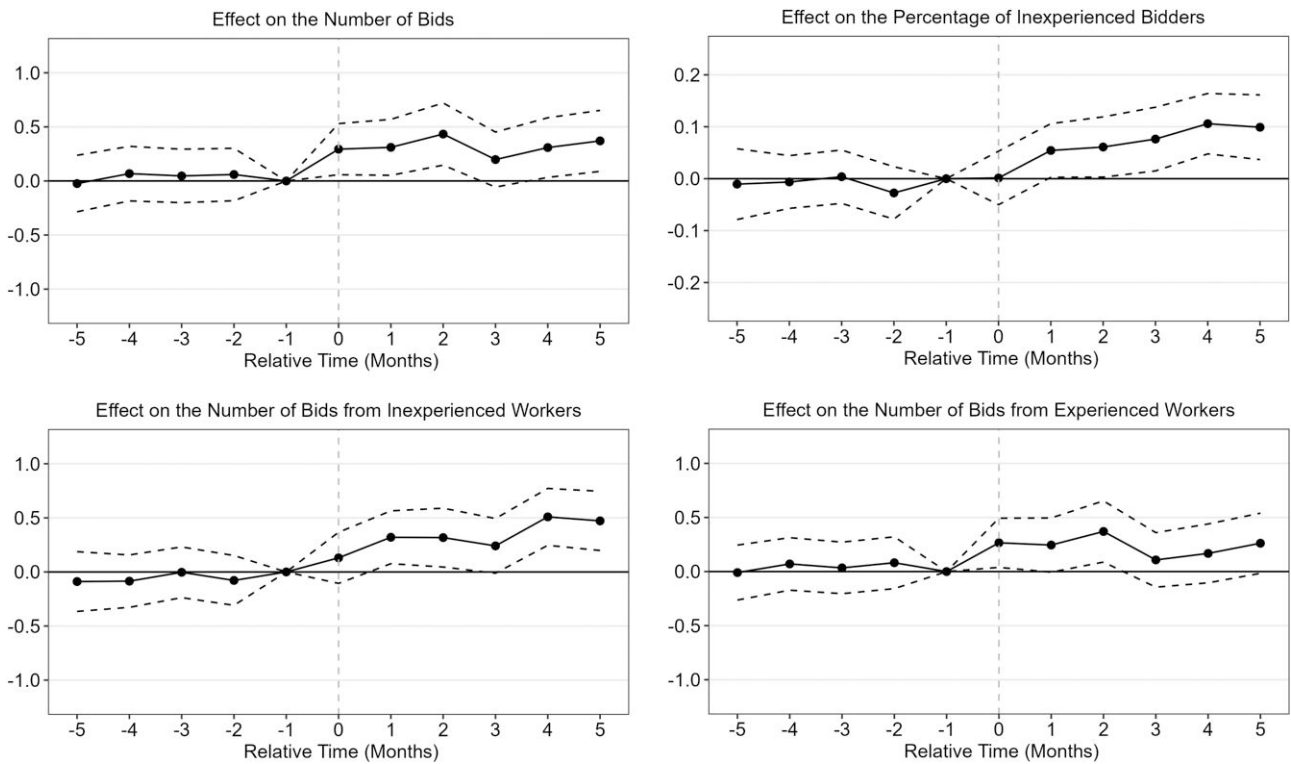
5.2. Endogenous Treatment Effects

In the prior analysis, we rely on the parallel trend assumption to identify the treatment effect of monitoring systems on inexperienced workers' bids for time-based projects. We report evidence for the parallel trend assumption and observe consistent results when we use covariate balancing IPTW and CEM to improve the comparability between two groups of projects. We next consider a model-based identification approach that identifies the treatment effect by functional form, that is, the endogenous treatment effect model (Greene 2012). Here, we consider a recursive two-stage framework to address the potential endogeneity of contract type. In the first stage, we estimate employers' contract choice decisions based on a probit model. In the second stage, we model how workers' aggregate entry decisions at the project level (e.g., number of bids and percentage of inexperienced bidders) with a linear regression model by allowing the unobserved project characteristics affecting employers' contract choice to be correlated with those affecting workers' entry decisions. The two-stage endogenous treatment effect model is specified as follows:

$$T_j = 1(\theta P_j + \mu E_{it(j)} + \pi_{t(j)} + \omega_j > 0), \quad (5)$$

$$Y_j = \beta_0 + \beta_1 T_j + \beta_2 T_j \times \text{After}_j + \delta P_j + \gamma E_{it(j)} + \tau_{t(j)} + \varepsilon_j, \quad (6)$$

Figure 6. Coefficients of the Monthly Dynamic Difference-in-Differences Estimates



Notes. The dashed vertical line denotes the month in which the platform officially introduced the monitoring system (February 2014). Error bars represent the 95% confidence intervals using clustered standard errors.

where T_j denotes the contract type of project j , which is equal to one if project j is time-based. The term Y_j denotes the dependent variables of our key interest (i.e., Log_bid_count , Pct_inexperienced , Log_inexperienced , and Log_experienced). In the first stage (Equation (5)), employers make the contract choice decisions based on project characteristics P_j (i.e., title length, description length, category), employer characteristics $E_{it(j)}$ (i.e., tenure, reputation), time effect $\pi_{t(j)}$, and the unobserved project characteristics leading employers to use the time-based contract type ω_j . In the second stage (Equation (6)), the workers' entry outcomes depend on the contract type T_j , whether the monitoring system was available $T_j \times \text{After}_j$, project characteristics P_j (i.e., title length, description length, project budget, category), employer characteristics $E_{it(j)}$ (i.e., tenure, reputation), time effect $\tau_{t(j)}$, and the unobserved project characteristics affecting workers' entries to projects ε_j . We allow the error term ω_j in the first stage and the error term ε_j in the second stage to be correlated and follow the following bivariate normal distribution (Equation (7)). Note that the endogenous treatment effect model does not allow us to include employer-level fixed effects. Therefore, we control for additional employer-related characteristics in both stages.

$$\begin{pmatrix} \omega_j \\ \varepsilon_j \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \quad (7)$$

As results show in Table 5, we find that all the results are highly consistent. The interaction term, $\text{Time_based} \times \text{After}$, is significantly positive in the second stage of the first three models, suggesting that there are more inexperienced workers (also a higher percentage of inexperienced workers) bidding for time-based projects after the introduction of the monitoring system.

In Online Appendix G, we further address the potential endogeneity of contract type choice with the instrumental variable approach by constructing the instrumental variable following the prior literature and explaining the validity of our proposed instruments. Whereas the preceding model-based identification approach relies on the parametric assumptions of the data generation process (Greene 2012), the instrumental variable approach relies on the relevancy and exclusion restriction assumptions. On the whole, the consistent results with the instrument-based and model-based identification approaches corroborate that our findings are not driven by or sensitive to specific modeling assumptions.

5.3. DR Estimator

The DR estimator can generate an unbiased treatment effect estimate if either the model for the treatment assignment (i.e., the propensity score model for the contract type) or the outcome regression model is correctly

Table 5. Estimation Results of the Endogenous Treatment Effect Model

Model	(1)	(2)	(3)	(4)
Stage 1	<i>Time_based</i>	<i>Time_based</i>	<i>Time_based</i>	<i>Time_based</i>
<i>Log_title_length</i>	0.090*** (0.033)	0.064** (0.033)	0.093*** (0.034)	0.092*** (0.034)
<i>Log_desc_length</i>	−0.133*** (0.017)	−0.125*** (0.014)	−0.142*** (0.015)	−0.138*** (0.016)
<i>Log_employer_tenure_month</i>	0.013 (0.017)	0.015 (0.016)	0.011 (0.017)	0.013 (0.017)
<i>Log_employer_overall_rating</i>	0.044 (0.028)	0.051** (0.025)	0.044 (0.027)	0.048* (0.028)
Category dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
Stage 2	<i>Log_bid_count</i>	<i>Pct_inexperienced</i>	<i>Log_inexperienced</i>	<i>Log_experienced</i>
<i>Time_based</i>	−0.592*** (0.099)	0.271*** (0.013)	0.245*** (0.041)	−0.537*** (0.097)
<i>Time_based</i> × <i>After</i>	0.079** (0.034)	0.064*** (0.008)	0.252*** (0.031)	0.025 (0.035)
<i>Log_budget_max</i>	0.180*** (0.008)	−0.009*** (0.001)	0.067*** (0.007)	0.194*** (0.008)
<i>Log_title_length</i>	−0.047** (0.023)	−0.005 (0.004)	−0.037** (0.017)	−0.055** (0.022)
<i>Log_desc_length</i>	0.118*** (0.012)	0.019*** (0.002)	0.088*** (0.008)	0.111*** (0.012)
<i>Log_employer_tenure_month</i>	0.010 (0.010)	−0.004* (0.002)	−0.003 (0.007)	0.013 (0.010)
<i>Log_employer_overall_rating</i>	0.010 (0.016)	−0.007* (0.004)	−0.006 (0.015)	0.018 (0.016)
Category dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
Employer dummies	No	No	No	No
Observations	12,462	12,462	12,462	12,462
LogLik	−22,315	−1,538	−20,809	−22,007

Notes. Robust standard errors clustered on employers are reported in parentheses. Note that we only include the project budget in δ_i in the second stage because the project budget is usually set after the contract type is chosen. Results are consistent regardless of whether we control for the project budget in the second stage or not.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

specified. Similar to the preceding two-stage model, we still use Y to denote the outcome variables and T to denote the contract type, which is equal to one if the project is time-based. For simplicity, we use X to refer to the observable project characteristics. Furthermore, the propensity score model for the treatment assignment $\hat{P}(X)$ and the outcome regression model $\hat{f}(X)$ are specified as follows:

$$\begin{aligned}\hat{P}(X) &= \Pr(T = 1|X) \text{ and } \hat{f}(X) = E[Y|X, T] \\ \hat{f}_1(X) &= E[Y|X, T = 1] \text{ and } \hat{f}_0(X) = E[Y|X, T = 0].\end{aligned}$$

Following the prior literature (Funk et al. 2011, Chernozhukov et al. 2017, Sant'Anna and Zhao 2020, Facure 2023), we have the following equation for the DR estimator ($\hat{\Delta}_{Dr}$):

$$\begin{aligned}\hat{\Delta}_{Dr} &= \frac{1}{N} \sum \left(\frac{T_j(Y_j - \hat{f}_1(X_j))}{\hat{P}(X_j)} + \hat{f}_1(X_j) \right) \\ &\quad - \frac{1}{N} \sum \left(\frac{(1 - T_j)(Y_j - \hat{f}_0(X_j))}{1 - \hat{P}(X_j)} + \hat{f}_0(X_j) \right).\end{aligned}$$

The first part is used to estimate $E[Y_1]$ when $T = 1$, and the second part is used to estimate $E[Y_0]$ when $T = 0$.

i. Case 1: The outcome regression model $\hat{f}(X)$ is correctly specified.

When $\hat{f}(X)$ is correctly specified, $T_j(Y_j - \hat{f}_1(X_j)) = 0$ because $Y_j - \hat{f}_1(X_j) = 0$ when $T_j = 1$. Similarly, we can have $(1 - T_j)(Y_j - \hat{f}_0(X_j)) = 0$.

$$\begin{aligned}\hat{\Delta}_{Dr} &= \frac{1}{N} \sum \hat{f}_1(X_j) - \frac{1}{N} \sum \hat{f}_0(X_j) \\ &= E[Y, T = 1] - E[Y, T = 0] = \widehat{ATE}.\end{aligned}$$

Therefore, $\hat{\Delta}_{Dr}$ is an unbiased estimator of the average treatment effect \widehat{ATE} .

ii. Case 2: The propensity score model $\hat{P}(X)$ is correctly specified.

When $\hat{P}(X)$ is correctly specified, we can rewrite the equation of $\hat{\Delta}_{Dr}$ in the following way:

$$\begin{aligned}\hat{\Delta}_{Dr} &= \frac{1}{N} \sum \left(\frac{T_j Y_j}{\hat{P}(X_j)} - \frac{T_j \hat{f}_1(X_j)}{\hat{P}(X_j)} + \hat{f}_1(X_j) \right) \\ &\quad - \frac{1}{N} \sum \left(\frac{(Y_j - \hat{f}_0(X_j))}{1 - \hat{P}(X_j)} - \frac{T_j(Y_j - \hat{f}_0(X_j))}{1 - \hat{P}(X_j)} + \hat{f}_0(X_j) \right) \\ \hat{\Delta}_{Dr} &= \frac{1}{N} \sum \left(\frac{T_j Y_j}{\hat{P}(X_j)} + \left(1 - \frac{T_j}{\hat{P}(X_j)} \right) \hat{f}_1(X_j) \right) \\ &\quad - \frac{1}{N} \sum \left(\frac{(1 - T_j) Y_j}{1 - \hat{P}(X_j)} + \left(1 - \frac{(1 - T_j)}{1 - \hat{P}(X_j)} \right) \hat{f}_0(X_j) \right).\end{aligned}$$

Given that $\hat{P}(X)$ is correctly specified, $1 - \frac{T_j}{\hat{P}(X_j)} = \frac{\hat{P}(X_j) - T_j}{\hat{P}(X_j)} = 0$ and $\frac{(1 - T_j)}{1 - \hat{P}(X_j)} = 1$. We have

$$\hat{\Delta}_{Dr} = \frac{1}{N} \sum \frac{T_j Y_j}{\hat{P}(X_j)} - \frac{1}{N} \sum \frac{(1 - T_j) Y_j}{1 - \hat{P}(X_j)}.$$

Given that $\hat{P}(X)$ is correct, the propensity score weighting estimator in the treatment group $\frac{T_j Y_j}{\hat{P}(X_j)} = \hat{f}_1(X_j)$, and that in the control group $\frac{(1-T_j)Y_j}{1-\hat{P}(X_j)} = \hat{f}_0(X_j)$. Therefore, $\hat{\Delta}_{DR}$ is still equal to \widehat{ATE} .

As shown, the DR estimator can provide an unbiased estimator of the average treatment effect as long as either the propensity score model or the outcome regression model is correct. Recently, Sant’Anna and Zhao (2020) further extend the DR estimator to the DID design when either panel data or repeated cross-sectional data are available. They also propose an improved doubly robust estimator, which is a combination of the inverse probability tilting estimator and weighted least squares. Table 6 reports the results from the traditional DR estimator and the improved DR estimator, both showing high consistency.

5.4. ITS Based on Time-Based Projects

By leveraging the exogenous change caused by the introduction of the monitoring system, we estimate the treatment effect on workers’ aggregate entry with the ITS approach (Zhang and Zhu 2011, Pu et al. 2020). In particular, following prior studies, we use time-based projects posted right before the system launch date as the control group and the time-based projects posted after that date as the treatment group.

$$Y_j = \beta_0 + \beta_1 \text{After}_j + \delta P_j + \gamma E_{it(j)} + \varepsilon_j. \quad (8)$$

The model specification is shown in Equation (8), where j indexes the project. We still use the same set of dependent variables (Y_j) and focus on the variable After_j , which denotes whether the time-based project j was posted after the system launch date. We also control for employer characteristic controls ($E_{it(j)}$) and project characteristic controls (P_j). With 90 days as the bandwidth on both sides near the system launch date, we further use the same set of covariates introduced in Section 4.3 and perform an IPTW-ITS estimate. In Table 7, we consistently find that the number of bids and the percentage of inexperienced bidders significantly increase after the introduction of monitoring systems.

5.5. Other Robustness Checks

We conduct a series of additional robustness checks. First, to alleviate the concern regarding potential composition changes in the cross-sectional DID, we further match time-based/fixed-price projects posted before and

after the system launch date based on propensity score and prune posttreatment pairs that are outside of the convex hull of the pretreatment (Keele et al. 2019). We find highly consistent findings (see Online Appendix F). Second, instead of relying on the parallel trend assumption, we use an instrument-based identification approach to account for the potential endogeneity of contract type and find results remain the same (see Online Appendix G). Third, to further alleviate the concern of potential spurious causality, we conduct two placebo tests based on treatment-assignment permutation and a pseudo treatment time (see Online Appendix H). Fourth, we conduct a sensitivity analysis to assess the robustness of our findings to the selection on unobservables. We find that our estimates are very unlikely to be driven by unobservables (see Online Appendix I). Fifth, our findings still hold when we use alternative model specifications (Poisson models with fixed effects and a fractional response model) (see Online Appendix J), exclude outliers with the blocked adaptive computationally efficient outlier nominators (Billor et al. 2000) (see Online Appendix K), or limit our sample to workers who bid for both types of projects (see Online Appendix L).

6. Additional Results and Post hoc Analysis of Project Outcomes

In this section, we conduct several additional analyses to better understand the impact of introducing a monitoring system to an online labor market platform. First, to characterize the heterogeneity of the treatment effect, we investigate whether there is a differential treatment effect by monitoring effectiveness (see Section 6.1). Second, to assess the benefit of monitoring to workers and employers, we conduct post hoc analysis on the following project outcomes: (1) the potential change in employers’ preference for inexperienced workers at the employment stage (see Section 6.2) and (2) the potential impact on labor costs and project delivery following employment (see Section 6.3).

6.1. Heterogeneous Treatment Effect (HTE)

Along with the overall impact of the monitoring system on workers’ entry decisions, we are interested in understanding what project characteristic(s) may moderate the effect of the monitoring system. In particular, the monitoring system that tracks workers’ working hours and work process essentially measures workers’ input instead of productivity (Zhao 2008). As suggested by the prior literature, whether input-based monitoring can

Table 6. Estimation Results of DR Estimators

Estimator	Log_bid_count	Pct_inexperienced	Log_inexperienced	Log_experienced
DR estimator	0.168*** (0.057)	0.056*** (0.012)	0.321*** (0.050)	0.104* (0.057)
Improved DR estimator	0.155*** (0.035)	0.056*** (0.008)	0.314*** (0.033)	0.089** (0.036)

Note. * $P < 0.1$; ** $P < 0.05$; *** $P < 0.01$.

Table 7. Results of the IPTW-ITS Estimate

Model	(1)	(2)	(3)	(4)
Dependent variable	<i>Log_bid_count</i>	<i>Pct_inexperienced</i>	<i>Log_inexperienced</i>	<i>Log_experienced</i>
<i>After</i>	0.130*** (0.042)	0.053*** (0.010)	0.221*** (0.040)	0.083* (0.043)
<i>Log_budget_max</i>	0.085** (0.036)	−0.046*** (0.009)	−0.054 (0.037)	0.138*** (0.036)
<i>Log_title_length</i>	−0.063 (0.046)	−0.007 (0.012)	−0.058 (0.045)	−0.059 (0.046)
<i>Log_desc_length</i>	0.124*** (0.023)	0.005 (0.006)	0.077*** (0.022)	0.115*** (0.024)
<i>Log_employer_tenure_month</i>	−0.031 (0.028)	−0.002 (0.006)	−0.017 (0.025)	−0.027 (0.029)
<i>Log_employer_overall_rating</i>	−0.023 (0.060)	0.002 (0.013)	−0.017 (0.076)	−0.022 (0.059)
Category dummies	Yes	Yes	Yes	Yes
Observations	1,661	1,661	1,661	1,661
<i>R</i> ²	0.119	0.149	0.120	0.115

Note. We calculate the dependent variables of the last two columns according to the following equations: $\text{Log_experienced} = \ln(\text{Bid_experienced} + 1)$ and $\text{Log_inexperienced} = \ln(\text{Bid_inexperienced} + 1)$.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

effectively alleviate moral hazard depends on the measurement error of workers' input (e.g., Prendergast 1999, Raith 2008). Specifically, if the project primarily requires cognitive skills (e.g., thinking of a slogan for a company or logo design), the measurement error could be nonnegligible because screenshots and tracked hours measured by the monitoring system may not well reflect workers' effort. In contrast, if the project primarily requires manual skills (e.g., translation or proofreading), monitoring records are more informative and the measurement error in terms of workers' input is expected to be smaller. Accordingly, the monitoring system is expected to be more effective in deterring moral hazard when tasks need more manual input.

Prior work on the labor force usually uses the manual task input index of job skills (Autor et al. 2003) to classify whether the job is cognitive or manual (e.g., Goos and Manning 2007, Dwyer 2013). Specifically, projects with a high manual task input index tend to be primarily manual, in which the measurement error of input-based monitoring is expected to be smaller than that in projects with a low manual task input index. Hence, we expect that the introduction of the monitoring system should have a relatively stronger impact on the entry barrier of inexperienced workers for projects with a high manual task input index than that for projects with a low manual index.

In the following section, we further investigate whether the impact of monitoring varies in a predictable way across projects with a high versus low manual index. We first search for the corresponding standard occupational classification (SOC) code for each project based on its primary required skill, and further find its American Community Survey Occupation Code and manual task input measure (David and Dorn 2013).¹⁵ This allows us to create a dummy variable, *High_manual*, which equals one if the manual task input of that project is greater than the median of all the projects in our sample and zero otherwise.

To test if there are heterogenous treatment effects on workers' bid entries, we add interactions between the

High_Manual dummy and three key independent variables (*Time_based*, *After*, and *Time_based* × *After*) into the DID models, respectively.¹⁶ Table 8 reports the results on bid entries. The coefficient of the interaction term (*Time_based* × *After*) is significantly positive in columns (2) and (3), suggesting that, for low-manual time-based projects, the percentage of bid entries from inexperienced workers slightly increases after the introduction of the monitoring system. More importantly, the coefficient of the three-way interaction (*Time_based* × *After* × *High_manual*) is also positive and significant. This implies that the impact of the monitoring system on attracting inexperienced workers' entries is stronger for high-manual projects. The significant heterogeneity of treatment effects lends support to our heterogeneity argument and underscores that the measurement error of input-based monitoring plays an important role in alleviating the cold-start problem with the monitoring system.

6.2. Employers' Preference

In the analysis of employer preference, we specify the econometric model at the bid level and estimate, after accounting for the effect of other bid and worker characteristics, how the impact of the worker experience (experienced versus inexperienced) on the probability of being hired changes after the introduction of monitoring systems. In particular, we use j to index the project and k to index the bid (worker) that is nested within each project. Here, the dependent variable denotes whether the bid (worker) k is awarded in project j as y_{jk} :

$$y_{jk} = \beta_1 I_{jk} + \beta_2 t_j I_{jk} + \beta_3 T_j I_{jk} + \beta_4 t_j T_j I_{jk} + \gamma B_k + \delta Z_{jk} + \varphi_j + \varepsilon_{jk}. \quad (9)$$

In Equation (9), t_j is the period dummy variable, which is set to one if project j is posted after the introduction of the monitoring system. Here, T_j is the project-type dummy variable, which is set to one if project j is time-based. The term I_{jk} denotes whether bidder k was inexperienced when bidding for project j .¹⁷ We have considered two inexperience cutoffs, that is, the median of the experience

Table 8. HTE of Bid Entries by Project Manual Index

Model	(1)	(2)	(3)	(4)
Dependent variable	<i>Log_bid_count</i>	<i>Pct_inexperienced</i>	<i>Log_inexperienced</i>	<i>Log_experienced</i>
<i>Time_based</i>	0.350*** (0.067)	0.008 (0.014)	0.154** (0.061)	0.338*** (0.063)
<i>Time_based</i> × <i>After</i>	0.146** (0.065)	0.059*** (0.014)	0.239*** (0.063)	0.096 (0.063)
<i>High_Manual</i>	−0.196*** (0.068)	−0.023 (0.014)	−0.162** (0.063)	−0.184*** (0.066)
<i>Time_based</i> × <i>High_Manual</i>	−0.038 (0.092)	0.022 (0.022)	0.005 (0.094)	−0.045 (0.089)
<i>High_Manual</i> × <i>After</i>	0.044 (0.085)	−0.016 (0.018)	−0.098 (0.079)	0.074 (0.081)
<i>Time_based</i> × <i>After</i> × <i>High_Manual</i>	0.073 (0.140)	0.062* (0.032)	0.284** (0.138)	−0.017 (0.137)
<i>Log_budget_max</i>	0.156*** (0.018)	−0.001 (0.003)	0.057*** (0.015)	0.162*** (0.017)
<i>Log_title_length</i>	−0.041 (0.037)	−0.001 (0.008)	−0.029 (0.035)	−0.046 (0.035)
<i>Log_desc_length</i>	0.174*** (0.019)	0.015*** (0.004)	0.105*** (0.016)	0.159*** (0.018)
Category dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
Employer dummies	Yes	Yes	Yes	Yes
Clusters (employers)	1,941	1,941	1,941	1,941
Observations	5,413	5,413	5,413	5,413
<i>R</i> ²	0.655	0.568	0.571	0.646

Note. Robust standard errors clustered on employers are reported in parentheses.

p* < 0.1; *p* < 0.05; ****p* < 0.01.

(no rating) for all workers and the median of the experience (five ratings) of workers with at least one project completed. Here, B_k captures bidder k 's characteristics, including bidder k 's country dummy, whether bidder k has a preferred worker badge, and whether bidder k works for local projects and passes various profile verifications. The term Z_{jk} represents a set of other project-bidder pair characteristics, including bidder k 's bidding price and whether bidder k has worked for this employer before. And φ_j captures project j 's fixed effect. Note that, because we control for project fixed effects, the effect of $t_j T_j$ is unidentifiable. The employer's hiring decision could be estimated with a linear probability model

(LPM) (Greenwood and Agarwal 2016) or a logit model. Given our focus on the existence of the treatment effect and the interaction effects,¹⁸ we follow the literature to use a LPM by clustering ε_{jk} at the project level (Hong and Pavlou 2017).

The results of the LPM are reported in Table 9. Before the introduction of the monitoring system, the coefficient of the *Inexperienced* dummy is more negative for time-based projects than for fixed-price projects (−0.061 versus −0.028), suggesting that employers show a stronger preference for experienced workers if workers are paid by the hour. After the platform introduces the monitoring system, the coefficient of the *Inexperienced* dummy

Table 9. Estimation Results of Employers' Preference

Dependent variable: <i>Bid_selected</i>	FE LPM	FE LPM
Definition of inexperienced	No rating	≤5 ratings
<i>Inexperienced</i>	−0.028*** (0.003)	−0.029*** (0.003)
<i>Inexperienced</i> × <i>Time_based</i>	−0.033*** (0.005)	−0.035*** (0.005)
<i>Inexperienced</i> × <i>After</i>	−0.002 (0.005)	0.001 (0.004)
<i>Inexperienced</i> × <i>Time_based</i> × <i>After</i>	0.013* (0.008)	0.019** (0.007)
<i>Hire_before</i>	0.540*** (0.014)	0.537*** (0.014)
<i>Log_bid_price</i>	−0.067*** (0.003)	−0.069*** (0.003)
<i>Preferred_worker</i>	0.019*** (0.003)	0.017*** (0.003)
<i>Local_worker</i>	−0.032*** (0.007)	−0.034*** (0.007)
<i>Payment_verified</i>	0.006*** (0.002)	0.006*** (0.002)
<i>Phone_verified</i>	0.008*** (0.003)	0.013*** (0.003)
<i>Identification_verified</i>	0.048*** (0.002)	0.043*** (0.002)
<i>Facebook_connected</i>	0.005*** (0.002)	0.006*** (0.002)
Project fixed effects	Yes	Yes
Bidder country dummies	Yes	Yes
Observations	139,336	139,336
Clusters (projects)	12,191	12,191
<i>R</i> ²	0.115	0.117

Notes. We limit our sample to those workers with at least one skill. Robust standard errors clustered on projects are reported in parentheses. Results are highly consistent if we exclude bids from workers that employers have hired before.

p* < 0.1; *p* < 0.05; ****p* < 0.01.

does not change significantly in fixed-price projects. In contrast, there is a relatively large decrease in employer preference for experienced workers (i.e., the coefficient of $Inexperienced \times Time_based \times After$ is significantly positive) in time-based projects in which the monitoring system is available. The results indicate that employers lower their preference for experienced workers following the introduction of the monitoring system, suggesting that the cold-start problem has been alleviated.

6.3. Hiring Price and Project Delivery

Having found the change in workers' entry decisions and employers' hiring preferences, we go further to investigate whether and how the monitoring system affects the hiring price of labor and the eventual project delivery. The answers to these questions have important implications not just for employers but for the platform as well. Table 10 details the impact of introducing the monitoring system on the hiring price. We use three different measures of hiring price (i.e., the log price, price normalized by the minimum bid, and price normalized by the minimum budget) and find that the interaction term $Time_based \times After$ is consistently negative. Based on column (1), on average, the hiring price decreases by 19.5% after the introduction of the monitoring system.

Furthermore, we examine various project outcome measures: completion status, employer ratings, on-budget, and on-time evaluations. As presented in Table 11, there is no significant change in project completion, employers' rating of workers' performance, on-budget delivery, and on-time delivery.

Combining project outcome results with those on worker entry and employer preference, we conclude that the lower employer preference for experienced workers and the increase in inexperienced workers' bid entries do not undermine project performance. This echoes Pallais's (2014) inefficient hiring statement that many

inexperienced workers do not lack the abilities or diligence to finish jobs, but rather the opportunities to demonstrate themselves. Overall, we find that the monitoring system lowers the hiring price of labor, while maintaining the quality of project delivery.

7. Discussion and Limitations

In this research, we present a stylized model and empirical evidence that the introduction of a monitoring system lowers entry barriers for inexperienced workers who lack platform reputation. Our empirical estimation exploits the differential availability of the monitoring system for similar projects that are either time-based or fixed-price to conduct a difference-in-differences analysis. We find that the introduction of the monitoring system attracts a larger number and a higher percentage of inexperienced workers. Consistently, we find supporting evidence that there is a decrease in employer preference for experienced workers after the introduction of a monitoring system. We also observe that monitoring helps employers save labor costs without compromising the quality of project delivery. Overall, our findings suggest that monitoring systems have the potential to alleviate the cold-start problem in online labor markets.

As reputation systems have been widely incorporated into online platforms to alleviate moral hazard (e.g., Delarocas 2006, Hui et al. 2016), an unintended consequence is that new workers find it excessively challenging to kick-start their careers on a particular platform, on which employers primarily screen workers based on their pre-existing platform reputation. This implies a high entry barrier for qualified new entrants, also known as the cold-start problem (Pallais 2014). This paper contributes to two streams of literature. First, our study contributes to the literature on online platforms (Kokkodis and Ipeirotis 2016, Horton 2019, Li and Zhu 2021),

Table 10. Estimation Results of the IPTW-DID Estimation on the Hiring Price

Model	(1)	(2)	(3)
Dependent variable	<i>Log_bid_price</i>	<i>Premium_min_bid</i>	<i>Premium_min_budget</i>
<i>Time_based</i>	−0.360*** (0.053)	0.181** (0.078)	−0.121 (0.143)
<i>Time_based</i> × <i>After</i>	−0.217*** (0.050)	−0.357*** (0.092)	−0.511*** (0.142)
<i>Log_budget_max</i>	0.721*** (0.025)	0.004 (0.027)	−0.015 (0.053)
<i>Log_title_length</i>	−0.071*** (0.027)	−0.081 (0.061)	−0.046 (0.093)
<i>Log_desc_length</i>	−0.023 (0.015)	0.012 (0.032)	0.074 (0.057)
Category dummies	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes
Employer dummies	Yes	Yes	Yes
Clusters (employers)	1,941	1,941	1,941
Observations	5,413	5,413	5,413
<i>R</i> ²	0.929	0.386	0.497

Notes. We calculate the dependent variables of the last two columns according to the following equations: $Premium_min_bid = (Bid_price - Min_bid)/Min_bid$ and $Premium_min_budget = (Bid_price - Min_budget)/Min_budget$. The results are highly consistent when we control for the week dummies instead of month dummies. Robust standard errors clustered on employers are reported in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

and particularly the emerging literature on the platform entry barrier for new workers (Pallais 2014, Hui et al. 2020) by theoretically considering and empirically showcasing evidence that monitoring systems can help to alleviate the cold-start problem. Notably, monitoring resolves the workers' moral hazard problem differently from reputation systems. Unlike worker reputation that accrues on specific platforms over time and creates an entry barrier for inexperienced workers, monitoring systems do not penalize inexperienced workers who lack platform experience. Second, as monitoring is increasingly important as remote positions become prevalent, it is essential to understand how monitoring may affect workers' job-seeking behaviors and employers' preferences for workers. Extending the prior work that focuses on the effect of monitoring on work performance after the employment arrangement is made (Pierce et al. 2015, Staats et al. 2017) and studies in a context in which monitoring is often costly (Brown and Potoski 2003, Chen and Bharadwaj 2009), we underscore the impact of automated monitoring systems on strategic behaviors of workers in their bidding decisions and employers in their hiring decisions, labor costs, and project delivery.

Our research also provides important managerial implications for platform design. Thus far, many small and medium-sized online labor platforms face challenges of sustained growth and fierce competition, that is, to attract new workers from rival platforms, partially because of the increasing difficulty for new workers to find buyers or employment, which discourages their entry to new platforms. Our paper suggests that monitoring systems play a pivotal role in propelling sustained growth for online labor platforms by alleviating the cold-

start problem with particular relevance to small and medium-sized platforms. For those platforms with primarily manual tasks that are easy to monitor, monitoring systems could be very beneficial. Further, even when platforms do not offer complimentary access to monitoring systems, employers are still able to resort to third-party monitoring solutions if the associated usage costs are outweighed by labor cost savings.

Finally, we acknowledge several limitations of this study, which open up avenues for future research. First, because of data limitations, employers' actual usage of records from the monitoring system is not available. Therefore, our study essentially provides an estimate of the intent to treat (ITT) (Ashraf et al. 2006), which serves as a realistic estimation of the policy change for the platform (Barrow et al. 2009). Prospective research, armed with comprehensive monitoring system usage data, could intricately estimate the treatment on the treated (TOT).¹⁹ Second, we only focus on examining how the introduction of the monitoring system influences workers' platform entries. Future research may consider exploring the long-term effect of monitoring on workers' skill investment. Finally, our study is conducted in the context of an online labor market, and our findings may be limited in their generalizability to other types of platforms. As the characteristics of products or services sold on online platforms vary, the design of monitoring systems should adapt accordingly, and the effect size of implementing such systems may vary based on factors such as the breadth of activities they track. Recently, some artificial intelligence (AI) monitoring software has emerged, and they are increasingly popular in the remote work context.²⁰ Further research should explore

Table 11. Estimation Results of the IPTW-DID Estimation on Project Delivery

Model	(1)	(2)	(3)	(4)
Dependent variable	<i>Completion</i>	<i>Rating_score</i>	<i>On_budget</i>	<i>On_time</i>
<i>Time_based</i>	−0.003 (0.002)	0.024 (0.063)	−0.017 (0.010)	−0.027** (0.012)
<i>Time_based</i> × <i>After</i>	−0.001 (0.005)	−0.004 (0.050)	0.004 (0.010)	0.015 (0.013)
<i>Log_budget_max</i>	−0.002 (0.002)	−0.008 (0.015)	−0.006* (0.003)	−0.008** (0.003)
<i>Log_title_length</i>	0.000 (0.002)	0.030 (0.035)	0.005 (0.007)	0.008 (0.008)
<i>Log_desc_length</i>	−0.001 (0.001)	−0.043*** (0.014)	−0.006* (0.003)	−0.010** (0.004)
Category dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
Employer dummies	Yes	Yes	Yes	Yes
Clusters (employers)	1,941	1,941	1,933	1,878
Observations	5,413	5,413	5,389	5,188
R ²	0.469	0.491	0.477	0.542

Notes. Robust standard errors clustered on employers are reported in parentheses. *Completion* is a dummy denoting whether the employer marked the project as completed or not; *Rating_score* is the overall rating given by the employer regarding the worker's work (five stars being the highest); *On_budget* is a dummy denoting whether the employer agrees that the worker completed the project for the agreed price; *On_time* is a dummy denoting whether the employer agrees that the worker completed the project on the agreed deadline. There are some missing values in *On_budget* and *On_time*. Therefore, the sample sizes in columns (3) and (4) are less than the first two columns.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

how to design monitoring systems to better suit other online platforms and the potential of AI in improving monitoring efficiency.

Acknowledgments

The authors thank the senior editor Xiaoquan (Michael) Zhang, the associate editor Beibei Li, and the anonymous reviewers for their constructive comments throughout the review process.

Endnotes

¹ The information asymmetry regarding workers' effort is commonly referred to as ex post information asymmetry because it arises after the employment contract is established.

² See more on <https://hbr.org/2014/09/breaking-down-the-freelance-economy>.

³ In fact, according to Upwork's Future Workforce Report, 60% of U.S. hiring managers reported difficulties in finding quality workers to fill open positions in online labor platforms. See more on <https://www.upwork.com/research/future-workforce-report-2022>.

⁴ See more on <https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic>.

⁵ In particular, in the gig economy context, Amazon Mechanical Turk monitors workers' time spent on tasks (Liang et al. 2023); Uber monitors drivers' driving behavior and trajectory (Liu et al. 2021); and Upwork monitors workers' working hours, computer screenshots, and keystrokes (Kuhn and Maleki 2017).

⁶ See <https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic>.

⁷ Note that the labor (time) allocated to the online platform L is not an exogenous variable. In fact, because of the market clearing assumption, the worker's labor output (L^α) is equal to the demand (D). Therefore, $L = D^{1/\alpha}(P)$, implying that L is a function of demand (D), which is determined by price (P). When $L > 0$, indicating $PL^\alpha + W(1 - L) > W$, it follows that $\log W < \log P - (1 - \alpha)\log L$. Utilizing the same notation in which $w = \log W$, $p = \log P$, $l = \log L$, and $d = \log D = \alpha l$, the condition can be expressed as $w < p - \frac{(1-\alpha)}{\alpha}d$.

⁸ As shown by Laouénan and Rathelot (2022), the theoretical predictions are similar even when the unobserved component of quality u follows a nonnormal distribution.

⁹ The release date of the monitoring system has been confirmed based on the platform's official blogs and Facebook page.

¹⁰ The application does not track time spent on fixed-price projects because workers can only find time-based projects rather than fixed-price projects in this application.

¹¹ Here, the selection on unobservables refers to the case when there may be some unobservables affecting employers' contract-type choices that also affect workers' bidding behavior. However, as explained soon after, those unobservables affecting employers' contract-type choices are unlikely to be known by workers and, thus, would unlikely affect workers' bidding behavior in online labor platforms.

¹² Because the dependent variable takes the log transformation, we transform the change in the coefficient with the exponential function to obtain the actual percentage change in the number of bids. $\text{Exp}(0.245) - 1 = 27.8\%$. Similar calculations are done for other effects.

¹³ Different from the model-free evidence, the difference in the number of bids from inexperienced workers between fixed-price

and time-based projects before the introduction of the monitoring system is not significant any more after controlling for the category dummies, time effect, and employer fixed effects.

¹⁴ As shown in Online Appendix D, the increase in number of bids from experienced workers is only marginally significant in the full sample and not significant in the CEM sample. In other robustness checks shown in Section 5, the increase in number of bids from experienced workers is only sporadically significant.

¹⁵ To find the corresponding SOC code for each project, we use its primary required skill name as the keyword to search for related occupations that require this skill in the O*Net database. After that, we manually verify whether the definition of the occupation is consistent with the skill definition and application.

¹⁶ Note that, even though we cannot identify the main effect of the *After* dummy in the DID model with month dummies, the interaction term *After* \times *High_manual* dummy can be identified.

¹⁷ We calculate worker i 's number of ratings until worker i 's bid for project j based on the full history of i 's ratings and the specific bidding time.

¹⁸ As noted by Horrace and Oaxaca (2006), a common critique of the LPM is that its predicted probability could be outside the $[0, 1]$ bound. Because our study is an existence-of-effect paper and our focus is not about prediction, this issue is relatively secondary. We also show that results are consistent if use a conditional logit model (see Online Appendix M).

¹⁹ TOT is another common estimate, which is equal to the ratio of ITT to the difference in the proportion treated (i.e., the adoption rate of monitoring in time-based projects). Inherently, the TOT definition implies that the ITT is a more conservative estimate than TOT.

²⁰ See <https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic>.

References

- Abadie A (2005) Semiparametric difference-in-differences estimators. *Rev. Econom. Stud.* 72(1):1–19.
- Adomavicius G, Gupta A, Sanyal P (2012) Effect of information feedback on the outcomes and dynamics of multisourcing multiattribute procurement auctions. *J. Management Inform. Systems* 28(4):199–230.
- Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Polit. Econom.* 113(1):151–184.
- Angelov D (2020) Top2vec: Distributed representations of topics. Preprint, submitted August 19, <https://arxiv.org/abs/2008.09470>.
- Angrist JD, Pischke JS (2008) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).
- Ashraf N, Karlan D, Yin W (2006) Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quart. J. Econom.* 121(2):635–672.
- Autor DH (2003) Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *J. Labor Econom.* 21(1):1–42.
- Autor DH, Levy F, Murnane RJ (2003) The skill content of recent technological change: An empirical exploration. *Quart. J. Econom.* 118(4):1279–1333.
- Bajari P, Tadelis S (2001) Incentives vs. transaction costs: A theory of procurement contracts. *RAND J. Econom.* 32(3):387–407.
- Banerjee AV, Duflo E (2000) Reputation effects and the limits of contracting: A study of the Indian software industry. *Quart. J. Econom.* 115(3):989–1017.

- Banker RD, Hwang I (2008) Importance of measures of past performance: Empirical evidence on quality of e-service providers. *Contemporary Accounting Res.* 25(2):307–337.
- Barrow L, Markman L, Rouse CE (2009) Technology's edge: The educational benefits of computer-aided instruction. *Amer. Econom. J. Econom. Policy* 1(1):52–74.
- Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.
- Billor N, Hadi AS, Velleman PF (2000) BACON: Blocked adaptive computationally efficient outlier nominators. *Comput. Statist. Data Anal.* 34(3):279–298.
- Brodsky A, Amabile TM (2018) The downside of downtime: The prevalence and work pacing consequences of idle time at work. *J. Appl. Psych.* 103(5):496–512.
- Brown TL, Potoski M (2003) Managing contract performance: A transaction costs approach. *J. Policy Anal. Management* 22(2):275–297.
- Bunch DS, Smiley R (1992) Who deters entry? Evidence on the use of strategic entry deterrents. *Rev. Econom. Stat.* 74(3):509–521.
- Butler JV, Carbone E, Conzo P, Spagnolo G (2020) Past performance and entry in procurement: An experimental investigation. *J. Econom. Behav. Organ.* 173(C):179–195.
- Che YK (1993) Design competition through multidimensional auctions. *RAND J. Econom.* 24(4):668–680.
- Chen Y, Bharadwaj A (2009) An empirical analysis of contract structures in IT outsourcing. *Inform. Systems Res.* 20(4):484–506.
- Chen PY, Hong Y, Liu Y (2017) The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Sci.* 64(10):4629–4647.
- Chen Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth vs. observational learning. *J. Marketing Res.* 48(2):238–254.
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W (2017) Double/debiased/Neyman machine learning of treatment effects. *Amer. Econom. Rev.* 107(5):261–265.
- David H, Dorn D (2013) The growth of low-skill service jobs and the polarization of the US labor market. *Amer. Econom. Rev.* 103(5):1553–1597.
- Dellarocas C (2006) How often should reputation mechanisms update a trader's reputation profile? *Inform. Systems Res.* 17(3):271–285.
- Dellarocas C, Dini F, Spagnolo G (2006) Designing reputation (feedback) mechanisms. Dimitri N, Piga G, Spagnolo G, eds. *Handbook of Procurement* (Cambridge University Press, Cambridge).
- Dewan S, Ho YJ, Ramaprasad J (2017) Popularity or proximity: Characterizing the nature of social influence in an online music community. *Inform. Systems Res.* 28(1):117–136.
- Duflo E, Hanna R, Ryan SP (2012) Incentives work: Getting teachers to come to school. *Amer. Econom. Rev.* 102(4):1241–1278.
- Dwyer RE (2013) The care economy? Gender, economic restructuring, and job polarization in the US labor market. *Amer. Sociol. Rev.* 78(3):390–416.
- Facure M (2023) *Causal Inference in Python*. (O'Reilly Media, Inc., Sebastopol, CA).
- Farrell J (1986) Moral hazard as an entry barrier. *RAND J. Econom.* 17(3):440–449.
- Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M (2011) Doubly robust estimation of causal effects. *Amer. J. Epidemiology* 173(7):761–767.
- Gao C, Zuzul T, Jones G, Khanna T (2017) Overcoming institutional voids: A reputation-based view of long-run survival. *Strategic Management J.* 38(11):2147–2167.
- Goos M, Manning A (2007) Lousy and lovely jobs: The rising polarization of work in Britain. *Rev. Econom. Statist.* 89(1):118–133.
- Greene W (2012) *Econometric Analysis*, 7th ed. (Prentice Hall, Upper Saddle River, NJ).
- Greenwood BN, Agarwal R (2016) Matching platforms and HIV incidence: An empirical investigation of race, gender, and socioeconomic status. *Management Sci.* 62(8):2281–2303.
- Gruca TS, Sudharshan D (1995) A framework for entry deterrence strategy: The competitive environment, choices, and consequences. *J. Marketing* 59(3):44–55.
- Hong Y, Pavlou PA (2017) On buyer selection of service providers in online outsourcing platforms for IT services. *Inform. Systems Res.* 28(3):547–562.
- Horrace WC, Oaxaca RL (2006) Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Econom. Lett.* 90(3):321–327.
- Horton JJ (2019) Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Sci.* 65(8):3518–3540.
- Hubbard TN (2000) The demand for monitoring technologies: The case of trucking. *Quart. J. Econom.* 115(2):533–560.
- Hui X, Liu Z, Zhang W (2020) Mitigating the cold-start problem in reputation systems: Evidence from a field experiment. Preprint, submitted January 19, 2021, <https://dx.doi.org/10.2139/ssrn.3731169>.
- Hui X, Saedi M, Shen Z, Sundaresan N (2016) Reputation and regulations: Evidence from eBay. *Management Sci.* 62(12):3604–3616.
- Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened exact matching. *Political Anal.* 20(1):1–24.
- Imai K, Ratkovic M (2014) Covariate balancing propensity score. *J. Roy. Statist. Soc. Ser. B Statist. Methodology* 76(1):243–263.
- International Labour Organization (2021) *World Employment and Social Outlook 2021: The Role of Digital Labour Platforms in Transforming the World of Work* (International Labour Organisation, Geneva).
- Jeong M, Minson J, Yeomans M, Gino F (2019) Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Sci.* 65(12):5813–5837.
- Jessmer SL, Anderson D (2001) The effect of politeness and grammar on user perceptions of electronic mail. *North Amer. J. Psych.* 3(2):331–346.
- Keele LJ, Small DS, Hsu JY, Fogarty CB (2019) Patterns of effects and sensitivity analysis for differences-in-differences. Preprint, submitted January 7, <https://arxiv.org/abs/1901.01869>.
- Kokkodis M, Ipeiotis PG (2016) Reputation transferability in online labor markets. *Management Sci.* 62(6):1687–1706.
- Kuhn KM, Maleki A (2017) Micro-entrepreneurs, dependent contractors, and instaservers: Understanding online platform workforces. *Acad. Management Perspect.* 31(3):183–200.
- Laouénan M, Rathelot R (2022) Can information reduce ethnic discrimination? Evidence from Airbnb. *Amer. Econom. J. Appl. Econom.* 14(1):107–132.
- Lee SY, Rui H, Whinston AB (2019) Is best answer really the best answer? The politeness bias. *Management Inform. Systems Quart.* 43(2):579–600.
- Li H, Zhu F (2021) Information transparency, multihoming, and platform competition: A natural experiment in the daily deals market. *Management Sci.* 67(7):4384–4407.
- Liang C, Hong Y, Chen PY, Shao BB (2022) The screening role of design parameters for service procurement auctions in online service outsourcing platforms. *Inform. Systems Res.* 33(4):1324–1343.
- Liang C, Peng J, Hong Y, Gu B (2023) The hidden costs and benefits of monitoring in the gig economy. *Inform. Systems Res.* 34(1):297–318.
- Liu M, Brynjolfsson E, Dowlatabadi J (2021) Do digital platforms reduce moral hazard? The case of Uber and taxis. *Management Sci.* 67(8):4665–4685.
- Ludwig S, Herhausen D, Grewal D, Bove L, Benoit S, De Ruyter K, Urwin P (2022) Communication in the gig economy: Buying and selling in online freelance marketplaces. *J. Marketing* 86(4):141–161.
- Moore S, Hayes LJB (2018) The electronic monitoring of care work—The redefinition of paid working time. *Humans Mach. Work*. (Springer, New York), 101–124.
- Moreno A, Terwiesch C (2014) Doing business with strangers: Reputation in online service marketplaces. *Inform. Systems Res.* 25(4):865–886.

- Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *J. Bus. Econom. Statist.* 37(2):187–204.
- Pallais A (2014) Inefficient hiring in entry-level labor markets. *Amer. Econom. Rev.* 104(11):3565–3599.
- Pich MT, Loch CH, Meyer AD (2002) On uncertainty, ambiguity, and complexity in project management. *Management Sci.* 48(8):1008–1023.
- Pierce L, Snow DC, McAfee A (2015) Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Sci.* 61(10):2299–2319.
- Powell D, Seabury S (2018) Medical care spending and labor market outcomes: Evidence from workers' compensation reforms. *Amer. Econom. Rev.* 108(10):2995–3027.
- Prendergast C (1999) The provision of incentives in firms. *J. Econom. Literature* 37(1):7–63.
- Pu J, Chen Y, Qiu L, Cheng HK (2020) Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Inform. Systems Res.* 31(2):297–322.
- Raith M (2008) Specific knowledge and performance measurement. *RAND J. Econom.* 39(4):1059–1079.
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.
- Sant'Anna PHC, Zhao J (2020) Doubly robust difference-in-differences estimators. *J. Econometrics* 219(1):101–122.
- Shapiro C (1982) Consumer information, product quality, and seller reputation. *Bell J. Econom.* 13(1):20–35.
- Shapiro C, Stiglitz JE (1984) Equilibrium unemployment as a worker discipline device. *Amer. Econom. Rev.* 74(3):433–444.
- Sheth JN (1976) Buyer-seller interaction: A conceptual framework. *Adv. Consum. Res.* (Association for Consumer Research, Cincinnati, OH), 382–386.
- Staats BR, Dai H, Hofmann D, Milkman KL (2017) Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Sci.* 63(5):1563–1585.
- Tadelis S (2016) Reputation and feedback systems in online platform markets. *Annual Rev. Econom.* 8(1):321–340.
- Wang C, Zhang X, Hann IH (2018) Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Inform. Systems Res.* 29(3):641–655.
- Williams KC, Spiro RL (1985) Communication style in the salesperson-customer dyad. *J. Marketing Res.* 22(4):434–442.
- Williamson OE (1981) The economics of organization: The transaction cost approach. *Amer. J. Sociol.* 87(3):548–577.
- Wood AJ, Graham M, Lehdonvirta V, Hjorth I (2019) Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work Employment Soc.* 33(1):56–75.
- Yeomans M, Kantor A, Tingley D (2018) Detecting politeness in natural language. *R J.* 10(2):489–502.
- Zhang XM, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.
- Zhao RR (2008) All-or-nothing monitoring. *Amer. Econom. Rev.* 98(4):1619–1628.