IT-enabled Monitoring in the Gig Economy

by

Chen Liang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2019 by the
Graduate Supervisory Committee:

Bin Gu, Co-Chair
Yili (Kevin) Hong, Co-Chair
Pei-yu Chen

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Two-sided online platforms are typically plagued by hidden information (adverse selection) and hidden actions (moral hazard), limiting market efficiency. Under the context of the increasingly popular online labor contracting platforms, this dissertation investigates whether and how IT-enabled monitoring systems can mitigate moral hazard and reshape the labor demand and supply by providing detailed information about workers' effort. In the first chapter, I propose and demonstrate that monitoring records can substitute for reputation signals such that they attract more qualified inexperienced workers to enter the marketplace. Specifically, only the effort-related reputation information is substituted by monitoring but the capability-related reputation information. In line with this, monitoring can lower the entry barrier for inexperienced workers on platforms. In the second chapter, I investigate if there is home bias for local workers when employers make the hiring decisions. I further show the existence of home bias from employers and it is primarily driven by statistical inference instead of personal "taste". In the last chapter, I examine if females tend to have a stronger avoidance of monitoring than males. With the combination of the observational data and experimental data, I find that there is a gender difference in avoidance of monitoring and the introduction of the monitoring system increases the gender wage gap due to genders differences in such willingness-to-pay for the avoidance of monitoring. These three studies jointly contribute to the literature on the online platforms, gig economy and agency theory by elucidating the critical role of IT-enabled monitoring.

ACKNOWLEDGMENTS

I feel so honored and privileged to have excellent advisors, a supportive family, and caring friends during the PhD journey.

First of all, I would like to express sincere appreciation to my mentors, Professor Bin Gu and Professor Yili (Kevin) Hong for your encouragement, patience, and guidance in helping to shape this dissertation. I am so blessed to have both of you as wonderful role models for my career. Meanwhile, I am very grateful for my committee member, Professor Peiyu Chen, for giving me important suggestions on my training and research. Moreover, I would like to thank Professor Raghu Santanam, Michael (Zhan) Shi, Benjamin Shao for profound discussions and valuable help of various types.

I would like to thank Jessica Pye, John Zhang, Sang Pil Han, Victor Benjamin, Nina Huang, Shinyi Wu, Uday Kulkarni, Robert D. St. Louis, and other faculty members in the IS department at ASU for their general support. Moreover, I thank Angelina Saric, Erica Grimwood, and Janeen Christensen for all the administrative help.

Additionally, many thanks to National Science Foundation Doctoral Dissertation Research Improvement Grants, NET Institute Grants, AWS Research Grants, and Center for the Study of Economic Liberty for the consecutive financial support.

Finally, I want to thank my parents, my grandparents, my sister and brother for their unflinching love and support. Thanks a lot to my friend Jing Peng, Lily Yuan, Jiamin Huang, and Wei Huang for your understanding and warm encouragement.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

INTRODUCTION

The "Gig" economy is thriving today with short-term jobs increasingly replacing traditional long-term jobs. Despite the great success of Gig economy in recent years, it is characterized by information asymmetries, including ex ante hidden information of workers and ex post hidden actions by workers, which undermine the efficiency of hiring outcomes. To mitigate information asymmetry, online monitoring, an IT-enabled technology, has gained popularity among online platforms. Specifically, with online monitoring, employers can observe workers' ex post actions through screenshots and webcams. While several studies have examined the impact of monitoring systems on workers' performance in offline contexts, few studies have considered the role of monitoring, particularly its potential interactions with other proxies of signals (e.g., reputation, nationality, and gender), which have important implications for competition and market efficiency in online employment. In my dissertation, I attempt to address the following three research questions:

1) Does monitoring substitute for reputation signals? If so, what type of reputation information is substituted by monitoring? Does monitoring lower the entry barrier for inexperienced workers?

2) What are the underlying mechanisms for home bias in online employment? Does home bias decrease with the introduction of monitoring?

3) Do females have a higher willingness-to-pay for the avoidance of monitoring than males do? Does monitoring increase the gender wage gap due to the gender difference in avoidance of monitoring?

I study these problems in three essays. Each of these essays deals with one of the major problems of the market: entry barrier, home bias and gender wage gap, with each piece contributing a different perspective to understand the impact of IT artifacts on the future of work.

My first essay investigates whether the introduction of an IT-enabled monitoring system mitigates moral hazard in online platforms by providing direct information on workers' effort. Our identification hinges on a quasi-natural experiment at Freelancer when the platform introduced an IT-enabled monitoring system for time-based projects but not for fixed-price projects. I find that IT-enabled monitoring systems can alleviate moral hazard, reduce the effect of effort-related reputation, and intensify supply-side competition.

My second essay studies the nature of home bias in online employment, wherein employers prefer workers hailing from the same home countries. Using a unique large-scale dataset from a major online labor market containing employers' consideration sets of workers and their ultimate selection of workers, I first empirically demonstrate that employers do exhibit home bias in their hiring decisions. Then, I use a quasi-natural experiment to examine the extent of statistical and taste-based home bias, respectively.

My third essay explores whether there exists a gender wage gap in the gig economy and examines to what extent the gap could be accounted for by gender differences in job application strategies. I find that females only earn around 81.4% of the hourly wage of their male counterparts. I further show that the gender wage gap can be largely explained by gender differences in job application strategies, including bid timing, job selection, and avoidance of monitoring. Overall, this study suggests the important role of job application strategies in the persistent gender wage gap.

CHAPTER 1

IT-ENABLED MONITORING AND LABOR CONTRACTING IN ONLINE PLATFORMS: EVIDENCE FROM A QUASI-NATURAL EXPERIMENT

Situated in the context of the increasingly popular online platforms for labor contracting (herein referred to as "online labor markets") where market efficiency is limited by information asymmetry, this paper investigates whether IT-enabled monitoring systems can mitigate moral hazard by providing detailed information about workers' efforts. Our identification hinges on a quasi-natural experiment at Freelancer, following the introduction of a monitoring system for time-based projects but not for fixed-price projects in February 2014. Based on a unique dataset comprising 5,383 fixed-price projects and 3,099 time-based projects matched on observable characteristics, we employ a difference-in-differences (DID) approach to identify the effect of the monitoring system on outcomes on both the demand side (i.e., employers' worker choices) and the supply side (i.e., workers' entry decisions). To that end, we decompose workers' reputations into two parts: effort-related and capability-related reputations. We observe that the introduction of the monitoring system decreases employer preference for bidders with high effort-related reputations for time-based projects, thus lowering the entry barrier for workers who have not yet established reputations. However, there is no significant change in employer preference for bidders with high capability-related reputations. Further, the introduction of the monitoring system increases the number of bids on time-based projects by 24.7% (primarily from bidders with no prior experience on the platform). Our results demonstrate a partial substitution relationship between reputation

systems and monitoring systems, and further suggest that IT-enabled monitoring systems have a significant effect on alleviating moral hazard, reducing agency costs, and intensifying supply-side platform competition.

Keywords: online labor market, moral hazard, monitoring systems, reputation systems, entry barrier

1.1 Introduction

Platform-based businesses are thriving in today's economy (Anderson et al. 2013; Parker et al. 2016; Eisenmann et al. 2006, 2011; Van Alstyne et al. 2016). Numerous new business models have been developed based on the platform paradigm, ranging from platforms that enable transactions of physical products (e.g., eBay, Taobao, Amazon), to ride sharing (e.g., Uber, Lyft) and short-term lodging (e.g, AirBnB, CouchSurfing). Online labor markets—two-sided platforms that connect employers with freelance workers—are at the forefront of this phenomenon. Over the past decade, online labor markets have experienced tremendous growth. As a prominent example, as of August 2018, about 29 million registered users have either posted (employers) or bid on (workers) millions of projects at Freelancer,[1] one of the major online labor markets.

Despite tremendous growth, online labor markets are plagued by two forms of information asymmetry—hidden information and hidden action—which can lead to significant agency problems. Hidden information refers to workers possessing ex ante private information about their capabilities and skills (Bolton and Dewatripont 2005; Horton 2017), which makes it difficult for employers to evaluate workers (Eisenhardt 1989). In such scenarios, employers tend to make contract decisions based on their beliefs about the distribution of capabilities and skills, which so-called adverse selection problems (Akerlo 1978; Hart and Holmstrom 1987; Greenwald 1986). In contrast, hidden action relates to ex post information asymmetry regarding workers' actual actions, such as the amount of time and effort spent on projects. Due to ex post information

[1] https://www.freelancer.com/community/articles/20-million-users-things-that-made-this-milestone-remarkable-for-freelancer-com

asymmetry, moral hazard occurs, when workers opportunistically misrepresent their effort levels to maximize their own utility, to the detriment of employers after the initiation of the contract (Pauly 1974; Holmstrom 1979,1982; Eisenhardt 1989).

To mitigate information asymmetry and both types of agency problems, a strategy commonly employed by employers is contract design. In general, two contract forms are available in online labor markets: time-based contracts and fixed-price contracts. For time-based contracts, compensation is based on the hourly wage set in the contract and the number of hours the worker has spent on the contracted project (Mani et al. 2012). While time-based contracts provide stronger incentives for high-quality project outcomes and a higher flexibility in renegotiation (Dey et al. 2010; Mani et al. 2012), they are more susceptible to moral hazard because workers' compensation is not directly linked to the project outcome (Dey et al. 2010; Mani et al. 2012). Specifically, when there is a low probability that shirking will be noticed (monitoring efficiency), workers tend to overreport their work hours (also known as hours-padding). In fixed-price contracts, workers' compensation is dependent on the outcome of a project, such that the worker receives payment only when the project has been completed (Mani et al. 2012). Therefore, fixed-price contracts provide enough incentive for workers to complete projects, suggesting a lower moral hazard risk in terms of cost-padding. However, fixed-price contracts can involve corner-cutting behavior, higher ex ante costs of contract design (Susarla et al. 2009), and higher ex post costs of maladaptation and renegotiation (Benaroch et al. 2016). The trade-off between two types of contracts is also referred as the "make-or-buy" decision in the Transaction Cost Economics (TCE) literature (e.g., Coase 1937; Williamson 1981; Walker and Weber 1984; Bajari and Tadelis 2001). The decision of contract forms is mostly dependent on employers' sensitivity to quality, task complexity, and uncertainty (Dey et al. 2010; Bajari and Tadelis 2001). Taken together, both time-based contracts and fixed-price contracts are limited in their efficacy for fully resolving information asymmetry in online labor markets.

To further help to alleviate agency problems, major online platforms have developed reputation systems and online monitoring systems, both of which are IT systems designed to reduce information asymmetry. First, reputation systems mitigate information asymmetry problems using positive externality derived from information sharing among users (e.g., Dellarocas 2006; Moreno and Terwiesch 2014). Specifically, reputation systems allow employers to share their experiences about workers, which help other employers screen for capable and diligent workers who are willing to expend commensurate effort for projects, thus mitigating both adverse selection and moral hazard. For one thing, reputation information regarding workers' capabilities (here referred as "capability-related reputation") lowers the likelihood that workers would misrepresent their capabilities to win contracts, which helps to alleviate adverse-selection problems. For another, reputation information reflecting workers' effort in previous projects (here referred as "effort-related reputation") serves as a sanctioning device that deters worker shirking behavior even if employers cannot observe workers' actual effort (Banker and Hwang 2008)—thus lowering the moral hazard risk. Taking reputation ratings as effective signals, employers use their beliefs about the capabilities and effort of workers to differentiate them. As a result, workers with high capability-related and effort-related reputations enjoy higher winning probabilities and price premiums (Ba and Pavlou 2002; Moreno and Terwiesch 2014). One unintended consequence of reputation systems, however, is that they create an entry barrier for qualified workers who have not yet established their reputations on a particular platform—known as the cold-start problem (Pallais 2014).

Second, IT-enabled monitoring systems, which have become increasingly popular among online platforms (Aron et al. 2007; Agrawal et al. 2014), serve as an effective means for employers to obtain detailed information on the actions of workers, thus mitigating moral hazard by addressing the hidden action issue (Bolton and Dewatripont 2005). The effectiveness of monitoring for increasing workers' effort and thus leading to better performance has been shown in multiple offline employment contexts, such as the trucking industry (Hubbard 2000), schools (Duflo et al. 2012), restaurants (Pierce et al. 2015), and hospitals (Staats et al. 2016). On online platforms where work is typically done remotely with the use of a suite of IT-enabled monitoring technologies, employers can observe workers' progress through screenshots, webcams, and even keystroke recordings from automatically archived log files, which offer firsthand information about workers' effort and can help alleviate employers' concerns about moral hazard among workers. However, these log files and tracked work hours are produced only after the contract is written; therefore, they are not useful for precontractual screening of worker capabilities and cannot alleviate the hidden information problem. In summary, monitoring is more effective for mitigating hidden action than for alleviating hidden information.

While a significant amount of research effort has been devoted to the design, evaluation, and optimization of reputation systems (Banker and Hwang 2008; Bockstedt and Goh 2011; Dellarocas 2006; Yoganarasimhan 2013) and the effectiveness of monitoring systems in mitigating moral hazard problems in offline settings (Drago 1991; Duflo et al. 2012; Hubbard 2000; Pierce et al. 2015), the effect of monitoring systems and reputation systems are usually investigated separately, without considering how they can jointly mitigate hidden action problems or whether they can serve as substitutes for or complements to each other (Demiroglu and James 2010; Diamond 1991). Therefore, a key research gap in the literature involves disentangling the roles of monitoring systems and reputation systems in addressing hidden action and hidden information problems. In particular, it is also important to theoretically and empirically differentiate the interaction between monitoring and capability-related reputation, versus the interaction between monitoring and effort-related reputation. To fill this research gap, we extend extant work, such as Lin et al.'s (2016), by isolating the moderating effect of the monitoring system, from that of contract type, on the effectiveness of reputation. Moreover, while there is a strand of literature investigating the complementarities between information technology (IT) and organizational practices (e.g., performance pay, human resource analytics practices, human capital) (Aral et al. 2012; Tambe and Hitt 2012; Brynjolfsson and Milgrom 2013), our study is distinct in that we focus on the effect of an important IT artifact on employers' hiring decisions and workers' bidding behaviors instead of focusing on the productivity of a relatively stable cohort of workers. The differences between our paper and these two related prior studies (i.e., Lin et al. 2016; Aral et al. 2012) are summarized in our theoretical supplement. Last but not least, given that monitoring systems may reduce employers' reliance on reputation for deterring moral hazard, an immediate follow-up question is whether monitoring systems can help to alleviate the cold-start problem (Pallais 2014), the unintended drawback of reputation systems. Specifically, we attempt to address the following two research questions:

- *How does IT-enabled monitoring moderate the effect of worker reputation on employer contracting decisions? Does the moderation effect vary for effort-related versus capability-related reputation?*
- *How does IT-enabled monitoring influence entry decisions by workers? Does it affect less experienced workers differently than more experienced workers?*

To answer the above research questions, our analyses leverage a quasi-natural experiment on Freelancer, when the platform first introduced an IT-enabled monitoring system on February 5, 2014. This quasi-natural experiment offers an appropriate research design for identifying the effects of IT-enabled monitoring systems in online labor markets. Our econometric identification hinges on the fact that monitoring was implemented for time-based projects, but not for fixed-price projects, which allows us to use time-based projects as the treatment group and fixed-price projects as the control group. Using a dataset from Freelancer.com, one of the leading online labor markets, we first performed propensity score matching (and also coarsened exact matching) to match fixed-price projects to time-based projects. The resulting matched sample of 5,383 fixed-price projects and 3,099 time-based projects are comparable in terms of any observable characteristic. We then use a difference-in-differences (DID) approach to identify the treatment effect of the introduction of the monitoring system on employers' worker choices and on worker entry decisions. Our analyses suggest that after the introduction of the IT-enabled monitoring system, employers place less weight on workers' effort-related reputation information, but not on capability-related reputation information. Further, using fixed-price projects as the baseline, the introduction of the monitoring system increases the number of bid entries in time-based projects by an average of 24.7%, primarily from bidders with no prior experience on the platform.

Our study contributes to the literature on IT-enabled monitoring on three fronts. First, most prior studies have extensively focused on the effect of monitoring on worker performance in offline contexts (e.g., Pierce et al. 2015; Staats et al. 2016; Ranganathan and Benson 2017), whereas this study focuses on the impact of an IT-enabled monitoring artifact on both demand-side (employer) preferences and supply-side (worker) entry barriers on online platforms. Second, our study advances prior literature on the interrelationship between monitoring systems and reputation systems for online platforms (Bakos and Dellarocas 2011; Lin et al. 2016) by showing that the introduction of monitoring systems reduces employers' preference for workers with high effort-related reputations, but does not affect preference for workers with high capability-related reputations. Building on recent work suggesting that the effect of reputation is less significant for time-based projects than for fixed-price projects (Lin et al. 2016), our study's setting allows us to identify the causal effect of implementing the monitoring system on both the supply and demand sides of an online labor market, and also allows us to disentangle effort-related reputation from capability-related reputation. Third, our study shows that the introduction of an IT-enabled monitoring system can lower the entry barrier for inexperienced workers, by reducing the need for the ex ante screening of effort-related reputation.

Table 1. A Summarization of the Significant Differences between Our Study and Highly-Related Prior Studies

Criterion	Lin et al. (2016)	Aral et al. (2012)	The present study	Specific differences and contributions
The impact of the IT-enabled monitoring system on hiring decisions	X	X	✓	This paper leverages a quasi-natural experiment for the causal identification of the introduction of the IT-enabled monitoring system on employers' hiring decision. In the research context of Lin et al. (2016), time-based projects always come with monitoring systems. Due to the lack of variation in the presence of monitoring systems, the authors are not able to isolate whether the reduced effectiveness of reputation is coming from the monitoring system or time-based projects Aral et al. (2012) focuses on the impact of human capital management (HCM) software adoption on the output (or productivity) instead of hiring decisions. This is also in line with the difference between online labor platforms and offline organizations. A unique advantage of online labor platforms is to help to increase the efficiency of the matching between employers and workers, especially for short-term contracts. As such, we underscore the impact of monitoring on employers' hiring decisions. On contrary, labor contracts in offline organizations are usually long-term and the workforce is relatively stable. Aral et al. (2012) mainly investigate how incentive plans and the adoption of human capital management (HCM) software influence the productivity of the relatively stable cohort of workers.
The impact of the IT-enabled monitoring system on entry barrier	X	X	✓	Hinging on a quasi-natural experiment, this paper theoretically proposes and empirically evaluates an important unintended benefit of monitoring systems: lowering the entry barrier for inexperienced workers.
The interaction between the IT-enabled monitoring system and reputation system	X	X	✓	This paper is the first to empirically investigate the interaction between the IT-enabled monitoring system and the reputation system.
The effect of reputation on hiring decisions in two contract forms	✓	X	✓	Similar to Lin et al. (2016), we also consider the effect of reputation on hiring decisions in both time-based and fixed-price contracts.
The differential effect of different dimensions of	X	X	✓	This paper extends Lin et al. (2016) by considering both the capability-related and effort-related

reputation on hiring decisions				reputation, and separating their effects on alleviating hidden information and hidden action.
The impact of the IT-enabled monitoring system on productivity	X	✓	✓	Aral et al. (2012) suggests that the three-way complementarities among information technology (IT), performance pay, and human resource analytics practices have a positive impact on the output (or productivity) of hired workers. This paper focuses on the impact of the IT-enabled monitoring system, a prime example of IT artifacts, on employers' hiring decisions and workers' entry decisions. Our additional analyses in Appendix C further explores the impact of the IT-enabled monitoring system on various measures of project outcomes.

1.2. Research Context and Hypotheses

1.2.1. Research Context

The research context of this study, online labor markets, is a web-based two-sided platform that facilitates contracting labor services around the world (Chan and Wang 2017; Lin et al. 2016; Horton and Golden 2015). In recent years, online labor markets have grown significantly. It is reported that 25 percent of jobs in the U.S. are outsourced offshore (Blinder and Kruger 2013), with a substantial portion delegated through online labor markets.[2] Because of spatial and temporal separations between employers and workers, workers' capabilities are difficult to observe and their actual effort is difficult to monitor. Therefore, information asymmetry is prevalent on these platforms (Hong and Pavlou 2017), making the agency problem and its mitigation major research topics in the literature on online labor markets. Our research context has two notable characteristics regarding its platform design and composition of participants: (1) The reputation system

[2] http://www.forbes.com/sites/groupthink/2014/10/21/the-next-big-thing-in-e-commerce-online-labor-marketplaces

used by our platform has been in place since the inception of the platform, whereas the IT-enabled monitoring system was implemented years later. Moreover, during our observation period, the reputation system presented multidimensional ratings of workers, provided by previous employers (if any). The variation in the presence of the monitoring system enables us to identify the potential interactions between the monitoring system and two distinct dimensions of reputation: capability and effort. (2) The platform approximates a free-flowing environment, which attracts incoming new workers with no prior ratings. This underscores the importance of addressing the cold-start problem facing new workers.

1.2.2. Hypothesis Development

We propose three hypotheses for this study. First, we propose a nuanced substitution effect between monitoring and reputation (H1). Second, we propose that the monitoring system attracts more bids (H2a) and lowers entry barriers for inexperienced workers (H2b) for time-based projects. To further justify our hypotheses, we also provide a stylized analytical model that investigates how increasing monitoring efficiency affects the value of effort-related reputation and the height of the entry barrier. This model is for illustration purpose and is provided in our theoretical supplement.

1.2.2.1. Nuanced Relationship between Monitoring and Reputation

Based on the previous literature, monitoring systems and reputation systems are two prevalent mechanisms for alleviating information asymmetry (Table 2). Specifically, monitoring systems are mainly found to effectively mitigate moral hazard and hidden action in offline employment contexts (Duflo et al. 2012; Hubbard 2000; Pierce et al. 2015; Ranganathan and Benson 2017; Staats et al. 2016). Meanwhile, reputation systems not only help mitigate moral hazard by deterring shirking behaviors, but they also help alleviate adverse selection by enabling precontractual screening. Given the differential effects of monitoring systems and reputation systems, instead of exploring the effect of general reputation, we segment reputation into two types, according to the specific type of information asymmetries that reputation can mitigate: (a) capability-related reputation, which helps to alleviate adverse selection (i.e., ex ante information asymmetry); and (b) effort-related reputation, which is effective in mitigating moral hazard (i.e., ex post information asymmetry). Additionally, we propose a nuanced substitution relationship between reputation and monitoring: monitoring can substitute for effort-related reputation by alleviating moral hazard, but cannot substitute for capability-related reputation.

Table 2. A Comparison between Reputation and Monitoring in Alleviating Information Asymmetry

Asymmetric information	Hidden information	Hidden action
Reputation systems	Provide precontractual screening in online service markets (e.g., Banker and Hwang 2008; Tadelis 1999)	Deter shirking in online trading and service markets (e.g., Dellarocas 2006; Bakos and Dellarocas 2011)
Monitoring systems	Not applicable	Mitigate moral hazard in multiple offline employment contexts (e.g., Duflo et al. 2012; Staats et al. 2016)

Different dimensions of reputation information tend to play differential roles in alleviating adverse selection and moral hazard. Specifically, capability-related reputation, or reputation information describing workers' capabilities (e.g., expertise, skills), serves as an effective quality signal that reflects a worker's capabilities. Employers will generally expect workers with better capability-related reputations to be more capable and to achieve better project outcomes. Accordingly, they tend to prefer workers with higher capability-related reputations, all else being equal. Meanwhile, reputation information based on workers' effort in completing previous projects—namely, effort-related reputation—disincentivizes workers from shirking because the potential negative feedbacks may be observable to future employers (Tadelis 2016). Therefore, effort-related reputation holds the potential to alleviate moral hazard.

The importance of effort-related reputation also depends on whether monitoring systems are in place. When monitoring systems are not available, the probability of workers' shirking going unnoticed is relatively high due to a lack of real-time information on effort, leading to employers' high reliance on effort-related reputation. In particular, due to the spatial and temporal separation between the workers and the employers, manual monitoring through instant audio or video communication tools provided by the platform is not cost-effective or practical for employers. Therefore, in order to mitigate moral hazard, employers tend to exploit effort-related reputation information. Specifically, we expect employers to prefer workers with better effort-related reputations for the following two reasons: First, effort-related reputations may reveal the workers' "commitment type," given that the reputation system serves as a sanctioning device that, to some extent, locks workers into choosing the "not shirking" strategy (Fudenberg and Levine 1989, 1992; Atakan and Ekmekci 2014). Although the stability of this strategy may depend on contextual factors, such as the possibility of receiving unfair ratings (Dellarocas 2006) and the stage of the workers' career life cycle

(Holmstrom 1999), employers tend to expect that workers with better effort-related reputations would be more likely to cooperate and exert sufficient effort to complete their projects on time. Second, employers tend to prefer workers with better effort-related reputations not only because they expect more effort from them, but also because it can lower cost uncertainty due to delay and risk of failure (Mani et al. 2012). As workers with better effort-related reputations are expected to be less likely to fall behind on preplanned work schedules, or even fail to complete projects, they will be perceived as less likely to cause a budget overrun for time-based projects.

Introducing an IT-enabled monitoring system allows employers to observe workers' effort more precisely and cost efficiently. In this way, such a monitoring system can reduce employers' reliance on effort-related reputation for deterring workers' shirking behaviors, leading employers to emphasize effort-related reputation less in their hiring decisions. First, monitoring efficiency significantly increases with the introduction of an IT-enabled monitoring system, which subsequently decreases the probability of workers' shirking going unnoticed. In such cases, workers' expected payoff from shirking decreases remarkably because they get little or no compensation if they are caught shirking. Therefore, a higher percentage of workers, including workers with lower effort-related reputations, will choose to cooperate and expend more effort after the introduction of a monitoring system. As such, when monitoring systems are in place, effort-related reputations become less informative, because they are less likely to be used to separate workers in terms of their commitment types. Second, monitoring systems offer employers real-time information about workers' performance (e.g., offer timely updates of project progress, workflow, etc.) and employers can thus terminate the project at the first sign of potential shirking behavior. Therefore, irrespective of the workers' effort-related reputations, employers can reduce cost uncertainty, due to potential delay in progress or project failure, by consistently monitoring worker performance, which reduces the disparities in expected productivity and cost uncertainty between workers with high effort-related reputations and those with low effort-related reputations.

We also argue that information generated by IT-enabled monitoring systems does not effectively alleviate the problem of hidden information. Therefore, monitoring cannot substitute for capability-related reputation. First, monitoring is implemented after the hiring decision. Therefore, despite a monitoring system in place, employers still need to rely on workers' capability-related reputations to infer worker capabilities in order to make informed hiring decisions. In fact, previous research suggests that the signaling effect of capability-related reputation exists even when moral hazard problems are completely resolved, since adverse selection problems continue to persist (Tadelis 1999). Second, the disparities between workers with high capability-related reputations and those with low capability-related reputations cannot be reduced by monitoring systems. After monitoring systems are implemented, the work quality of low-capability workers will still be inferior to that of the high-capability workers, even given the same level of effort. Therefore, implementing monitoring systems has little effect on employers' preference for workers with high capability-related reputations, given that monitoring cannot effectively mitigate the problem of adverse selection. Therefore, we propose:

H1: Introduction of an IT-enabled monitoring system leads employers to place less emphasis on workers' effort-related reputations for time-based projects, but not on their capability-related reputations.

1.2.2.2. Monitoring, Entry Barrier and Worker Competition

Now we consider the effect of IT-enabled monitoring systems on the worker (supply) side. Absent such monitoring systems, employers must rely on reputation information for purposes of precontractual screening and moral hazard mitigation. This leads to a significant advantage for workers who enter the platform early, as they are able to accrue platform-specific work experience and establish a reputation. Specifically, due to ex ante information asymmetry, employers tend to mitigate adverse selection issues by inferring workers' actual capabilities based on capability-related reputations. Given the lack of capability-related reputation information about inexperienced workers, employers can only infer workers' capabilities based on their beliefs about the distribution of workers' capabilities in a specific market. Thus, employers will tend to prefer workers with high capability-related reputations to inexperienced workers, even though some of the latter are, in fact, highly capable. Similarly, due to ex post information asymmetry and the lack of effort-related reputations for inexperienced workers, employers tend to infer that they would expend an average level of effort in the market. As such, employers would tend to prefer to hire workers with high effort-related reputations, instead of inexperienced workers, because the workers with better reputations invoke less uncertainty about "commitment type" and presumably present lower probabilities of budget overrun. As employers tend to prefer hiring workers with platform reputations (Pallais 2014), inexperienced workers, who have not yet established their reputations for capability and effort, are less likely to participate in the market because they are less likely to land contracts. Moreover, inexperienced workers will presumably only be considered if they propose and accept poorer treatment than reputable workers—namely, less compensation

(Friedman and Resnick 2001). Consequently, the high entry barrier created by the reputation system discourages inexperienced workers from participating in the market.

However, the introduction of an IT-enabled monitoring system allows employers to observe workers' effort based on procedural track records, rather than relying solely on workers' effort-related reputations. In particular, by increasing monitoring efficiency, monitoring systems deter moral hazard by lowering the probability of shirking going unnoticed; thus, workers are incentivized to expend sufficient effort and disincentivized from padding hours, delaying progress or even failing to complete projects. Thus, the introduction of the monitoring system increases the expected market-average effort level—which is also equivalent to the expected effort level of inexperienced workers—and thus decreases the disparities in expected effort level between workers with high effort-related reputations and inexperienced workers. Therefore, although the entry barrier due to capability-related reputation is not affected by the introduction of the monitoring system, the entry barrier due to accumulated effort-related reputation decreases (Demiroglu and James 2010) in time-based contracts, relative to fixed-price contracts. Because of the decreased entry barrier into time-based projects with IT-enabled monitoring, more workers will be likely to bid for time-based contracts when monitoring is in place. Specifically, we expect that the lower entry barrier in time-based projects will disproportionately attract more bids from workers who are qualified but have not yet established their effort-related reputations. Bearing the above in mind, we propose:

H2a: Introduction of an IT-enabled monitoring system leads more bidders to bid for time-based projects.

H2b: Introduction of an IT-enabled monitoring system leads to a higher percentage of workers with no platform experience bidding for time-based projects.

1.3. Data

1.3.1. Data Source

We obtained our data from www.freelancer.com (Freelancer), one of the largest online labor market platforms. At Freelancer, an employer can post a project with a description, estimated budget, and required skills. The employer can choose between two contract types: fixed price contract (Figure 1-a) for which the employer provides the estimated budget for the entire project; or time-based contract (Figure 1-b) for which the employer provides the estimated hourly budget for the project in dollars per hour.

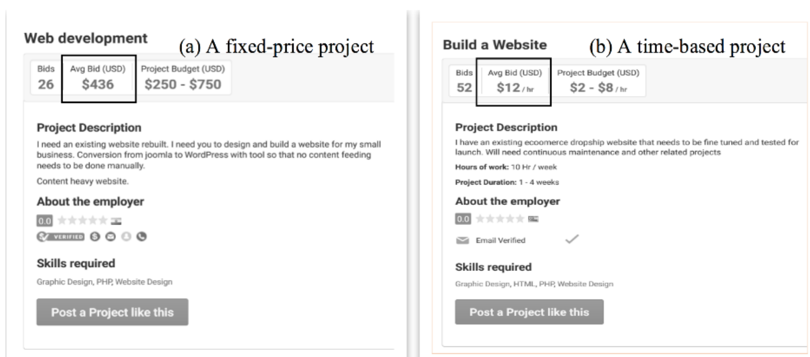


Figure 1. Screenshots of Web Pages for a Fixed-Price versus a Time-Based Project

Typically, a project is open for bidding for one week and any worker is interested in the project can bid on it. For fixed-price projects, each bidder (worker) submits a bid amount for the entire project, whereas for time-based projects, each bidder submits a bid in terms of hourly rate. At the end of the bidding period, the employer reviews bidders' information, including bid amount, former employer ratings, and past project experience. Additionally, sorting tools are available to enable the employer to sort bidders according to their reputations. Once the employer finds a bidder who best satisfies his or her requirements, the employer can award that worker a contract.

1.3.2. Sample and Variables

We obtained a unique archival dataset from Freelancer that includes detailed project information and worker information from September 1, 2013 to August 31, 2014. Following Lin et al. (2016), we construct a matched sample from fixed-price projects for time-based projects. To construct the matched sample, we consider contingent factors that previous studies suggest are associated with contract decisions (Dey et al. 2010; Bajari and Tadelis 2001). Specifically, we limited our sample to awarded projects reflecting realistic labor demand without the contamination of resubmitted projects. Further, to reduce possible selection bias and the association between various pretreatment covariates and contract choices, we matched fixed-price and time-based projects (Abadie 2005; Ho et al. 2007) based on distributions of important covariates suggested by the previous literature (details reported in Table 6) using Propensity Score Matching (PSM). Our final sample includes 5,383 fixed-price projects and 3,099 time-based projects. The dataset includes the following attributes: (1) project-level information (e.g., project description, project budget, contract type, number of bidders, average bid price); (2) worker-level information (e.g., ratings, the amount of reviews); (3) bid-level information (e.g., bid price). The descriptive statistics of the aforementioned dataset variables are shown in Table 3, Table 4 and Table 5, respectively.

Table 3. Definitions and Summary Statistics of Project-Level Variables

Variable	Variable definition	Mean	SD	Min	Max

Budget_min	The minimum of project budget set by the employer	45.17	117.46	0.00	1965.00
Budget_max	The maximum of project budget set by the employer	127.81	212.22	0.00	3000.00
Bid_min	The minimum of bid prices for each project	56.92	122.94	2.00	1965.00
Bid_max[3]	The maximum of bid prices for each project	200.26	562.43	2.00	22272.00
Time-based	A dummy variable; =1 if the project is a time-based project; =0 if the project is a fixed-price project	0.33	0.47	0.00	1.00
Bid_count	Total number of bids received by the project	12.64	13.53	1.00	137.00
Bid_mean	Average bid price for each project	94.70	174.75	2.00	2670.68
Paid_amount	Amount of dollars paid by the employer after the project was completed	129.26	250.04	0.00	2000.00
Project_title_length	Number of words in the project title	5.64	3.21	1.00	40.00
Project_desc_length	Number of characters in the project description	375.06	386.54	1.00	4088.00

Note: Summary statistics are calculated based on the matched sample. We dropped outliers with the Stata command "bacon", using the top 10th quantile of the Mahalanobis distance as a cutoff.

Table 4. Definitions and Summary Statistics of Worker-Level Variables for Dual-Typed Bidders

Variable	Variable definition	Mean	SD	Min	Max
Quality	Average quality rating given by all the employers (ranging from 0 to 5)	4.82	0.34	1.00	5.00
Communication	Average communication rating given by all the employers (ranging from 0 to 5)	4.83	0.34	1.00	5.00
Expertise	Average expertise rating given by all the employers (ranging from 0 to 5)	4.82	0.34	1.00	5.00
Professionalism	Average professionalism rating given by all the employers (ranging from 0 to 5)	4.84	0.33	1.00	5.00
Hire-again rating	Average hire-again rating given by all the employers (ranging from 0 to 5)	4.82	0.36	1.00	5.00

[3] The large variation in Bid_max is driven by outliers. In rare cases, workers asked for unreasonably high prices.

Overall	Average overall employer-entered ratings for the worker	4.83	0.33	1.00	5.00
Review_count	Total number of reviews which were written by previous employers	89.24	202.63	1.00	4128.00
Completion_rate	Percentage of awarded projects which were successfully completed as scheduled	0.78	0.19	0.02	1.00

Note: Summary statistics are calculated based on the matched sample wherein the bids are submitted by workers who bid for both fixed-price and time-based projects (named as "dual-typed workers") (Lin et al. 2016). We dropped outliers with the Stata command "bacon", using the top 10th quantile of the Mahalanobis distance as a cutoff.

Table 5. Definitions and Summary Statistics of Bid-level Variables

Variable	Variable definition	Mean	SD	Min	Max
Bid_price	Bid price submitted by the worker	114.94	263.04	2.00	22272.00
Hire_before	A dummy variable; =1 if the worker has been hired by the employer before	0.30	0.46	0.00	1.00
No_rating	A dummy variable; =1 if the worker has not received any ratings when he/she submitted the bid	0.04	0.20	0.00	1.00
Bidder_tenure_Month	The worker's tenure at *Freelancer* measured in months	35.49	25.85	1.00	178.00
Bid_rank	The bidder's ranking among all the candidates, *Freelancer* automatically sorts all the bidders according to its own ranking algorithm which is mainly based on bidders' employer-entered reviews	13.02	13.31	1.00	132.00
Bid_order_rank	The sequence order in which the bidders' bids were submitted	14.45	14.43	1.00	135.00
Preferred_freelancer	A dummy variable; =1 if the worker gets a special Preferred Freelancer badge because of their workmanship and customer service abilities	0.21	0.41	0.00	1.00
Local_freelancer	A dummy variable; =1 if the worker works for offline jobs nearby	0.02	0.14	0.00	1.00

Note: Summary statistics are calculated based on the matched sample. We dropped outliers with the Stata command "bacon", using the top 10th quantile of the Mahalanobis distance as a cutoff.

1.4. Research Methodology

1.4.1. Identification: A Quasi-Natural Experiment

While a field experiment with the random assignment of contract types is the most ideal design, the difficulty in persuading employers to make contract choices without altering the employers' and bidders' would-be choices makes it almost impossible to implement such a large-scale experiment in the field. As such, following the prior studies on policy change (e.g., Autor 2003; Chan and Ghose 2014; Chen et al. 2017), we combine a quasi-natural experiment design based on the observed panel data with DID estimation, matching methods, and a series of robustness checks, which is a reasonable design for causal inference (Atasoy et al. 2016; Hong 2013; Hirano et al. 2003; Bergemann et al. 2009).

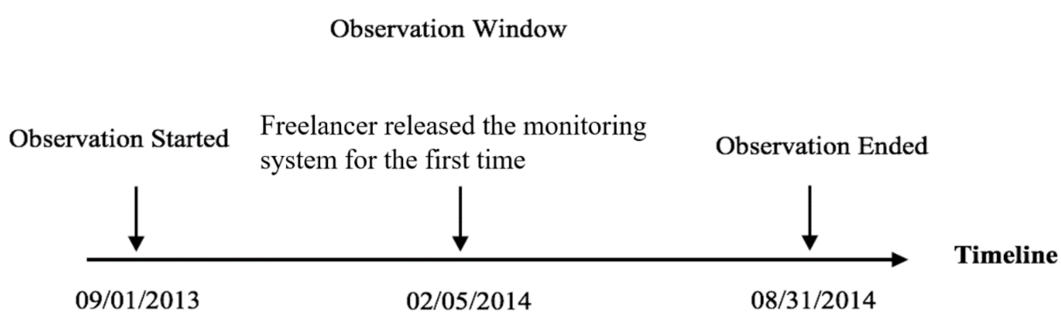


Figure 2. A Timeline of Our Observation Window

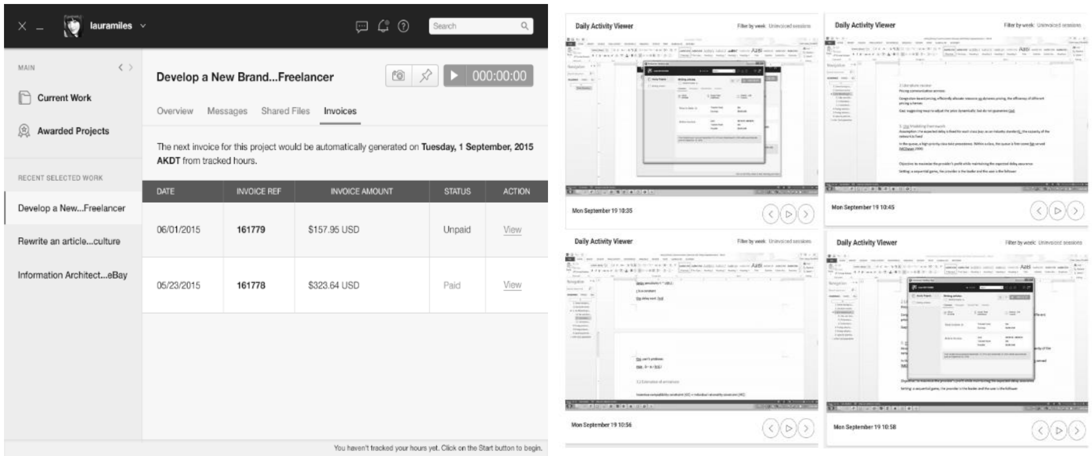


Figure 3. Screenshots of the Freelancer Monitoring System

Specifically, we leverage a quasi-natural experiment based on Freelancer's initial release of its monitoring system on February 5, 2014. Note that the monitoring system is only available for time-based contracts. The monitoring system is a software application that allows employers to effortlessly monitor freelancers. Freelancer encourages workers with time-based contracts to download and install the application on its Facebook, Twitter, and official blog. Once installed,[4] the monitoring system randomly takes several screenshots about every ten minutes, and continuously tracks the number of minutes the worker has spent on each time-based project.[5] Specifically, it automatically tracks when and for how long the worker has worked, the accumulated compensation the worker has earned, and the corresponding screenshots with precise timestamps. Therefore, it effectively keeps a detailed record of the workers' effort, providing the employer with

[4] If workers who work on time-based projects do not use this monitoring application to track their work hours, they are not guaranteed to get paid for their work. Moreover, it is worth noting that whether employers install this monitoring application or not, they can always check the monitoring records from the Freelancer website.

[5] The application does not track time spent on fixed-price projects, because workers can only find time-based contracts, rather than fixed-price contracts, through this application.

up-to-date information on the progress of the project. The employer can file a dispute to the platform regarding the worker's effort or claimed hours with the detailed monitoring records as evidence of the worker's shirking behavior. Figure 3 is a screenshot of the monitoring application provided by Freelancer.

Since the use of the monitoring system is advocated for all time-based projects and not used for fixed-price projects, this provides us with a unique research opportunity. In this study, we leverage a difference-in-differences (DID) design with fixed-price projects as the control group to examine the effect of the IT-enabled monitoring system on time-based projects, relative to fixed-price projects, comparing employers' hiring decisions and workers' entry behaviors across the two types of projects before and after the introduction of the monitoring system. The DID model is used extensively in IS research when exogenous changes are available (e.g., Chan and Ghose 2014, Huang et al. 2017; Wang et al. 2018; Zhang and Li 2017).

1.4.2. Econometric Analyses

1.4.2.1. Propensity Score Matching

In order to satisfy the common support requirement and reduce potential disparities across time-based projects and fixed-price projects, we use the PSM method to generate a comparable sample. The PSM approach for matching has been widely applied in the information systems literature (Hong et al. 2016; Lin et al. 2016; Xu et al. 2016; Xue et al. 2010). First, following related prior literature (Banerjee and Duflo 2000; Gopal and Sivaramakrishnan 2008; Lin et al. 2016; Roels et al. 2010), we identify project characteristics and employer characteristics that might correlate with employers' choices of contract type (Table 6). Moreover, to better match employers' needs and their task requirement, we generate the project-skill matrix with the top 20 skills ranked by frequency. In addition, to further match the two types of projects on the aspects of employers' uncertainty and their quality sensitivity (Dey et al. 2010; Bajari and Tadelis 2001), we ranked the importance score of each token in predicting the contract type by applying the gradient boosting algorithm to the subsample posted prior to the introduction of the monitoring system with the Text2vec and Xgboost packages in R.[6] Further, we reduce the dimension of the project-term matrix by limiting the list to the 20 tokens with highest information gains. Then we predict the propensity scores and match fixed-price projects with time-based projects. Furthermore, we compare the distribution of the propensity score and perform a balance check for all observed covariates (Xu et al. 2016). As Table A1 in the Empirical Appendix A shows, the matching process significantly reduces the difference between the control and treatment groups, and the means of all covariates are not statistically different across the two types of projects after the matching process. Based on the full sample with 12,467 projects posted on Freelancer, we generate our final matched sample, which includes 5,383 fixed-price projects and 3,099 time-based projects.

Table 6. Pre-treatment Covariates Used to Adjust for Potential Selection Bias

Dimension	Variable	Variable Description
Task complexity, risk of project (Gopal and Sivaramakrishnan 2008)	Project category dummies	Dummy variables for various project categories, including software, design, marketing, administrative, etc.

[6] We use the decision-tree boosting method provided by the Xgboost package in R to conduct the text-mining analysis. To reduce the possibility of the importance score being affected by the introduction of monitoring systems, we only include projects posted prior to the introduction date.

Project title length (Lin et al. 2016)	Project_title_length	Number of characters in the project title
Project description length (Lin et al. 2016)	Project_desc_length	Number of characters in the project description shown on the project page
Project description	Description token dummies	The description token dummies which are used to control the employer's uncertainty about his/her project need
Skill requirements	Skill requirement dummies	The skill requirement dummies describing project task requirement
Client level of knowledge (Lin et al. 2016)	Employer_tenure_month; Employer_overall_rating	Employer's tenure at Freelancer measured in months, which is also a proxy of employers' experience and relevant knowledge; employers' overall rating indicating employers' reputations

1.4.2.2. Principal Component Analysis for Dimension Reduction

Freelancer employs a multidimensional reputation system, which prominently displays multiple indicators when the cursor hovers over the bidder's username. We collapse the six dimensions of reputation information into a few informative scalars in order to capture the effect of reputation in reducing employers' uncertainty. As high correlations are observed among some rating dimensions, we employ Principal Component Analysis (PCA) for dimension reduction, which generates two principal components by using ~1 as the cutoff for eigenvalues and 80% as the threshold of the cumulative variance explained, as shown in Table 7. The first component (PC1) comprises dimensions of ratings entered by previous employers after the transactions, which largely helps to reduce future employers' uncertainty regarding workers' capability. The second component (PC2) has high loadings on the workers' project completion rate, which was computed by the system based on the percentage of projects completed out of all contracted projects. This component largely indicates a worker's effort at work because project incompletion is typically due to workers' lack of effort in completing the milestones set by the employer on time.[7] Therefore, the second component helps to alleviate employers' uncertainty regarding workers' effort.

In summary, the five items with significant loadings on the first component help to mitigate ex ante information asymmetry (hidden information), whereas the one item with significant loading on the second component helps to alleviate ex post information asymmetry (hidden action). In addition, to better understand how employers perceive workers' reputation signals, we analyze data from a survey of employers with the aid of Freelancer's management team, which confirms our PCA results indicating that employers are generally concerned about freelancers' capabilities and service effort. Therefore, guided by the two dimensions of employers' uncertainty about workers and the item loadings of raw reputation information, we label PC1 as "Capability" and PC2 as "Effort." This label assignment is further confirmed by interviewing a number of Freelancer employers on how they perceive the reputation signals of workers. We report the item loadings, and eigenvalues/cumulative variance explained in Table 7 and Table 8, respectively. In addition, the results are highly consistent when we use two raw measures of reputation information—namely, the overall rating and completion rate.

---

[7] According to Freelancer.com, "Projects are marked as completed once the freelancer is paid in full with Milestone Payments. If a freelancer has a high Completion Rate, employers will have the security of knowing that their projects will be completed and will not be abandoned by an unreliable freelancer." This suggests that the worker needs to spend sufficient effort and finish milestones following the schedule set by the employer to get the project marked as "completed" (source: https://www.freelancer.pl/faq/topic.php?id=2).

Table 7. Item Loadings of Two Principal Components with Varimax Rotation

Variable	Eigenvectors	
	1	2
Quality	0.449	-0.003
Communication	0.432	0.008
Expertise	0.451	-0.004
Professionalism	0.452	0.001
Hire-again rating	0.451	-0.002
Completion Rate	0.000	1.000

Table 8. Eigenvalues and Variance Explained by Two Principal Components

Label	Component	Eigenvalue	Diff	Proportion of variance explained	Cumulative variance explained
Capability	1	4.374	3.381	0.729	0.729
Effort	2	0.994	0.741	0.166	0.895

1.4.2.3. Estimating Employer Preference

To estimate employer preference for workers (H1) in terms of their observable characteristics, we formulate a model for each worker's hiring outcome within each project. Specifically, we estimate the employer's hiring decision regarding whether bidder k is awarded in project j as Pr_{jk}.

$$Pr_{jk} = \beta_1 t_j C_{jk} + \beta_2 T_j C_{jk} + \beta_3 t_j T_j C_{jk} + \beta_4 t_j E_{jk} + \beta_5 T_j E_{jk} + \beta_6 t_j T_j E_{jk} + \beta_7 t_j P_{jk} + \beta_8 T_j P_{jk} + \beta_9 t_j T_j P_{jk} + \gamma B_k + \delta Z_{jk} + \varepsilon_{jk} \quad (1)$$

In Equation (1), t_j is the period dummy variable, which is set to 1 if project j is posted after the introduction of the monitoring system. T_j is the contract type dummy variable, which is set to 1 if project j is a time-based project. C_{jk} denotes bidder k's reputation related to his or her capabilities based on the principal component analysis. E_{jk} denotes bidder k's reputation related to his or her effort based on principal component analysis. P_{jk} denotes the bid price submitted by bidder k. Z_{jk} represents a set of other project-bidder paired characteristics, including the bidder k's ranking based on his/her reputation and experience among all the competitors, the order of bidder k's bid based on the sequence of all the bids were submitted, whether the bidder *k* has worked for this employer before. B_k captures bidder k's individual characteristics, including whether bidder *k* has received any ratings or not (or the number of ratings entered by bidder k's previous employers), whether bidder k gets a special "Preferred Freelancer" badge, and whether bidder *k* also works for local projects.[8] The employer's hiring decision could be estimated with a linear probability model (Heckman and Snyder 1997; Greenwood and Agarwal 2015) or a logit model (Lin et al. 2016; Liu et al. 2015). Given our focus on the existence of the treatment effect, in our main analyses, we use a linear probability model by clustering ε_{jk} at the project level. We also estimate a conditional logit model and observe highly consistent results.

1.4.2.4. Difference-in-Differences Models

To assess workers' entry decisions (for H2), we estimate standard difference-in-differences models (Bertrand et al. 2004; Angrist and Pischke 2008):

$$Log\_bid\_count_{ij} = \alpha + \beta_1 Time\_based_j + \beta_2\, Time\_based_j \times After_j + \gamma_i + \delta_j + \tau_t + \varepsilon_{ij} \quad (2)$$

[8] Based on our review data, workers' average ratings were basically constant during our observation period. In particular, the median changes in workers' cumulative average rating within a quarter was only 0.009. This low variation suggests that for workers with high reputations, the negative impact of their potential shirking behaviors might be small given the large number of total projects they have completed. This is also in line with one of the motivations of our study (i.e., reputation systems are not the perfect tool for deterring moral hazard).

$$Pct\_no\_rating_{ij} = \alpha + \beta_1 Time\_based_j + \beta_2\, Time\_based_j \times After_j + \gamma_i + \delta_j + \tau_t + \varepsilon_{ij} \quad (3)$$

In equation (2), the dependent variable is the log transformation of the total number of bids for each project j posted by employer i, $Log\_bid\_count_{ij}$. In equation (3), the dependent variable $Pct\_no\_rating_{ij}$ denotes the percentage of inexperienced bidders (i.e., bidders without ratings) in project j posted by employer i. The contract type is indicated by $Time\_based_j$, which equals 1 if project j is a time-based project, and 0 if it is a fixed-price project. $After_j$ is the dummy variable indicating whether project j is awarded after the introduction of the monitoring system. The coefficient of the interaction term $Time\_based_j \times After_j$ (β_2) thus identifies the effect of the introduction of the IT-enabled monitoring system on time-based projects relative to fixed-price projects. To control for project heterogeneity, we also add other project characteristic controls (δ_j), a vector of employer fixed-effects (γ_i), and a vector of time fixed-effects (τ_t) into the DID model and ε_{ij} denotes the robust standard errors clustered on employers.

1.4.3. Empirical Results

1.4.3.1. Employer Preference Estimation

The results of the linear probability model are reported in Table 9.[9] We observe that, before and after the IT-enabled monitoring system was implemented, the coefficients for

[9] Given that the linear model helps ensure consistency of the estimation results and provides a meaningful interpretation of coefficients for the interaction terms (Greenwood and Agarwal 2015), we estimate the change in employer preference with the linear probability model. The results are highly consistent with those of the conditional logit model.

the reputation of workers' capabilities, $Capability\_of\_worker$, remain unchanged at 0.006. This finding indicates that employers' preference for workers with high capability-related reputations does not change due to the presence of the monitoring system. Notably, we observe a different pattern regarding the coefficients for the worker's effort-related reputation, $Effort\_at\_work$. As Table 9 attests, the coefficient of the effort-related reputation is 0.026 and 0.051 for fixed-price and time-based projects, respectively, before the introduction of the monitoring system, suggesting that employers prefer to hire workers with high effort-related reputations. After the introduction of the system, for fixed-price projects, the employer preference remains at a similar level, since the coefficient of $Effort\_at\_work * After$ is insignificant. In contrast, for time-based projects, there is a relatively large decrease in employer preference (i.e., the coefficient of $Effort\_at\_work * Time\_based * After$ is significantly negative), indicating that employers' preference for workers with high effort-based reputations decreases after the monitoring system was introduced. This significant decrease in employers' emphasis on workers' effort-related reputations and the insignificant change in the importance of workers' capability-related reputations for employers suggest that the introduction of the monitoring system helps mitigate moral hazard and lower the wage premium acquired by workers with high effort-related reputations, but has a limited effect on alleviating adverse selection, as predicted in *H1*. We also find that the magnitude of employers' price-sensitivity for time-based projects increases (from $|-0.053|$ to $|-0.109|$) after the introduction of the monitoring system, while that for fixed-price projects remains unchanged, further confirming that the alleviation of moral hazard problems makes employers more sensitive to bid prices. Overall, the findings suggest that there exists a partial substitution relationship between monitoring and reputation such that monitoring substitutes for some of the effort-related reputation but not for the capability-related reputation.

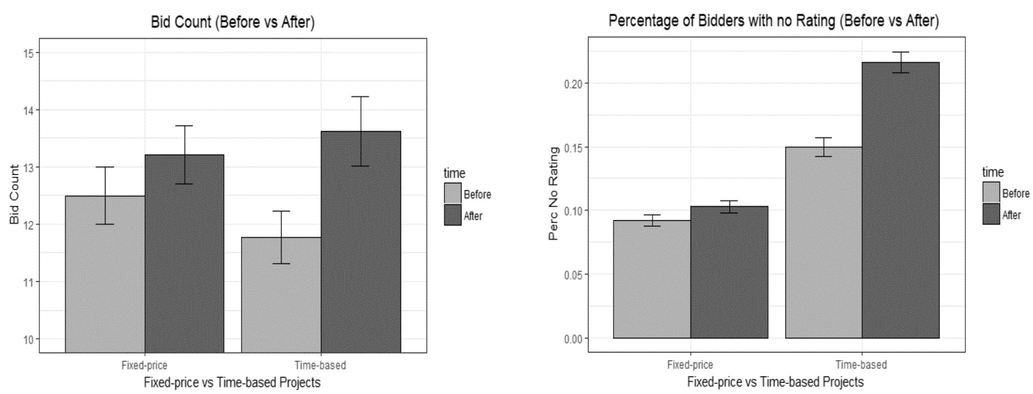
Table 9. Estimation Results of the Linear Probability Model

	Dependent variable: Bid_selected
Capability_of_worker	0.006***(0.001)
Capability_of_worker* Time_based	0.001 (0.001)
Capability_of_worker*After	0.002 (0.001)
Capability_of_worker* Time_based *After	-0.002 (0.002)
Effort_at_work	0.026***(0.002)
Effort_at_work* Time_based	0.025***(0.003)
Effort_at_work* After	0.000 (0.003)
Effort_at_work* Time_based *After	-0.015***(0.005)
Log_bid_price	-0.087***(0.004)
Log_bid_price* Time_based	0.034***(0.006)
Log_bid_price*After	0.006 (0.006)
Log_bid_price*Time_based *After	-0.056***(0.010)
Hire_before	0.538***(0.015)
No_rating	-0.044***(0.004)
Log_bidder_rank	-0.020***(0.001)
Log_bid_order_rank	0.014***(0.001)
Preferred_freelancer	0.007** (0.003)
Local_freelancer	-0.027***(0.007)
Observations	69,975
Clusters (projects)	5,694
R-squared	0.115
Log likelihood	-1278

Notes: (a) We limit our sample to those projects with more than one bid and awarded to only one worker. Our results are based on all the workers who bid for both fixed-price and time-based projects (named as “dual-typed workers”) (Lin et al. 2016). (b) Since we do not have any capability-related or effort-related reputation information for workers who have not received any ratings from employers, we add the No_rating dummy and set their capability-related and effort-related reputation component scores as zeros. We also estimate the model with only those workers with reputations and add the Review_count variable instead of the No_rating dummy. (c) Results are highly consistent when we estimate the treatment effect with conditional logit models. (d) Robust standard errors clustered on projects are reported in parentheses. (e) * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

1.4.3.2. Bidding Behavior and Entry Barrier

As employers are less willing to pay high wage premiums to workers with high effort-related reputations when the monitoring system is in place, we expect that the entry barrier for inexperienced (or new) workers into the time-based project will become lower, leading to more bids for a given time-based project. Specifically, our analysis regarding workers' entry decisions proceeds as follows. Before we report our DID estimates, based on the matched sample, we first present some model-free evidence of the change in both dependent variables (i.e., the number of bids and percentage of bidders with no rating). As Figure 4 shows, both dependent variables significantly increase for time-based projects but not for fixed-price projects without controlling for the effect of project characteristics and employer characteristics.



Note: The matched sample is used. The bars represent the average number of bids (Bid_count) and the average percentage of bidders with no rating (Pct_no_rating) in fixed-price projects and time-based projects before and after the introduction of the monitoring system. Error bars represent the 95% confidence intervals of standard errors.

Figure 4. Model-free Evidence of the Change in Dependent Variables Among the Matched Sample

Furthermore, the DID regression results reported in Columns (1) and (2) of Table 10 still show a consistent result. Column (1) of Table 10 reports the results based on the DID analysis of the effect of the monitoring system on the number of bids. We find the coefficient (β_3) of the interaction term $After_j \times Time\_based_j$ to be significantly positive, which suggests that the introduction of the monitoring system significantly increases the number of bids (Bid_count_{ij}) for time-based projects. Further, the coefficient of the interaction term of 0.221 translates to a 24.7% increase in number of bids,[10] supporting *H2a*.

We further assess the conjecture that the monitoring system reduces the entry barrier for new bidders into time-based projects. We compare the percentage of inexperienced workers (workers with no reputation score) among all the bidders in time-based contracts before and after the introduction of the IT-enabled monitoring system. We create a binary variable, *No_rating*, denoting whether the worker has received any ratings (Lin et al. 2016). Then we use the percentage of workers ($Pct\_no\_rating_{ij}$) who haven't accumulated any reputation records from employers, as a proxy for the entry barrier for inexperienced workers. We include employer-level fixed effects and project characteristics to control for unobserved heterogeneity across employers and the heterogeneity across projects. The estimation results are reported in Table 10. The marginal effect of the $Time\_based_j$ dummy is insignificant, indicating that the

[10] Based on the estimation results in Column (1) of Table 10, before the introduction of monitoring systems, the partial correlation $Time\_based_j$ dummy and Log_bid_count is 0.349. This partial coefficient becomes 0.570 after the introduction. Since 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. Exp(0.221) - 1=24.7%.

percentage of inexperienced bidders for time-based projects is roughly the same as the percentage of inexperienced bidders for fixed-price projects before the introduction of the monitoring system. However, after the introduction of the monitoring system, the coefficient of $Time\_based_j$ increases significantly (by 0.085). This increase suggests that, all other things being equal, the percentage of workers with no ratings increases more for time-based projects than for fixed-price projects. Specifically, the marginal effect estimate based on the delta method indicates that the percentage increases by 8.50%. The fact that disproportionately more inexperienced workers participate in time-based projects after the introduction of the monitoring system validates *H2b* that the monitoring system lowers the entry barrier for inexperienced workers.

Overall, the results regarding the number of bids and percentage of inexperienced bidders provide support for our hypothesis that the monitoring system attracts more bids by lowering the entry barrier for inexperienced workers and alleviates the cold-start problem. Both *Hypothesis 2a* and *Hypothesis 2b* are thus supported.

Table 10. Estimation Results of the DID Models

Model	(1)	(2)
Dependent Variable	Log_bid_count	Pct_no_rating
Time_based	0.349***(0.105)	0.013 (0.015)
Time_based*After	0.221** (0.091)	0.085***(0.015)
Log_budget_max	0.147***(0.027)	-0.003 (0.004)
Log_title_length	-0.075 (0.056)	0.003 (0.009)
Log_desc_length	0.132***(0.030)	0.001 (0.005)
Category dummies	Yes	Yes
Month dummies	Yes	Yes
Employer dummies	Yes	Yes
Clusters (employers)	1,261	1,261
Observations	2,976	2,976
R-squared	0.314	0.106

Notes: (a) The results are highly consistent when we control for the week dummies instead of month dummies. (b) Robust standard errors clustered on employers are reported in parentheses. (c) The results are consistent when we use the top 1, 5, or 10 quantile of the Mahalanobis distance as a threshold to separate outliers from nonoutliers. (d) * p<0.1, ** p<0.05, *** p<0.01.

1.5. Robustness Checks

Given that the contract types of projects are not randomly assigned in the field, the control group and the treatment group may differ in terms of both observables and unobservables. To further evaluate the credibility of our result, we conduct a series of robustness checks to evaluate and address these potential identification concerns (Table 11). Since our fixed-effects specification is immune to time-invariant selection issues, we focus on addressing potential problems regarding time-varying selection on observables and unobservables.

First, to further alleviate the issue of selection on observables, we employ alternative matching algorithms (e.g., coarsened exact matching and matching in causal inference) and obtain highly consistent results.

Second, we use two strategies to address the issue of selection on unobservables. The first strategy uses an alternative quasi-experiment design by identifying two different subgroups from the time-based projects that could be used as the treatment group and control group respectively. Given that both the new control group and the treatment group are derived from the time-based projects, the design of these quasi-natural experiment settings tends to alleviate the concern of selection on unobservables (Manchanda et al. 2015). The results from these alternative quasi-natural experiment designs are highly consistent. The second strategy employs the instrumental variable approach to estimate the local average treatment effect, which again lends support to our main finding.

Third, we conduct several additional robustness checks to further rule out the possibility of a spurious relationship. For instance, we show that our data satisfies the parallel trend assumption. Additionally, we conduct two placebo tests, one regarding the placebo treatment time and the other regarding the placebo treatment assignment. Furthermore, we find that our results are highly consistent when we rule out outliers or use alternative measures of reputation.

Table 11. Overview of the Analyses

Section	Analysis	Objective
Section 1.4.3.	Selection on observables: Baseline DID estimation in the matched sample	Quantifying the treatment effect; controlling for selection on observables;
Section 1.5.1.	Selection on observables: DID estimation with coarsened exact matching	Controlling for selection on observables;
Section 1.5.2.	Selection on unobservables: DID estimation with a subgroup of time-based projects as the control group	Alleviating potential issues regarding selection on unobservables and double counting the treatment effect
Section 1.5.3.	Selection on unobservables: IV Estimation	Control for potential biases due to employers' self-selection[11]
Section 1.5.4.	Placebo test/shuffling (placebo treatment time and placebo treatment assignment)	Checking the assumption of DID models; avoiding spurious causality with alternative variance-covariance specifications
Appendix B	Parallel trend assumption	Checking the assumption of DID models
Appendix C	Additional analysis on project outcomes	Exploring other impacts on the platform

Note: Except for those analyses reported in Sections 1.4 and 1.5, all the other robustness checks listed in the above table are included in our Online Empirical Supplementary Appendices.

[11] It's worth noting that employers' self-selection may not be a concern if those unobservables affecting selection are not revealed to bidders, which is likely the case in our context given the anonymity of employers.

1.5.1. Selection on Observables: Alternative Matching Method

For our main analysis, we perform Propensity Score Matching to generate the matched sample that is balanced on the distributions of observed characteristics between the treatment and control groups. To further alleviate the concern of selection on observables, we employ another matching algorithm— Coarsened Exact Matching (CEM)—to regenerate a comparable sample (Iacus et al. 2012; Blackwell et al. 2009; Subramanian and Overby 2016). CEM enables us to explicitly match fixed-price projects with time-based projects within the same category and with similar skill requirements and descriptions. As such, CEM increases the homogeneity between the two types of projects from a multivariate perspective and lends support to the causality of our findings. We rerun the DID models on the CEM-matched samples and report the results in Table 12 and Table 13. Overall, the results based on the CEM-matched sample are consistent with our main results. Again, we find that after the introduction of the monitoring system, employers placed less emphasis on workers' effort-related reputations, and that the number of bids and the percentage of inexperienced bidders significantly increase. Additionally, we retest our results using two alternative matching strategies: Inverse Probability of Treatment Weighting method (Blackwell 2013) and pruning posttreatment pairs to alleviate the potential concern of composition change in time-based projects (Keele et al. 2016). Both produce consistent findings.

Table 12. Linear Estimation of Employers' Preference with Time-based Projects based on the CEM-Matched Sample

Variable	Bid_Selected	
Capability_of_worker	0.004***	(0.001)
Capability_of_worker* Time_based	0.002	(0.001)

Capability_of_worker*After	0.004**	(0.002)
Capability_of_worker* Time_based *After	-0.002	(0.003)
Effort_at_work	0.029***	(0.003)
Effort_at_work* Time_based	0.023***	(0.004)
Effort_at_work* After	-0.001	(0.005)
Effort_at_work* Time_based *After	-0.013**	(0.007)
Log_bid_price	-0.098***	(0.006)
Log_bid_price * Time_based	0.050***	(0.009)
Log_bid_price *After	-0.002	(0.008)
Log_bid_price *Time_based *After	-0.046***	(0.014)
Hire_before	0.583***	(0.018)
No_rating	-0.045***	(0.005)
Log_bidder_rank	-0.021***	(0.002)
Log_bid_order_rank	0.013***	(0.002)
Preferred_freelancer	0.006	(0.004)
Local_freelancer	-0.023**	(0.009)
Observations	40,742	
Clusters(projects)	3,479	
R-squared	0.131	
Log likelihood	-1252	

Notes: (a) We limit our sample to those projects with more than one bid and awarded to only one worker. Our results are based on all the workers who bid for both fixed-price and time-based projects (named as "dual-typed workers") (Lin et al. 2016). (b) Since we do not have any capability-related or effort-related reputation information for those workers who have not received any ratings from employers, we add the No_rating dummy and set their capability-related and effort-related reputation component scores as zeros. (c) Robust standard errors clustered on projects are reported in parentheses. d) * p<0.1, ** p<0.05, *** p<0.01.

Table 13. Estimation Results of the DID Models based on the CEM-Matched Sample

Model	(1)		(2)	
Dependent Variable	Log_bid_count		Pct_no_rating	
Time_based	0.147	(0.141)	-0.010	(0.020)
Time_based*After	0.252**	(0.121)	0.112***	(0.021)
Log_budget_max	0.072	(0.045)	-0.008	(0.006)
Log_title_length	0.108	(0.087)	0.006	(0.013)
Log_desc_length	0.165***	(0.042)	-0.000	(0.006)
Category dummies	Yes		Yes	
Month dummies	Yes		Yes	
Employer dummies	Yes		Yes	
Clusters (employers)	719		719	
Observations	1,601		1,601	
R-squared	0.291		0.115	

Notes: (a) Robust standard errors clustered on employers are reported in parentheses. (b) * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

1.5.2. Selection on Unobservables: DID Estimation with a Subgroup of Time-based Projects as a Control Group

Instead of using matched fixed-price projects as counterfactual, we consider an alternative approach, and identify a subgroup of the time-based projects as the control group. This approach leverages project category heterogeneity, and is less likely to be susceptible to the selection on unobservables issue (Manchanda et al. 2015). Specifically, we investigate the treatment effect of the introduction of the monitoring system by leveraging the variation of the efficacy of behavior-based controls (e.g., monitoring systems) across project categories. According to the literature on organizational theory, the efficacy of behavior-based controls, such as monitoring, is dependent on outcome measurability (Ouchi 1979; Eisenhardt 1985, 1989). Outcome measurability refers to the extent to which the project performance can be reliably, validly and easily measured (Ouchi 1979; Eisenhardt 1985, 1989). When the value of project outcome mainly depends on the quantity (i.e., count of units the worker finished), its outcome measurability is high (Wüllenweber et al. 2009). In contrast, projects focusing on product quality instead of unit count tend to have low outcome measurability (Wüllenweber et al. 2009). Moreover, for projects with high outcome measurability, such as administrative projects (e.g., customer service, HR service, accounting service) and marketing projects (e.g., adding Facebook fans, voting), employers can deter workers' moral hazard by checking the count of small tasks finished by workers (e.g., number of calls, number of replies, number of votes). Thus, the monitoring system, a prime example of a behavior-based control tool, does not add much value for projects with outcomes that are relatively easy to measure. Conversely, behavior-based control tools are more effective for projects with outcomes that cannot be easily measured by counting, such as design or html jobs. Therefore, we expect that for projects with highly measurable outcomes, the impact of implementing the monitoring system will be relatively lower than for projects with low outcome measurability. As such, among all the time-based projects, we consider projects with high outcome measurability (i.e., administrative projects and marketing projects) as the control group and the other projects (i.e., software, writing, translation, design, and others) as the treatment group.

Estimating the DID models identical to our main analysis, we find that the introduction of the monitoring system significantly lowers employers' preference for workers with high effort-related reputations (Table 14), and that it lowers the entry barrier and increases the number of bids for time-based projects with low outcome measurability (Table 15). These results lend support to our main analysis.

Moreover, we perform another robustness check with a "Regression-Discontinuity" style control group in which date is the running variable. In particular, we use time-based projects posted before the introduction date of the monitoring system as the control group and time-based projects posted after the introduction date as the treatment group. Our results are highly consistent.

Table 14. Linear Estimation of Employers' Preference with Time-based Projects with High Outcome Measurability as the Control Group

Variable	Bid_Selected	
Capability_of_worker	0.007**	(0.003)
Capability_of_worker* Low_outcome_measurability	-0.000	(0.003)

Capability_of_worker*After	-0.003	(0.004)
Capability_of_worker* Low_outcome_measurability *After	0.004	(0.005)
Effort_at_work	0.042***	(0.007)
Effort_at_work* Low_outcome_measurability	0.009	(0.008)
Effort_at_work* After	0.001	(0.012)
Effort_at_work* Low_outcome_measurability *After	-0.022*	(0.013)
Log_bid_price	-0.041***	(0.016)
Log_bid_price * Low_outcome_measurability	-0.027	(0.017)
Log_bid_price *After	-0.067**	(0.026)
Log_bid_price * Low_outcome_measurability *After	0.019	(0.029)
Hire_before	0.564***	(0.022)
No_rating	-0.065***	(0.005)
Log_bidder_rank	-0.003	(0.003)
Log_bid_order_rank	0.017***	(0.002)
Preferred_freelancer	0.021***	(0.006)
Local_freelancer	-0.047***	(0.012)
Observations	23,639	
Clusters (projects)	2,159	
R-squared	0.141	
Log likelihood	-1301	

Notes: (a) We limit our sample to those projects with more than one bid and awarded to only one worker. Our results are based on all the workers who bid for both fixed-price and time-based projects (named as "dual-typed workers") (Lin et al. 2016). (b) Since we do not have any capability-related or effort-related reputation information for those workers who have not received any ratings from employers, we add the No_rating dummy and set their capability-related and effort-related reputation component scores as zeros. (c) Robust standard errors clustered on projects are reported in parentheses. (d) * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 15. Estimation Results of the DID Models with Time-based Projects with High Outcome Measurability as the Control Group

Model	(1)		(2)	
Sample	Time-based projects		Time-based projects	
Dependent Variable	Log_bid_count		Pct_no_rating	
Low_outcome_measurability	-0.342	(0.301)	-0.032	(0.071)
Low_outcome_measurability *After	0.320*	(0.173)	0.145***	(0.050)
Log_budget_max	0.162***	(0.062)	-0.003	(0.015)
Log_title_length	0.028	(0.080)	0.008	(0.017)
Log_desc_length	0.080	(0.050)	-0.023***	(0.008)
Category dummies	Yes		Yes	

Month dummies	Yes	Yes
Employer dummies	Yes	Yes
Clusters (employers)	435	435
Observations	1,126	1,126
R-squared	0.118	0.097

Notes: (a) Robust standard errors clustered on employers are reported in parentheses. (b) * p<0.1, ** p<0.05, *** p<0.01.

1.5.3. Selection on Unobservables: IV Estimation

Within the quasi-natural experiment settings, our DID estimation may be biased in the context of time-varying unobservables that simultaneously affect employers' contract choices and dependent variables regarding entry behavior. It's worth noting that selection on time-varying unobservables may not be a concern if the unobservables affecting selection are not revealed to bidders, which appears to be a reasonable condition given the anonymity of employers. To assess and alleviate concerns regarding the selection on time-varying unobservables that are also revealed to bidders, we employ the instrumental variable (IV) approach to estimate the local average treatment effect (Angrist and Imbens 1994; Angrist and Krueger 2001)—i.e., the causal effect of the monitoring system on bidding behavior and entry barrier. Specifically, we need instruments that are associated with the employer's contract choice for project j ($Time\_based_j$) but not with the error term (ε_{ij}) in bidders' decisions to bid for project j. In particular, we employ the "residual" type IV (Dobbie et al. 2017; Arnold et al. 2018) and the Hausman type IV (Hausman et al. 1994; Hausman 1996; Schneider 2010; Ghose et al. 2012).

The first instrument, i.e., the "residual" type IV, is the mean of the residuals from the prediction of employers' contract choices in their previous projects. We predict employers' contract choices with various project characteristics and time dummies, and then estimate the residual for each project. The "residual" type IV is a combination of the employer's time-invariant tendency to use time-based contracts and the effect of some idiosyncratic unobserved features of previous projects on previous contract choices, which correlate with this employer's contract choice for the current project (relevance). But this is unlikely to correlate with the effect of time-varying unobservables on bidders' entry decisions for project *j* after controlling employers' fixed effects (exogeneity). Additionally, when the time-invariant tendency to use time-based contracts are stronger or the idiosyncratic unobserved features of previous projects that are more likely to nudge employers to use time-based projects, we expect them to monotonically increase the probability of employers using time-based contracts for project *j* (monotonicity).

The second instrument is the Hausman type IV, which is the percentage of time-based projects posted by other employers from the same country within the same week of project *j* (Schneider 2010). The Hausman type IV has been well applied in the previous literature (Hausman et al. 1994; Hausman 1996; Schneider 2010; Ghose et al. 2012). In our data, a high variation exists in the percentage of time-based projects across employer countries. We suspect that the leave-out average percentage of time-based projects probably correlates with the contract choice of the employer of project *j* due to the common economic environment in their countries (e.g., the short-term interest rate) or common cultural background (relevance). Moreover, as suggested by a prior study (Schneider 2010), this leave-out average percentage of time-based projects does not correlate with bidders' entry decisions, after controlling for fixed effects on time and employers. This instrument seems to be exogenous to any platform-wide variation in unobserved bidding preference that correlates with two dependent variables (i.e., number of bids and percentage of bidders with no rating), such as workers' idiosyncratic distaste for a specific project requirement (exogeneity). Additionally, these common environmental factors are likely to affect the contract decisions of the employers from the same country in a similar way (monotonicity).

With the linear model framework, we employ the 2SLS method into the DID estimation. In our model for bidders' entry decisions, there are two endogenous variables, i.e., $Time\_based_j$, $After_j \times Time\_based_j$. Since the timing of when the platform decided to implement monitoring systems ($After_j$) is an exogenous factor, we can have four instrumental variables: $Residual\_leaveout\_timebased_i$, $After_j \times Residual\_leaveout\_timebased_j$, $Pct\_timebased\_week_j$, and $After_j \times Pct\_timebased\_week_j$ (Wooldridge 2010). Similarly, we also construct instruments for six potential endogenous variables in the linear model for employers' preference. For instance, regarding the endogenous variable related to workers' capability-related reputations, $C_{jk} \times Time\_based_j$, we use $C_{jk} \times Residual\_leaveout\_timebased_i$ and $C_{jk} \times Pct\_timebased\_week_j$ as its instruments.

As Table 16 shows, the Kleibergen-Paap rk Wald F statistics of all the models are higher than Stock-Yogo weak IV test critical values, which suggests that we can firmly reject the null hypothesis of weak instruments. Moreover, the Hansen *J* statistic indicates that we cannot reject the null hypothesis that the overidentifying restriction is valid. As Table 16 and Table 17 show, the results regarding bidders' entry decisions and employers' preference are highly consistent with our main results.

Table 16. IV Estimation of the DID Models

Model	(1)		(2)		(3)		(4)	
Instrument(s)	"Residual" IV		"Residual" IV & "BLP" IV		"Residual" IV		"Residual" IV & "BLP" IV	
Dependent Variable	Log_bid_count		Log_bid_count		Pct_no_rating		Pct_no_rating	
Time_based	0.495*	(0.271)	0.495*	(0.271)	0.030	(0.032)	0.032	(0.031)
Time_based*After	0.328**	(0.144)	0.328**	(0.144)	0.074***	(0.026)	0.076***	(0.026)
Log_budget_max	0.194***	(0.057)	0.194***	(0.057)	0.001	(0.007)	0.001	(0.007)
Log_title_length	-0.048	(0.065)	-0.048	(0.065)	0.000	(0.011)	-0.002	(0.011)
Log_desc_length	0.058*	(0.035)	0.058*	(0.035)	-0.005	(0.005)	-0.006	(0.006)
Hansen J statistic			0.374	(0.829)			0.374	(0.829)
Kleibergen-Paap rk Wald F statistic	25.001		12.756		25.001		12.756	
Category dummies	Yes		Yes		Yes		Yes	
Month dummies	Yes		Yes		Yes		Yes	
Employer dummies	Yes		Yes		Yes		Yes	
Clusters(employers)	820		806		820		806	
Observations	2,234		2,185		2,234		2,185	
R-squared	0.052		0.055		0.079		0.082	

Note: (a) Robust standard errors clustered on employers are reported in parentheses. (b) The significance levels and standard errors of all the coefficients are also very consistent if we calculate the standard errors with 100 bootstrap cycles. (c)* p<0.1, ** p<0.05, *** p<0.01.

Table 17. IV Estimation of Employers' Preference with the Linear Probability Model

Variable	Bid_Selected		Bid_Selected	
Instrument(s)	"Residual" IV		"Residual" IV & "BLP" IV	
Capability_of_worker	0.002	(0.001)	0.002	(0.002)
Capability_of_worker* Time_based	0.005**	(0.002)	0.005**	(0.003)
Capability_of_worker*After	0.004	(0.003)	0.004	(0.003)
Capability_of_worker* Time_based *After	-0.003	(0.005)	-0.004	(0.005)
Effort_at_work	0.021***	(0.004)	0.020***	(0.004)
Effort_at_work* Time_based	0.027***	(0.007)	0.027***	(0.007)
Effort_at_work* After	0.006	(0.006)	0.007	(0.006)
Effort_at_work* Time_based *After	-0.039***	(0.010)	-0.039***	(0.010)
Log_bid_price	-0.072***	(0.006)	-0.072***	(0.006)
Log_bid_price * Time_based	0.032***	(0.012)	0.033***	(0.012)
Log_bid_price *After	-0.007	(0.009)	-0.009	(0.009)
Log_bid_price *Time_based *After	-0.055***	(0.021)	-0.052**	(0.021)
Hire_before	0.519***	(0.016)	0.517***	(0.016)
No_rating	-0.049***	(0.005)	-0.048***	(0.005)

Log_bidder_rank	-0.007*** (0.002)	-0.007*** (0.002)
Log_bid_order_rank	0.015*** (0.002)	0.015*** (0.002)
Preferred_freelancer	0.019*** (0.004)	0.020*** (0.004)
Local_freelancer	-0.042*** (0.009)	-0.042*** (0.009)
Hansen J statistic		8.209 (0.223)
Kleibergen-Paap rk Wald F statistic	28.122	16.352
Observations	30,824	30,239
Clusters (projects)	2,695	2,647
R-squared	0.178	0.177
Log likelihood	1080	1017

Notes: (a) We limit our sample to those projects with more than one bid and awarded to only one worker. Our result is based on all the workers who bid for both fixed-price and time-based projects (named as "dual-typed workers") (Lin et al. 2016). (b) Since we do not have any capability-related or effort-related reputation information for those workers who have not received any ratings from employers, we add the No_rating dummy and set their capability-related and effort-related reputation component scores as zeros. (c) Robust standard errors clustered on projects are reported in parentheses. (d) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

1.5.4. Placebo Tests

To assess the parallel trends assumption of the DID models and rule out the possibility of a spurious causal relationship, we conduct a series of placebo tests. First, we reassign the intervention to the middle of our pre-treatment period (November 1st, 2013) and check for the existence of a pretreatment tendency in the observation window before the actual introduction of the monitoring system. As the placebo treatment does not exist, we do not expect to observe a significant effect from that placebo treatment. As Table 18 shows, the interaction between the placebo treatment time ($After\_placebo$) and the contract type ($Time\_based$) is insignificant.

Second, following Abadie et al. (2015), we conduct another placebo test by randomly reassigning the treatment to projects within our sample. Again, since only projects that are actually treated (time-based projects) should be affected by the introduction of the monitoring system, if we randomly assign treatment to projects, we should not see a treatment effect. We simulate this permutation procedure 1000 times and capture the distribution of the placebo effects based on the randomly assigned placebo treatments. Table 19 summarizes the permutation result. After comparing the estimated coefficient of the actual treatment to the distribution of placebo effects, the probability of a similar size treatment effect happening by chance is close to zero (outside the 99% confidence interval), indicating that the significant finding is robust to alternative variance-covariance specifications.

Lastly, in another related robustness check, we conduct a dynamic DID analysis, reported in Empirical Appendix B. We observe that all the relative time parameters are insignificant prior to the introduction while some of the relative time parameters in two models are significant after February 2014 when Freelancer introduced the IT-enabled monitoring system. As such, the results of the relative-time model lend further support to the validity of the parallel trend assumption and also to our main findings.

Table 18. Estimation Results of the DID Models based on Placebo Treatment Time

Model	(1)		(2)	
Dependent Variable	Log_bid_count		Pct_no_rating	
Time_based	-0.141	(0.196)	-0.003	(0.031)
Time_based* After_placebo	0.173	(0.155)	-0.009	(0.024)
Log_budget_max	0.045	(0.054)	-0.001	(0.007)
Log_title_length	0.117	(0.095)	0.016	(0.014)
Log_ desc_length	0.180***	(0.055)	-0.003	(0.008)
Category dummies	Yes		Yes	
Month dummies	Yes		Yes	
Employer dummies	Yes		Yes	
Clusters (employers)	510		510	
Observations	1,159		1,159	
R-squared	0.220		0.081	

Notes: (a) Robust standard errors clustered on employers are reported in parentheses. (b) * p<0.1, ** p<0.05, *** p<0.01.

Table 19. Placebo Effects of Random Assignment Model

Dependent Variable	(1) Log_bid_count	(2) Pct_no_rating
Placebo effects (mean)	0.003	0.003
Placebo effects (st.d.)	0.083	0.013
Actual treatment effects	0.221** (0.091)	0.085***(0.015)
Replication	1000	1000
z-score (H_0: placebo = actual effect)	2.635	6.236
p-value	0.008	0.000

1.6. General Discussion

In this research, we report evidence that demonstrates that the introduction of an IT-enabled monitoring system reduces employers' preference for workers with high effort-related reputations and lowers entry barriers for inexperienced workers. Our estimation results are based on a unique quasi-natural experiment at Freelancer that implemented a monitoring system for time-based projects but not for fixed-price projects. This allows us to use the DID framework to estimate the causal effects of implementing a monitoring system. We report two main findings. First, after the introduction of the IT-enabled monitoring system, while employers' preference for the capability-related reputation for both fixed-price and time-based projects remains unchanged, employers place less emphasis on effort-related reputation for time-based projects (but not for fixed-price projects). Second, the introduction of the IT-enabled monitoring system lowers the entry barrier for inexperienced workers and attracts more bids for time-based projects. This finding suggests a nuanced substitution relationship between monitoring and reputation such that monitoring partially substitutes for effort-related reputation but cannot substitute for capability-related reputation. It further suggests that IT-enabled monitoring can alleviate the cold-start problem.

Our study contributes to several streams of IS research. First, this is the first large-scale empirical research to examine the causal impact of deploying an IT system on the mitigation of moral hazard on a leading online labor platform. Specifically, we examine the role of the IT-enabled monitoring system in matching the demand and supply of online labor. Unlike the previous literature, which mainly examines the effect of monitoring systems in a firm setting (Gopal and Koka 2010; Pierce et al. 2015; Ranganathan and Benson 2017), we analyze the impact of a monitoring system on a two-sided online labor platform, which enables us to identify unique aspects of online platforms and systematically study the effect of the IT-enabled monitoring system on both the demand and supply sides of the online labor platform. Second, our study extends the previous research on the effect of reputation systems on digital platforms (Ba and Pavlou 2002; Bockstedt and Goh 2011; Dellarocas 2005, 2006; Lin et al. 2016; Moreno and Terwiesch 2014). The previous literature on reputation systems commonly views reputation as a signal of workers' competence (Banker and Hwang 2008), which motivates workers to expend more effort (Dellarocas 2006). This paper adds to the understanding of reputation by underscoring the distinct impacts of capability-related reputation and effort-related reputation. Our results suggest that while IT-enabled monitoring has no significant impact on the importance of capability-related reputation, it can serve as a substitute for the signaling effect of effort-related reputation, which alleviates moral hazard by providing more precise and timely information about workers' actions (Agrawal et al. 2014; Pierce et al. 2015). This suggests that future research on reputation systems should also take the availability of monitoring systems into account as a critical contingency factor. Third, our research suggests that the IT-enabled monitoring system is not simply a partial substitution for reputation systems. By substituting for effort-related reputation, IT-enabled monitoring systems reduce agency costs by lowering the entry barrier for workers who have no prior experience on a focal platform. Therefore, our finding underscores a critical role of IT-enabled monitoring in overcoming a significant limitation of reputation systems that has hitherto been ignored in the IS literature: they create entry barriers for qualified workers who have not yet established reputations on a platform. Notably, IT-enabled monitoring could lower the entry barrier due to the effort-related reputation.

Our research also provides important managerial implications for the design of online labor markets (Hong et al. 2016) and online platforms in general (Ghasemkhani 2017). There is a large body of research suggesting that reputation helps to mitigate moral hazard by acting as both a stimulus for high effort (Horton and Golden 2015) and a sanctioning mechanism (Dellarocas 2006). Meanwhile, it has been suggested that monitoring systems are highly effective in improving agents' performance (Duflo et al. 2012; Hubbard 2000; Pierce et al. 2015). Our study suggests that there is a nuanced substitution relationship between monitoring and reputation. Specifically, monitoring partially substitutes for effort-related reputation but does not substitute for capability-related reputation. Hence, our study deepens our understanding of the optimal design of online labor platforms (Hong et al. 2016) by emphasizing the potential interaction effect between effort-related reputation and monitoring.

We acknowledge a number of limitations of this study, which open up avenues for future research. First, we note that due to data limitations, employers' actual usage of records from monitoring systems is not available. Second, we only focused on testing the effect of the IT-enabled monitoring system on employers' preference and workers' bidding behaviors. Future research should consider exploring the long-term effect of the IT-enabled monitoring system on workers' skill investment. Third, though we show that the sign and the existence of the treatment effect are robust to various alternative specifications (e.g., an alternative control group), the point estimates of the treatment effects in our main analysis may be amplified to some extent due to limited enforceability (Cooley et al. 2014). Finally, our study is conducted in the context of online labor markets and our findings may be limited in their generalizability to other online platforms. Although moral hazard is a universal issue on most online platforms, the IT artifact examined in this study—a monitoring system—may not be applicable to platforms focusing on the transaction of physical products, such as eBay. Further research should explore the effects of other monitoring systems that may be suitable for other online platforms.

1.7. Concluding Remark

Using a large-scale dataset from one of the major platforms that facilitate labor contracting, we utilize matching methods in tandem with a quasi-natural experimental difference-in-differences analysis to identify and quantify the effects of implementing an IT-enabled monitoring system on employers' preference and workers' entry decisions. Our results demonstrate a nuanced substitution relationship between monitoring and reputation, such that monitoring partially substitutes for effort-related reputation but not for capability-related reputation. Our findings further suggest that implementing a monitoring system lowers the entry barrier for inexperienced workers with no prior reputation and thus attracts more bids. Overall, our results provide support for the effectiveness of IT-enabled monitoring in mitigating moral hazard and alleviating the cold-start problem in online labor markets and carry important implications for designing two-sided digital platforms.

CHAPTER 2

HOME BIAS IN ONLINE EMPLOYMENT: EVIDENCE FROM AN ONLINE LABOR MARKET

We study the nature of home bias in online employment—an employer preference for workers from his or her home country. Using a unique large-scale data set from a major online labor market containing employers' consideration sets of workers and their ultimate selection of workers, we first estimate employers' home bias in their online employment decisions. Further, we disentangle two types of home bias—statistical and taste-based home bias—using a quasi-natural experiment based on the introduction of a monitoring system on an online employment platform, which enables employers to easily observe workers' effort on time-based projects. After matching comparable fixed-price projects as a control group using coarsened exact matching, our difference-in-differences estimations indicate that home bias in online employment is partially driven by statistical discrimination. Finally, we study heterogeneity in home bias across employers from different countries with a post-treatment sample with minimal statistical discrimination. We find that, consistent with the in-group favoritism literature, employers from countries with stronger traditional values, lower cultural diversity, and smaller populations tend to have a stronger home bias. Taken together, these findings shed light on the coexistence of statistical and taste-based home bias.

Keywords: home bias, employment, statistical discrimination, taste-based discrimination, quasi-natural experiment, gig economy, online labor market

2.1. Introduction

Employers often make hiring decisions under significant uncertainty due to information asymmetry (Kugler and Saint-Paul 2000). Hiring uncertainty leads employers to rely on observed worker characteristics (e.g., race and gender) that are not directly related to their capabilities or diligence (Hendricks et al. 2003; Fryer and Jackson 2008). The global expansion of online employment platforms has created employment arrangements that allow the employer and the worker to come from different parts of the world (Hong and Pavlou 2017), leading to the potential for another form of discrimination: *home bias*, according to which employers might prefer hiring workers from the same country. [12] Anecdotes have suggested that this form of discrimination exists even in offline employment. For example, managers at Oracle, predominantly Asians, were recently involved in a lawsuit because of their discrimination against qualified white workers and in favor of applicants who mostly immigrated to the U.S. from Asian countries.[13]

The literature has documented home bias in a variety of contexts (Hortaçsu et al. 2009; Lin and Viswanathan 2015; Chan and Ghose 2014) such as portfolio management (Coval and Moskowitz 1999), peer-to-peer lending (Lin and Viswanathan 2015), and product purchase (Hortaçsu et al. 2009). While this topic has not yet been formally examined, the existence of home bias in online employment settings is potentially

---

[12] Note that such geographic based preference can be rational or irrational. Following the previous literature, we use the term "home bias" instead of geographic-based preference. In the previous literature, home bias refers to the phenomenon of individuals preferring to conduct transactions with counterparts within a shorter geographic distance (Hortaçsu et al. 2009; Lin and Viswanathan 2015).

[13] http://fortune.com/2017/01/18/labor-department-oracle/

harmful to the platforms because it may decrease the potential for global labor arbitrage (Roach 2003; Gefen and Carmel 2008) and may thus lead to significant inefficiencies (Gong et al. 2018). Furthermore, home bias may be one type of bias that makes employment decisions unfair to some workers (Bertrand et al. 2005) and thus leads to entry losses for labor market platforms. Important as this question is, there is no study on home bias in employment settings yet. Prior studies on employment predominantly focus on offline settings where home bias is unlikely to be an issue because labor contracting in such settings often takes place locally. Further, even when there are ample variations in workers' home countries, recruiting data is difficult to obtain, due to the proprietary and confidential nature of recruiting. In addition, the recruiters are oftentimes not the employers who will be directly working with prospective workers, further compounding the challenges in using offline employment data to analyze home bias. In contrast, owing to the global nature of online labor markets with fine-grained worker data enabled by web-based information technology and direct observations of the employers' consideration sets and hiring choices, the online employment context offers an excellent venue for exploring and identifying home bias in employment decisions.

It is important to understand the mechanisms responsible for home bias in online employment. Previous literature offers explanations for home bias in investment portfolios and international trade contexts (Obstfeld and Rogoff 2000). Generally, the identification of overall home bias relies simply on the estimation of preferences for transactions with partners from the same country (Lin and Viswanathan 2015; Hortaçsu et al. 2009).[14] However, it is the understanding of the mechanisms that drive home bias that offers important theoretical and policy implications. In fact, home bias can be driven by either the statistical discrimination mechanism (Cooper and Kaplanis 1994; Helliwell 2000) or the taste-based discrimination mechanism (Lewis 1999; Lin and Viswanathan 2015). For example, in the investment example, statistical discrimination refers to investors' preferences for domestic portfolios or trade because of the associated higher expected returns based on signal extraction from the group-level characteristics and the product-specific characteristics (Phelps 1972; Arrow 1973). In other words, investors tend to expect that domestic portfolios or trade will perform better than foreign ones with the same observable characteristics, due to their higher trust or confidence in domestic ones. In contrast, taste-based discrimination arises from a priori liking for domestic portfolios or trade, which is not related to the signal extraction or utility function (Becker 1971). This line of literature suggests that home bias can be driven either by statistical discrimination for rational reasons—such as established institutional factors and the possibility of direct contract enforcement (French and Poterba 1991; Hortaçsu et al. 2009)—or by the taste-based discrimination based on irrational reasons or prejudicial tastes, such as individuals' reluctance to share risks with foreigners (Lewis 1999).

Due to the heterogeneous nature of different countries, taste-based home bias could also be heterogeneous. As workers' locations (countries) are highlighted in many leading global online labor markets (e.g. Upwork, Freelancer), employers may use geographic

[14] It should be noted that some previous studies define and explore the home bias phenomenon at a more granular level (e.g., the state level or city level) (Lin and Viswanathan 2015; Hortaçsu et al. 2009). In our paper, we focus on employers' home bias at the country level because it is a salient cue for the employers during the hiring process. In the robustness check section, we also show that employers have additional home bias at the city level too.

information as a salient cue for social categorization when they evaluate workers. In doing so, employers may have a higher nonpecuniary utility when hiring local workers who are the same social groups with them due to in-group favoritism (Chen and Li 2009). While the literature on in-group favoritism generally finds it to be positive, the magnitude of such irrational biases may vary according to social-environmental factors, such as group norms (Sagiv and Schwartz 1995), within-group similarity (or lack of diversity) (Luijters et al. 2008), and group size (Brewer and Kramer 1986; Simon and Hamilton 1994). In a similar vein, the strength of employers' home bias may depend on contingent factors such as norms in the country (e.g., traditional values), diversity in the country (e.g., cultural diversity within national population), and group size (e.g., size of country population).

Home bias in employment decisions bears some similarities to international trade in terms of potential mechanisms. It is, however, also distinct from international trade in terms of the significant hidden-action issues involved in employment decisions. Specifically, unlike trade decisions concerning standardized products or commodities, involving only ex ante information asymmetry of hard-to-observe qualities, online employment involves noncontractible elements of labor, and thus imposes severe information asymmetry between the workers and employers (Hong and Pavlou 2017)—especially the ex post information asymmetry on hard-to-observe worker efforts. On the one hand, information asymmetry and the consequences thereof—for example, unpredictable project performance—may induce employers to more carefully contemplate potential economic outcomes, rather than relying solely on the home country heuristic. On the other hand, asymmetric information might make the employment decision more effortful and resource-consuming, thus exacerbating the employers' reliance on their preexisting liking of local workers.[15] More importantly, the interpersonal nature of employment relationship may reinforce employers' optimism bias (Strong and Xu 2003) on local workers. Therefore, it is not clear which mechanism would primarily drive home bias in online employment settings. Bearing the above in mind, we seek to extend the previous literature on home bias by examining employment decisions in online labor markets; specifically, we address the following three research questions:

- *Q1 (existence): Does home bias exist for employment decisions in online labor markets?*
- *Q2 (mechanism): Which mechanism (statistical versus taste-based) drives home bias in online labor markets?*

To answer these questions, we obtained a unique, large-scale data set from an online labor market (Freelancer.com), in which we are able to reliably observe both the employer and workers' countries (and other attributes), the employers' consideration sets of workers who applied for projects, and the employers' hiring choices. Since online labor markets are global, this research setting allows us to examine home bias in online employment because it offers the desired variation in the workers' countries of origin. We first quantify home bias in our sample and then disentangle the mechanisms for home bias by leveraging a quasi-natural experiment—the introduction of a monitoring system on Freelancer.com. The introduction of the monitoring system constitutes a significant event serving as an exogenous shock to the level of information asymmetry between employers and workers. By contrasting the theoretical predictions of the statistical versus

[15] In our paper, we use "local workers" to refer to workers residing in the same country as employers.

taste-based mechanisms, we identify the underlying mechanism of the observed home bias. Specifically, because the monitoring system lowers employers' reliance on group-specific signal extraction by providing information on ex post individual-specific effort, we expect that the introduction of the monitoring system will lower home bias driven by statistical discrimination. At the same time, we believe that taste-based home bias will remain unchanged, as it will not be affected by the change in the availability of individual-specific information. Our econometric identification further hinges on the fact that the monitoring system is applicable only to time-based projects, and not to fixed-price projects, which allows us to use a difference-in-differences (DID) framework for causal analyses. Additionally, based on the sample during the post-treatment period, which is characterized by very little statistical home bias, we explore whether there is heterogeneity of home bias across employers, in order to examine the existence of taste-based home bias. Based on our analyses, we observe three key findings. First, there is a robust observation of the existence of home bias after controlling for language, time-zone, and currency differences. Second, after the Coarsened Exact Matching (CEM) of comparable fixed-price projects as a control group for the time-based projects that received the exogenous information shock, our difference-in-differences estimations show that the home bias in online employment is partially driven by statistical discrimination. Lastly, consistent with the in-group favoritism literature, we find that employers from countries with high traditional values, lower diversity, and a smaller user base (or population size), tend to have a stronger home bias. This lends support to the existence of taste-based home bias. As a whole, our study provides compelling evidence of the coexistence of statistical and taste-based home bias.

Our paper contributes to several related streams of literature. First, our study contributes to the home bias literature, being among the first studies to investigate the existence and mechanisms of home bias in the employment setting with a quasi-natural experiment. Our study extends previous home bias research in financial and trade transactions—which has focused mainly on decisions under ex ante information asymmetry—to the employment decision, which is threatened by both ex ante and ex post information asymmetry. Moreover, since most of the previous discrimination literature on labor markets does not distinguish statistical discrimination due to ex ante information asymmetry (e.g. worker capabilities) or ex post information asymmetry (e.g. worker effort), we add to this strand of literature by specifically examining whether alleviating ex post information asymmetry helps to reduce home bias. Further, given that some expect online employment to soon comprise the majority of the U.S. workforce,[16] this paper advances the employment discrimination research by demonstrating the impact of the home country affiliation between employers and workers using detailed data at the individual level and precise information about the employer's consideration set. Additionally, beyond confirming the existence of home bias, our study differentiates statistical home bias from taste-based home bias and further highlights that heterogeneity in taste-based home bias can be explained by in-group favoritism. Second, our paper adds to the understanding of discrimination in the online gig economy and, to the best of our knowledge, represents the first attempt to investigate the potential of monitoring systems in attenuating statistical home bias, and more broadly, statistical discrimination. While

[16] See more on https://www.upwork.com/press/2017/10/17/freelancing-in-america-2017/; https://www.weforum.org/agenda/2017/08/why-the-future-of-work-could-lie-in-freelancing

the gig economy seems to provide a frictionless avenue of low-entry barrier for the two-sided matching, some emerging research suggests that it can also develop into a breeding ground of discrimination (Ge et al. 2016; Edelman et al. 2017; Chan and Wang 2017), which is legally prohibited yet not enforced in the online gig economy (Todisco 2014; Edelman et al. 2017; Belzer and Leong 2018). One easy solution is to restrict the availability of information that employers have shown to discriminate against. However, this does not eliminate statistical discrimination but only shift it from one type of information (e.g., criminal records) to another (e.g., race) (Agan and Starr 2016; Doleac and Hansen 2016). Our study suggests that monitoring can effectively alleviate statistical discrimination.

2.2. Theoretical Background

2.2.1. Home Bias

Home bias is a phenomenon that is well-documented in the literature on financial markets (Forman et al. 2009, 2012; Sorenson and Stuart 2001; Lin and Viswanathan 2015) and international trade (Brunetti et al. 1997; Ghani et al. 2014; Helliwell 2000; Hortaçsu et al. 2009). Studies on home bias have primarily focused on offline contexts (Obstfeld and Rogoff 2001). For instance, Lewis (1999) finds that the reluctance to share the risks associated with international equity helps to explain the observed equity-home-bias, characterized by investors who have a much higher percentage of equity in domestic assets than the optimal ratio which has the minimum variance for investment return. Moreover, Coval and Moskowitz (1999) suggest that investors also strongly prefer the stocks of firms that are geographically closer when making domestic investment decisions.

As online trade and online financial markets emerge, recent work starts to explore the geography-based preference in online settings. These related studies focus on how rational explanations and irrational factors potentially lead to home bias. On the one hand, some studies conclude that the preference for shorter geographic distance is driven by rational considerations. For instance, Hortaçsu et al. (2009) find that contract enforcement and localized consumption of goods jointly contribute to home bias in the online trade context (such as eBay). Other rational explanations regarding home bias in traditional financial markets include established institutional factors (French and Poterba 1991; Helliwell 2000; Brunetti et al. 1997) and investors' rational desires to hedge specific sources of risk (Cooper and Kaplanis 1994). On the other hand, a few studies on online financial markets suggest that the geography-based preference is more consistent with taste-based preference. For example, Lin and Viswanathan (2015) explore the home bias in online peer-to-peer markets and identify that part of home bias is driven by lenders' taste-based preferences. Ghani et al. (2014) find that Indians show ethnic discrimination when making outsourcing decisions, and that this discrimination is also more consistent with the prediction of taste-based preferences. Overall, the evidence regarding the mechanisms of home bias is mixed (Hortaçsu et al. 2009; Ghani et al. 2014; Lin and Viswanathan 2015). Additionally, despite the rich literature on home bias in offline and online financial markets and trade, no research has explored home bias in the employment setting. In financing or trade contexts, assets or products for sale are usually standardized and ex ante information asymmetry regarding worker capabilities is the only potential trigger for statistical home bias. In contrast, for the employment context, employers' benefits depend on both the capabilities of workers and the effort expended by workers, which involves both ex ante information asymmetry and ex post information asymmetry. The unique information structure in the employment context may cause the strength and mechanism of home bias to be different than in other contexts explored by previous literature (Hortaçsu et al. 2009; Ghani et al. 2014; Lin and Viswanathan 2015).

For the identification of home bias, scholars employ different empirical approaches according to their levels of analysis. When researchers only have access to macro-level data (e.g., country pairs or city pairs), a commonly used approach is the gravity equation approach (Bergstrand 1985). Modeled after the gravity equation in physics, the gravity equation for international trade is a power function of the inverse of the distance between two parties, the economy volume of each party (e.g., GDP measurements) (Wolf 2000), and other associated factors such as the "remoteness" of two parties in relation to other parties (Anderson 1979) and the observed quality of trade (Burtch et al. 2014). The gravity model assumes identical expenditure functions among all parties, with smaller parties naturally modeled as having stronger home bias, given the small ratio of their GDP to the global GDP. As such, the estimator in the gravity model tends to be biased (Anderson 1979). Therefore, when microlevel data are accessible, researchers typically prefer alternative methods such as choice models (Ghani et al. 2014) or the potential-dyads approach (Lin and Viswanathan 2015). When the decision makers' consideration sets are not well specified, potential-dyads analysis considers all available alternatives in the model to explore whether the decision makers have a stronger preference for transaction partners from their home countries by assuming that all potential alternatives are included in the consideration set. However, the potential-dyads approach is threatened by the nonindependence concern, given that each dyad is also directly or indirectly associated with other dyads, leading to the common-actor effect (Lincoln 1984; Stuart 1998). Moreover, this approach may also be biased by the inflated number of dyads, which are highly unlikely in reality (Stuart 1998), and which may influence the estimate of all coefficients and usually cannot be fully corrected by adding dyad-specific fixed effects (Sorenson and Stuart 2001). Hence, when the decision makers' consideration sets are well specified, choice models are preferable. Owing to the granular data about each employer's consideration sets and selection choices, we employ choice models instead of potential-dyads analysis without imposing strong and untestable assumptions.

2.2.2. Overview of the Home Bias Mechanisms

Due to limited worker information or prejudicial distaste, employers often rely on heuristics based on the workers' identity characteristics to extrapolate individual workers' capabilities and expected effort, and to evaluate the potential utility of hiring these workers. This process tends to result in discrimination. Discrimination refers to employers' systematic differential treatment of workers based on their group or demographic characteristics that are not directly related to productivity (Arrow 1973). These characteristics include, for example, race (Altonji and Blank 1999; Bertrand and Mullainathan 2004; Fryer and Levitt 2004; Arceo-Gomez and Campos-Vazquez 2014; Ge et al. 2016), gender (Chan and Wang 2017; Bertrand and Duflo 2017; Neumark et al. 1996; Goldin and Rouse 2000; Edelman et al. 2017), and immigrant identity (Åslund et al. 2014). Multiple mechanisms have been proposed to explain the sources of discrimination. Based on its mechanism, discrimination can be classified as rational/statistical discrimination or as taste-based discrimination (Bertrand and Mullainathan 2004). Specifically, statistical discrimination assumes that employers are rational and use group identity to infer individual workers' capabilities or effort (Arrow 1973). For instance, if the employer learns that local workers are more skilled and diligent based on his or her private information, the employer might use location information (e.g., local vs. foreign) as a signal to infer workers' capabilities or effort. Conversely, taste-based discrimination is purely based on employers' prejudicial distaste of foreign workers, which does not involve the rational inference of worker capabilities or effort that would affect utility for the employer (Becker 1971; Heckman 1998; Ghani et al. 2014).

A key challenge in this research stream is to empirically disentangle statistical discrimination from taste-based discrimination. In general, there are two ways of identifying statistical discrimination, namely, the static and dynamic approaches (Rubineau and Kang 2012). The static approach measures the static difference among between-group pairs after accounting for other observable productivity characteristics (Bertrand and Duflo 2017). Statistical discrimination diminishes among between-group parties when there is more information or stronger signals concerning productivity characteristics. However, Heckman and Siegelman (1993) suggest that the differences among between-group parties are difficult to measure or control for. As such, the static approach is usually plagued by omitted variable bias and relies heavily on assumptions about the distribution of unobservable characteristics.

The dynamic approach, in contrast, that measures how the discrimination of between-group pairs changes with information shocks that address or alleviate information asymmetry (Rubineau and Kang 2012), serves as a better means of identification. When there is a significant information change, individuals will update their beliefs, which alleviates information asymmetry and its consequence, statistical racial discrimination. However, researchers observe that students tend to show stronger discrimination after a year of training, which suggests that it is not statistical discrimination that drives racial disparities (Rubineau and Kang 2012). Based on the dynamic predictions of statistical discrimination, information changes such as removing gender information (Goldin and Rouse 2000) or criminal background information (Doleac and Hansen 2016) will lead to a change in the magnitude of discrimination. In summary, the dynamic approach verifies whether the pattern change of observations is consistent with the expectation of statistical discrimination to provide a reliable way of identifying statistical discrimination.

As List (2004) suggests, information asymmetry tends to drive statistical discrimination, and information changes can influence statistical discrimination by reducing ex ante information asymmetry (e.g. worker capabilities) or ex post information asymmetry (e.g. worker effort). Here, ex ante information asymmetry refers to unobserved or hard-to-observe worker capabilities or productivity characteristics, and ex post information asymmetry denotes unobserved or hard-to-observe actions that are related to worker productivity, such as their effort. However, given that most of the previous studies either primarily focus on ex ante information asymmetry (Rubineau and Kang 2012) or do not distinguish between ex ante and ex post information asymmetry (Goldin and Rouse 2000; Doleac and Hansen 2016), it is still unknown whether ex post information asymmetry plays a critical role in employment discrimination. Furthermore, examining and demonstrating whether marketplace designs helping to mitigate ex post information asymmetry (e.g., monitoring systems) can alleviate employment discrimination have important managerial implications for digital platforms.

2.2.3. In-group Favoritism

The home bias phenomenon is also related to the literature on in-group favoritism (Chen and Li 2009), wherein individuals prefer in-group members over out-group members (Allen and Wilder 1975; Efferson et al. 2008; DiDonato et al. 2011). In-group favoritism takes a few forms, including in-group bias (Chen and Li 2009), in-group altruism (Brewer and Kramer 1986; Sun et al. 2015), in-group trust (Falk and Zehnder 2013), and out-group comparison (Reynolds et al. 2000). In the context of online labor markets, countries of residence are typically included in public information saliently shown on workers' profiles. Therefore, country of residence, as a group characteristic of workers, is expected to serve as a basis for social categorization. In particular, employers may consider local workers as in-group members and show a preference for them, as compared to foreign workers with similar characteristics. Moreover, despite the positive effects for in-group favoritism generally reported in the literature, the magnitude of such favoritism may vary with multiple group-level contingent factors, including group norms (Sagiv and Schwartz 1995), within-group similarity (or lack of diversity) (Luijters et al. 2008), and group size (Brewer and Kramer 1986; Simon and Hamilton 1994). Specifically, as in-group social norms and the conformity with norms become stronger, in-group favoritism tends to be stronger. Moreover, the strength of in-group favoritism is weaker in more diverse groups versus homogeneous groups. In addition to within-group diversity, group size also matters. Individuals tend to show stronger in-group favoritism in smaller groups versus larger groups.

2.2.4. Online Labor Markets

Online labor markets facilitate the procurement of on-demand labor services across the borders of cities or countries. Recently, online labor markets have experienced tremendous growth and are projected to play a prominent role in the U.S. labor market.[17] Due to the low barrier to entry for workers from various countries and well-established arbitration systems, online labor markets enable employers to access a broad set of prospective workers by reducing search and transaction costs (Chen and Horton 2016; Chan and Wang 2017). That being said, online labor markets are also limited because of their impersonal nature. Specifically, unlike traditional labor markets in which employers can assess and verify workers' capabilities through field interviews and their effort through manual monitoring, due to spatial and temporal separations, online labor markets have a higher degree of information asymmetry between workers and employers (Kokkodis and Ipeirotis 2015). In general, there are two forms of information asymmetry in online labor markets: ex ante information asymmetry associated with hard-to-observe worker capabilities or other productivity characteristics (Fong Boh et al. 2007; Huang and Zhang 2016), and ex post information asymmetry associated with hard-to-observe actions or effort. In many cases, employers make hiring decisions based on the limited information provided by the platform, such as reputation and workers' countries of residence, and make inferences about workers' quality and effort. Given that the worker

[17] http://www.forbes.com/sites/groupthink/2014/10/21/the-next-big-thing-in-e-commerce-online-labor-marketplaces/#5f62eb9c6117

country is a salient information cue, like reputation and wages, employers may consider it a cue for social categorization and use it to form their expectations about workers' capabilities or effort expenditures, which will thus affect hiring decisions. Employers may therefore overvalue their expectations of local workers' capabilities or effort due to higher trust in them or merely show home bias resulting from in-group bias.

Moreover, home bias in online labor markets could be driven by either a statistical or taste-based mechanism. Specifically, the taste-based mechanism is due to employers' inherent tastes or to stereotypes rooted in the cultural environment offering employers a higher nonpecuniary utility derived from working with in-group members (i.e., local workers) versus out-group members (i.e., foreign workers), whereas the statistical mechanism is mostly due to the aforementioned lack of information on worker capabilities and effort. Since employers' taste-based home bias tends to be stable and persistent (Becker 1971), a feasible way of reducing the inefficiency costs associated with home bias would be to target employers' statistical home bias with information shocks.

2.3. Research Context and Data

Most online labor markets follow a reverse, buyer-determined hiring mechanism (Hong et al. 2015). To hire workers in online labor markets, an employer first posts a project on a web-based platform such as Upwork, Freelancer, or Guru. Detailed information about the project such as requirements and budget are provided in the dedicated webpage for the project. Workers who are interested in the job opportunity then bid on the project. After that, the employer makes a hiring decision based on the bid prices and workers' characteristics (e.g., reputation, country) (Ye et al. 2014).[18]

Compared to conventional offline labor markets, online labor markets offer a more suitable context for exploring employers' home bias because of the following reasons. First, a key confounding factor in the estimation of employers' home bias is the more common referrals given for local employers. For example, in offline labor markets, local employers may be more likely to hire local workers because of a potential direct social relationship between them or due to the higher reliability of job referrals from common acquaintances. Therefore, it is difficult to disentangle home bias from the unobserved differences in the number and reliability of referrals for local workers versus those for foreign workers. Owing to the impersonal nature of the online labor market, a social relationship between employers and prospective workers would be highly unlikely, which helps reduce concerns about the confounding influence of referrals. Further, while local employers can obtain more private information regarding prospective workers through field interviews, interviews in the offline setting tend to be proprietary and confidential, and thus usually unobservable to researchers. By contrast, because of the general lack of field interviews for online employment, employers' hiring choices are generally attributed mainly to observable variables, which are also available to researchers. Moreover, in most cases, precise information about employers' consideration sets and their hiring decisions in offline labor markets is not available to researchers, which forces researchers to rely on other untestable assumptions (e.g. independence of irrelevant

[18] In rare cases a project can have multiple winners. In our main analysis, projects with more than one winner are dropped. The results of our analysis are consistent if we keep the projects with more than one winner.

alternatives (IIA), random missing values in decisions). Given the apparent advantages of online labor markets for the estimation of home bias, we opt to use a data set from Freelancer (www.freelancer.com), one of the largest online labor market platforms.

On Freelancer, the employer may specify the project as a fixed-price project or a time-based project, and an employer pays a fixed amount or hourly wages, respectively, to the hired worker. Workers can browse active or ongoing projects on the website and selectively bid on them. Freelancer imposes a limit on the number of bids each worker can submit each month.[19] It is therefore in the interest of the worker to bid on projects for which they are likely to be hired that maximize their expected total compensation.

To rule out the effect of the auction format on employers' choices, we limit our analysis to projects using the most common public, open-bid auction format. As such, special projects, such as those with NDA, featured projects, sealed bid projects, and full-time projects are dropped from our sample. Further, to construct a homogenous sample of projects, we focus on projects in the most popular category, i.e., "IT, Software & Website." In addition, in order to avoid the potential disproportionate influence of observations from several small countries, we restrict our sample to projects posted by employers from the top 25 employer countries, which account for 83.8% of total projects on Freelancer. It's worth noting that our sample comprises 96.8% of projects including at least one bidder from the employer country. The definition and basic statistics of the key variables in our final sample are provided in Table 20. As Table 20 shows, on average, only 5% of bids are submitted from workers from the employer's home country.

[19] Free members could submit 8 bids per month. Gold members could submit more. However, the percentage of gold members in our data set is less than 0.1%.

Moreover, bidders vary greatly in terms of account tenure, project experience, and rating. Additionally, the percentage of bidders whose self-reported primary language is the same as the employer's is as high as 76%, which is probably due to the dominance of English on Freelancer. Meanwhile, the percentage of bidders who reside in the same time zone and use the same currency as the employers is 2% and 45%, respectively.

Table 20. Definitions and Summary Statistics of Key Variables

Variable	Variable definition	Mean	SD	Min	Max
A. Bids' Characteristics					
Bid price	The bid price posted by the worker	306.70	491.00	2.00	5000.00[20]
Milestone percentage	A feature provided by Freelancer, it denotes the percentage of controlled payments paid to the worker during the project	73.72	33.31	0.00	120.00[21]
Bidder tenure	The worker's tenure at Freelancer measured in months	31.60	28.70	0.00	183.00
Homecountry	A dummy variable (0,1), =1 if the worker and the employer live in the same country	0.05	0.22	0.00	1.00
Bid order rank	The sequence order of the worker's bid	19.63	20.07	1.00	263.00
Preferred freelancer	A dummy variable (0,1), =1 if the worker won the "Preferred Freelancer" badge	0.21	0.41	0.00	1.00
Review count	The number of reviews entered by previous employers	81.67	175.14	0.00	3937.00
Same language	A dummy variable (0,1), =1 if the employer's primary language is the same as that of the worker on this platform	0.76	0.43	0.00	1.00
Avg rating	The average overall employer-entered ratings for the bidder	4.00	1.60	0.00	5.00
Same time zone	A dummy variable (0,1), =1 if the time zone in which the employer lived is the same as that of the worker. Inferred, based on the IP address	0.02	0.14	0.00	1.00
Same currency	A dummy variable (0,1), =1 if the employer's primary currency is the	0.45	0.50	0.00	1.00

[20] Because there are unreasonably large values in maximum bid prices and the win rate of those bids are all zeros, we dropped the top 1%. The results of our main analysis are consistent if we keep all the bids in our full sample.

[21] For time-based projects, there are a few cases in which the milestone percentage is larger than 100. This means that employers pay more than one hourly salary to workers when workers finish part of work.

	same as that of the worker on the platform				
B. Project Characteristics					
Time-based	A dummy variable (0,1), =1 if the project is a time-based project; =0 if it is a fixed-price project	0.09	0.29	0.00	1.00
Log (employer overall rating)	The average overall worker-entered ratings for the employer (log-transformed)	1.77	0.15	0.00	1.79
Language Eng	A dummy variable (0,1), =1 if the project is described in English	0.88	0.32	0.00	1.00
Log (paid amount)	Amount (in USD) paid by the employer after the project was completed (log-transformed)	4.44	1.31	0.00	11.00
Log (budget max)	The maximum of bid prices for this project set by the employer (log-transformed)	4.74	1.44	0.69	11.78
Log (title length)	Number of characters in the project title (log-transformed)	1.61	0.52	0.00	3.85
Log (description length)	Number of characters in the project description posted by the employer (log-transformed)	2.77	0.30	0.00	3.26
Employer developed	A dummy variable (0,1), =1 if the employer comes from a developed country	0.80	0.40	0.00	1.00
C. Country Characteristics					
Trad value	The average traditional value reported in the World Value Survey (Inglehart and Welzel 2010)	-0.36	0.59	-1.58	1.48
Cultural diversity	The cultural diversity index measured by Fearon and Laitin (2003)	0.26	0.15	0.02	0.67
Log popu	The log-transformed population size (in thousands) reported by the World Bank[22]	11.39	1.37	8.41	14.13

2.4. Empirical Evidence of Home Bias

2.4.1. Model-free Evidence of Employers' Home Bias

To provide some model-free evidence of the existence of home bias, we summarize the basic statistics for the employment choice made by employers from the top 25

[22] https://data.worldbank.org/indicator/SP.POP.TOTL?page=2

countries in Table 21.[23] For each country, we first identify all employers from that country and then calculate the ratio of their projects assigned to bidders from his/her home country for each employer. Meanwhile, we also calculate the average ratio of projects assigned to bidders from each country on a whole. By checking the difference between these two ratios (the last column of Table 21), we find that most employers tend to prefer bidders from their home countries. This finding suggests that employers may have a home bias. Next, we will present our econometric identification strategies.

Table 21. Country Distribution of International Employers' Employment Decisions

Employer's country	The awarded bidder's country								Share of projects assigned to bidders from that country	Diff
	country1	country2	country3	country4	country5	country6	country7	country8		
country1	4.11%	1.21%	0.43%	1.02%	38.26%	0.24%	0.04%	0.20%	2.61%	1.49%
country2	2.41%	2.09%	0.28%	1.05%	40.48%	0.14%	0.05%	0.16%	1.22%	0.87%
country3	2.05%	1.26%	2.57%	0.88%	43.12%	0.23%	0.02%	0.23%	0.53%	2.04%
country4	3.59%	1.03%	0.39%	2.36%	36.77%	0.00%	0.10%	0.16%	0.95%	1.41%
country5	1.69%	0.46%	0.09%	0.40%	59.65%	0.09%	0.06%	0.12%	41.58%	18.07%
country6	2.72%	1.44%	0.15%	0.45%	36.61%	1.51%	0.00%	0.00%	0.24%	1.28%
country7	1.85%	1.15%	0.09%	0.44%	42.03%	0.35%	0.18%	0.18%	0.05%	0.13%
country8	1.64%	0.70%	0.47%	0.00%	47.25%	0.00%	0.00%	0.23%	0.18%	0.05%

Note: "Diff" refers to the difference between the share of projects assigned to bidders from country X among all the projects posted by employers from country X and the share of projects assigned to bidders from country X among all the projects.

Table 21. Country Distribution of International Employers' Employment Decisions

(Cont'd)

[23] Employment decision is a two-sided matching process, if no domestic workers bid on the project, employers' hiring decision for that specific project will not influence the identification of employers' home bias.

Employer's country	The awarded bidder's country								Share of projects assigned to bidders from that country	Diff
	country 9	country10	country11	country 12	country 13	country14	country15	country16		
country9	3.57%	0.12%	0.00%	0.00%	0.12%	0.12%	0.12%	12.33%	0.21%	3.37%
country10	0.00%	2.23%	0.12%	0.00%	0.50%	0.25%	0.12%	8.44%	0.22%	2.02%
country11	0.00%	0.45%	0.00%	0.00%	0.45%	0.15%	0.00%	9.08%	0.10%	-0.10%
country12	0.16%	0.16%	0.16%	0.16%	0.65%	0.00%	0.00%	8.87%	0.04%	0.13%
country13	0.14%	0.00%	0.00%	0.00%	4.53%	0.14%	0.14%	7.82%	0.59%	3.93%
country14	0.45%	0.75%	0.00%	0.00%	0.60%	3.00%	0.15%	9.30%	0.19%	2.81%
country15	0.38%	0.19%	0.00%	0.00%	0.19%	0.00%	0.58%	11.73%	0.08%	0.50%
country16	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.91%	10.99%	10.92%

Note: "Diff" refers to the difference between the share of projects assigned to bidders from country X among all the projects posted by employers from country X and the share of projects assigned to bidders from country X among all the projects.

Table 21. Country Distribution of International Employers' Employment Decisions (Cont'd)

Employer's country	The awarded bidder's country									Share of projects assigned to bidders from that country	Diff
	country17	country18	country19	country20	country21	country22	country23	country24	country25		
country17	0.37%	0.00%	0.56%	0.93%	0.19%	0.00%	3.74%	3.93%	0.75%	0.10%	0.27%
country18	0.00%	1.02%	0.61%	2.25%	0.20%	0.00%	1.02%	3.68%	0.82%	0.10%	0.92%
country19	0.18%	0.36%	1.61%	2.14%	0.18%	0.00%	2.86%	6.61%	1.61%	0.40%	1.21%
country20	0.24%	0.24%	0.71%	9.95%	0.00%	0.00%	2.61%	4.03%	0.71%	1.96%	8.00%
country21	0.00%	0.00%	0.48%	2.14%	2.38%	0.48%	2.62%	5.95%	1.90%	0.43%	1.95%
country22	0.30%	0.00%	0.30%	1.48%	0.30%	0.30%	1.18%	6.51%	0.59%	0.14%	0.16%
country23	0.00%	0.00%	0.86%	0.58%	0.00%	0.29%	12.39%	3.75%	0.00%	2.37%	10.02%
country24	0.00%	0.00%	0.00%	2.34%	0.26%	0.00%	1.56%	29.09%	0.00%	5.07%	24.02%
country25	0.00%	0.00%	0.00%	0.49%	0.00%	0.00%	1.46%	3.88%	2.91%	1.01%	1.91%

Note: "Diff" refers to the difference between the share of projects assigned to bidders from country X among all the projects posted by employers from country X and the share of projects assigned to bidders from country X among all the projects.

2.4.2. Identification Challenges

In our context, we can precisely observe employers' consideration sets, their location, reputation, and other individual-specific information, as well as detailed workers' information. This offers a more fine-grained identification, and the omitted variable bias tend to be a smaller concern, compared to the gravity-equation based home bias analysis in the prior literature (Hortaçsu et al. 2009). We further control for the effect of static unobserved factors by employing models with project-specific fixed effects. In addition, we address the following identification challenges: First, geographical distance is confounded with some productivity-related factors for online employment that may add to friction between the employer and the worker (Hong and Pavlou 2017). To address this concern, we control for key differences between the bidder home country and the employer home country (language, time zone, currency). Second, workers from some countries with low GDPs are more desired, because they demand lower wages, and those countries (e.g., Pakistan) may be very different from countries where the majority of employers reside. To tackle this potential confounding factor, we control for bidders' country effects via dummy variables. Moreover, in the robustness check section, we also control for the country-month two-way fixed effect to account for the potential time-varying variations in terms of competitiveness and "market tightness"[24] from a worker's fellow countrymen in different projects.

[24] We assume that as the number of bidders increase, the market competition becomes fiercer and the market tightness increases.

2.4.3. Identifying the Existence of Home Bias

Following Ghani et al. (2014) and Lin et al. (2016), we estimate the extent of home bias with a conditional logit model, as well as a linear probability model (LPM) with project-level fixed-effects. Taking the conditional logit model as an example, the utility that the employer of project *i* obtains from hiring bidder *j* is constructed as follows:

$$\mathrm{U}(Project_i\_by\_bidder_j) = \alpha_i + \beta_1 Home\mathrm{country}_{ij} + controls(Bidder_{ij}) + \varepsilon_{ij} \quad (4)$$

where α_i represents the project-level fixed effect, which nests the employer-level fixed effects, since every project only has one employer. The focal variable, $Home\mathrm{country}_{ij}$, denotes whether the employer of project *i* and bidder *j* are from the same country. In addition, $controls(Bidder_{ij})$ includes various characteristics related to bidder j and his or her bid—such as bidders' review count, cumulative average rating, tenure, bid order rank, bid price, "Preferred Freelancer" badge, and whether bidder j shares the same language, uses the same currency, or is located in the same time zone as the employer, as well as bidder country dummies.[25] It is assumed that ε_{ij} follows the type-I extreme value distribution (Train 2009). A significant positive effect of $Home\mathrm{country}_{ij}$ (captured by $\hat{\beta}_1$) suggests that, on average, employers in the focal online labor market hold a home bias.

As mentioned before, there are two potential threats to our identification of the above model. First, time-invariant differences may exist across countries. For example, workers from some countries may be more competitive or require lower wages systematically. In response to this issue, we include 25 dummies for the top 25 countries in our sample and

[25] Within data, a worker's average rating is almost constant during our observational period. Therefore, we don't treat the worker rating as a time-variant variable here.

one dummy for the rest of the bidder countries. Second, the productivity of workers may suffer from differences in languages and time-zones, which may result in a spurious home bias. Moreover, employers may also prefer local workers due to currency exchange frictions. To alleviate these concerns, we use additional dummies to control, respectively, for whether employers and workers speak the same primary language, use the same currency, and reside in the same time zone.

As Model (1) in Table 22 suggests, without controlling for the similarities in language, currency, and time zone, employers show a preference for local workers. Employers' home bias slightly decreases after we control for the language, currency, and time zone effects (see Model 3), but the magnitude of home bias is still quite significant. To better understand the economic impact of home bias, we use Equation (2) to compute its monetary value, a common approach adopted by previous studies (Leung 2017; Dahl and Sorenson 2010). According to the results estimated by the conditional logit model in Model (3) of Table 22, employers are willing to pay local workers 24.97% more than foreign workers. [26]

$$\Delta Bid\ price = \ exp^{\frac{\beta_{homecountry}}{\beta_{log\ bid\ price}}} \tag{5}$$

Table 22. Estimation Results of Employers' Home Bias

Sample	Full sample		Full sample	
Model	(1) Logit	(2) LPM	(3) Logit	(4) LPM
Homecountry	0.516***(0.041)	0.042***(0.004)	0.387***(0.047)	0.032***(0.004)
Same language			0.416***(0.026)	0.018***(0.001)
Same currency			0.062***(0.021)	0.004***(0.001)
Same time zone			0.255***(0.057)	0.025***(0.005)
Log bid price	-1.735***(0.018)	-0.090***(0.001)	-1.736***(0.018)	-0.090***(0.001)
Log milestone percentage	-0.068***(0.016)	-0.003***(0.001)	-0.067***(0.016)	-0.003***(0.001)
Log review count	0.099***(0.007)	0.005***(0.000)	0.096***(0.007)	0.005***(0.000)

[26] Exp(0.387/1.736)-100%=24.97%.

Log avg rating	0.102***(0.009)	0.003***(0.000)	0.103***(0.009)	0.003***(0.000)
Log bid order rank	-0.327***(0.013)	-0.017***(0.001)	-0.313***(0.014)	-0.016***(0.001)
Preferred freelancer	0.499***(0.018)	0.027***(0.001)	0.471***(0.018)	0.026***(0.001)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	371,968	371,968	371,968	371,968
R-squared	0.486	0.043	0.494	0.044
LogLik	-47,740		-47,557	
AIC	95,542		95,182	
BIC	95,877		95,550	
Number of projects	23,943	23,943	23,943	23,943

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) R-squared in the logit model is calculated based on the maximum likelihood R-squared. f)* p<0.1, ** p<0.05, *** p<0.01.

2.4.4. Robustness Checks

Controlling for worker countries dummies in the main analyses reduces the potential existing differences of worker competitiveness across countries, and time-varying or project-specific contingent factors may influence the estimate of home bias. Here we use a multitude of measures to check the robustness of the results. First, we consider the potential time-varying variation in terms of competitiveness and "market tightness" from a worker's fellow countrymen in different projects as potential confounding factors. We assume that as the number of bidders increases, the market competition becomes fiercer and the market tightness increases. Therefore, we calculate and control for the number of workers and the average rating of workers from each country within the employer's specific consideration set. We find all the analysis, including the existence of home bias and the heterogeneity of home bias, to be highly consistent. Additionally, to further control for the potential time-varying worker competitiveness or cost difference across countries, we control for the country-month two-way fixed effect and reexamine our analysis. For instance, workers from some countries may have a labor cost advantage owing to the lower purchasing power parity (PPP) of their countries or the lower exchange rate of local currencies. In such cases, the country-month two-way fixed effects should help to control for the time-varying difference in workers from different countries. Again, the results are highly consistent.

Table 23. Robustness Results of Employers' Home Bias

Sample	Full sample		Full sample	
Model	(1) Logit	(2) LPM	(3) Logit	(4) LPM
Homecountry	0.388***(0.046)	0.030***(0.004)	0.376***(0.047)	0.030***(0.004)
Same language	0.413***(0.026)	0.017***(0.001)	0.418***(0.027)	0.017***(0.001)
Same currency	0.064***(0.021)	0.004***(0.001)	0.067***(0.022)	0.004***(0.001)
Same time zone	0.255***(0.057)	0.025***(0.005)	0.254***(0.057)	0.025***(0.005)
Log bid price	-1.735***(0.018)	-0.090***(0.001)	-1.744***(0.018)	-0.090***(0.001)
Log milestone percentage	-0.068***(0.016)	-0.003***(0.001)	-0.072***(0.016)	-0.003***(0.001)
Log review count	0.097***(0.007)	0.005***(0.000)	0.098***(0.007)	0.005***(0.000)
Log avg rating	0.102***(0.010)	0.001***(0.000)	0.101***(0.010)	0.001***(0.000)
Log bid order rank	-0.312***(0.013)	-0.016***(0.001)	-0.315***(0.014)	-0.016***(0.001)
Preferred freelancer	0.475***(0.018)	0.026***(0.001)	0.476***(0.018)	0.026***(0.001)
log avg country rating	0.000 (0.011)	0.009***(0.001)	0.005 (0.012)	0.009***(0.001)
log country bidder	-0.093***(0.015)	-0.002* (0.001)	-0.077***(0.015)	-0.001 (0.001)
Bidder country dummies	Yes	Yes	--	--
Bidder country and month two-way fixed effects	--	--	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	371,968	371,968	371,968	371,968
R-squared	0.490	0.045	0.487	0.048
LogLik	-47,538		-359,447	
AIC	95,148		720,063	
BIC	95,538		726,386	
Number of projects	23,943	23,943	23,943	23,943

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is only limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) R-squared in the logit model is calculated based on the maximum likelihood R-squared. f)* p<0.1, ** p<0.05, *** p<0.01.

2.4.5. Additional Analysis: Home Bias at the City Level

In line with the previous literature suggesting that consumers also have home bias at the state and city levels (Hortaçsu et al. 2009; Lin and Viswanathan 2015), we reexamine whether employers show additional bias toward local workers by leveraging the fact that Freelancer provides detailed users' location information. As Table 24 shows, after controlling for the differences in language, currency, time zone, and country, employers have a significant positive preference for workers located in the same city as them. This city-level home bias may be mainly related to statistical discrimination, as living in the same city opens up the possibility of direct contract enforcement (Hortaçsu et al. 2009).

Table 24. Estimation Results of City-level Home Bias

Sample	Full sample	
Model	(1) Logit	(2) LPM
	DV: whether the bidder is awarded	
Same city	0.898***(0.142)	0.120***(0.022)
Homecountry	0.363***(0.047)	0.029***(0.004)
Same language	0.417***(0.026)	0.018***(0.001)
Same currency	0.062***(0.021)	0.004***(0.001)
Same time zone	0.239***(0.057)	0.024***(0.005)
Log bid price	-1.736***(0.018)	-0.090***(0.001)
Log milestone percentage	-0.066***(0.016)	-0.003***(0.001)
Log review count	0.096***(0.007)	0.005***(0.000)
Log avg rating	0.104***(0.009)	0.003***(0.000)
Log bid order rank	-0.313***(0.014)	-0.016***(0.001)
Preferred freelancer	0.471***(0.018)	0.026***(0.001)
Country dummy	Yes	Yes
Project fixed effects	Yes	Yes
Observations	371,968	371,968
R-squared	0.486	0.044
LogLik	-47,539	
AIC	95,148	
BIC	95,527	
Number of projects	23,943	23,943

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) R-squared in the Logit model is calculated based on the maximum likelihood R-squared. f)* p<0.1, ** p<0.05, *** p<0.01.

In summary, we find that employers show a significant positive preference for local workers, even after controlling for the effects of language, currency, and time zone. Moreover, employers from a country with stronger traditional values, a smaller user base (or population size), and lower diversity, tend to have a stronger home bias. The analysis of heterogeneity deepens our understanding of potential country characteristics associated with home bias, and it provides some correlational insights into the existence of the taste-based home bias mechanism. Since the mechanism by which home bias manifests provides actionable implications about potential approaches to alleviate worker discrimination and loss of market efficiency, it is crucial to further understand the underlying driver of home bias in a causal framework. Therefore, to identify the mechanism of employers' home bias, we take advantage of a quasi-natural experiment in which ex post information asymmetry regarding workers' effort was exogenously reduced following the introduction of an IT-enabled monitoring system. We explain the research setting and our identification strategy in the next section.

2.5. Exploring the Mechanisms for Home Bias

2.5.1. Identification Strategy Regarding the Mechanisms of Home Bias

2.5.1.1. Testing Statistical Home Bias: A Quasi-Natural Experiment with Information Shock

In this study, guided by the dynamic approach, we propose a new identification strategy by examining changes in information that reduce ex post information asymmetry by providing detailed monitoring records reflecting each individual worker's effort. On February 5, 2014, Freelancer started implementing a monitoring system, enabling employers to conveniently monitor the progress of time-based projects. This monitoring system automatically takes screenshots and keeps track of workers' effort input (Figure 5). This monitoring system potentially affects employers in the following two ways: 1) the monitoring system can ensure the production of high-quality work, especially when hiring foreign workers. The monitoring system automatically takes a screenshot every few minutes and allows employers to provide detailed instructions or comments regarding any step in the work process. As such, the monitoring system can improve efficiency for employers working with freelancers in an online setting. 2) The monitoring system allows employers to keep track of each individual worker's work, so that employers have access to more verified information about individual worker efforts. Note that, according to the platform policy, the monitoring system is obligatory for all time-based projects, but is not applicable to fixed-price projects. Given that there is an exogenous change in the availability of ex post information asymmetries among time-based projects alone, we use time-based projects and fixed-price projects as the treatment and control groups, respectively, to investigate whether this information change caused by the monitoring system reduces employers' home bias, which subsequently enables inferring whether the statistical discrimination mechanism accounts for home bias. Our rationale for this identification approach is that it explicitly anchors on examining whether monitoring decreases the role of worker country in shaping employers' expectations about worker effort and can potentially lower employers' statistical home bias in online labor markets.

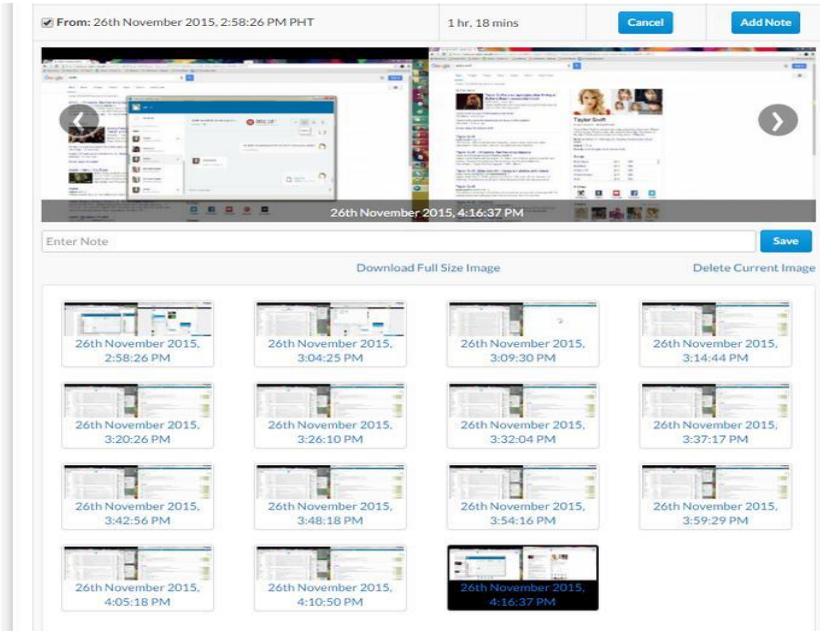


Figure 5. Screenshots of the Monitoring System[27]

According to the previous literature on statistical discrimination (Arrow 1973) and taste-based discrimination (Becker 1971), we make two distinct predictions about the underlying mechanisms of home bias following the introduction of the monitoring system. First, the introduction of an effective monitoring system, as a vital change in technology applied to online labor markets to alleviate ex post information asymmetry, will reduce employers' reliance on trust as a means of mitigating workers' opportunistic behaviors (Gulati 1995), which will thus attenuate employers' statistical home bias. Without a monitoring system in place, employers tend to have limited information regarding workers' effort and use trust to deter moral hazard (Barney and Hansen 1994; Lazzarini et al. 2008). In particular, the previous literature defines trust as one party's confidence in the other party's future benevolent behavior (Ring et al. 1992; Pavlou and

[27] https://www.freelancer.com/community/articles/what-you-need-to-make-remote-collaboration-work

Dimoka 2006; Lado et al. 2008) and low likelihood of choosing opportunistic behavior (Gulati 1995), implying low expected moral hazard risk (i.e., higher expected effort) (Barney and Hansen 1994). Employers tend to have higher trust in local workers due to perceived familiarity (Gulati 1995; Ba and Paul 2002) and their beliefs in the societal norms and cultural values of their own countries (Hofstede 1980; Doney et al. 1998). As such, employers expect local workers to be less likely than foreign workers to exhibit shirking behaviors, and they therefore show statistical home bias for local workers. However, if an effective monitoring system is in place, employers can prevent employee shirking, regardless of workers' home countries. This suggests that in the presence of an effective monitoring system, the difference in expected effort levels between local and foreign workers—and thus the level of statistical home bias—will be minimal. We thus predict that the information change imposed by the monitoring system will decrease employers' statistical home bias.

Our second prediction concerns taste-based discrimination, which is irrelevant to the availability of information and expected productivity, since it is based on personal likes or tastes, (Becker 1971; Rubineau and Kang 2012). Models of taste-based discrimination usually assume that employers show a constant distaste for foreign workers irrelevant to the unobservable characteristics of these workers (Becker 1971). If employers' home bias is taste-based, they discriminate against foreign workers simply on the basis of their animus or prejudice toward them, rather than because of any beliefs about foreign workers' higher probabilities of opportunistic behavior (Levitt 2004). Therefore, we predict that the introduction of an effective monitoring system would affect statistical discrimination but not taste-based discrimination (Becker 1971; Fang and Moro 2010). If statistical discrimination is at play, we will expect to observe a significant decrease in the level of home bias. Therefore, we propose the following predictions (Table 25):

Table 25. Types of Discrimination and Predictions

Forms of Discrimination	Dynamic Predictions about the Change in Home Bias
Statistical discrimination	After the introduction of the monitoring system, compared to the control group, employers' ex post information asymmetry related to worker effort decreases in the treatment group (i.e., time-based projects), leading to a lower home bias
Taste-based discrimination	After the introduction of the monitoring system, compared to the control group, employers' home bias remains unchanged in the treatment group (i.e., time-based projects)

2.5.1.2. Testing Taste-Based Home Bias: Heterogeneity Analysis Based on In-group Favoritism

Unlike statistical home bias which varies according to information shocks, taste-based home bias, similar to other types of in-group bias, tends to be stable and shaped by the long-existing cultural environment within each group. Building on the literature regarding in-group favoritism (Chen and Li 2009), we intend to explore the potential heterogeneity of home bias across different employer countries with various strengths of norms, diversities, and population sizes. By investigating whether the strength of home bias varies as predicted by in-group favoritism, we can better understand the heterogeneity of home bias and further infer whether employers' home bias is partially driven by the taste-based discrimination mechanism. Specifically, since employers' statistical home bias is expected to be small following the introduction of the monitoring system, we use only the matched sample during the post-treatment period to examine the heterogeneity of employers' home bias. Below we expound on how the literature on in-group favoritism predicts the heterogeneity of taste-based home bias.

First, in-group social norms encourage in-group collaboration as well as discouraging out-group collaboration. According to Sagiv and Schwartz (1995), in-group favoritism is positively related to conformity with norms and the importance of traditional values. Following this logic, in the context of online labor markets, employers from countries with a deep-rooted nationalistic public mindset and a strong tendency toward conformity to cultural norms are more likely to show in-group favoritism and a preference for workers from the same group (country). Therefore, we expect that employers residing in countries emphasizing traditional and nationalistic values will tend to have a home bias, as they are subject to a strong influence of social norms and national identity. Regarding the measure of the emphasis on social norms and nationalistic mindset across countries, a widely-used measure is the traditional values[28] reported in the World Values Survey (WVS).[29]

[28] According to the World Values Survey (WVS), country culture can be categorized based on the extent to which a society emphasizes traditional rather than secular values. Countries with high traditional values emphasize the importance of religion, deference to authority, traditional family values and have high levels of national pride (Inglehart and Welzel 2010). Compared with other measures such as those defined by Hofstede (1991, 2001), the WVS measure is based on more representative samples, and it has been employed to measure country culture in the IS literature (e.g. Burtch et al. 2014; Hong and Pavlou 2017). In this paper, we employ the traditional value estimated by WVS to measure the importance of social norms and national identity.

[29] WVS is a worldwide survey which mainly relies on face-to-face interviews. By now, WVS has been conducted in almost 100 countries comprising around 90 percent of the global population. It covers multiple thematic subsections, such as social values and stereotypes, societal well-being, trust and organizational membership, and economic values. It mainly adopts the full probability sampling on primary sampling units (PSU) and requires no replacements. Since the WVS is a longitudinal data set and the traditional values of different countries may slightly fluctuate, we calculate the average traditional values (denoted as "Trad value") for each country by combing multiple waves of WVS.

Second, within-group diversity tends to weaken the importance of shared social identity and reduce in-group favoritism. As Luijters et al. (2008) suggest, individuals' perceived levels of cultural-value similarity are correlated with their levels of identification with the group. As a result, we expect that employers from countries with greater cultural diversity would exhibit less home bias. Moreover, the previous literature provides a measure for cultural diversity among residents within each country—i.e, the country-specific cultural diversity index (Fearon 2003; Fearon and Laitin 2003). Specifically, this cultural diversity index measures the probability that every two individuals, randomly drawn from a country, speak a similar language (Fearon 2003). High levels of cultural diversity imply the potential of a low resemblance between employers and workers from this country, and thus a low taste-based home bias.

Third, in-group favoritism may be influenced by group size. The literature has documented evidence that as the group size decreases, individuals tend to be more prosocial toward other in-group members in their transactional relationships (Brewer and Kramer 1986; Simon and Hamilton 1994), leading to a stronger in-group favoritism. Along these lines, home bias is likely weaker for employers from countries that are more populous.

In summary, we build on prior research on in-group favoritism (Allen and Wilder 1975; Chen and Li 2009; Efferson et al. 2008; DiDonato et al. 2011) to explore how employers' home biases vary according to the traditional values, diversity, and population size of their countries in online labor markets. If the heterogeneity of employers' home bias following the introduction of an effective monitoring system is consistent with our predictions based on in-group favoritism literature, this lends support to the existence of a taste-based mechanism.

2.5.2. Main Results Regarding the Mechanisms of Home Bias

2.5.2.1. Coarsened Exact Matching

To balance the distribution of observables between the treatment and control group, we conduct CEM to generate a comparable sample (King et al. 2010; Iacus et al. 2012). CEM is a matching approach based on the Monotonic Imbalance Bounding (MIB) method which prunes observables to increase the balance of sample distribution between the treated and control groups (Stuart 2010). Moreover, unlike the Propensity Score Matching (PSM) approach which matches samples merely based on the expected probability of outcome variable, CEM is designed to balance the distribution of multiple covariates which are related to the treatment assignment between two groups (Iacus et al. 2012). As such, CEM has two advantages—that is, the lower model dependence and the better balance among various coarsened levels of covariates (Iacus et al. 2012). Specifically, by using CEM, we explicitly match the fixed-price projects with time-based projects based on all the observables that might affect employers' choice of contract type (Wu et al. 2012), including the length of project description, the length of title, the number of bids, the size of project, employers' experience and reputation, and the exact submit month of the project. CEM allows us to match two types of projects posted within the exactly same month, of similar size, and with similar level of information disclosure and level of competition, without being burdened by the curse-of-dimensionality issues of one-to-one exact matching (King et al. 2009). By matching fixed-price projects with time-based projects from a multivariate perspective, CEM helps demonstrate the robustness of our findings within a balanced sample. Consistent findings using an alternative matching method, PSM, are explicated in Appendix F.

2.5.2.2. DID Estimation Results

To measure the decrease in employers' home bias after the introduction of the monitoring system, we construct the differences-in-differences estimation in both the conditional logit model and the linear probability model with the project-specific fixed effects based on the matched sample. Using the logit model as an example, the DID specification is given by:

$$\mathrm{U}(Project_i\_award\_bidder_j) = \alpha_i + \beta_1 Home\mathrm{country}_{ij} + \beta_2 Home\mathrm{country}_{ij} \times Time - based_i + \beta_3 After_i \times Home\mathrm{country}_{ij} + \beta_4 After_i \times Home\mathrm{country}_{ij} \times Time\_based_i + controls(Bidder_j) + \varepsilon_{ij} \quad (6)$$

A significantly positive effect of $Home\mathrm{country}_{ij}$ prior to the introduction of monitoring systems (captured by $\hat{\beta}_1 + \hat{\beta}_2$) suggests that employers previously held home bias. Moreover, based on our previous discussions, if $\hat{\beta}_4$ is significantly negative, it implies that employers adjust their home bias according to available information provided by the monitoring system, which is known as statistical discrimination (Rubineau and Kang 2012).

As expected, the coefficient of the $After_i \times Home\mathrm{country}_{ij} \times Time\_based_i$ ($\hat{\beta}_4$) is significantly negative, which suggests that, for time-based projects, employers' additional preferences for bidders from their home countries decrease as the monitoring system makes more ex post individual-specific information available. The decrease in employers' home bias due to the introduction of the monitoring system suggests that employers' home bias cannot be attributed to taste-based discrimination. This lends support to the role of the statistical discrimination mechanism in shaping employers' home bias. To better understand the strength and the economic value of home bias, we next examine the sizes of related coefficients based on the full sample. Specifically, we focus on the coefficients based on the conditional logit model instead of the linear model, because the logit model better accounts for the interdependence among hiring decisions on all the bids for the same project. Before the introduction of the monitoring system, the total effect of $Home\text{country}_{ij}$ is 0.953 ($\hat{\beta}_1 + \hat{\beta}_2 = 0.219 + 0.759 = 0.978$) while the coefficient of log (bid price) is -1.736. In this sense, the change in the bid price required to reach parity in workers' likelihood of winning projects from foreign employers versus local employers is 1.757 (exp(0.978/1.736)=1.757). Given that the average hourly wage of foreign workers in time-based projects is USD 19.44, the effect of home bias translates to a premium of USD 14.708[30] for local workers. However, after deploying the monitoring system, the effect of $Home\text{country}_{ij}$ is reduced to 0.085, implying that local workers have to charge a lower price premium after the monitoring system was implemented, all else being equal. In other words, the economic value of $Home\text{country}_{ij}$ decreases to 0.776 dollars.[31] Since only the level of statistical discrimination decreases due to the availability of ex post individual-specific information, our bootstrap results suggest that roughly 89.94% of home bias is driven by statistical discrimination.[32] Given

[30] 19.44*[exp(0.978/1.736)-1]=14.708

[31] 19.44*[exp(0.085/1.736)-1]=0.976

[32] We calculate the percentage of statistical home bias for each bootstrap sample. Based on 1,000 bootstrap samples, we find that on average the statistical bias percentage is 89.94%, implying that at least 89.94% of home bias is driven by statistical discrimination.

that monitoring is very likely to be imperfect and that it is not capable of alleviating ex ante information asymmetry, the statistical discrimination is not likely to completely disappear after the introduction of the monitoring system. Therefore, this number is a conservative estimate. In fact, our finding that statistical discrimination is the primary driver is consistent with Arrow's statement (1971) which implies that perfect competition tends to drive out taste-based discrimination. Since online labor markets are prime examples of competitive two-sided markets, taste-based discrimination should play a relatively small role in hiring decisions in this context.

Table 26. DID Estimation of Employers' Home Bias

Sample	Full sample		Matched sample	
Model	(1) Logit	(2) LPM	(3) Logit	(4) LPM
Homecountry	0.219***(0.076)	0.016***(0.006)	0.307** (0.141)	0.041** (0.016)
Time-based×Homecountry	0.759***(0.175)	0.088***(0.020)	0.709***(0.223)	0.081***(0.028)
After×Homecountry	0.223***(0.083)	0.020***(0.007)	0.169 (0.151)	0.015 (0.018)
Time-based×After×Homecountry	-1.116***(0.232)	-0.119***(0.025)	-1.073***(0.288)	-0.129***(0.034)
Same language	0.415***(0.026)	0.018***(0.001)	0.385***(0.043)	0.026***(0.003)
Same currency	0.063** (0.021)	0.004***(0.001)	0.084** (0.035)	0.008***(0.003)
Same time zone	0.250***(0.057)	0.025***(0.005)	0.404***(0.091)	0.052***(0.010)
Log bid price	-1.736***(0.018)	-0.090***(0.001)	-1.849***(0.032)	-0.139***(0.002)
Log milestone percentage	-0.066***(0.016)	-0.003***(0.001)	-0.198***(0.026)	-0.016***(0.002)
Log review count	0.096***(0.007)	0.005***(0.000)	0.092***(0.012)	0.007***(0.001)
Log avg rating	0.103***(0.009)	0.003***(0.001)	0.080***(0.014)	0.003***(0.001)
Log bid order rank	-0.313***(0.014)	-0.016***(0.001)	-0.328***(0.024)	-0.025***(0.001)
Preferred freelancer	0.471***(0.018)	0.026***(0.001)	0.450***(0.032)	0.040***(0.003)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	371,968	371,968	86,840	86,840
R-squared	0.489	0.044	0.486	0.071
LogLik	-47,545		-14,823	
AIC	95,163		29,720	
BIC	95,564		30,067	
Number of projects	23,943	23,943	9,028	9,028

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) R-squared in the Logit model is calculated based on the maximum likelihood R-squared. f)* *p*<0.1, ** *p*<0.05, *** *p*<0.01.

In line with the previous literature (Autor 2003; Burtch et al. 2018; Chen et al. 2017), we explicitly test the parallel trend assumption of the DID model (Angrist and Pischke 2008) by checking whether the control group (fixed-price projects) has the same trend as the treatment group (time-based projects). Accordingly, we estimate the time-varying change in employers' home bias for time-based projects based on the following equation:

$$\mathrm{U}(Project_i\_award\_bidder_j) = \alpha_i + \beta_1 Home\mathrm{country}_{ij} + \beta_2 Home\mathrm{country}_{ij} \times Time\_based_i + \rho\tau_t \times Home\mathrm{country}_{ij} + \mu(\tau_t \times Time\_based_j \times Home\mathrm{country}_{ij}) + controls(Bidder_j) + \varepsilon_{ij} \quad (6)$$

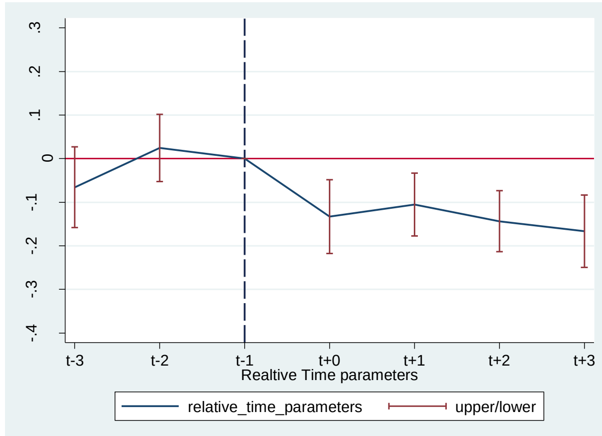
where τ_t represents a vector of time dummies and $\{\mu\}$ denotes the matrix of relative time parameters of employer i' home bias for bidder j estimated at time t. Estimating the treatment effect at different time periods enables us to examine the potential pretreatment trend. Specifically, given that the monitoring system was implemented on February 5, 2014, we use the quarter prior to the actual treatment (from October 2013 to January 2014) as the baseline (Autor 2003). According to Table 27 and Figure 6, all the relative time parameters are insignificantly positive prior to the introduction, while most of the relative time parameters in both the conditional logit model and the linear probability model turn out to be negative. In summary, the results of such an event study design suggest that a preexisting downward trend is unlikely to exist prior to the introduction of the monitoring system.

Table 27. Estimation Results of the Relative Time Model

Sample	Full sample		Full sample	
Model	(1) Logit		(2) LPM	
Homecountry	0.260**	(0.110)	0.025**	(0.010)

Time-based×Homecountry	0.783***	(0.295)	0.100***	(0.035)
Quarter_{t-3} ×Homecountry	0.044	(0.184)	-0.004	(0.015)
Quarter_{t-2} ×Homecountry	-0.179	(0.165)	-0.025*	(0.013)
Quarter_{t-1} ×Homecountry		Omitted baseline		
Quarter_{t+0} ×Homecountry	0.024	(0.146)	0.008	(0.014)
Quarter_{t+1} ×Homecountry	0.188	(0.137)	0.015	(0.012)
Quarter_{t+2} ×Homecountry	0.229*	(0.128)	0.008	(0.011)
Quarter_{t+3} ×Homecountry	0.276	(0.172)	0.017	(0.016)
Quarter_{t-3} ×Time-based×Homecountry	-0.671	(0.462)	-0.066	(0.056)
Quarter_{t-2} ×Time-based×Homecountry	0.469	(0.413)	0.025	(0.047)
Quarter_{t-1} ×Time-based×Homecountry		Omitted baseline		
Quarter_{t+0} ×Time-based×Homecountry	-0.995**	(0.439)	-0.133***	(0.051)
Quarter_{t+1} ×Time-based×Homecountry	-0.776**	(0.385)	-0.105**	(0.044)
Quarter_{t+2} ×Time-based×Homecountry	-1.413***	(0.394)	-0.144***	(0.043)
Quarter_{t+3} ×Time-based×Homecountry	-1.896***	(0.608)	-0.167***	(0.051)
Same language	0.413***	(0.026)	0.018***	(0.001)
Same currency	0.063***	(0.021)	0.004***	(0.001)
Same time zone	0.251***	(0.057)	0.025***	(0.005)
Log bid price	-1.737***	(0.018)	-0.090***	(0.001)
Log milestone percentage	-0.065***	(0.016)	-0.003***	(0.001)
Log review count	0.096***	(0.007)	0.005***	(0.000)
Log avg rating	0.103***	(0.009)	0.003***	(0.000)
Log bid order rank	-0.313***	(0.014)	-0.016***	(0.001)
Preferred freelancer	0.471***	(0.018)	0.026***	(0.001)
Country dummy	Yes		Yes	
Project fixed effects	Yes		Yes	
Observations	371,968		371,968	
R-squared	0.490		0.044	
LogLik	-47,537			
AIC	95,168			
BIC	95,677			
Number of projects	23,943		23,943	

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price to the model. d) The results are highly consistent if we limit our sample to the sample matched by the CEM method. e) Robust standard errors clustered by projects are reported in parentheses. f) R-squared in the logit model is calculated based on the maximum likelihood R-squared. g)* p<0.1, ** p<0.05, *** p<0.01.



Note: This graph is plotted by quarter-level relative time parameters. The dash vertical line denotes the quarter when *Freelancer* first implemented the monitoring system (from October 2013 to January 2014). Error bars denote the 90% confidence intervals calculated based on clustered standard errors.

Figure 6. Coefficients of the Relative Time DID Estimates of the Treatment Effect

2.5.2.3. Heterogeneity of Home Bias

As the model-free evidence in Table 21 shows, the preference of employers toward workers from home countries tends to vary across countries. However, since the model-free evidence might be confounded by multiple differences across worker countries (e.g., lower labor costs), the heterogeneities of home bias require formal analyses. Therefore, we further investigate how the strength of home bias may be associated with the traditional values, diversity, and population size of the employer's home country with the conditional logit model and the LPM.

Based on the results reported in Tables 28-31, we find that employers show home bias before and after the introduction of the monitoring system. Specifically, we focus on the post-treatment sample because the statistical homebias is expected to be minimal with the accessibility of more detailed individual information. We find that the strength of employers' home bias 1) is positively related to the traditional values of the employer country, such that employers from countries with a strong nationalistic outlook among residents have a stronger home bias (Inglehart and Welzel 2010); 2) is negatively associated with the country-specific cultural diversity (Fearon and Laitin 2003); 3) is negatively related to the population size of the employer country after the introduction of the monitoring system. Moreover, the results remain highly consistent after controlling for the competition among workers from the same country. The marginal effects of these interactions are visualized in Figure 7. Based on the marginal effect estimation, one standard deviation increase from the average traditional values is associated with a 0.0006 increase in the probability of being hired. Similarly, an increase of one standard deviation from both the average cultural diversity and the average log-transformed population size is associated with a decrease in the probability of being hired of 0.0005 and 0.0006, respectively. Overall, the heterogeneous effects indicate that employers' home bias is influenced by context-contingent factors, which seems to suggest that at least a portion of employers' home bias is related to their stereotyped liking and preferences.

Table 28. Estimation Results of Heterogeneity with Conditional Logit Model (Pre-monitoring)

Sample	Pre-monitoring sample			
Homecountry	0.837***(0.129)	0.697***(0.165)	2.206***(0.724)	2.416***(0.615)
Trad value×Homecountry	0.478***(0.163)			0.352***(0.111)
Cultural diversity×Homecountry		-0.659* (0.374)		-1.086***(0.363)
Log popu size×Homecountry			-0.140** (0.057)	-0.113** (0.057)
Same language	0.363***(0.052)	0.355***(0.048)	0.355***(0.048)	0.433***(0.035)
Same currency	0.022 (0.042)	0.039 (0.039)	0.036 (0.039)	0.068** (0.027)
Same time zone	0.350***(0.109)	0.249** (0.103)	0.240** (0.104)	0.286***(0.070)
Log bid price	-1.927***(0.037)	-1.926***(0.034)	-1.925***(0.034)	-1.656***(0.024)

Log milestone percentage	-0.054* (0.028)	-0.038 (0.026)	-0.038 (0.026)	-0.078***(0.023)
Log review count	0.093***(0.014)	0.089***(0.013)	0.090***(0.013)	0.099***(0.009)
Log avg rating	0.134***(0.018)	0.141***(0.017)	0.141***(0.017)	0.089***(0.012)
Log bid order rank	-0.418***(0.025)	-0.413***(0.023)	-0.413***(0.023)	-0.267***(0.018)
Preferred freelancer	0.508***(0.035)	0.496***(0.032)	0.496***(0.032)	0.475***(0.024)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	95,030	110,794	110,794	95,030
R-squared	0.535	0.534	0.534	0.536
LogLik	-12,685	-14,767	-14,765	-12678
AIC	25,441	29,604	29,601	25,429
BIC	25,772	29,940	29,938	25,779
Number of projects	6,937	8,032	8,032	6,937

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared is calculated based on the maximum likelihood R-squared; f) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table 29. Estimation Results of Heterogeneity with Linear Probability Model

(Pre-monitoring)

Sample	Pre-monitoring sample			
Homecountry	0.077***(0.015)	0.056***(0.014)	0.229***(0.073)	0.259** (0.113)
Trad value×Homecountry	0.059***(0.022)			0.057** (0.025)
Cultural diversity×Homecountry		-0.057* (0.033)		-0.078 (0.063)
Log popu size×Homecountry			-0.015***(0.006)	-0.012 (0.011)
Same language	0.014***(0.003)	0.014***(0.002)	0.014***(0.002)	0.015***(0.003)
Same currency	0.004 (0.002)	0.005** (0.002)	0.004* (0.002)	0.004 (0.002)
Same time zone	0.035***(0.009)	0.026***(0.009)	0.024***(0.009)	0.035***(0.009)
Log bid price	-0.108***(0.002)	-0.107***(0.002)	-0.107***(0.002)	-0.108***(0.002)
Log milestone percentage	-0.004** (0.002)	-0.003** (0.001)	-0.003**(0.001)	-0.004** (0.002)
Log review count	0.005***(0.001)	0.005***(0.001)	0.005***(0.001)	0.005***(0.001)
Log avg rating	0.003***(0.001)	0.004***(0.001)	0.004***(0.001)	0.003***(0.001)
Log bid order rank	-0.025***(0.002)	-0.024***(0.001)	-0.024***(0.001)	-0.025***(0.002)
Preferred freelancer	0.031***(0.003)	0.030***(0.003)	0.030***(0.003)	0.031***(0.003)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	95,030	110,794	110,794	95,030
R-squared	0.056	0.055	0.055	0.056
Number of projects	6,937	8,032	8,032	6,937

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table 30. Estimation Results of Heterogeneity with Conditional Logit Model

(Post-monitoring)

Sample	Post-monitoring sample			
Homecountry	0.598***(0.084)	0.881***(0.108)	2.932***(0.467)	2.416***(0.615)
Trad value×Homecountry	0.414***(0.109)			0.352***(0.111)
Cultural diversity×Homecountry		-1.369***(0.244)		-1.086***(0.363)
Log popu size×Homecountry			-0.205***(0.037)	-0.113** (0.057)
Same language	0.442***(0.035)	0.440***(0.032)	0.441***(0.032)	0.433***(0.035)
Same currency	0.068** (0.027)	0.067***(0.026)	0.066**(0.025)	0.068** (0.027)
Same time zone	0.295***(0.071)	0.228***(0.068)	0.217***(0.068)	0.286***(0.070)
Log bid price	-1.654***(0.024)	-1.654***(0.022)	-1.654***(0.022)	-1.656***(0.024)
Log milestone percentage	-0.077***(0.023)	-0.091***(0.021)	-0.091***(0.021)	-0.078***(0.023)
Log review count	0.099***(0.009)	0.101***(0.009)	0.102***(0.009)	0.099***(0.009)
Log avg rating	0.089***(0.012)	0.087***(0.011)	0.086***(0.011)	0.089***(0.012)
Log bid order rank	-0.266***(0.018)	-0.263***(0.017)	-0.262***(0.017)	-0.267***(0.018)
Preferred freelancer	0.474***(0.024)	0.466***(0.022)	0.465***(0.022)	0.475***(0.024)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	217,135	261,174	261,174	217,135
R-squared	0.472	0.473	0.490	0.473
LogLik	-27,586	-32,671	-32,672	-27564
AIC	55,241	65,413	65,414	55,202
BIC	55,601	65,779	65,780	55,582
Number of projects	13,557	15,911	15,911	13,557

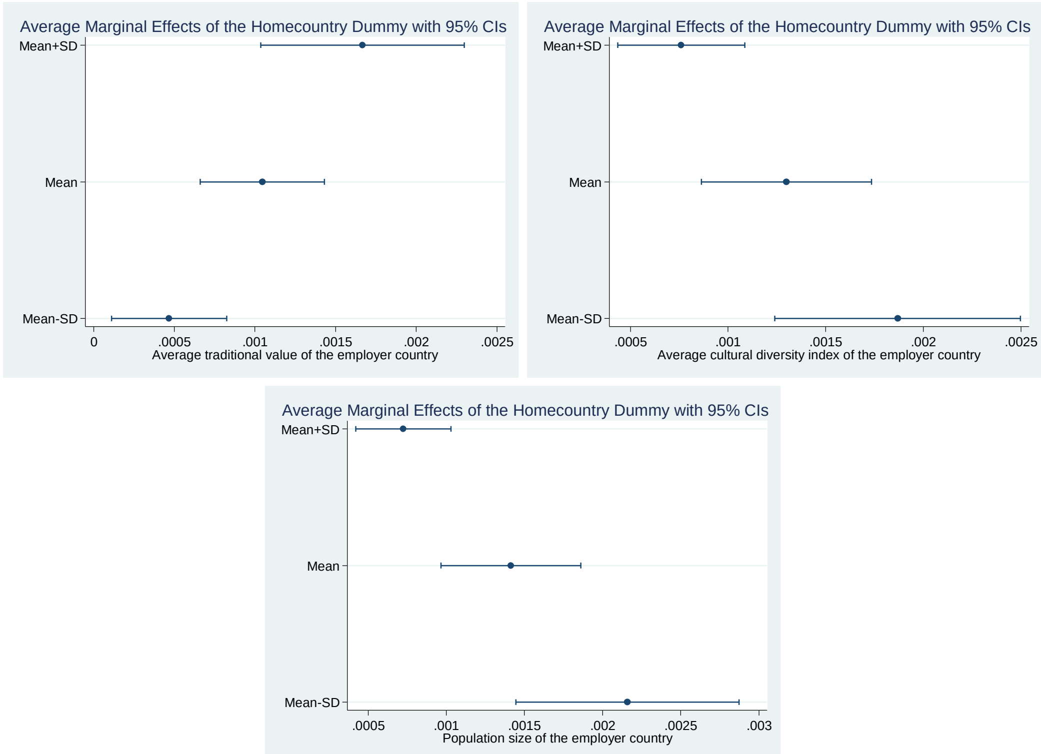
Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared is calculated based on the maximum likelihood R-squared; f) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table 31. Estimation Results of Heterogeneity with Linear Probability Model

(Post-monitoring)

Sample	Post-monitoring sample			
Homecountry	0.053***(0.009)	0.079***(0.010)	0.271***(0.045)	0.217***(0.059)
Trad value×Homecountry	0.039***(0.012)			0.036***(0.013)
Cultural diversity×Homecountry		-0.129***(0.022)		-0.116***(0.034)
Log popu size×Homecountry			-0.019***(0.004)	-0.009* (0.005)
Same language	0.019***(0.001)	0.019***(0.001)	0.019***(0.001)	0.019***(0.001)
Same currency	0.004** (0.001)	0.003** (0.001)	0.003**(0.001)	0.004** (0.001)
Same time zone	0.030***(0.006)	0.024***(0.006)	0.023***(0.006)	0.030***(0.006)
Log bid price	-0.085***(0.001)	-0.083***(0.001)	-0.083***(0.001)	-0.085***(0.001)
Log milestone percentage	-0.003***(0.001)	-0.004***(0.001)	-0.004***(0.001)	-0.003***(0.001)
Log review count	0.005***(0.000)	0.005***(0.000)	0.005***(0.000)	0.005***(0.000)
Log avg rating	0.003***(0.001)	0.003***(0.000)	0.002***(0.000)	0.003***(0.001)
Log bid order rank	-0.013***(0.001)	-0.012***(0.001)	-0.012***(0.001)	-0.013***(0.001)
Preferred freelancer	0.026***(0.002)	0.025***(0.001)	0.025***(0.001)	0.026***(0.002)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	217,135	261,174	261,174	217,135
R-squared	0.041	0.040	0.040	0.042
Number of projects	13,557	15,911	15,911	13,557

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) * $p<0.1$, ** $p<0.05$, *** $p<0.01$.



Note: All the marginal plots are generated with the assumption that fixed effect is zero.

Figure 7. Heterogeneity of Marginal Effects of the Homecountry Dummy Based on Logit Model

2.6. Robustness Checks regarding the Mechanisms for Home Bias

In the research design for quasi-natural experiments, the identification challenges mainly come from two sources: 1) the violation of the differences-in-differences model assumption (Autor 2003; Abadie 2005), and 2) selection on observables (Abadie 2005; Altonji et al. 2005) and selection on unobservables (Dale et al. 2002; Altonji et al. 2005; Oster 2016). Next, we will explain how we deal with each of these identification challenges.

First, to further examine whether the treatment group (time-based projects) and the control group (fixed-price projects) follow the parallel trend assumption, we provide two related tests in our robustness check section. In our main analysis, we show that the control group follows the same trend as the treatment group prior to the introduction of the monitoring system. This result suggests that the control group serves as a valid counterfactual of the treatment group. Furthermore, we conduct a placebo test by testing whether our DID estimate is robust to alternative variance-covariance specifications. Again, the results shown in Section 2.6.3 lend support to the significant treatment effect on employers' home bias.

Second, we address the *selection on observables* issue by matching fixed-price projects with time-based projects, and we tackle the issue of *selection on unobservables* by employing instrumental variables (IV) and estimating the coefficient stability. Specifically, we employ both the CEM and PSM algorithms to generate matched samples and still find similar treatment effects on home bias in Appendix F. Further, we employ the IV method to estimate the local average treatment effect (LATE) in Section 2.6.1 and further demonstrate that our results are unlikely to be explained by selection on unobservables in Section 2.6.2. Therefore, our results are robust to the omitted variable bias.

2.6.1. Instrumental Variables Analysis

With the introduction of the monitoring system that enables us to explore the mechanism of home bias through a quasi-natural experiment, there are two potential endogeneity concerns when unobserved characteristics might be correlated with employers' hiring decisions: 1) there may be potential unobserved variables that affect both employers' preference for contract type and hiring preference; 2) bidders may infer employers' preferences and determine their bid prices accordingly.[33] To alleviate these concerns, we employ both the 2SLS and the conditional logit model with the control function method to estimate the local average treatment effect (Angrist et al. 1996; Angrist 2004)—i.e., the causal effect of the monitoring system on the margin of home bias.

First, regarding the potential endogenous contract type, we need instruments that are associated with the contract type but not with the error term (ε_{ij}) for employers' hiring decisions. Since employers' time-invariant preference for time-based contracts is nested within the project-specific fixed-effects, we only need to instrument employers' time-varying preference for time-based contracts. Specifically, we employ two instruments: 1) the "residual" type IV (Dobbie et al. 2018; Arnold et al. 2018): a residualized, leave-out employers' tendency to use time-based contracts that accounts for selection bias (Dahl et al. 2014; Dobbie et al. 2018); 2) the "Hausman" type IV (Hausman et al. 1994; Schneide 2010; Ghose et al. 2012). Since the employers' tendency to use time-based contracts might be related to employers' characteristics, the simple leave-out mean of employers' contract types or and the lagged term of employers' contract types (the employer's last-used contract type) may still be influenced by the selection concern. Therefore, we predict employers' contract type choices with employer-specific fixed-effect, observable project characteristics—i.e., time dummies—and calculate the residuals. Here, the residuals may capture the specific unobserved project characteristics or the match

[33] Since employers' hiring preference is estimated given the bid prices submitted by workers, this will not be a concern if we are only interested in the extent to which the statistical discrimination mechanism could explain employers' home bias. We instrument for it to better estimate the monetary value of employers' home bias.

between the employers' monitoring cost function and the specific project characteristics. Since these residuals capture the idiosyncratic features of the specific project, they are unlikely to correlate with the hiring decisions of a different project in the LPM with project-specific fixed effects or with the conditional logit setting. Therefore, we use the leave-out mean of residuals as the instrument for contract type. Moreover, following the previous literature (Hausman et al. 1994; Hausman 1996; Schneide 2010; Ghose et al. 2012), we use the percentage of time-based contract types in other rivals' projects which are submitted in the same week of the focal project as the instrument for employers' contract type choice (Ghose et al. 2012). This variable is theoretically a valid instrument for the following reasons. First, similar to Hausman's (1996) approach, after controlling for the project-specific fixed effect, worker characteristics, and bid characteristics, we assume that employer-specific utility from hiring worker *j* for project *i* is uncorrelated with other employers' contract type choices (exogeneity). Second, the contract type choice of the employer for project *i* is correlated with other employers' contract type choices within the same week because the employer's choice is likely influenced by the common platform environment or the preference of other employers' contract types given the common labor supply force (relevance).

With regard to the second potential endogenous variable, workers' bid price, we take advantage of the exogenous "cost-shifter" from the supply side (the exogenous variation in the exchange rate of different currencies relative to the U.S. dollar) as the instrumental variable (Nevo 2000; Hong and Pavlou 2017). Since the exchange rate of local currencies against the U.S. dollar is negatively correlated with the actual purchasing power of the final payment and workers' reservation wage,[34] we expect the exogenous variation of normalized exchange rates of various currencies would be negatively related to workers' bid prices.

To estimate the treatment effect according to both the linear probability model and the conditional logit model, we employ the 2SLS and control function method, respectively. First, in the linear probability setting, there are three endogenous variables in the LPM, including $Home\text{country}_{ij} \times \widehat{Time\_based}_i$, $After_i \times Home\text{country}_{ij} \times \widehat{Time\_based}_i$ and bidders' bid price $\widehat{Bidprice}_{ij}$. Since bidder j's bidding decision ($Home\text{country}_{ij}$) and the platform's decision to implement monitoring systems ($After_i$) are exogenous factors, we can have four instrumental variables by assuming that the error term in predicting project i's contract type (η_{ki}) is uncorrelated with the error term in employers' hiring decision equation (ε_{ij}), including $Home\text{country}_{ij} \times residual\_leaveout\_timebased_i$, $Home\text{country}_{ij} \times perc\_timebased\_week_i$, $After_i \times Home\text{country}_{ij} \times residual\_leaveout\_timebased_i$ and $After_i \times Home\text{country}_{ij} \times perc\_timebased\_week_i$. In the logit model estimation scenario, we employ the control function to estimate the treatment effect. Specifically, following the previous literature (Petrin and Train 2009; Polyakova 2016), we assume the linearity and additive separability of the unobservables, and include the residuals of the first stage of the 2SLS model into the control function of the conditional logit model.

Based on the first-stage of 2SLS, we find that all the four instrument variables related to contract types are significantly correlated with time-based contract dummy.

[34] The final contract price is measured in the currency set by the employer. To rule out the unobserved workers' preference for currencies, we rule out those projects whose currencies are not U.S. dollar.

Specifically, when there are more time-based projects posted by other employers within a certain week, it is less likely that employers will choose time-based projects. Moreover, the higher residualized, leave-out employers' tendency to use time-based contracts implies that employers will be more likely to choose time-based projects. Additionally, when the exchange rate of local currencies against the U.S. dollar is higher, workers tend to bid lower prices. We also conduct the weak identification test and find that the Cragg-Donald Wald F statistic is 78.97 and the Kleibergen-Paap Wald rk F statistic is 65.48, which are above the Stock and Yogo (2005) suggested cut-off values. Therefore, the week instrument issue is not a concern in our study. Our over-identification test statistic (Hansen J statistic is 0.552, Chi-sq(2) *p*-value is 0.7589) suggests that the failure to reject the null hypothesis that the instruments used are exogenous.

Additionally, we adjust the standard error for the conditional logit model with the control function by calculating the bootstrap error clustering at the project level. As Table 32 suggests, employers reduce their home bias after the introduction of the monitoring system. The results are highly consistent after employing the IV estimation on the matched sample.

Table 32. IV Estimation of Employers' Home Bias in the Quasi-Natural Experiment

Sample Model	Full sample (1) 2SLS		Full sample (2) Logit Control Function	
Homecountry	0.010	(0.008)	0.103	(0.104)
Time-based×Homecountry	0.108***	(0.029)	0.989**	(0.275)
After×Homecountry	0.016*	(0.009)	0.243*	(0.113)
Time-based×After×Homecountry	-0.156***	(0.037)	-1.437**	(0.370)
Same language	-0.123***	(0.033)	0.307**	(0.095)
Same currency	0.030***	(0.006)	0.050	(0.032)
Same time zone	0.005**	(0.002)	0.238**	(0.077)
Log bid price	0.024***	(0.006)	-1.737**	(0.031)
Log milestone percentage	-0.005***	(0.002)	-0.090**	(0.020)
Log review count	0.005***	(0.000)	0.103**	(0.008)
Log avg rating	0.003***	(0.001)	0.098**	(0.011)

Log bid order rank	-0.020***	(0.005)	-0.310**	(0.015)
Preferred freelancer	0.027***	(0.002)	0.464**	(0.023)
Bidder country dummy	Yes		Yes	
Project fixed effects	Yes		Yes	
Observations	263,752[35]		234,736	
R-squared	0.041		0.466	
LogLik			-33,257	
AIC			66,595	
BIC			67,009	
Number of projects	20,255		17,903	

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) the significance levels and standard errors of all the coefficients in the control function are calculated after 1,000 bootstrap cycles. f) R-squared in the logit model is calculated based on the maximum likelihood R-squared; g) * p<0.1, ** p<0.05, *** p<0.01.

2.6.2. Selection on Unobservables

In the previous subsection, we employed the CEM approach and instrumental variables to address the concern that unobserved project features may drive both the contract type and employers' home bias. It is possible that the average treatment effects estimated based on the matched sample, or the local average treatment effects estimated with instrumental variables, may tend to overestimate the treatment effect because the average causal effect for the "complier" group estimated by the IV approach may be higher than the average treatment effect for the whole population (Angrist et al. 1996). In order to assess the robustness of our findings among the more general population, we employ another method to alleviate the omitted variable bias concern—that is, *selection*

[35] Because we obtain the monthly short-term interest rate of each worker country from the OECD data website: https://data.oecd.org/interest/short-term-interest-rates.htm, those workers whose home countries' interest rate information is not provided by this website are ruled out of the IV estimation.

on unobservables (Altonji et al. 2005; Oster 2016). Though it is difficult to consider and control for all potential unobservables that simultaneously correlate with the project contract type and the employers' hiring decisions in a systematic way, we use the selection on unobservables approach to estimate the sensitivity of our estimated home bias and treatment effect in terms of the selection issue (Altonji et al. 2005; Oster 2016). This method offers a way of assessing the degree of selection on unobservables based on the degree of selection on observables. In other words, we can evaluate the sensitivity of our estimated coefficients to the relevant unobservables by inferring from the movement of coefficients and R-squared.

By employing the selection on unobservables approach, we assess the sensitivity of our findings to the omitted variable bias and generate the lower/upper bounds of reported home bias and treatment effect. First, we evaluate the possibility that the estimated home bias and treatment effect of monitoring may be driven by selection on unobservables. Following the previous literature (Dale and Krueger 2002; Altonji et al. 2005; Oster 2016), we assess the minimum of selection on unobservables which can explain away the home bias and treatment effect found in the previous analysis. Specifically, we use parameter δ to denote the ratio of selection on unobservables to selection on observables. We find that the selection on unobservables needs to be at least twice as strong as the selection of observables ($|\delta_1| \geq 3.414$ for the reported home bias; $|\delta_2| \geq 2.601$ for the estimated treatment effect) in order for the selection bias to completely explain away the observed home bias and treatment effect of monitoring. Further, since the previous literature (Dale and Krueger 2002; Altonji et al. 2005; Oster 2016) suggests that, at most, equal selection ($|\delta| \leq 1$) on unobservables and observables is a well-accepted assumption, our results ($|\delta_1| \geq 3.414$ and $|\delta_2| \geq 2.601$) imply that it is very unlikely that the estimated home bias and treatment effect are driven by the omitted variable bias. The result with "$|\phi| \geq 1$" indicates that the estimate is not sensitive to the effect of unobservables. Specifically, $|\phi| = 1$ refers to an extreme scenario—observables are randomly selected from all factors that can affect the outcome (Altonji et al. 2005). In our paper, we select observables that tend to have substantial explanatory power in contract choice and hiring choice—such as project-level fixed effects, project size, employers' reputation and experience, and bidders' reputation. Therefore, we have strong reasons for believing that the selection on unobservables is weaker than the selection on observables (Altonji et al. 2005). In other words, our results suggest that the reported home bias and the treatment effect of monitoring are not sensitive to the effect of unobservables.

Second, as suggested by Oster (2016), we construct the lower bound and the upper bound of the estimated home bias and treatment effect, using $\delta = 0$ (when there is no selection on unobservables) and $\delta = 1$ (when the amount of selection on unobservables is equal to that of selection on observables) as the boundaries. As shown in Table 33, we find that both the upper bound and the lower bound of the coefficient of the "Homecountry" dummy are greater than zero, and that those bounds of the estimated treatment effect are negative. This result lends support to the existence of home bias, especially the statistical home bias.

Table 33. Sensitivity Analysis of the Coefficients using the Method of Selection on Unobservables

$\theta_{Homecoutnry}$	Home bias	$\theta_{Homecoutnry \times After \times Time-Based}$	Treatment effect
$|\delta_1|$ when $\theta_{Homecoutnry}$ =0	$|\delta_1|$ = 3.414	$|\delta_2|$ when $\theta_{Homecoutnry \times After \times Time-Based}$ =0	$|\delta_2|$ =2.601

Lower bound	0.023	Lower bound	-0.258
Upper bound	0.027	Upper bound	-0.098

2.6.3. Placebo Tests

To reinforce the credibility of our main findings, we conduct two placebo tests. First, we assign a placebo intervention to the middle of our pretreatment period (August 1, 2013) and check whether a pretreatment tendency existed prior to the actual introduction of the monitoring system. As Table 34 shows, the interaction between the “pseudo” $After$ dummy and the $Time\_based$ dummy is insignificant. Second, following Abadie et al. (2015), we randomly reassign the treatment to the projects and run the same model with the placebo treatment assignment. We replicate the analysis 1000 times and generate the distribution of the placebo treatment effects based on the “pseudo” treatments of the monitoring intervention (Greenwood and Wattal 2015; Ranganathan and Benson 2017).[36] By comparing the actual estimated coefficient of three key covariates to the whole distribution of “placebo” treatment effects (Table 35), we find that it would be very unlikely to observe a similar size of treatment effect by chance, which implies that our findings are robust to alternative variance-covariance specifications.

Table 34. Estimation Results based on the “Placebo” Treatment Time

Sample	Full sample				Matched sample			
Model	(1) Logit		(2) LPM		(3) Logit		(4) LPM	
Homecountry	0.196	(0.135)	0.010	(0.009)	0.197	(0.255)	0.024	(0.024)
Time-based×Homecountry	0.734**	(0.287)	0.078**	(0.031)	0.786**	(0.367)	0.078*	(0.043)
Afterplacebo×Homecountry	0.183	(0.152)	0.020*	(0.011)	0.338	(0.295)	0.042	(0.031)
Time-based×Afterplacebo ×Homecountry	0.111	(0.377)	0.023	(0.041)	-0.161	(0.483)	-0.006	(0.057)

[36] We employ both the LPM with project-specific fixed effects and the conditional logit model to estimate the placebo treatment effects and find highly consistent results. We report conditional logit model results in Table 35.

Same language	0.349***	(0.049)	0.014***	(0.002)	0.328***	(0.086)	0.020***	(0.007)
Same currency	0.045	(0.040)	0.005**	(0.002)	0.077	(0.069)	0.010	(0.006)
Same time zone	0.240**	(0.107)	0.024***	(0.009)	0.480***	(0.174)	0.065***	(0.021)
Log bid price	-1.938***	(0.035)	-0.108***	(0.002)	-2.011***	(0.066)	-0.164***	(0.005)
Log milestone percentage	-0.034	(0.027)	-0.003*	(0.002)	-0.163***	(0.045)	-0.016***	(0.004)
Log review count	0.089***	(0.013)	0.005***	(0.001)	0.081***	(0.022)	0.007***	(0.002)
Log avg rating	0.142***	(0.017)	0.004***	(0.001)	0.123***	(0.028)	0.006**	(0.002)
Log bid order rank	-0.414***	(0.024)	-0.024***	(0.001)	-0.379***	(0.046)	-0.035***	(0.004)
Preferred freelancer	0.507***	(0.033)	0.031***	(0.003)	0.559***	(0.061)	0.057***	(0.008)
Log avg_country_rating	0.196	(0.135)	0.010	(0.009)	0.197	(0.255)	0.024	(0.024)
Log country_bidder	0.734**	(0.287)	0.078**	(0.031)	0.786**	(0.367)	0.078*	(0.043)
Bidder country dummy	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	104,653		104,653		21,797		21,797	
R-squared	0.539		0.056		0.515		0.087	
LogLik	-13,933				-3,931			
AIC	27,940				7,936			
BIC	28,294				8,232			
Number of projects	7,602		7,602		2,625		2,625	

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses. e) R-squared in the logit model is calculated based on the maximum likelihood R-squared. f) * p<0.1, ** p<0.05, *** p<0.01.

Table 35. Placebo Effects of Random Assignment Model

Variables	Homecountry	Time-based×Homecountry	Time-based ×After ×Homecountry
μ of placebo β	0.338	0.001	-0.004
σ of placebo β	0.020	0.215	0.270
Estimated β	0.218	0.735	-1.120
Replication	1000	1000	1000
Z-score	-15.289	3.520	-4.113
P-value	p<0.001	p<0.001	p<0.001

Notes: a) The result of the placebo test based on the full sample is reported. b) All the bids which are submitted by bidders having previous collaboration experience with the employer before are dropped. Moreover, our sample is only limited to projects with only one winner. c) Conditional logit model with project FE and bidder country dummies are included in the model; LPM provides consistent results. d) Robust standard errors clustered by projects are reported in parentheses.

2.6.4. Other Robustness Checks

To further check the robustness of our conclusions, we conduct additional analyses that are reported in the Appendices. First, we rerun the model with a shorter-range observational window (six months before and after) and still find consistent results based on the full sample and matched sample (see Appendix D). Second, to ensure the workers are comparable and similar between the treatment group and the control group, we limit our sample to the bids submitted by those workers who bid on both fixed-price and time-based projects. The results of the restricted sample are still highly consistent with our main findings (see Appendix E). Third, to show the robustness of our findings, we employ an alternative matching algorithm, PSM, to regenerate a matching sample, and our results are still consistent (see Appendix F). Fourth, similar to the robustness checks concerning the existence of home bias, we further control for the time-varying or project specific contingent factors influencing home bias. In particular, we control for the number of workers and the average rating of workers from each country within the employer's specific consideration set and the country-month two-way fixed effect (see Appendix G). On the whole, all our robustness checks are consistent with our main findings.

2.7. Additional Analysis: Is the Impact of Monitoring a Function of Task Routineness?

We further explore whether the impact of the introduction of monitoring systems on home bias varies in a predictable way across job subcategories with different routines task levels. Assessing whether the impact of monitoring varies in a theoretically predictable way improves our confidence in our findings and provides a better understanding of the mechanism of employers' home bias and the channel for reducing their statistical home bias. Specifically, as suggested by Ranganathan and Benson (2017), the effectiveness of monitoring tends to be stronger for routine tasks. Therefore, we expect that employers' home bias will be less influenced by the introduction of monitoring systems when projects are less routine or more abstract.

Following Autor et al. (2003), we employ the routine task-intensity (RTI) as a comprehensive proxy of the measure of task routineness. RTI refers to the ratio of routine task inputs to nonroutine task inputs (such as highly manual and abstract tasks) in each occupation. We calculate it based on the following equation (7). And it increases as routine tasks become more important for the specific occupation. In order to calculate RTI, we first search for the routine task input index, the manual task input index and the abstract task input index for each project based on the definition of its specific project subcategory. Specifically, we match the project subcategory list to the standard occupational classification (SOC) system,[37] and then find the corresponding abstract, manual, routine task inputs for each occupation.

$$\mathrm{RTI}_k = \ln\left(T_{k,1980}^{Routine}\right) - \ln\left(T_{k,1980}^{Manual}\right) - \ln\left(T_{k,1980}^{Abstract}\right) \tag{7}$$

Moreover, to show the heterogeneous treatment effect of monitoring, we rerun the DID model for those projects with high RTI and those with low RTI. Here, "high RTI" means the RTI of that job subcategory is higher than the mean RTI of the overall

[37] To find the corresponding SOC code for each job subcategory, we put the subcategory name into the search field of the O*Net database (https://www.onetonline.org/find/quick?s=), and search for related occupations within the "IT, software & website" area. Further, we manually verify whether the definition of the occupation is consistent with the definition of the job subcategory. Based on SOC codes, we further find the corresponding 2000 ACS Occupation Codes (OCC) and then 1990 ACS Occupation Codes (OCC). Next, based on the 1990OCC codes, we find the corresponding occupational task data from Autor and Dorn (2013), which includes the abstract, manual, routine task inputs for each occupation.

software category and “low RTI” indicates the opposite scenario. As Table 36 shows, for those highly nonroutine and abstract projects, monitoring does not significantly affect employers’ home bias. However, for projects with high routine-task intensity, monitoring significantly decreases employers’ home bias. The result based on the full-sample also confirms the heterogeneous treatment effect of monitoring on projects with different RTI levels.

Table 36. Estimation Results of Low RTI Sample versus High RTI Sample

Sample	Low RTI sample, before		Low RTI sample, after		High RTI sample before		High RTI sample after	
Model	(1) LPM		(2) LPM		(3) LPM		(4) LPM	
Homecountry	0.039**	(0.016)	0.021*	(0.011)	0.012***	(0.005)	0.022***	(0.003)
Time-based×Homecountry	0.035	(0.045)	0.025	(0.031)	0.064***	(0.014)	-0.029***	(0.007)
Same language	0.026***	(0.006)	0.024***	(0.003)	0.006***	(0.001)	0.011***	(0.001)
Same currency	-0.001	(0.005)	0.008**	(0.003)	0.004**	(0.001)	0.001	(0.001)
Same time zone	0.025	(0.020)	0.027**	(0.014)	0.016***	(0.006)	0.016***	(0.003)
Log bid price	-0.106***	(0.005)	-0.087***	(0.003)	-0.060***	(0.001)	-0.051***	(0.001)
Log milestone percentage	-0.002	(0.004)	-0.002	(0.003)	-0.002*	(0.001)	-0.003***	(0.001)
Log review count	0.006***	(0.002)	0.005***	(0.001)	0.003***	(0.000)	0.003***	(0.000)
Log avg rating	0.003	(0.002)	0.003**	(0.001)	0.002***	(0.000)	0.001***	(0.000)
Log bid order rank	-0.025***	(0.003)	-0.018***	(0.002)	-0.013***	(0.001)	-0.007***	(0.001)
Preferred freelancer	0.038***	(0.007)	0.025***	(0.004)	0.017***	(0.002)	0.016***	(0.001)
Bidder country dummy	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	20,999		48,471		162,829		351,339	
R-squared	0.060		0.049		0.031		0.024	
Number of projects	2,358		4,924		12,423		22,699	

Notes: a) The results are estimated based on the matched sample with the CEM approach. Our results are highly consistent if we estimate the model based on the full sample. b) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. c) Log(bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log(bidder tenure) instead of Log(count rating). d) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker’s fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as “dual-type workers”) (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. e) Robust standard errors clustered by projects are reported in parentheses. f) R-squared in the logit model is calculated based on the maximum likelihood R-squared. g) * p<0.1, ** p<0.05, *** p<0.01.

Table 37. Estimation Results of Low RTI Sample versus High RTI Sample After Matching

Sample	Low RTI sample				High RTI sample			
Model	(1) Logit		(2) LPM		(3) Logit		(4) LPM	
Homecountry	0.317	(0.294)	0.048	(0.043)	0.299*	(0.158)	0.040**	(0.017)
After×Homecountry	-0.050	(0.325)	-0.010	(0.048)	0.244	(0.172)	0.020	(0.019)
Time-based×Homecountry	0.568	(0.485)	0.079	(0.071)	0.747***	(0.252)	0.081***	(0.030)
Time-based×After ×Homecountry	-0.089	(0.616)	-0.030	(0.088)	-1.368***	(0.327)	-0.150***	(0.037)
Same language	0.421***	(0.093)	0.037***	(0.008)	0.369***	(0.049)	0.023***	(0.003)
Same currency	0.043	(0.075)	0.006	(0.008)	0.092**	(0.040)	0.008**	(0.003)
Same time zone	0.450**	(0.200)	0.067**	(0.029)	0.385***	(0.102)	0.048***	(0.011)
Log bid price	-1.870***	(0.072)	-0.171***	(0.006)	-1.844***	(0.036)	-0.131***	(0.002)
Log milestone percentage	-0.213***	(0.058)	-0.021***	(0.005)	-0.195***	(0.029)	-0.015***	(0.002)
Log review count	0.106***	(0.025)	0.012***	(0.003)	0.085***	(0.013)	0.006***	(0.001)
Log avg rating	0.088***	(0.030)	0.002	(0.003)	0.079***	(0.016)	0.004***	(0.001)
Log bid order rank	-0.380***	(0.052)	-0.036***	(0.006)	-0.317***	(0.028)	-0.024***	(0.002)
Preferred freelancer	0.443***	(0.073)	0.045***	(0.009)	0.459***	(0.035)	0.040***	(0.004)
Bidder country dummy	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	15,647		15,647		71,193		71,193	
R-squared	0.501		0.104		0.485		0.065	
LogLik	-2,982				-11,813			
AIC	6,038				23,699			
BIC	6,321				24,039			
Number of projects	2,153		2,153		6,875		6,875	

Notes: a) The results are estimated based on the matched sample with the CEM approach. The results are highly consistent if we estimate the model based on the full sample. b) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. c) Log(bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log(bidder tenure) instead of Log(count rating). d) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. e) Robust standard errors clustered by projects are reported in parentheses. f) R-squared in the logit model is calculated based on the maximum likelihood R-squared. g) * p<0.1, ** p<0.05, *** p<0.01.

2.8. General Discussion

Using a large-scale proprietary data set from a leading online labor market, this paper examines the existence of and mechanisms associated with home bias, a type of discrimination based on the closeness or similarity between the employer and worker.

First, our estimation results suggest that a home bias against foreign workers does exist in online employment. Second, based on a quasi-natural experiment following the introduction of a monitoring system for time-based projects on an online employment platform, we explore the change in employers' preference for local workers after the introduction of the monitoring system. The introduction of the monitoring system reduces the ex post information asymmetry regarding hidden actions and lowers employers' home bias. Based on the different predictions from taste-based discrimination and statistical discrimination mechanisms, we suggest that employers' home bias is primarily driven by statistical discrimination. Third, to further examine the existence of a taste-based mechanism, we investigate whether the potential heterogeneity of home bias across different employer countries is consistent with our predictions based on the literature on in-group favoritism. We find that employers from countries with high traditional values, lower cultural diversity, and smaller size of country population, tend to hold a stronger home bias, which lends support to the existence of taste-based home bias. Our study suggests that home bias in online employment is driven by both statistical and taste-based mechanisms.

This paper makes several important contributions to the related literature. First, our study is among the first to formally examine the existence and mechanisms of home bias in an employment setting. Despite the rich literature on home bias in equity and trade, the existence of home bias in employment settings remains an open question. We extend prior studies by controlling for additional confounding factors and examining whether home bias exists in employment settings. We control for the effect of common language, time zone, and currency, which can lead to the overestimation of home bias. Also, unlike prior studies (e.g., Åslund et al.'s (2014)), which fail to control for potential supply-side bias due to data limitations, we estimate employers' preferences with precise data regarding their consideration sets and final choices. Second, we contribute to the emerging stream of research on discrimination studies that are based on the dynamic approach and quasi-experiments (Goldin and Rouse 2000; Rubineau and Kang 2011). By now, the most popular method applied in discrimination studies is the correspondence study design (Bertrand and Duflo 2017). However, Bertand and Duflo (2017) express the concern that correspondence studies suffer from the following limitation: With fictitious similar applicants, correspondence studies can only be used to test the discrimination at the first stage based on callback rate, but not at later stages (i.e., interview and hiring). But quasi-natural experiments (e.g. Goldin and Rouse 2000) can provide more information regarding discrimination in the full hiring process, especially regarding the final hiring decisions. As Rubineau and Kang (2012, P662) state, *"The key to identifying statistical discrimination lies in scrutinizing its dynamic rather than static predictions."* By empirically examining the consistency between the predictions based on the statistical discrimination assumption and the actual observed result, we establish a robust causal relationship between information change and the dynamic change in discrimination, and subsequently identify the mechanism of discrimination in a real-world setting. Most notably, the information changes employed by the existing literature (such as gender or criminal background information) do not differentiate between the ex ante worker capabilities and the ex post effort of workers (Goldin and Rouse 2000; Doleac and Hansen 2016). Our approach directly operates on information changes that relate only to worker effort and isolates the ex post hidden action mechanism that causes employer uncertainty. Thus, our study suggests that the asymmetric information about worker effort plays an important role in explaining employers' home bias and potentially other discriminatory behavior, such as gender discrimination (Sterling and Fernandez 2018). Third, our paper also contributes to the recent research on the discrimination phenomenon in the gig economy by suggesting the role of monitoring systems for attenuating discrimination behavior. It has been found that racial discrimination exists in the on-demand sharing economy, such as in accommodation sharing (Edelman et al. 2017) and on-demand e-hailing services (Ge et al. 2016). We contribute to this stream of discrimination literature by showing that discrimination based on the similarity of the home country is a type of discrimination prevalent in the gig economy. Our study further suggests that monitoring systems, can increase the fairness of the gig economy without reducing the market efficiency.

Meanwhile, we acknowledge several limitations of our study, opening up avenues for follow-up studies. First, we note that our sample is limited to projects within the IT category. It is possible that taste-based preferences may play a more important role in certain design or data analysis tasks that tend to have a lower RTI than IT projects. Second, we conduct our study in the online employment setting and our findings may not be directly generalizable to offline labor markets and other online platforms. In offline labor markets, for example, employment contracts for long-term collaboration may serve per se as an effective incentive mechanism to motivate workers' effort, thus reducing the risk of shirking due to ex post information asymmetry and reducing the reliance on statistical home bias. Furthermore, our study also provides insightful implications for other online platforms. In particular, our study suggests the coexistence of the statistical and taste-based home bias in online labor markets. Future research can explore whether home bias on other platforms can also be partially attributed to the statistical discrimination or taste-based mechanisms.

Using a unique large-scale data set from one of the prevalent online labor markets, we investigate the existence and mechanism of home bias in the online employment setting for the first time. Moreover, owing to the quasi-natural experiment design, we conclude that the introduction of the monitoring system substantially lowers employers' home bias. Our results suggest that when information is limited, employers might employ statistical discrimination and prefer to hire workers from their home countries. This kind of discrimination could be alleviated without the loss of market efficiency if platforms implement monitoring systems and reduce ex post information asymmetry. Overall, our study offers strong implications for the marketplace design by underscoring the value of monitoring systems in increasing the fairness and efficiency of online platforms.

CHAPTER 3

GENDER DIFFERENCES IN AVOIDANCE OF MONITORING AND GENDER WAGE GAP IN ONLINE GIG ECONOMY

We explore whether there exists a gender wage gap in the gig economy and examine to what degree gender differences in job application strategy could account for the gap. With a large-scale dataset from a leading online labor market, we show that females only earn around 81.4% of the hourly wage of their male counterparts. We further investigate three main aspects of job application strategy, namely bid timing, job selection, and avoidance of monitoring. After matching males with females using the propensity score matching method, we find that females tend to bid later and prefer jobs with a lower budget. In particular, the observed gender difference in bid timing can explain 7.6% of the difference in hourly wage, which could account for 41% of the gender wage gap (i.e. 18.6%) observed by us. Moreover, taking advantage of a quasi-natural experiment wherein the platform rolled out the monitoring system, we find that females are less willing to bid for monitored jobs than males. To further quantify the economic value of the gender difference in avoidance of monitoring, we run a field experiment on Amazon Mechanical Turk (AMT), which suggests that females tend to have a higher willingness to pay (WTP) for the avoidance of monitoring. The gender difference in WTP for the avoidance of monitoring can explain 8.1% of the difference in hourly wage, namely, 44% of the observed gender wage gap. Overall, our study reveals the important role of job application strategies in the persistent gender wage gap.

Keywords: gender wage gap, job application strategy, gig economy, quasi-natural experiment

3.1. Introduction

There is a growing literature documenting the gender wage gap in the labor market. As the previous literature suggests, while employers exhibit less discrimination against females in the hiring process, females still earn a lower wage than males in the same positions (Goldin and Rouse 2000; Kuhn and Shen 2012). Therefore, an emerging school of thought is that "*gender wage gap is caused mainly by women's choice, not discrimination*."[38] In the same vein, more studies are suggesting that the gender wage gap is partially attributable to motherhood penalty, gender differences in career plans, or preferences for non-monetary attributes in a job, such as flexibility (Mas and Pallais 2017), work-from-home (Mas and Pallais 2017), and workplace competitiveness (Niederle and Vesterlund 2007,2011; Flory et al. 2014).

Given that gender pay gap is a longstanding phenomenon, the new gig economy, which is thriving in many industries (e.g., ridesharing, temporary lodging, outsourcing), seems to provide an efficient way to reduce the gender wage gap. Owing to the market openness and the emphasis on spot-market based short-term employment in gig economy, many scholars predict that gender differences in career development, as well as the gender wage gap, will be smaller in the gig economy (e.g. Goldin 2014). Specifically, it's predicted but not empirically confirmed that workers tend to have more flexible work hours and locations in the gig economy, making motherhood penalty less likely to become an obstacle to career development. As the booming gig economy is projected to

[38] https://www.campusreform.org/?ID=9827
This report reads "*The American Association of University Women (AAUW) has finally admitted that the "gender pay gap" is caused primarily by women's choices, not discrimination. In fact, the AAUW's own research suggests that only about 7% of the observed pay gap can be attributed to discrimination, with simple economic factors accounting for the remainder.*"

comprise a large portion of the future of work[39], it is imperative to examine whether there is a gender wage gap in the gig economy. Moreover, given that females tend to have much more flexibility in gig work than in the traditional workplace, the gig economy also provides us an unprecedented opportunity to explore factors other than motherhood penalty or compensation differential for flexibility that might influence the gender wage gap, which is critical to policy prescription to further narrow the gender wage gap. In particular, as the gig economy, especially the online gig economy platform, enables workers from all over the world to seek a wider diversity of remote jobs posted by employers from various countries, this provides a unique setting to dig into potential gender differences in job application strategy, which is hitherto little explored. To this end, with the advantage of the availability of large-scale micro-level granular data in the online gig economy (Hong and Pavlou 2017), we attempt to explore several critical aspects of gender differences in job application strategy and their impact on the gender wage gap. Specifically, we examine whether there are gender differences in avoidance of monitoring and to what extent such gender differences can account for the gender wage gap in the gig economy (if any). In particular, we are interested in the following questions:

1) *Is there a gender wage gap in the gig economy?*

2) *Whether and to what extent the gender wage gap is driven by gender differences in avoidance of monitoring?*

[39] "Independent work: Choice, necessity, and the gig economy" https://www.mckinsey.com/featured-insights/employment-and-growth/independent-work-choice-necessity-and-the-gig-economy

In this paper, we take advantage of a comprehensive dataset from a leading gig economy platform, a quasi-natural experiment, and a supporting field experiment to answer the above research questions. First, we infer workers' genders based on their profile images[40] with human labeling. We find that there is a gender wage gap based on the historical hiring data. This result is consistent when we control for various workers' characteristics. We find that, on average, females earn 81.4% of the hourly wage of their male counterparts.

Second, we recover each worker's consideration set of jobs based on our comprehensive dataset. It is notable that although there are a few studies analyzing employers' preference for workers in the online labor market (Chan and Wang 2017), workers' behaviors are yet to be explored, e.g., gender differences in avoidance of monitoring, likely due to the lack of data regarding workers' consideration sets. In our study, because the platform restricts workers to only bid for jobs with at least one skill requirement matched with their own skill sets, we are able to reconstruct the whole list of contemporaneous jobs which were available for workers to bid. Based on the recovered consideration sets, we find that females prefer to bid jobs without monitoring based on a quasi-natural experiment and that females tend to have a higher willingness to pay (WTP) for the avoidance of monitoring through a field experiment. Specifically, hinging on the exogenous shock when the platform implemented the monitoring system on all the hourly jobs, we observe that females are less willing to bid for monitored jobs based on a difference-in-differences (DID) estimation and difference-in-difference-in-differences (DDD) estimation. In particular, we take fixed-price jobs as the control group and

[40] We find consistent results when we use the first name of workers to infer gender.

incorporate the interaction of the monitoring treatment with contractual forms across jobs and the worker's gender. To further quantify the economic value of the gender difference in WTP for the avoidance of monitoring than males, we conduct a randomized field experiment on Amazon Mechanical Turk (AMT). We randomly provide two hourly jobs for workers on AMT (Turkers), in which only one requires monitoring. We also randomize the wage premium offered by the job with the monitoring requirement, which varies between $-2 and $5. The result suggests that females have a higher WTP for the avoidance of monitoring than males, which lends support to our finding from the quasi-natural experiment. In fact, the gender difference in WTP for the avoidance of monitoring can explain roughly 8.13% of the hourly wage, which is equivalent to 43.71% of the observed gender wage gap.

Our paper contributes to three related strands of literature. First, our study contributes to the literature on gender wage gap (Blau and Kahn 2017; Mas and Pallais 2017; Wiswall and Zafar 2015, 2017) by providing new explanations for the gender wage gap that are unrelated to gender discrimination, i.e., gender differences in the avoidance of monitoring. Second, this study also contributes to the literature on the online labor market by showing the importance of workers' job preferences. Although employers' preference of workers has been recently explored (Chan and Wang 2017; Hong and Pavlou 2017), there is little research exploring the preference from the supply side (i.e., workers' preference for jobs). Our study advances the previous literature on online labor markets by documenting gender differences in job application strategy and how they may explain the gender wage gap. Lastly, this paper also contributes to the literature on compensation differential (Bonhomme and Jolivet 2009; Mas and Pallais 2017). Our study takes advantage of both a quasi-natural experiment and a field experiment to show the gender difference in WTP for the avoidance of monitoring, a non-wage aspect which has hardly been explored in the compensation differential literature.

3.2. Theoretical Background

3.2.1 Gender Wage Gap

The gender wage gap has been established long ago. According to the estimates from the Institute for Women's Policy Research, women are still paid 20% less than their male counterparts in the same position in 2015[41]. In fact, based on the statistics from the Census Bureau, the female-to-male earnings ratio, which has not been updated since 2007, is 0.805[42]. The persistence of gender pay gap is difficult to explain because the explanations for the wage gap provided by the previous literature, such as gender differences in occupation choice and preference for flexibility, seem to be less relevant in today's society, especially in gig economy. For instance, even in the IT industry, which tends to provide workers with a relatively flexible work schedule, women are still systematically paid less than men and are promoted more slowly.

There is a large body of literature exploring the causes of the gender wage gap. First, discrimination from the demand side has found to be one of the key explanations. Regarding the mechanisms of discrimination, the findings from the previous literature are still mixed. Some studies suggest that only statistical discrimination (Gupta and Smith

---

[41] https://iwpr.org/publications/the-gender-wage-gap-2015-annual-earnings-differences-by-gender-race-and-ethnicity

[42] https://iwpr.org/wp-content/uploads/2017/09/C459_9.11.17_Gender-Wage-Gap-2016-data-update.pdf

2012; Castillo et al. 2013) contributes to the gender wage gap while some other papers lend support to the taste-based discrimination explanation (Goldin and Rouse 2000; Marom et al. 2016). Second, a growing literature suggests that gender differences in worker confidence and compensation differential also help to account for the gender wage gap, which will be discussed below.

3.2.2 Gender Wage Gap and the Gig Economy

The emerging gig economy is expected to help to decrease the gender wage gap by increasing work schedule flexibility and reducing the motherhood penalty (Goldin and Rouse 2000). According to a report from Hyperwallet, a gig-work payment platform, 86% of females believe that they can earn equal pay to males in the gig economy, while only 41% of females think so in the traditional workplace[43]. Moreover, Chan and Wang (2017) found that employers prefer to hire female workers in feminine-typed jobs and even gender-neutral jobs in an online gig economy platform, which suggests that discrimination is less likely to be a serious obstacle to females. That being said, females are still found to pay an invisible cost owing to gender differences in preference-based characteristics, such as females' lower willingness to work more hours in the car-hailing service industry when the hourly wage is high (Cook et al. 2018). However, it is still unknown whether females still earn less than males in online gig economy platforms wherein the hourly wage is less dependent on the working time and location. Given that the effect of discrimination in online gig economy platforms has already been explored in the prior study (Chan and Wang 2017), in this paper, we will focus on examining key

[43] "The Future of Gig Work is Female," available at www.hyperwallet.com

factors that contribute to the gender wage gap other than the gender discrimination in online gig economy platforms.

3.2.3 Gender Wage Gap and Gender differences in Confidence and Avoiding Uncertainty

Gender differences in worker confidence and avoidance of uncertainty are found to be key contributing factors to the gender wage gap. First, gender differences in confidence may lead to gender differences in competitiveness and the wage gap. For instance, Niederle and Vesterlund (2007) identified gender differences in competitiveness in a lab experiment. They found that although there are no significant gender differences in performance, women show less preference for the competitive tournament. Further, they explained that gender differences in competitiveness were caused by the differences in confidence and attitudes toward competition instead of gender differences in risk aversion (Niederle and Vesterlund 2011). In line with this study, Flory et al. (2014) found that gender differences in preferences for uncertainty and competition jointly drive gender differences in job-entry choices. Moreover, some contingent factors influence the size of gender differences, including whether the job involves teamwork or has overt gender associations, and his/her age, etc. (Flory et al. 2014). Inspired by this stream of literature, we expect that there might exist gender differences in job application strategy due to gender differences in confidence and avoidance of uncertainty suggested in the previous literature and explore the subsequent impact on the gender wage gap.

3.2.4 Gender Wage Gap and Gender difference in Compensation Differential/Preference

Meanwhile, the gender wage gap can also be caused by compensation differential. Research in this space has focused on how gender differences in preference for various non-wage job characteristics may account for the gender wage gap. This is also referred to as cross-gender compensation differential. Cross-gender compensation differential means that females and males may have different WTP for different nonwage job attributes (Arnould and Nichols 1983), which subsequently leads to their different job choices and wages. For example, gender differences in work flexibility have been found to help to explain the gender wage gap. Marini and Fan (1997) found that gender differences in worker characteristics (including occupational aspirations, job-related skills, and credentials) explain roughly 30% of the gender wage gap. More recently, Wiswall and Zafar (2017) found that females show a stronger preference for work flexibility and job stability whereas males prefer potential earnings growth. Moreover, such gender differences in preference also indirectly lead to gender differences in college major choices and subsequent income (Wiswall and Zafar 2017). In the same vein, Mas and Pallais (2017) find a significant gender difference in WTP for working from home but an insignificant gender difference in WTP for scheduling flexibility in their large-scale field experiment. Given that most jobs in the gig economy tend to have high scheduling flexibility and allow working-from-home, we focus on potential gender differences in WTP for the avoidance of monitoring, which has become increasingly important with the popularity of online, IT-enabled monitoring systems.

3.3. Research Methodology

3.3.1 Research Framework

Gender wage gap is a longstanding phenomenon. There is a large body of literature exploring the causes of the gender wage gap from the demand (employer) side, e.g., the gender hiring bias (Goldin and Rouse 2000; Kuhn and Shen 2012) or promotion bias in the workplace (Budig and England 2001; Anderson et al. 2002). We propose an alternative explanation from the supply (worker) side, i.e., gender differences in avoidance of monitoring (Figure 8).

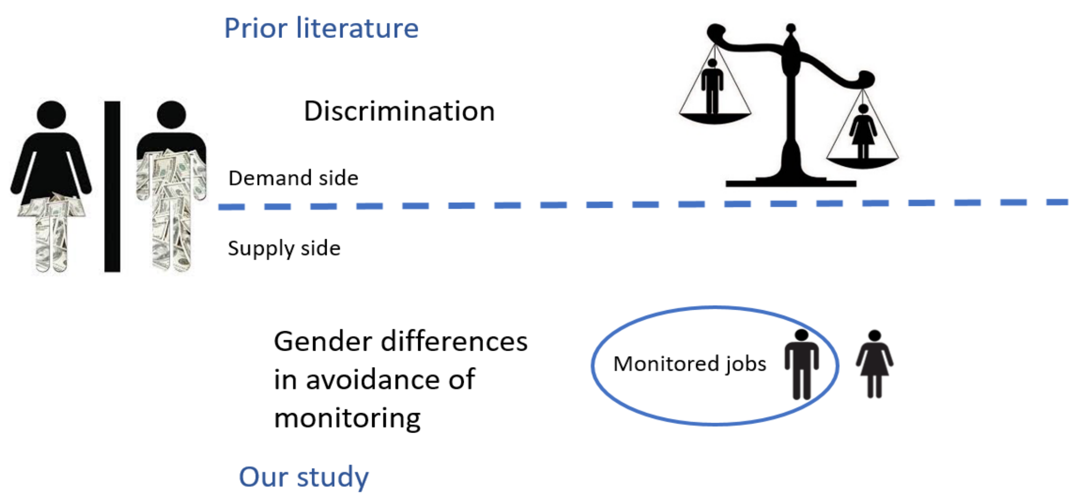


Figure 8. Research Questions of this Study

Specifically, we first explore whether a gender wage gap exists in the gig economy. Then, we examine whether there are gender differences in job application strategy and how these differences may contribute to the gender wage gap. Table 38 summarizes our research framework. Next, we explain them in turn.

Table 38. Research Agenda and Empirical Identification Strategy

Key concepts	Research questions	Data source	Empirical model
Gender wage gap	Is there a gender wage gap in the gig economy?	Observational data from Freelancer.com	Fixed-effect model with the worker country and month two-way fixed effects
Gender differences in job application strategy	Do females prefer to bid jobs without monitoring?	Observational data from Freelancer.com with a quasi-natural experiment	Propensity score matching between female and male workers; propensity score matching between fixed-price and hourly jobs; Differences-in-Differences and triple differences estimator based on a quasi-natural experiment (with the control

	A field experiment on AMT	for the fixed effects on the consideration set) logit model

3.3.2 Observational Data

Our data for the main analysis were collected from Freelancer.com, one of the leading online gig economy platforms. In Freelancer, all jobs are awarded based on a reverse auction mechanism wherein employers post jobs first and workers bid for those jobs of their interest. When posting the job, the employer provides the project title, project description, required skills, and project budget. To reduce the potential confounding effects of various job requirements, we limit our sample to the “IT, software & website” category, which is the most popular category in Freelancer, in terms of number of jobs and transactions. Given that we attempt to explore the gender difference in job preference, we focus on jobs that can be done remotely.[44] Our final dataset includes a majority of the IT jobs posted in Freelancer between October 2013 and November 2014. Users of Feelancer.com come from over 100 countries. Before making the first bid on the platform, they are required to list those skills they acquired and upload their profile images. Our dataset includes various job- and user- level characteristics as reported in Table 39.

Table 39. Definitions and Summary Statistics of Related Variables

Variable	Variable definition	Mean	SD	Min	Median	Max
Bid	A dummy variable (0,1); =1 if the worker bids for the job or not	0.005	0.069	0.000	0.000	1.000
Employer_norating	A dummy variable (0,1); =1 if the employer has not any reviews written by previous workers hired by him/her	0.080	0.272	0.000	0.000	1.000

[44] Local jobs only accounts for less than 0.01% of all the jobs posted on the platform.

Log(employer_review)	The number of reviews for the employers entered by previous workers (log-transformed)	3.255	1.207	0.000	3.178	6.071
Employer_rating	The average overall ratings for the employer (in the range of [0-5])	4.884	0.470	0.000	4.996	5.000
Log(budget)	The maximum of the hourly wage for this job set by the employer (log-transformed)	2.300	1.275	0.000	2.197	5.994
Female	A dummy variable (0,1); =1 if the worker is a female	0.146	0.353	0.000	0.000	1.000
Log(title_length)	Number of characters in the job title (log-transformed)	3.472	0.412	2.398	3.466	4.796
Log(desc_length)	Number of characters in the job description (log-transformed)	5.312	0.918	2.773	5.242	8.101
Log(skills_count)	Number of necessary skills listed by the employer (log-transformed)	1.509	0.315	0.693	1.609	1.792
Featured_job	A dummy variable; =1 if this job is featured prominently on the job catalog page	0.005	0.071	0.000	0.000	1.000
NDA	A dummy variable; =1 if this job requires NDA (Non-Disclosure Agreement)	0.000	0.012	0.000	0.000	1.000
Log(remain_days)	Number of days between the bid date and the date when the auction is closed (log-transformed)	1.361	0.594	0.000	1.386	3.135
Log(auction_duration)	Number of days wherein the job is open for bid (log-transformed)	2.089	0.106	0.693	2.079	3.135
Log(hourly_wage)	The hourly wage of the awarded bid (log-transformed)	2.247	0 .983	0 .693	2.303	7.600

Notes: a) Due to the overdispersion in the "*log(budget)*" variable, we dropped the outliers based on 99^{th} percentile cutoff; b) Given that the consideration set of each worker's bid decision is very large (close to 200 jobs), the mean value of "bid" is relatively low. If "bid "is equal to 0.005, it means that the worker chooses one job to bid among all the 200 jobs for which s/he could bid. c)We label the gender of each user based on his/her profile image. we hired student workers and MTurk workers to label workers' genders based on their profile images. For each image, there are at least two persons to label them. For those images we could not identify their genders based on the profile images or there is some inconsistency between the labels of the same image, we label their genders as "unknown". We find consistent results when we use the first name of workers to infer gender.

3.3.3 Construction of Workers' Consideration Sets

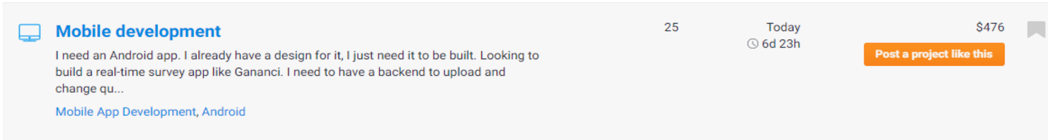
To explore workers' job application strategy, we compile the whole dataset and reconstruct each worker's consideration set based on the platform regulation policy

(Figure 9). In general, there are two main restrictions imposed on the workers' job selection. First, the job should be open for bids at that time. Second, the worker has at least one skill matched with the skill requirements of the project. As such, we take advantage of the comprehensiveness of our dataset, which includes both the detailed auction duration, job skill requirements and all workers' skill sets, and further construct workers' consideration sets as follows:

To begin with, we first find a list of active workers and their bids during our observation window. Specifically, the worker j is considered as an active worker at day t only if s/he bid at least once on that day. Further, we find all the IT jobs which were open to bidding when s/he made the bid decision. Lastly, we check whether the worker has at least one skill matched with the job skill requirements to finalize his/her consideration set. According to the platform regulation, the worker could bid for all the jobs satisfying with these two restrictions. In essence, we examine female and male workers' revealed preference for job characteristics based on the actual bid decisions they made, given all the open jobs fitted with their skills.[45]

[45] To ensure that workers can bid for all the jobs in the consideration set, we only limit to those jobs which do not use sealed auctions and are described in English. Additionally, since the "hireme" jobs are posted for targeted workers, we also rule out these jobs from our sample.

When browsing a job without any matched skills:



When browsing a job with at least one matched skill, the “bid now” button is displayed:

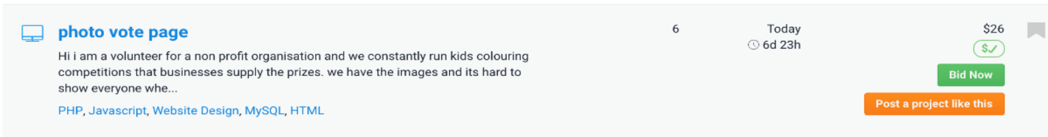


Figure 9. A Screenshot from Freelancer.com for Jobs with/without Matched Skills

3.3.4 Experimental Data for the Analysis of Gender Differences in WTP

We conduct a field experiment on Amazon Mechanical Turk. In total, we have recruited 300 participants, among which 276 have completed the experiment. The experiment follows a between-subject design with 15 treatments by varying wage premium of the job with monitoring (Table 40). For each treatment group, participants will be provided a short introduction of the monitoring system and two job options shown randomly. The order of available job positions is also randomized to reduce the potential concern of the anchoring effect of the first option (Strack and Mussweiler 1997). When the participant is choosing between two hourly choices with different wages, his/her WTP to avoid monitoring can only be driven by his/her distaste for monitoring. To ensure the internal validity of randomization, we ensure the comparability of participants in different treatment groups across various wage premium cases.

Table 40. Treatment Design of the Field Experiment

Single choice question	Job option design	Wage premium of the job with monitoring
	An hourly job without monitoring or an hourly job with monitoring	Wage premiums ∈[-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5]

3.4. Measures and Models

3.4.1 Measuring the Gender Wage Gap

To measure the gender wage gap in the gig economy, we explore whether female workers systematically earn a lower hourly wage in all the hourly job transactions made on Freelancer.com. Specifically, we use the log-transformed hourly wage based on those awarded bids as the dependent variable and the $Female_i$ dummy is the key independent variable of our interest. We employ the following linear regression model to estimate the effect of gender on hourly wage:

$$\log(wage_{it}) = \alpha_{it} + \beta_1 Female_i + controls(\text{Worker}_i) + \varepsilon_{it} \quad (8)$$

According to the literature on the gender wage gap, we attempt to calculate the adjusted gender wage gap which needs to be corrected for differences in payment due to country or occupation, differences in period, and differences in human capital (Freeman and Oostendorp 2000; O'Neill 2003; Oostendorp 2004; Blau and Kahn 2017). To adjust for the country or period differences, we control for the worker country and month two-way fixed effects and cluster standard errors accordingly. Given that our observations come from the same type of jobs (online IT jobs), the occupation differences among our sample is relatively small. To correct for human capital, we assume that the worker's rating, experience, and tenure can serve as good proxies for the worker's human capital. Accordingly, we further add the control for various time-varying covariates regarding worker i, such as the number of reviews entered by previous employers, the average rating, the tenure measured in the month unit, the primary language set by worker i, verification measures and the length of the tagline on worker i's profile, etc. A significant coefficient of the dummy $Female_i$ suggests that there is a gender wage gap in the gig economy.

3.4.2 Gender Differences in Avoidance of Monitoring in the Quasi-Natural Experiment

We estimate gender differences in avoidance of monitoring by taking advantage of the exogenous shock when the platform introduced the monitoring system for workers in hourly jobs. Meanwhile, this monitoring system is not available for fixed-price jobs. The monitoring system can automatically take screenshots of the workers' laptops and share those with employers.[46]

By exploiting the different availability of monitoring across two types of jobs, we employ the DID estimation and the DDD estimation to check whether females are less willing to work under monitoring. First, in the DID estimation framework, we are interested in the coefficient of the interaction term (β_2), which denotes that whether female workers are less willing to bid for hourly jobs after the introduction of monitoring systems by taking the fixed-price jobs as the control group. Here, we employ the propensity score matching to control for the selection on observables among job types and only use highly comparable fixed-price jobs as the counterfactual.

[46] To protect the privacy of workers, workers can delete a few screenshots if they don't feel comfortable to be seen other others. However, the short time interval logged along with these sensitive screenshots may not be guaranteed to get paid.

$$Worker_{it}\_choice\_on\_job_j = \alpha_{it} + \beta_1 Hourly_j + \beta_2 After_t \times Hourly_j + controls(Job_j) + \varepsilon_{itj} \quad (9)$$

Further, we also match male and female workers based on their reputation and various profile information. Based on the comparable females and males within the matched sample, we explore the difference in the treatment effect of monitoring on males and females in term of job preference:

$$Worker_{it}\_choice\_on\_job_j = \alpha_{it} + \beta_1 Hourly_j + \beta_2 Female_i \times Hourly_j + \beta_3 After_t \times Hourly_j + \beta_4 Female_i \times After_t \times Hourly_j + controls(Job_j) + \varepsilon_{itj} \quad (10)$$

We compare the difference in preference for hourly jobs for males before and after the introduction of monitoring systems (DD_{male}) with the difference in preference for hourly jobs for females before and after (DD_{female}). In other words, $\mathrm{DDD} = \mathrm{DD}_{female}$-$\mathrm{DD}_{male}$, which will be captured by the coefficient of the triple interaction (β_4). Note that, compared to the traditional DDD estimation, the term $After_{it} \times Female_i$ is omitted because it is nested in the time-varying fixed effect α_{it}. If we observe a significantly negative coefficient of DDD (β_4), it suggests that females tend to have a stronger avoidance of monitoring than their male counterparts, which means that females prefer to bid for jobs without monitoring.

3.4.3 Gender Differences in WTP for Avoidance of Monitoring in the Field Experiment

Following the modeling framework of Mas and Pallais (2017), the probability of workers choosing a job with monitoring when the wage premium of the monitored job is $\Delta W = W_{monitoring} - W_{none}$ as follows:

$$P(Y_i = 1) = \Pr\left(W_{monitoring,i} + \delta X - Z_i > W_{none,i} + \delta X\right) = \Pr(\Delta W_i - Z_i > 0) \quad (11)$$

where X is a vector of various job characteristics other than the hourly wage and the monitoring condition, Z_i is the disutility for worker i if s/he works under monitoring. $\Delta W_i - Z_i$ is the utility of worker i choosing a monitored job with the utility for a job without monitoring normalized to zero. Further, we can get the likelihood function of the above probability is ln $\prod_i \left(P(Y_i = 1)\right)^{Y_i} \left(1 - P(Y_i = 1)\right)^{1-Y_i}$ and use the maximum likelihood estimation to identify μ and σ, which represent the mean and standard deviation of the distribution of WTP, respectively. In our robustness checks, we also check if our result is consistent when the probit model is employed.

3.5. Results Regarding the Existence of Gender Wage Gap

As Table 41 shows, the coefficient of the "female" dummy is significantly negative, which suggests that females systematically earn a lower wage than males. We control for workers' reputation and experience in Model 1 and additional characteristics of their profiles in Model 2. The result is highly consistent. Based on the result of Model 2, on average, females can only earn 81.4% of the wage of their male counterparts, which is very close to the gender wage gap found in the general fulltime job in the US (i.e., 80%)[47].

Table 41. Evidence of Gender Wage Gap in the Gig Economy

Dependent variable: log(hourly_wage)				
Model	(1)		(2)	
Job type	hourly		hourly	
Female	-0.208**	(0.099)	-0.205**	(0.101)
Log(bidder_rating)	0.021	(0.094)	0.020	(0.091)
Log(bidder_reviews)	0.055*	(0.029)	0.055*	(0.032)

[47] Based on the report from American Association of University Women (AAUW), females working in full-time jobs usually get paid 80% of the wage earned by males (source: https://www.aauw.org/research/the-simple-truth-about-the-gender-pay-gap/).

Bidder_primary_language_eng	0.020	(0.252)	0.020	(0.248)
Log(bidder_tenure_month)	0.150***	(0.036)	0.162***	(0.037)
Log(tagline_length)			0.085	(0.059)
Identity_verified			-0.050	(0.068)
Phone_verified			0.024	(0.297)
Preferred_freelancer			0.041	(0.078)
Log(milestone_percentage)			0.009	(0.034)
Observations	1,300		1,288	
R-squared	0.047		0.053	
Bidder country dummy	yes		yes	
Month fixed effects FE	yes		yes	

Notes: a) Here, log(tagline_length) denotes the length of the tagline on worker i's profile, which can be considered as the short headline of the self-introduction on the profile page; b) Robust standard errors clustered by the bidder country and month two-way fixed effects are reported in parentheses; c) * p<0.1, ** p<0.05, *** p<0.01.

3.6. Results Regarding Gender Differences in Job Application

3.6.1 Sample Matching

To ensure the similarity between females and males, we employ the propensity score matching method to match females with males, and match fixed-price jobs with hourly jobs. As suggested in Table 42, we match males and females based on their reputation, experience, verification, primary language, primary currency and whether they have the "preferred freelancer" badges, most of which serve as proxies for their human capital and the credibility of their identity or work. The balance check result and the density distribution of the propensity score suggest that after the matching, females and males are highly comparable in most of the observable characteristics displayed to the employers.

Table 42. Balance Check for Propensity Score Matching between Females and Males

Variable	Sample	Mean		%bias	% reduced |bias|	t-test	
		Female	Male			t	p>|t|
Registration_month	Unmatched	635.910	629.490	27.400		7.070	0.000
	Matched	635.910	636.200	-1.200	95.500	-0.300	0.768
Bidder_reviews	Unmatched	14.587	15.650	-1.500		-0.430	0.669
	Matched	14.587	15.798	-1.700	-14.000	-0.260	0.795
Bidder_rating	Unmatched	2.082	2.207	-5.200		-1.440	0.149
	Matched	2.082	1.953	5.400	-3.800	1.150	0.251

Payment_verified	Unmatched	0.006	0.009	-4.600		-1.770	0.076
	Matched	0.006	0.006	0.000	100.000	0.000	1.000
Identity_verified	Unmatched	0.004	0.011	-7.800		-2.870	0.004
	Matched	0.004	0.004	0.500	93.100	0.240	0.808
Phone_verified	Unmatched	0.001	0.001	-1.600	-70.600	-0.610	0.542
	Matched	0.001	0.002	-2.700		-0.820	0.414
Preferred_freelancer	Unmatched	4.916	4.940	-8.300	95.500	-3.340	0.001
	Matched	4.916	4.917	-0.400		-0.100	0.918

Notes: 1) Results of Nearest Neighbor (1) Matching Method without replace are presented. 2) Due to length limitation, results regarding some variables are omitted, including the "primary_language_Eng" and "primary_currency_US" dummies. The means of both variables are not significantly different across groups.

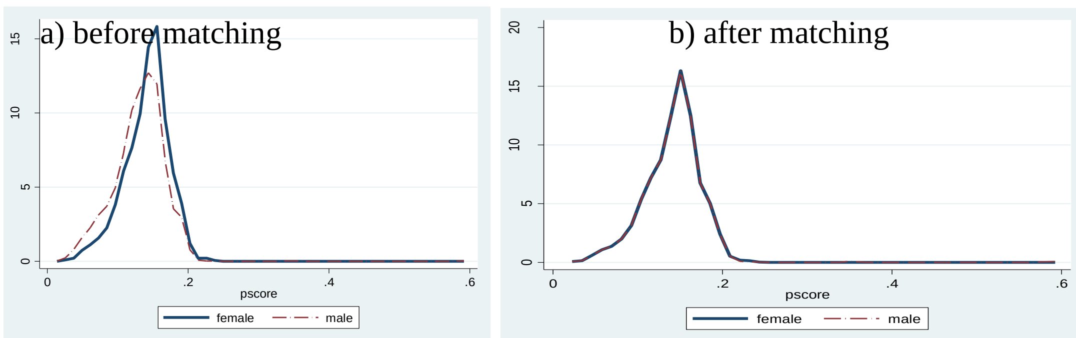


Figure 10. Density of Propensity Score of Being Female (before and after Matching)

Similarly, given that we use fixed-price jobs as the control group in our analysis for the quasi-natural experiment wherein Freelancer.com rolled out its monitoring system for hourly jobs, we deploy the propensity score matching method to match two types of jobs. We match two types of jobs based on various characteristics which are suggested to be correlated with the contract type by the previous literature (Banerjee and Duflo 2000; Gopal and Sivaramakrishnan 2008; Lin et al. 2016; Roels et al. 2010), such as employers' reputation, project size (the total amount of project), the complexity of job (the number of skills required), whether employers have a concrete idea of the job (the length of job title and description), and so on. As suggested by Figure 10 and Table 43, the density of the propensity score and the mean of all observable covariates are highly comparable in two groups after matching.

Table 43. Balance Check for PSM Between Fixed-Price Jobs and Hourly Jobs

Variable	Sample	Mean		%bias	% reduced |bias|	t-test	
		Hourly	Fixed-price			t	p>|t|
Employer_developed	Unmatched	0.381	0.762	-83.300		-35.970	0.000
	Matched	0.381	0.379	0.600	99.300	0.190	0.852
Title_length	Unmatched	31.968	35.346	-20.400		-8.100	0.000
	Matched	31.968	32.164	-1.200	94.200	-0.420	0.674
Job_desc_length	Unmatched	270.970	455.200	-41.600		-15.500	0.000
	Matched	270.970	285.680	-3.300	92.000	-1.490	0.136
Employer_tenure_ month	Unmatched	25.497	32.570	-25.300	95.200	-9.630	0.000
	Matched	25.497	25.158	1.200		0.460	0.645
Employer_ rating	Unmatched	4.916	4.940	-8.300	95.500	-3.340	0.001
	Matched	4.916	4.917	-0.400		-0.100	0.918
Primary_language_ Eng	Unmatched	0.947	0.902	17.000	85.700	6.570	0.000
	Matched	0.947	0.953	-2.400		-0.970	0.331
Auction_duration	Unmatched	7.996	7.646	7.400	99.500	2.480	0.013
	Matched	7.996	7.994	0.000		0.010	0.995
Total_paid_amount of_project (/$100)	Unmatched	1.764	2.752	-6.700	67.100	-2.410	0.016
	Matched	1.764	2.090	-2.200		-1.240	0.214
Skills_count	Unmatched	3.530	3.317	15.300	79.100	6.410	0.000
	Matched	3.530	3.486	3.200		1.070	0.287

Notes: 1) Results of Nearest Neighbor (1) Matching Method without replace are presented. 2) Due to length limitation, results regarding some covariates are omitted, including the “featured_job”, “urgent”, “NDA”, and “payment_verified” dummies. The means of all these variables are not significantly different across groups.

a) Before matching b) After matching

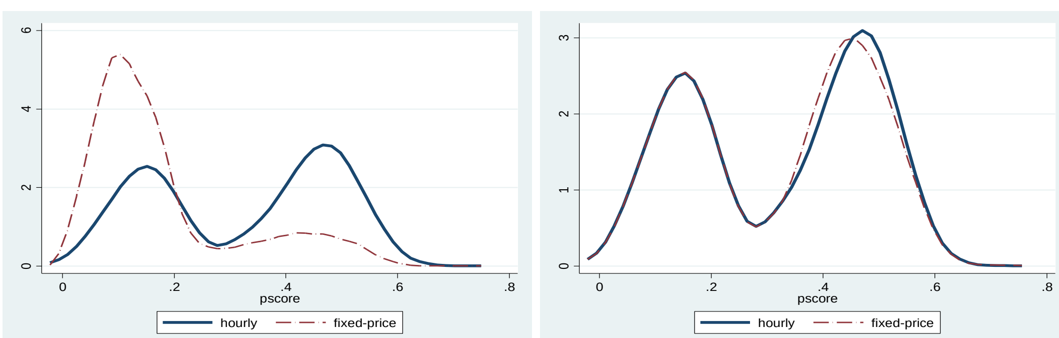


Figure 11. Density of Propensity Score of Being an Hourly Job

(before and after Matching)

3.6.2 Results on Gender Differences in Avoidance of Monitoring

Another gender difference of our key interest is workers' avoidance of monitoring. Specifically, if females have a stronger avoidance of monitoring, they may be less willing to bid for hourly jobs or accept a lower wage job which does not require monitoring in other platforms or markets, which subsequently lowers their labor participation or average hourly wage in the gig economy. Based on the result of Model 1 and Model 2 with the DID estimation, females are significantly less willing to bid for hourly jobs after the introduction of the monitoring system, with the trend in their preference of the fixed-price jobs as the counterfactual. Moreover, we further explore gender differences in avoidance of monitoring with the DDD estimation by taking the difference between the differences-in-differences (DD) observed in the female sample and the DD observed male sample. As the result of Model 3 in Table 44 shows, females are less willing to bid for hourly jobs after the introduction of monitoring systems. Given that monitoring systems are advocated for all hourly jobs on Freelancer.com after the introduction and it is difficult to observe the outside option for most female workers, we turn to a field experiment to observe gender differences in WTP for the avoidance of monitoring and infer its impact on the gender wage gap accordingly.

Table 44. Gender Differences in Avoidance of Monitoring

	Dependent Variable: Bid or Not					
Sample	Full sample		Female, matched jobs		Matched sample	
Model	(1)		(2)		(3)	
Hourly	0.003**	(0.002)	0.004**	(0.002)	-0.003	(0.002)
After*hourly	-0.003**	(0.002)	-0.004**	(0.002)	0.002	(0.002)
Hourly*female					0.007***	(0.003)
After*hourly*female					-0.006**	(0.003)
Employer_norating	0.001	(0.001)	0.001	(0.001)	-0.000	(0.001)

Log(employer_review)	0.000	(0.000)	0.000	(0.000)	-0.000	(0.000)
Employer_rating	0.001**	(0.000)	0.001	(0.000)	0.001*	(0.000)
Log(budget)	0.000***	(0.000)	0.000	(0.000)	0.000***	(0.000)
Log(title_length)	-0.000	(0.001)	-0.001	(0.001)	-0.001*	(0.001)
Log(desc_length)	0.000	(0.000)	0.000	(0.000)	0.001*	(0.000)
Log(skills_count)	-0.001	(0.001)	-0.001	(0.001)	0.000	(0.001)
Featured_job	-0.011***	(0.004)	-0.01*	(0.004)	-0.014***	(0.004)
NDA	-0.02***	(0.009)	omitted		omitted	
Log(remain_days)	0.008***	(0.000)	0.007***	(0.001)	0.009***	(0.000)
Log(auction_duration)	-0.009***	(0.001)	-0.005***	(0.002)	-0.009***	(0.001)
Consideration set FE	yes		yes		yes	
Employer country dummy	yes		yes		yes	
Observations	105,479		52,221		101,420	
R-squared	0.089		0.188		0.159	
Adjusted R-squared	0.035		0.092		0.062	
Residual Std. Error	0.064		0.057		0.065	

Notes: a) Model 1 is estimated based on all the hourly and fixed-price job choices made by all the female workers; Model 2 is estimated based on the matched hourly and fixed-price job choices made by the matched female workers; Model 3 is estimated based on the matched hourly and fixed-price job choices made by the matched female and male workers; b) Robust standard errors clustered by the consideration set of each bid decision are reported in parentheses; c) * p<0.1, ** p<0.05, *** p<0.01.

3.6.3 Results on Gender Differences in WTP for Avoidance of Monitoring

To further investigate whether there exists gender difference in avoidance of monitoring, we conducted a field experiment by providing all participants with two hourly job options and asking them to choose the one they preferred. Following the previous literature, we estimated the mean WTP of males and females with a logit model (Mas and Pallais 2017). Specifically, we estimated the distribution of WTP among all the participants. In particular, based on the difference in the probability of choosing a job with monitoring as the wage premium of the job with monitoring changes, we estimate the mean and standard deviation of female and male participants' willingness to pay for the avoidance of monitoring. As the result of the maximum likelihood logit model in Table 45 shows, an average female is willing to pay $1.779 for the avoidance of monitoring while an average male is only willing to pay $1.276. The gender difference in WTP for avoidance of monitoring is around $0.503, which is significant at the 0.05 significance level based on 1,000 bootstrap samples. In particular, according to the prior study on AMT, the mean hourly wage is $6.19 for all those paid work (Hara et al. 2018). Therefore, the gender difference in WTP for the avoidance of monitoring is equivalent to 8.13% of the average hourly wage on AMT. In other words, females are willing to accept an hourly job without monitoring by offering 8.13% discount on their hourly wage.

Table 45. Gender Differences in WTP for Avoidance of Monitoring

Willingness to pay for the avoidance of monitoring			
	female	male	difference
Mean (μ)	\$ 1.779 (0.138)	\$ 1.276 (0.188)	\$0.503 (0.227)
SD (σ)	\$ 1.223 (0.135)	\$ 0.891 (0.169)	

Note: Standard errors are calculated based on 1000 bootstrap samples.

3.7. Robustness Checks

3.7.1 Alternative Measure

To assess the robustness of our result, we use an alternative measure to show females' preference to bid later than males. Specifically, we construct another measure, *log(passed_days)*, which represents the number of days between the start date of the auction and the bid decision date. We again find a negative coefficient for the main effect of *log(passed_days)* and a positive coefficient for the interaction term between *log(passed_days)* and the gender dummy, which suggests that females tend to bid later than males (Table 46).

Table 46. Gender Differences in Job Application Strategy for Hourly Jobs

Dependent Variable: Bid or Not								
Sample	Matched sample (1)		Matched sample (2)		Matched sample (3)		Full sample (4)	
Log(passed_days)	-0.016***	(0.001)	-0.019***	(0.001)	-0.019***	(0.001)	-0.017***	(0.001)
Log(passed_days)* female			0.006***	(0.002)	0.006***	(0.002)	0.005***	(0.002)
Log(budget)	0.000	(0.000)	-0.000*	(0.000)	0.000	(0.000)	0.000	(0.000)
Log(budget)* female	-0.001**	(0.000)			-0.001*	(0.000)	-0.001**	(0.000)
Employer_norating	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
Log(employer_review)	0.001**	(0.000)	0.001**	(0.000)	0.001**	(0.000)	0.000**	(0.000)
Employer_rating	0.002	(0.002)	0.002	(0.002)	0.002	(0.002)	0.002	(0.001)
Log(auction_duration)	0.021***	(0.003)	0.021***	(0.003)	0.021***	(0.003)	0.016***	(0.002)
Log(title_length)	-0.002*	(0.001)	-0.002*	(0.001)	-0.002*	(0.001)	-0.001***	(0.001)
Log(desc_length)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)	0.001**	(0.000)
Log(skills_count)	0.003**	(0.001)	0.003**	(0.001)	0.003**	(0.001)	0.000	(0.001)
Featured_job	-0.021***	(0.007)	-0.020***	(0.007)	-0.02***	(0.007)	-0.006	(0.004)
Consideration set FE	yes		yes		yes		yes	
Employer country dummy	yes		yes		yes		yes	
Observations	42,545		42,545		42,545		150,706	
R-squared	0.302		0.302		0.302		0.281	
Adjusted R-squared	0.097		0.098		0.098		0.080	
Residual Std. Error	0.062		0.062		0.062		0.064	

Notes: a) Robust standard errors clustered by the consideration set of each bid decision are reported in parentheses; b) The “NDA” dummy is omitted because of the lack of variation. Among all the hourly jobs in the matched workers’ consideration set, all jobs do not require NDA. c)Because we control for the fixed effect of the consideration set of each bid, the worker’s fixed effect is omitted. d) The dependent variable, “bid”, the dummy denoting whether the worker chose to bid for the job or not) is relatively small (its mean is 0.005). As such, even the magnitude of the coefficient is small, its marginal effect measured with percentage change can be large. e)* p<0.1, ** p<0.05, *** p<0.01.

Moreover, instead of merely inferring the workers’ genders based on their profile images, we predict each worker’s gender based on his/her first names by taking advantage of the Facebook profile name database (Tang et al. 2013; Chan and Wang 2017). Following the previous literature (Chan and Wang 2017), we limit to those first names with a gender probability higher or equal to 95%, based on which we can reliably infer the worker’s gender. Further, we rerun all the models with the sample of those workers whose genders can be consistently predicted with both profile images and first names. As Table 47 shows, the results are highly consistent with our main finding.

Table 47. Gender Differences in Avoidance of Monitoring

	Dependent Variable: Bid or Not					
Sample	Full sample		Female, matched jobs		Matched sample	
Model	(1)		(2)		(3)	
Hourly	0.006***	(0.003)	0.005**	(0.003)	-0.005**	(0.003)
After*hourly	-0.006***	(0.003)	-0.005**	(0.003)	0.006***	(0.003)
Hourly*female					0.011***	(0.004)
After*hourly*female					-0.013***	(0.004)
Employer_norating	0.000	(0.001)	0.000	(0.001)	-0.000	(0.001)
Log(employer_review)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
Employer_rating	0.001	(0.000)	0.001	(0.000)	0.001	(0.001)
Log(budget)	0.000*	(0.000)	0.000	(0.000)	0.000	(0.000)
Log(title_length)	-0.000	(0.001)	-0.001	(0.001)	-0.001**	(0.001)
Log(desc_length)	0.000	(0.000)	-0.000	(0.000)	0.000	(0.000)
Log(skills_count)	-0.002**	(0.001)	-0.002	(0.001)	0.000	(0.001)
Featured_job	-0.016***	(0.007)	-0.011	(0.005)	-0.011*	(0.004)
NDA	-0.025***	(0.017)	-0.002	(0.002)	-0.002	(0.002)
Log(remain_days)	0.007***	(0.001)	0.006***	(0.001)	0.008***	(0.001)
Log(auction_duration)	-0.007***	(0.002)	-0.004*	(0.002)	-0.01***	(0.002)
Consideration set FE	yes		yes		yes	
Employer country dummy	yes		yes		yes	
Observations	47,041		32,064		70,000	
R-squared	0.088		0.130		0.119	
Adjusted R-squared	0.029		0.049		0.046	
Residual Std. Error	0.058		0.054		0.064	

Notes: a) Model 1 is estimated based on all the hourly and fixed-price job choices made by all the female workers; Model 2 is estimated based on the matched hourly and fixed-price job choices made by the matched female workers; Model 3 is estimated based on the matched hourly and fixed-price job choices made by the matched female and male workers; b) Robust standard errors clustered by the consideration set of each bid decision are reported in parentheses; c) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

3.7.2 Alternative Specification

To further evaluate the credibility of our result, we also try to check our result is still consistent if we adopt an alternative specification. For one thing, we explore gender differences in job application strategy by controlling the bidder-day pair-specific fixed effects instead of the consideration set fixed effects. We still find highly consistent results. The result in Table 48 suggests that females show a stronger avoidance of monitoring.

Table 48. Gender Differences in Avoidance of Monitoring

	Dependent Variable: Bid or Not					
Sample	Full sample		Female, matched jobs		Matched sample	
Model	(1)		(2)		(3)	
Hourly	0.002*	(0.002)	0.003*	(0.002)	-0.002	(0.002)
After*hourly	-0.003*	(0.002)	-0.003*	(0.002)	0.002	(0.002)
Hourly*female					0.006**	(0.003)
After*hourly*female					-0.005**	(0.003)
Employer_norating	0.001	(0.001)	0.001	(0.001)	-0.000	(0.001)
Log(employer_review)	0.000	(0.000)	0.000	(0.000)	-0.000	(0.000)
Employer_rating	0.001**	(0.000)	0.001	(0.000)	0.001*	(0.000)
Log(budget)	0.000***	(0.000)	0.000	(0.000)	0.000***	(0.000)
Log(title_length)	-0.000	(0.001)	-0.001	(0.001)	-0.001*	(0.001)
Log(desc_length)	0.000	(0.000)	0.000	(0.000)	0.001*	(0.000)
Log(skills_count)	-0.001	(0.001)	-0.001	(0.001)	0.000	(0.001)
Featured_job	-0.011***	(0.004)	-0.01*	(0.004)	-0.014***	(0.003)
NDA	-0.020***	(0.009)	omitted		omitted	
Log(remain_days)	0.008***	(0.000)	0.007***	(0.001)	0.009***	(0.000)
Log(auction_duration)	-0.009***	(0.001)	-0.005***	(0.002)	-0.009***	(0.001)
Bidder-day pair FE	yes		yes		yes	
Employer country dummy	yes		yes		yes	
Observations	105,479		52,221		101,420	
R-squared	0.082		0.175		0.143	
Adjusted R-squared	0.036		0.095		0.063	
Residual Std. Error	0.064		0.057		0.065	

Notes: a) Model 1 is estimated based on all the hourly and fixed-price job choices made by all the female workers; Model 2 is estimated based on the matched hourly and fixed-price job choices made by the matched female workers; Model 3 is estimated based on the matched hourly and fixed-price job choices made by the matched female and male workers; b) Robust standard errors clustered by the consideration set of each bid decision are reported in parentheses; c) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

For another, we also use another model, i.e., the probit model to estimate the gender difference in WTP for the avoidance of monitoring. As Table 49 suggests, the gender difference in WTP for the avoidance of monitoring is still $0.503 and significantly larger than zero based on the bootstrapped standard errors, which is highly consistent with our main result.

Table 49. Gender Differences in WTP for Avoidance of Monitoring Estimated with Probit

Willingness to pay for the avoidance of monitoring			
	female	male	difference
Mean (μ)	\$ 1.773 (0.143)	\$ 1.283 (0.213)	\$0.503 (0.257)
SD (σ)	\$ 1.208 (0.173)	\$ 0.828 (0.312)	

Note: Standard errors are calculated based on 1000 bootstrap samples.

3.7.3 Parallel Trends Assumption

We further test the parallel trend assumption between the matched female and male sample using the approach proposed by Autor (2003). Specifically, we estimate the time-varying change in females' avoidance of hourly jobs with the matched males as the counterfactual based on the following equation:

$$Worker_{it}\_choice\_on\_job_j = \alpha_{it} + \beta_1 Hourly_j + \beta_2 Female_i \times Hourly_j + \beta_3 \delta_t \times Hourly_j + \beta_4 Female_i \times \delta_t \times Hourly_j + controls(Job_j) + \varepsilon_{itj} \qquad (12)$$

where δ_t represents a vector of time dummies and $\{\beta_4\}$ denotes the matrix of relative time parameters of females' avoidance of hourly jobs estimated at time t. Given that the monitoring system was implemented on February 5th, 2014, we use the month prior to the policy change (January 2014) as the baseline (Autor 2003). We find that all the relative time coefficients are not significant prior to the introduction of the monitoring system and roughly half of the relative time coefficients are significantly negative after the introduction. This implies that the pre-existing treatment trend is not an issue in our study, which lends support to the causality of our findings.

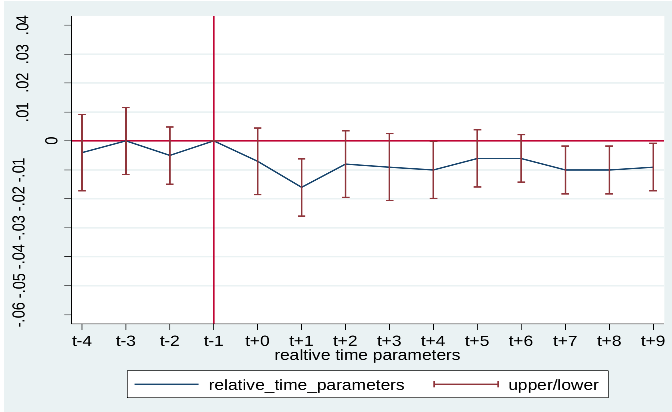


Figure 12. Coefficients of the Monthly DID Estimates of the Treatment Effect

Note: The dash vertical line denotes when *Freelancer* first implemented the monitoring system (February 2014).

3.7.4 Alternative Matching

To further alleviate the concern of incomparability between fixed-price jobs and hourly jobs, we employ alternative matching methods, including Coarsened Exact Matching (CEM) (Iacus et al. 2012; Blackwell et al. 2009) and the propensity score matching with five nearest neighbors, to match males with females, and fixed-price jobs with hourly jobs. In Table 50, we summarize the result based on the matched sample with the CEM approach, which is highly consistent with our main result.

Table 50. Gender Differences in Avoidance of Monitoring

Dependent variable: whether the worker chose to bid for the job or not						
Sample	Full sample		Female, matched jobs		Matched sample	
Model	(1)		(2)		(3)	
Hourly	0.003**	(0.002)	0.004**	(0.002)	-0.003	(0.002)
After*hourly	-0.003**	(0.002)	-0.004***	(0.002)	0.003*	(0.002)
Hourly*female					0.008***	(0.003)
After*hourly*female					-0.009***	(0.003)
Employer_norating	0.001	(0.001)	0.000	(0.001)	0.001	(0.001)
Log(employer_review)	0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
Employer_rating	0.001**	(0.000)	0.002**	(0.000)	0.001*	(0.000)
log(budget)	0.000***	(0.000)	0.000**	(0.000)	0.000**	(0.000)
Log(title_length)	-0.000	(0.001)	-0.001	(0.001)	-0.001	(0.001)
Log(desc_length)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)

Log(skills_count)	-0.001	(0.001)	-0.000	(0.001)	-0.000	(0.001)
Featured_job	-0.011***	(0.004)	-0.008*	(0.003)	-0.010*	(0.003)
NDA	-0.020***	(0.006)	-0.005	(0.006)	-0.003	(0.006)
Log(remain_days)	0.008***	(0.000)	0.008***	(0.000)	0.008***	(0.000)
Log(auction_duration)	-0.009***	(0.001)	-0.008***	(0.002)	-0.010***	(0.001)
Consideration set FE	yes		yes		yes	
Employer country dummy	yes		yes		yes	
Observations	105,479		71,825		90,139	
R-squared	0.089		0.129		0.117	
Adjusted R-squared	0.035		0.054		0.045	
Residual Std. Error	0.064		0.062		0.064	

Notes: a) Model 1 is estimated based on all the hourly and fixed-price job choices made by all the female workers; Model 2 is estimated based on the matched hourly and fixed-price job choices made by the matched female workers; Model 3 is estimated based on the matched hourly and fixed-price job choices made by the matched female and male workers; b) Robust standard errors clustered by the consideration set of each bid decision are reported in parentheses; c) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Overall, all the robustness checks lend support to our finding that females tend to bid later, prefer jobs with lower wage budget, and have a higher WTP for the avoidance of monitoring than males.

3.8. Discussion

In this paper, we explore whether there is a gender wage gap in the gig economy and examine whether there are gender differences in job application strategy which could account for the persistent gender wage gap. First, we show that females can only earn around 81.4% of the hourly wage of their male counterparts. Second, we find that females tend to bid later and prefer jobs with a smaller hourly wage budget based on both the model-free evidence and the empirical results of the linear probability model with the consideration set fixed-effect. We further find that the observed gender difference in bid timing can lead to a decrease of 7.58% in hourly wage, which could roughly account for 40.75% of the gender wage gap (i.e. 18.6%) observed by us. Third, we examine the gender difference in avoidance of monitoring with a quasi-natural experiment and a field experiment. We find that females are less willing to bid for hourly jobs than males and tend to have a higher willingness to pay for the avoidance of monitoring. The gender difference in WTP for the avoidance of monitoring can explain roughly 8.13% of the hourly wage, which is equivalent to 43.71% of the observed gender wage gap. On the whole, our study underscores the important impact of gender differences in job application strategy on the gender wage gap.

Our paper contributes to several streams of literature. First, our paper contributes to several streams of literature. First, our study adds to the literature on gender wage gap and highlights new explanatory factors for the gender wage gap other than gender discrimination, i.e. gender differences in bid timing, job selection, and avoidance of monitoring. The existing literature mainly focusing on the traditional employment relationship suggests that discrimination, WTP for flexibility (Mas and Pallais 2017), motherhood penalty and career choices (Blau and Kahn 2017) could help to explain the gender wage gap. On top of that, some scholars predict that the gig economy is an emerging labor market design which helps to narrow the gender wage gap owing to the flexibility and remoteness of its on-demand employment relationship (Goldin and Rouse 2000; Goldin 2014). In contrast, a recent study on the gig economy suggests that the gender wage gap still exists. Using a large-scale dataset from a gig economy platform which provides offline car-hailing service (i.e. Uber), Cook et al. (2018) find that, gender differences in experience and willingness to work extra hours when the hourly wage is high, mainly explain the gender wage gap. However, given that workers tend to have limited freedom to choose jobs in the car-hailing platform, the existence and potential impact of gender differences in job application strategy is hitherto little explored in their study. Given that the freedom of choosing jobs based on preference is such a common primary feature shared by most gig economy platforms, our study focuses on potential gender differences in job application strategy and points out that gender differences in bid timing, job selection, and WTP for the avoidance of monitoring help to explain the gender wage gap in gig-economy.

Second, this paper contributes to the literature on the online labor market by providing a framework to recover workers' consideration sets and underscores the importance of workers' job preference. Though employers' preference of workers has been recently explored (Chan and Wang 2017; Hong and Pavlou 2017), there is little research exploring the preference from the supply side (i.e. workers' preference for jobs). We extend this prior work by taking advantage of a comprehensive dataset and the platform policy to recover workers' consideration sets. We further demonstrate gender differences in job application strategy from three aspects, including bid timing, job budget preference and avoidance of monitoring. Our study advances the previous literature on online labor markets by documenting gender differences in job application strategy, which has strong academic and managerial implications for the online labor market.

Lastly, our study also expands the literature on compensation differential. Prior studies have found compensation differential in several non-wage job amenities in traditional employment relationship (Bonhomme and Jolivet 2009), such as flexibility (Mas and Pallais 2017), unemployment benefits (Hall and Mueller 2015), and non-wage job value (Sorkin 2017). Given that online monitoring is prevalent in most online labor markets, we focus on potential compensation differential in avoidance of monitoring, a non-wage aspect which has hardly been explored in the previous compensation differential literature. Taking the introduction of the monitoring system as an exogenous shock, we find that females are less willing to bid for jobs with monitoring, compared to males. We further conduct a field experiment on AMT to explicitly estimate gender differences in WTP for the avoidance of monitoring. Our finding suggests that gender differences in WTP for the avoidance of monitoring are likely to persistently contribute to the gender wage gap.

Meanwhile, we acknowledge several limitations of this study. For instance, we note that our results are limited by the IT job sample and it should be cautious to generalize the results to other job categories, especially those feminine-typed jobs. Further, it might not be appropriate to generalize the results to other offline labor markets until sufficient evidence is available. Last but not least, although our analysis points to a strong relationship between these gender difference in job preference and the gender wage gap, we admit that we cannot rule out all the possible unobserved factor influencing both the gender difference in job application strategy and the gender wage gap. We believe our study helps to suggest the potential ways to reduce the wage gap instead of concluding the drivers of the gender wage gap.

CONCLUSION

My research contributes to the literature on Gig economy and IT-enabled monitoring on four fronts. First, most prior studies focus on the performance effect of monitoring in offline contexts (Duflo et al. 2012; Hubbard 2000; Ranganathan and Benson 2017), whereas my research focuses on the impact of an IT-enabled monitoring artifact on both demand-side (employer) preference and supply-side (worker) competition in online platforms.

Second, my research advances the prior literature on the relationship between monitoring systems and reputation systems in online platforms by showing that the introduction of monitoring systems only reduces employers' preference for workers with high effort-related reputation but not those with high capability-related reputation (Demiroglu and James 2010; Diamond 1991; Lin et al. 2016). Given that my study's setting allows me to identify the causal effects of the introduction of the monitoring system on both the supply and demand sides of an online labor market, I found that IT-enabled monitoring facilitates market competition by lowering the entry-barrier in terms of reputation.

Third, my research contributes to the home bias literature, as it is among the first to investigate the existence of home bias in the employment setting that explored the mechanisms with a quasi-natural experiment. It extends the previous home bias research in contexts of equity or trade, which mainly focuses on decisions under ex ante information asymmetry, to the employment decision threatened by both ex ante and ex post information asymmetry.

Last but not least, my research also contributes to the literature on compensation differentials and the gender wage gap. There is an emerging literature investigating workers' WTP for non-wage job amenities (Bonhomme and Jolivet 2009) and gender wage gap, such as flexibility (Mas and Pallais 2017), unemployment benefits (Hall and Mueller 2015), and non-wage job value (Sorkin 2017). My study takes advantage of both a quasi-natural experiment and a field experiment to show the gender difference in WTP for the avoidance of monitoring, a non-wage aspect which has hardly been explored in the compensation differential literature.

REFERENCES

Abadie, A. (2005) Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*. 72(1): 1-19

Abadie, A., Diamond, A., & Hainmueller, J. (2015) Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*. *59*(2): 495-510.

Agan, A. Y., & Starr, S. B. (2016). Ban the box, criminal records, and statistical discrimination: A field experiment.

Agrawal, A., Lacetera, N., and Lyons, E. (2014) Does Standardized Information in Online Markets Disproportionately Benefit Job Applicants from Less Developed Countries?. Mimeo.

Akerlof, G. A. (1978). The Market for "Lemons": Quality Uncertainty and The Market Mechanism. In *Uncertainty in Economics* (pp. 235-251).

Allen, V.L. and Wilder, D.A. (1975) Categorization, belief similarity, and intergroup discrimination. *Journal of Personality and Social Psychology*, *32*(6), 971.

Altonji, J.G. and Blank, R.M. (1999) Race and gender in the labor market. *Handbook of labor economics*, *3*, 3143-3259.

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005) Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *Journal of political economy*, *113*(1), pp.151-184.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, *69*(1), 106-116.

Anderson, D. J., Binder, M., & Krause, K. (2002). The motherhood wage penalty: Which mothers pay it and why?. *American Economic Review*, 92(2), pp.354-358.

Anderson Jr, E.G., Parker, G.G. and Tan, B. (2013) Platform Performance Investment in the Presence of Network Externalities. *Information Systems Research.* 25(1): 152-172.

Angrist, J.D. (2004) Treatment effect heterogeneity in theory and practice. *The Economic Journal*, *114*(494).

Angrist, J. and Imbens, G.(1995). Identification and estimation of local average treatment effects.

Angrist, J.D., Imbens, G.W. and Rubin, D.B.(1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association, 91*(434), 444-455.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, *15*(4): 69-85.

Angrist, J. D., & Pischke, J. S. (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science*, *58*(5): 913-931.

Arceo-Gomez, E.O. and Campos-Vazquez, R.M. (2014). Race and marriage in the labor market: a discrimination correspondence study in a developing country. *The American Economic Review*, *104*(5), 376-380.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, *133*(4), 1885-1932.

Arnould, R.J. and Nichols, L.M., (1983). Wage-risk Premiums and Workers' Compensation: A Refinement of Estimates of Compensating Wage Differential. *Journal of Political Economy*, (91:2), pp.332-340.

Aron, R., Jayanty, S. and Pathak, P. (2007) Impact of Internet-based Distributed Monitoring Systems on Offshore Sourcing of Services. *ACM Transactions on Internet Technology (TOIT)*. 7(3): 16.

Arrow, K.J.(1973). The theory of discrimination.

Åslund, O., Hensvik, L., & Skans, O. N. (2014). Seeking similarity: how immigrants and natives manage in the labor market. *Journal of Labor Economics, 32*(3), 405-441.

Atakan, A. E., & Ekmekci, M. (2014). *Reputation in Repeated Moral Hazard Games* (No. 54427). University Library of Munich, Germany.

Atasoy, Hilal, Rajiv D. Banker, and Paul A. Pavlou. (2016) On the Longitudinal Effects of IT Use on Firm-Level Employment. *Information Systems Research*, 27.1: 6-26.

Autor, D.H., (2003). The economics of discrimination theory. *Lecture Note of Labor Economics, MIT*, *14*.

Autor, D. H. (2003) Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*. *21*(1): 1-42.

Azmat, G. and Petrongolo, B., (2014). "Gender and the labor market: What have we learned from field and lab experiments?". *Labour Economics, 30,* pp.32-40.

Ba, S. and Pavlou, P.A. (2002) Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly*. 26(3): 243-268.

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

Bakos, Y., & Dellarocas, C. (2011). Cooperation without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms. *Management Science*, *57*(11): 1944-1962.

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

Banker, R.D. and Hwang, I. (2008) Importance of Measures of Past Performance: Empirical Evidence on Quality of E-Service Providers. *Contemporary Accounting Research.* 25(2): 307-337.

Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal*, *15*(S1), 175-190.

Becker, G. S. (1971) *The economics of discrimination*. University of Chicago Press Economics Books.

Benaroch, M., Lichtenstein, Y., & Fink, L. (2016) Contract Design Choices and The Balance of ex ante and ex post Transaction Costs in Software Development Outsourcing. *MIS Quarterly*. 40(1): 57-82.

Belzer, A. and Leong, N.(2016). The New Public Accommodations

Bergemann, Annette, Bernd Fitzenberger, and Stefan Speckesser. (2009) Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-differences. *Journal of Applied Econometrics,* 24(5): 797-823.

Bergstrand J 1985. The gravity equation in international trade: some microeconomic foundations and empirical evidence. *Rev. Econom. Statist, 67*(3),474–481.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: the new car market. *Journal of Political Economy*, *112*(1), 68-105.

Bertrand, M., & Mullainathan, S. (2004) Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, *94*(4), 991-1013.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005) Implicit discrimination. *American Economic Review*, 94-98.

Bertrand, M., & Duflo, E. (2017) Field experiments on discrimination. *Handbook of Economic Field Experiments*, *1*, 309-393.

Bertrand, M., & Mullainathan, S. (2004) Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, *94*(4), 991-1013.

Blackwell, M., Iacus, S. M., King, G., & Porro, G. (2009) CEM: Coarsened Exact Matching in Stata. *The Stata Journal*. *9*(4): 524-546.

Blackwell, M. (2013). A Framework for Dynamic Causal Inference in Political Science. *American Journal of Political Science*. *57*(2): 504-520.

Blau, F.D. and Kahn, L.M., (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, (*55*:3), pp.789-865.

Blinder, A. S., & Krueger, A. B. (2013) Alternative Measures of Offshorability: A Survey Approach. *Journal of Labor Economics*. 31(2): 97-128.

Bockstedt, J. and Goh, K.H. (2011) Seller Strategies for Differentiation in Highly Competitive Online Auction Markets. *Journal of Management Information Systems.* 28(3): 235-268.

Bolton, P., & Dewatripont, M. (2005). *Contract theory*. MIT press.

Bonhomme, S. and Jolivet, G., (2009). The pervasive absence of compensating differentials" *Journal of Applied Econometrics*, (*24*:5), pp.763-795.

Brynjolfsson, E., & Milgrom, P. (2013). Complementarity in Organizations. *The Handbook of Organizational Economics*, 11-55.

Brewer, M.B. and Kramer, R.M. (1986) Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology*, *50*(3), 543.

Brunetti, A., Kisunko, G., & Weder, B. (1997) *Institutional obstacles to doing business: region-by-region results from a worldwide survey of the private sector* (Vol. 1759). World Bank Publications.

Budig, M.J. and England, P., (2001). The wage penalty for motherhood. *American Sociological Review*, pp.204-225.

Burtch, G., Ghose, A., & Wattal, S. (2014) Cultural differences and geography as determinants of online prosocial lending. *MIS Quarterly*, *38*(3), 773-794.

Burtch, G., Carnahan, S. and Greenwood, B.N. (2018). Can you gig it? an empirical examination of the gig economy and entrepreneurial activity. *Management Science*, forthcoming.

Castillo, M., Petrie, R., Torero, M. and Vesterlund, L., (2013). Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics*, (*99)*, pp.35-48.

Chan, J. and Wang, J. (2017) Hiring preferences in online labor markets: evidence of a female hiring bias. *Management Science*. Articles in Advance, 1-22.

Chan, J. and Ghose, A. (2014) Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission. *MIS Quarterly*. 38(4): 955-976.

Chen, Hailiang, Yu Jeffrey Hu, Michael D. Smith. (2017). The Impact of eBook Distribution on Print Sales: Analysis of a Natural Experiment. *Management Science* (forthcoming).

Chen, P.Y., Hong, Y. and Liu, Y. (2017) The value of multidimensional rating systems: evidence from a natural experiment and randomized experiments. Management Science, forthcoming.

Chen, D. L., & Horton, J. J. (2016). Research note—Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research*, *27*(2), 403-423.

Chen, Y. and Li, S.X. (2009) Group identity and social preferences. *The American Economic Review*, *99*(1), 431-457.

Coase, R. H. (1937). The Nature of the Firm. *Economica*, *4*(16): 386-405.

Cook, C., Diamond, R., Hall, J., List, J.A. and Oyer, P., (2018). *The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers* (No. repec: ecl: stabus: 3637).

Cooley, T., Marimon, R., & Quadrini, V. (2004). Aggregate Consequences of Limited Contract Enforceability. *Journal of Political Economy*, *112*(4): 817-847.

Cooper, I., & Kaplanis, E. (1994) Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *Review of Financial Studies*, *7*(1), 45-60.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: local equity preference in domestic portfolios. *The Journal of Finance, 54*(6), 2045-2073.

Dahl, M.S. and O. Sorenson. (2010) The social attachment to place. *Social Forces*, 89(2) 633–658.

Dale, S.B. and Krueger, A.B. (2002) Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables. *The Quarterly Journal of Economics*, *117*(4), 1491-1527.

Dellarocas, C. (2005) Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard. *Information Systems Research.* 16(2): 209-230.

Dellarocas, C. (2006) Reputation Mechanisms. *Handbook on Economics and Information Systems:* 629-660.

Demiroglu, C., and James, C. M. (2010) The Role of Private Equity Group Reputation in LBO Financing. *Journal of Financial Economics.* 96(2): 306-330.

Dey, D., Fan, M., and Zhang, C. (2010) Design and Analysis of Contracts for Software Outsourcing. *Information Systems Research.* 21(1): 93-114.

Diamond, D. W. (1991) Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt. *Journal of Political Economy.* 99(4): 689-721.

DiDonato, T.E., Ullrich, J. and Krueger, J.I. (2011) Social perception as induction and inference: an integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology*, *100*(1), 66.

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review*, *108*(2), 201-40.

Doleac, J. L., & Hansen, B. (2016) *Does "ban the box" help or hurt low-skilled workers? statistical discrimination and employment outcomes when criminal histories are hidden* (No. w22469). National Bureau of Economic Research.

Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, *23*(3), 601-620.

Drago, R. (1991) Incentives, Pay, and Performance: a Study of Australian Employees. *Applied Economics.* 23(9): 1433-1446.

Duflo, E., Hanna, R., and Ryan, S. P. (2012) Incentives Work: Getting Teachers to Come to School. *The American Economic Review.* 102(4): 1241-1278.

Edelman, B., Luca, M. and Svirsky, D. (2017) Racial discrimination in the sharing economy: evidence from a field experiment. *American Economic Journal: Applied Economics*, *9*(2), 1-22.

Efferson, C., Lalive, R. and Fehr, E. (2008) The coevolution of cultural groups and ingroup favoritism. *Science*, *321*(5897), 1844-1849.

Eisenhardt, K.M., (1985). Control: Organizational and Economic Approaches. *Management science*, *31*(2):134-149.

Eisenhardt, K. M. (1989) Agency Theory: an Assessment and Review. *Academy of Management Review.* 14(1): 57-74.

Eisenmann, T., Parker, G. and Van Alstyne, M.W. (2006) Strategies for Two-Sided Markets. *Harvard Business Review*. 84(10): 92.

Eisenmann, T., Parker, G., and Van Alstyne, M. (2011) Platform Envelopment. *Strategic Management Journal.* 32(12): 1270-1285.

Falk, A. and Zehnder, C. (2013) A city-wide experiment on trust discrimination. *Journal of Public Economics*, *100*, 15-27.

Fang, H. and Moro, A. (2010) *Theories of statistical discrimination and affirmative action: a survey* (No. w15860). National Bureau of Economic Research.

Fearon, J.D. (2003) Ethnic and cultural diversity by country. *Journal of Economic Growth*, *8*(2), 195-222.

Fearon, J.D. and Laitin, D.D. (2003) Ethnicity, insurgency, and civil war. *American Political Science Review*, *97*(1), 75-90.

Flory, J. A., Leibbrandt, A., & List, J. A. 2014. "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions," The Review of Economic Studies, (82:1), pp.122-155.

Fong Boh, W., Slaughter, S. A., & Espinosa, J. A. (2007). Learning from experience in software development: A multilevel analysis. *Management Science*, *53*(8), 1315-1331.

Forman, C., Ghose, A. and Goldfarb, A. (2009) Competition between local and electronic markets: how the benefit of buying online depends on where you live. *Management science*, *55*(1), pp.47-57.

Forman, C., Goldfarb, A. and Greenstein, S. (2012) The internet and local wages: a puzzle. *The American Economic Review*, *102*(1), pp.556-575.

French, K. R., & Poterba, J. M. (1991) *Investor diversification and international equity markets* (No. w3609). National Bureau of Economic Research.

Friedman, E. J., & Resnick, P. (2001). The Social Cost of Cheap Pseudonyms. *Journal of Economics & Management Strategy*, 10(2): 173-199.

Fryer Jr., R.G. and Levitt, S.D. (2004) The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, *119*(3), 767-805.

Fryer, R. and Jackson, M.O. (2008) A categorical model of cognition and biased decision-making. *The BE Journal of Theoretical Economics*, *8*(1), 1935-1704.

Fudenberg, D., & Levine, D. K. (1989). Reputation and Equilibrium Selection in Games with a Single Long-Run Player. *Econometrica*, *57*.

Fudenberg, D., & Levine, D. K. (1992). Maintaining a Reputation When Strategies Are Imperfectly Observed, *Review of Economic Studies*, 59(3): 561-579.

Ge, Y., Knittel, C.R., MacKenzie, D. and Zoepf, S. (2016) *Racial and gender discrimination in transportation network companies* (No. w22776). National Bureau of Economic Research.

Gefen, D. and Carmel, E. (2008) Is the world really flat? a look at offshoring at an online programming marketplace. *MIS Quarterly*, pp.367-384.

Goldin, C., & Rouse, C. (2000) *Orchestrating impartiality: the impact of "blind" auditions on female musicians*. *American Economic Review, 90*(4), 715-741.

Gong, J., Hong, Y. and Zentner, A., (2018) Vanishing borders in the internet age: the income elasticity of the supply of foreign labor in virtual versus physical markets. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Ghani, E., Kerr, W. R., & Stanton, C. (2014). Diasporas and outsourcing: evidence from Odesk and India. *Management Science, 60*(7), 1677-1697.

Ghasemkhani, H., Li, Y.M., Moinzadeh, K. and Tan, Y. (2017) Contracting Models for P2P Content Distribution. *Production and Operations Management* (forthcoming).

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

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), 493-520.

Goldin, C. and Rouse, C., (2000). Orchestrating impartiality: The impact of 'blind' auditions on female musicians, American Economic Review, (90:4), pp.715-741.

Goldin, C., (2014). A grand gender convergence: Its last chapter. *American Economic Review*, (*104*:4), pp.1091-1119.

Gopal, A., and Koka, B. R. (2010) The Role of Contracts on Quality and Returns to Quality in Offshore Software Development Outsourcing. *Decision Sciences.* 41(3): 491-516.

Gopal, A., and Sivaramakrishnan, K. (2008) Research Note-On Vendor Preferences for Contract Types in Offshore Software Projects: The Case of Fixed Price vs. Time and Materials Contracts. *Information Systems Research.* 19(2): 202-220.

Greenwald, B. C. (1986). Adverse Selection in The Labour Market. *The Review of Economic Studies*, *53*(3): 325-347.

Greenwood, B.N. and Agarwal, R. (2015). Matching Platforms and HIV Incidence: An Empirical Investigation of Race, Gender, and Socioeconomic Status. *Management Science.* 62(8): 2281-2303.

Gupta, N.D. and Smith, N., (2002). Children and career interruptions: The family gap in Denmark. *Economica*, (*69*:276), pp.609-629.

Hall, R. and Mueller, A.I., (2015). Wage Dispersion and Search Behavior (No. 9527)". IZA Discussion Papers.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C. and Bigham, J.P., (2018). A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM*, pp.449-463.

Hart, O., & Holmstrom, B. (1987). *The Theory of Contracts*, Advances in Economic Theory, Cambridge, U.K.: University Press, 294-351.

Heinz, M., Normann, H.T. and Rau, H.A., (2016). How competitiveness may cause a gender wage gap: Experimental evidence. *European Economic Review*, (*90)*, pp.336-349.

Hausman, J., Leonard, G., & Zona, J. D. (1994). Competitive Analysis with Differentiated Products. *Annals of Economics and Statistics*, (34), 143-157.

Hausman, J.A., 1996. Valuation of New Goods Under Perfect and Imperfect Competition. In the Economics of New Goods (pp. 207-248). University of Chicago Press.

Heckman, J. J. (1998). detecting discrimination. *The Journal of Economic Perspectives*, *12*(2), 101-116.

Heckman, J. J., & Siegelman, P. (1993). The urban institute audit studies: their methods and findings.

Heckman, J. J., & Snyder Jr, J. M. (1997). Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators. *RAND Journal of Economics*. 28: 142-189.

Helliwell, J. F. 2000. *How much do national borders matter?*. Brookings Institution Press.

Hendricks, W., DeBrock, L. and Koenker, R. (2003). Uncertainty, hiring, and subsequent performance: the NFL draft. *Journal of Labor Economics*, *21*(4), 857-886.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica* ,71(4), 1161-1189.

Ho, D.E., Imai, K., King, G. and Stuart, E.A. (2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis.* 15(3):199-236.

Hong, S. H. (2013). Measuring the Effect of Napster on Recorded Music Sales: Difference-in-Differences Estimates Under Compositional Changes. *Journal of Applied Econometrics*, *28*(2), 297-324.

Hong, Y., Wang, C., & Pavlou, P. A. (2016) Comparing Open and Sealed Bid Auctions: Evidence from Online Labor Markets. *Information Systems Research.* 27(1): 49-69.

Hong, Y. and Pavlou, P.A. (2017) On Buyer Selection of Service Providers in Online Outsourcing Platforms for IT Services. *Information Systems Research*. 28(3): 547-562.

Hölmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 74-91.

Hölmstrom, Bengt. 1982. Moral Hazard in Teams. *The Bell Journal of Economics* pp. 324–340.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The review of Economic studies*, *66*(1), 169-182.

Hofstede, G. (1980). Culture's consequences. Beverly Hills, CA: Sage.

Hofstede, G. (1991). Cultures and organizations. New York, NY: McGraw-Hill,

Hofstede, G. (2001). Culture's consequences, 2nd ed. Sage Publications, Thousand Oaks, CA.

Hong, Yili, Chong Wang, and Paul A. Pavlou. (2015). Comparing open and sealed bid auctions: evidence from online labor markets. *Information Systems Research*, *27*(1), pp.49-69.

Hong, Y., Pavlou, P. A. (2017). On buyer selection of service providers in online outsourcing platforms. *Information Systems Research*, forthcoming.

Hortaçsu, A., Martínez-Jerez, F., & Douglas, J. (2009). The geography of trade in online transactions: evidence from ebay and mercadolibre. *American Economic Journal: Microeconomics*, *1*(1), 53-74.

Horton, J. J. (2017) Buyer Uncertainty about Seller Capacity: Causes, Consequences, and A Partial Solution. *Management Science* (forthcoming).

Horton, J. J., and Golden, J. M. (2015) Reputation Inflation: Evidence from an Online Labor Market. Working paper, New York University, New York.

Huang, N., Hong, Y., Burtch, G. (2017) Social Network Integration and User Content Generation: Evidence from Natural Experiments. *MIS Quarterly*. 41(4):1035–1058.

Huang, P., & Zhang, Z. (2016). Participation in Open Knowledge Communities and Job-Hopping: Evidence from Enterprise Software. *Management Information Systems Quarterly*, *40*(3), 785-806.

Hubbard, T. N. (2000) The Demand for Monitoring Technologies: the Case of Trucking. *Quarterly Journal of Economics.* 115(2): 533-560.

Iacus, S. M., King, G., Porro, G., & Katz, J. N. (2012) Causal Inference without Balance Checking: Coarsened Exact Matching. *Political analysis*. 20: 1-24.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.

Keele, L., Small, D. S., & Hsu, J. Y. (2016) Patterns of Effects and Sensitivity Analysis for Differences-in-Differences. Working paper.

Kuhn, P. and Shen, K., 2012. “Gender discrimination in job ads: Evidence from china,” The Quarterly Journal of Economics, (128:1), pp.287-336.

Inglehart, R., C. Welzel. (2010). Changing mass priorities: the link between modernization and democracy. *Perspectives on Politics, 8*(2) 551-567.

Iacus, S. M., King, G., Porro, G., & Katz, J. N. (2012). *Causal inference without balance checking: Coarsened Exact Matchin*g. Political Analysis, 1-24.

King, G., Blackwell, M., Iacus, S., & Porro, G. (2010). CEM: Coarsened Exact Matching in Stata.

Kokkodis, M., & Ipeirotis, P. G. (2015). Reputation transferability in online labor markets. *Management Science*, *62*(6), 1687-1706.

Kugler, A.D. and Saint-Paul, G. (2000). *Hiring and firing costs, adverse selection and long-term unemployment* (No. 134). IZA.

Lado, A. A., Dant, R. R., & Tekleab, A. G. (2008). Trust-opportunism paradox, relationalism, and performance in interfirm relationships: evidence from the retail industry. *Strategic Management Journal*, *29*(4), 401-423.

Lazzarini, S. G., Miller, G. J., & Zenger, T. R. (2008). Dealing with the paradox of embeddedness: the role of contracts and trust in facilitating movement out of committed relationships. *Organization Science*, *19*(5), 709-728.

Leung, M.D., (2017). Learning to hire? hiring as a dynamic experiential learning process in an online market for contract labor. *Management Science*, forthcoming.

Levitt, S.D., (2004). Testing theories of discrimination: evidence from weakest link. *The Journal of Law and Economics*, *47*(2), 431-452.

Lewis, K. K. (1999). Trying to explain home bias in equities and consumption. *Journal of Economic Literature*, *37*(2), 571-608.

Lin, M., & Viswanathan, S. (2015). Home bias in online investments: an empirical study of an online crowdfunding market. *Management Science*, *62*(5), 1393-1414.

Lin, M., Liu, Y., & Viswanathan, S. (2016). Effectiveness of reputation in contracting for customized production: evidence from online labor markets. *Management Science*.

Lincoln, J. R. (1984). Analyzing relations in dyads: problems, models, and an application to interorganizational research. *Sociological Methods & Research*, *13*(1), 45-76.

Liu, D., Brass, D., Lu, Y., & Chen, D. (2015) Friendships in Online Peer-to-Peer Lending: Pipes, Prisms, and Relational Herding. *MIS Quarterly*. 39(3): 729-742.

Luijters, K., Van der Zee, K. I., & Otten, S. (2008). Cultural diversity in organizations: enhancing identification by valuing differences. *International Journal of Intercultural Relations*, *32*, 154–163, http://dx.doi.org/10.1016/j.ijintrel.2007.09.003.

Manchanda, P., Packard, G., & Pattabhiramaiah, A. (2015). Social dollars: The Economic Impact of Customer Participation in A Firm-Sponsored Online Customer Community. *Marketing Science*, *34*(3), 367-387.

Mani, D., Barua, A., and Whinston, A. B. (2012) An Empirical Analysis of the Contractual and Information Structures of Business Process Outsourcing Relationships. *Information Systems Research.* 23(3): 618-634.

Marini, M.M. and Fan, P.L., (1997). The gender gap in earnings at career entry. *American Sociological Review*, pp.588-604.

Marom, D., Robb, A. and Sade, O., (2016). Gender dynamics in crowdfunding (Kickstarter): Evidence on entrepreneurs, investors, deals and taste-based discrimination". Working Paper.

Mas, A. and Pallais, A., (2017). Valuing alternative work arrangements, American Economic Review, (107:12), pp.3722-59.

Moreno, A. and Terwiesch, C. (2014) Doing Business with Strangers: Reputation in Online Service Marketplaces. *Information Systems Research. 25*(4): 865-886.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand, *Journal of Economics & Management Strategy*, *9*(4), 513-548.

Neumark, D., Bank, R. J., & Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: an audit study. *The Quarterly Journal of Economics*, *111*(3), 915-941.

Niederle, M. and Vesterlund, L., (2007). Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, (*122*:3), pp.1067-1101.

Niederle, Muriel, and Lise Vesterlund, (2011). Gender and competition. *Annual Review of Economics* (3:1), pp. 601-630.

Obstfeld, M. and Rogoff, K., (2000). The six major puzzles in international macroeconomics: Is there a common cause?. *NBER Macroeconomics Annual*, *15*, 339-390.

Oster, E. (2016). Unobservable selection and coefficients stability: theory and evidence. *Journal of Businnes Economics and Statistics*, Forthcoming.

Ouchi, W.G., (1979). A Conceptual Framework for the Design of Organizational Control Mechanisms. In *Readings in Accounting for Management Control* (pp. 63-82). Springer, Boston, MA.

Pallais, A. (2014) Inefficient Hiring in Entry-Level Labor Markets. *The American Economic Review*. 104(11): 3565-3599.

Parker, G., Van Alstyne, M., and Choudary, S. (2016) *Platform Revolution*. W.W. Norton & Company, New York, NY.

Pauly, M. V. (1974). Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. *The Quarterly Journal of Economics*. 88(1): 44-62.

Pavlou, P.A. and Dimoka, A. (2006) The nature and role of feedback text comments in online marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research.* 17(4): 392-414.

Petrin, A. and Train, K., (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Tesearch*, *47*(1), 3-13. Phelps, E.S., 1972. The Statistical Theory of Racism and Sexism. *The American Economic Review*, *62*(4), 659-661.

Phelps, E.S., (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659-661.

Pierce, L., Snow, D. C., and McAfee, A. (2015) Cleaning House: The Impact of Information Technology Monitoring on Employee Theft and Productivity. *Management Science.* 61(10): 2299-2319.

Polyakova, M., (2016). Regulation of insurance with adverse selection and switching costs: evidence from medicare part D. *American Economic Journal: Applied Economics*, 8(3), 165-195.

Ranganathan, A. and Benson, A. (2017) Hemming and Hawing over Hawthorne: Work Complexity and The Divergent Effects of Monitoring on Productivity. *Working Paper*.

Reynolds, K.J., Turner, J.C. and Haslam, S.A., (2000). When are we better than them and they worse than us? a closer look at social discrimination in positive and negative domains. *Journal of Personality and Social Psychology*, *78*(1), p.64.

Ring, P. J. (2012). Trust: A challenge for private pension policy. *Journal of Comparative Social Welfare*, *28*(2), 119-128.

Roach, S., 2003. Coping with the global labor arbitrage. In *Morgan Stanley Global Economic Forum (available at http://www. morganstanley.com)*.

Roels, G., Karmarkar, U.S. and Carr, S., (2010). Contracting for collaborative services. *Management Science,* (*56*:5), pp.849-863.

Rubineau, B., & Kang, Y. (2012). Bias in white: A longitudinal natural experiment measuring changes in discrimination. *Management Science*, *58*(4), 660-677.

Roels, G., Karmarkar, U. S., and Carr, S. (2010) Contracting for Collaborative Services. *Management Science.* 56(5): 849-863.

Sagiv, L., & Schwartz, S. H. (1995). Value priorities and readiness for out-group social contact. *Journal of Personality and Social Psychology*, *69*(3), 437.

Schneider, H. (2010). Moral hazard in leasing contracts: Evidence from the New York City taxi industry. *The Journal of Law and Economics*, *53*(4), 783-805.

Simon, B. and Hamilton, D.L., (1994). Self-stereotyping and social context: the effects of relative in-group size and in-group status. *Journal of Personality and Social Psychology*, *66*(4), 699.

Staats, B.R., Dai, H., Hofmann, D. and Milkman, K.L. (2016) Motivating Process Compliance through Individual Electronic Monitoring: An Empirical Examination of Hand Hygiene in Healthcare. *Management Science* (forthcoming).

Sterling, A.D. and Fernandez, R.M., (2018). Once in the door: gender, tryouts, and the initial salaries of managers. *Management Science (forthcoming)*.

Strong, N., & Xu, X. (2003). Understanding the equity home bias: Evidence from survey data. *Review of Economics and Statistics*, *85*(2), 307-312.

Stuart, T. E. (1998). Network positions and propensities to collaborate: an investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly*, 668-698.

Stuart, E.A., (2010). Matching methods for causal inference: a review and a look forward. *statistical science: A review. Journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Sorenson, O. and Stuart, T.E., (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, *106*(6), 1546-1588.

Sorkin, I., (2017). Ranking firms using revealed preference. *The Quarterly Journal of Economics*. (132:1), pp:1-63.

Stock, J. and Yogo, M., (2005). *Asymptotic distributions of instrumental variables statistics with many instruments* (Vol. 6). Chapter.

Strack, F. and Mussweiler, T., (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*(3), p.437.

Subramanian, H. and Overby, E., 2016. Electronic Commerce, Spatial Arbitrage, and Market Efficiency. *Information Systems Research*. 28(1): 97-116.

Sun, T., Gao, G. G., & Jin, G. Z. (2015). Mobile Messaging for Offline Group Formation in Prosocial Activities: A Large Field Experiment. SSRN.

Susarla, A., Barua, A., and Whinston, A. B. (2009) A Transaction Cost Perspective of the" Software as a Service" Business Model. *Journal of Management Information Systems*. 26(2): 205-240.

Tambe, P., & Hitt, L. M. (2012). The Productivity of Information Technology Investments: New Evidence from IT Labor Data. *Information Systems Research*, *23*(3-part-1), 599-617.

Tang J, Gao H, Hu X, Liu H (2013). Exploiting homophily effect for trust prediction, *Proc. 6th ACM Internat. Conf. Web Search Data Mining* (ACM, New York), pp.53–62.

Tadelis, S. (1999). What's in a Name? Reputation as a Tradeable Asset. *American Economic Review*, *89*(3), 548-563.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, *8*, 321-340.

Todisco, M., (2014). Share and share alike: considering racial discrimination in the nascent room-sharing economy. *Stanford Law Review. Online*, *67*, 121.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Van Alstyne, M., Parker, G., Choudary, S. (2016) Pipelines, Platforms, and the New Rules of Strategy. *Harvard Business Review*. 94(4): 54.

Wang, A., Zhang, M. and Hann, I.H .(2018) Socially nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings. *Information Systems Research* (forthcoming).

Walker, G., & Weber, D. (1984). A Transaction Cost Approach to Make-Or-Buy Decisions. *Administrative Science Quarterly*, 373-391.

Williamson, O. E. (1981). The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology*, *87*(3), 548-577.

Wiswall, M. and Zafar, B., (2014). Determinants of college major choice: Identification using an information experiment, *The Review of Economic Studies*, (*82*:2), pp.791-824.

Wiswall, M. and Zafar, B., (2015). How do college students respond to public information about earnings?. *Journal of Human Capital*, (*9*:2), pp.117-169.

Wiswall, M. and Zafar, B., (2017). Preference for the workplace, investment in human capital, and gender, The *Quarterly Journal of Economics*, (133:1), pp.457-507.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press, P942.

Wolf, H. C. (2000). Intranational home bias in trade. *Review of Economics and Statistics*, *82*(4), 555-563.

Wu, D. J., Ding, M., & Hitt, L. M. (2012). IT implementation contract design: Analytical and experimental investigation of IT value, learning, and contract structure. *Information Systems Research*, *24*(3), 787-801.

Wüllenweber, Kim, et al. (2009). The impact of process standardization on business process outsourcing success. *Information Systems Outsourcing*. Springer, Berlin, Heidelberg, 527-548.

Xu, K., Chan, J., Ghose, A. and Han, S.P. (2016) Battle of the Channels: The Impact of Tablets on Digital Commerce. *Management Science* (forthcoming).

Xue, M., Hitt, L. M., & Chen, P. Y. (2010) Determinants and Outcomes of Internet Banking Adoption. *Management Science.* 57(2): 291-307.

Yoganarasimhan, H. (2013) The Value of Reputation in an Online Freelance Marketplace. *Marketing Science.* 32(6): 860-891.

Zafar, B., (2011). How do college students form expectations?. *Journal of Labor Economics*, (*29*:2), pp.301-348.

Zhang, Z. and Li, B., (2017) A Quasi-experimental Estimate of the Impact of P2P Transportation Platforms on Urban Consumer Patterns. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1683-1692.

APPENDIX A

BALANCE CHECK FOR PROPENSITY SCORE MATCHING

Table A1. Balance Check for Propensity Score Matching[48]

Variable	Sample	Mean		%bias	% reduced |bias|	T-test	
		Treated	Control			t	p>|t|
Desc_length (/100)	Unmatched	3.51	4.42	-20.50		-9.18	0.00
	Matched	3.51	3.64	-2.90	86.00	-1.38	0.17
Title_length	Unmatched	5.67	5.58	3.10		1.48	0.14
	Matched	5.67	5.67	0.20	94.90	0.06	0.95
Software	Unmatched	0.32	0.33	-1.10		-0.55	0.58
	Matched	0.32	0.33	-0.70	42.80	-0.26	0.80
Design	Unmatched	0.09	0.09	-1.70		-0.79	0.43
	Matched	0.09	0.09	-0.60	66.20	-0.22	0.83
Writing	Unmatched	0.15	0.12	7.70		3.79	0.00
	Matched	0.15	0.14	0.30	96.30	0.11	0.91
Marketing	Unmatched	0.05	0.04	3.10		1.53	0.13
	Matched	0.05	0.04	1.10	64.50	0.43	0.67
Administrative	Unmatched	0.06	0.04	11.50		5.89	0.00
	Matched	0.06	0.07	-0.70	93.90	-0.25	0.81
Translation	Unmatched	0.02	0.03	-0.90		-0.44	0.66
	Matched	0.02	0.03	-1.00	-12.90	-0.41	0.68
Employer_tenure_month	Unmatched	30.21	30.18	0.10		0.06	0.95
	Matched	30.21	30.25	-0.10	-0.80	-0.05	0.96
Employer_overall_rating	Unmatched	4.92	4.92	-0.50		-0.22	0.82
	Matched	4.92	4.92	-0.90	-198.5	-0.34	0.73
Article writing	Unmatched	0.11	0.06	18.30		9.48	0.00
	Matched	0.11	0.12	-2.40	86.70	-0.84	0.40
Php	Unmatched	0.23	0.23	1.30		0.62	0.53
	Matched	0.23	0.24	-1.60	-23.60	-0.62	0.54
Article rewriting	Unmatched	0.05	0.03	8.70		4.44	0.00
	Matched	0.05	0.05	-0.70	92.10	-0.25	0.81
Ghost writing	Unmatched	0.04	0.02	9.40		4.87	0.00
	Matched	0.04	0.04	-0.80	91.90	-0.26	0.79
Video services	Unmatched	0.01	0.01	-2.90		-1.33	0.18
	Matched	0.01	0.01	0.80	73.60	0.33	0.74
Blog	Unmatched	0.03	0.01	9.30		4.85	0.00
	Matched	0.03	0.03	-1.50	84.30	-0.49	0.62
Website design	Unmatched	0.15	0.13	5.90		2.87	0.00
	Matched	0.15	0.16	-2.30	61.20	-0.86	0.39
Technical writing	Unmatched	0.04	0.02	7.90		4.04	0.00
	Matched	0.04	0.04	-0.30	96.40	-0.10	0.92
Cms	Unmatched	0.00	0.00	-6.60		-2.75	0.01
	Matched	0.00	0.00	-1.00	84.40	-0.64	0.52

[48] We match fixed-price projects with time-based projects by using the Nearest Neighbor (4) matching method. In order to construct a more homogenous sample, we limit our sample to projects with the common public auction format. Therefore, those projects which require NDA contracts, are featured or sealed, are fulltime jobs, use a non-dollar currency, are not written in English are dropped.

Aftereffects	Unmatched	0.00	0.01	-4.10		-1.85	0.07
	Matched	0.00	0.00	1.10	72.00	0.57	0.57
Shopping carts	Unmatched	0.01	0.01	-0.80		-0.38	0.71
	Matched	0.01	0.01	-0.60	21.90	-0.24	0.81
Report writing	Unmatched	0.01	0.02	-2.30		-1.10	0.27
	Matched	0.01	0.01	1.80	20.90	0.79	0.43
Action script	Unmatched	0.00	0.00	0.90		0.43	0.67
	Matched	0.00	0.00	0.00	100.00	0.00	1.00
Adobe flash	Unmatched	0.01	0.01	-0.50		-0.25	0.81
	Matched	0.01	0.01	-0.20	60.70	-0.08	0.94
Xml	Unmatched	0.00	0.00	0.00		0.01	0.99
	Matched	0.00	0.00	-0.50	-2461.80	-0.19	0.85
Ajax	Unmatched	0.01	0.01	1.40		0.68	0.49
	Matched	0.01	0.01	-1.20	14.20	-0.44	0.66
Captcha	Unmatched	0.00	0.00	-0.40		0.18	0.86
	Matched	0.00	0.00	0.90	-132.70	0.39	0.70
Grow	Unmatched	0.00	0.00	2.00		-0.91	0.36
	Matched	0.00	0.00	0.40	78.40	0.21	0.83
Overview	Unmatched	0.00	0.01	8.10		-3.27	0.00
	Matched	0.00	0.00	0.90	88.80	-0.73	0.47
Solid	Unmatched	0.00	0.01	2.50		-1.09	0.28
	Matched	0.00	0.00	0.20	91.60	-0.10	0.92
Blank	Unmatched	0.00	0.01	2.90		-1.27	0.21
	Matched	0.00	0.01	1.40	51.60	-0.63	0.53
Voic	Unmatched	0.01	0.01	1.50		-0.73	0.47
	Matched	0.01	0.01	1.50	3.00	-0.58	0.56
Upgrad	Unmatched	0.01	0.01	3.70		-1.50	0.13
	Matched	0.01	0.01	0.40	90.40	-0.32	0.75
Drop	Unmatched	0.01	0.02	3.60		-1.63	0.10
	Matched	0.01	0.01	0.50	85.10	-0.23	0.82
Load	Unmatched	0.01	0.02	8.40		-3.58	0.00
	Matched	0.01	0.01	0.80	90.40	-0.47	0.64
Team	Unmatched	0.03	0.02	3.90		1.86	0.06
	Matched	0.03	0.02	2.50	35.20	1.00	0.32
Valu	Unmatched	0.02	0.02	3.20		-1.42	0.15
	Matched	0.02	0.01	1.40	57.20	0.61	0.54
Full	Unmatched	0.03	0.04	3.70		-1.74	0.08
	Matched	0.03	0.03	0.80	77.70	0.36	0.72
Menu	Unmatched	0.02	0.04	7.40		-3.20	0.00
	Matched	0.02	0.02	0.20	97.90	0.09	0.93
Market	Unmatched	0.03	0.04	3.80		-1.67	0.10
	Matched	0.03	0.03	0.10	97.70	0.04	0.97
Written	Unmatched	0.04	0.05	5.70		-2.62	0.01
	Matched	0.04	0.04	1.20	79.10	-0.52	0.60
Part	Unmatched	0.04	0.05	5.60		-2.48	0.01
	Matched	0.04	0.04	0.80	85.50	-0.37	0.71
Field	Unmatched	0.03	0.06	7.20		-3.30	0.00
	Matched	0.03	0.03	0.20	97.90	-0.08	0.94

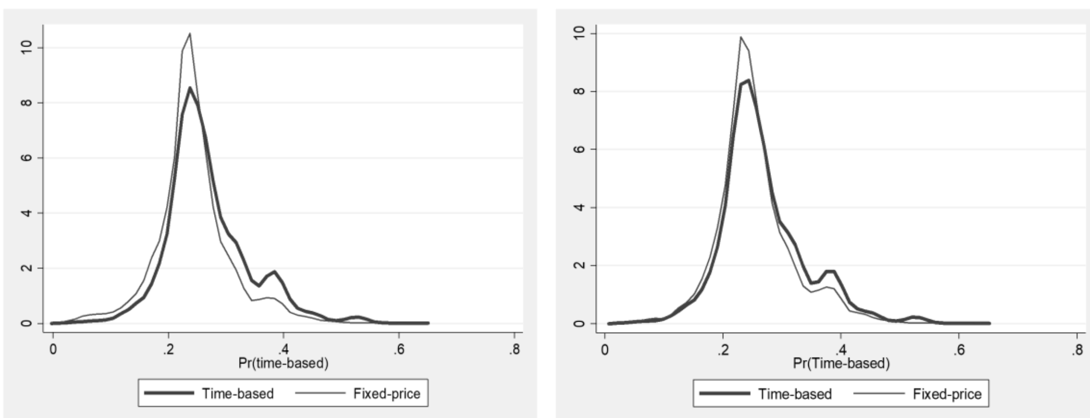
Check	Unmatched	0.05	0.06	2.10		-0.98	0.33
	Matched	0.05	0.05	0.20	92.60	0.06	0.95
Address	Unmatched	0.04	0.07	7.00		-3.08	0.00
	Matched	0.04	0.04	0.20	97.80	-0.08	0.94
Excel	Unmatched	0.06	0.07	0.90		-0.41	0.68
	Matched	0.06	0.06	0.40	58.60	-0.15	0.88

Notes: (a) Results of Nearest Neighbor (4) Matching Method are presented. We also conducted robustness checks with other matching algorithms in the additional analysis section. The result is qualitatively consistent. (b) Within the matched sample, the group means of all the month dummies are not significantly different between time-based projects and fixed-price projects.

Figure A13. Distribution of Propensity Scores for Time-based Projects and Fixed-price Projects

(a) Before Matching

(b) After Matching



APPENDIX B

RELATIVE TIME MODEL

In order to further test the parallel trend assumption of the DID model (Angrist and Pischke 2008), we employ the relative time model test to assess whether time-based projects and fixed-price projects have a common trend during the pre-treatment period. This analysis also allows us to check at what time the effects start to emerge. We specify the relative time model as follows:

$$y_{ij} = \alpha + \rho\tau_t + \mu Time\_based_j + \beta(\tau_t \times Time\_based_j) + \gamma_i + \delta_j + \varepsilon_{ij} \tag{A1}$$

where y_{ij} represents the dependent variables of our interest, including Bid_Count_{ij} and $Pct\_no\_rating_{ij}$. τ_t represents a vector of time dummies and $\{\beta\}$ denotes the matrix of relative time parameters to be estimated for project j posted by employer i whose auction duration ends at time t. If there exists a pre-treatment trend, we should observe significant relative time parameters before the introduction of the monitoring system. Following Autor (2003)'s approach, we use projects whose auctions end at the week before the change (the last week of January 2014) as the baseline since the monitoring system introduction happened on February 5th, 2014. We visualize the results in Figure A2. The analysis shows that all the relative time parameters are insignificant prior to the introduction while some of the relative time parameters in these two models are significant after February 2014 wherein *Freelancer* introduced the IT-enabled monitoring system. As such, the result of the relative time model lends further support to the validity of the parallel trend assumption and also to our main findings.

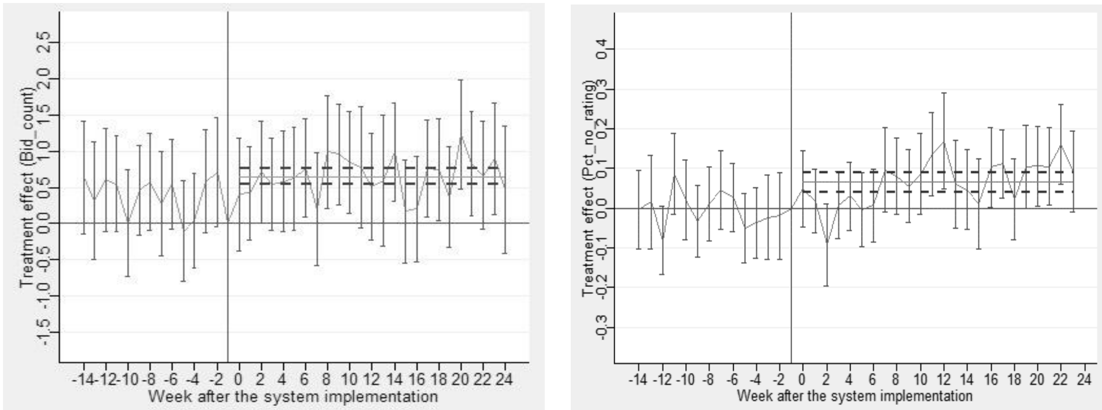


Figure A2. Coefficients of the Weekly Dynamic Difference-in-Differences Estimates

Note: The dash vertical line denotes the week in which *Freelancer* first introduced the monitoring system (February 2014). Error bars represent the 90% confidence intervals using clustered standard errors.

APPENDIX C

ADDITIONAL ANALYSIS REGARDING PROJECT OUTCOMES

Given our main analyses suggesting that the introduction of the monitoring system reduces employers' preference for effort-related reputation and lowers the entry barrier, an interesting follow-up questions of the key interest of policy makers is that what is the impact of this policy change on project outcomes. Especially, if project outcomes are worse than before, then the long-run impact of the introduction may have a negative impact on the growth of the platform. In fact, there is a stream of literature on the traditional workplace suggesting that reducing privacy can demotivate workers and subsequently reduce their productivity (Bernstein 2012, 2014). It is interesting to test whether the introduction of the monitoring system cannot influence the matching between employers and workers in the online labor market, but also affect the eventual project outcomes, which is important to the long-turn growth of the platform. Therefore, we employ the DID model to the same sample used in our main analysis and investigate the impact of monitoring on various measures of project outcomes, such as whether the project is completed, the completion time, and various rating measures. In particular, given that we do not have an accurate measurement of completion time, we use the time gap between the date when the awarded bid was submitted and the date when the employer wrote the review as the proxy measure of completion time. As Table A2 and Table A3 show, there is no significant change in various measures of project outcomes, including project completion, completion time, or any rating measures. Overall, the result suggests that although the entry barrier has been decreased by the introduction of the monitoring system, this policy change does not seem to have a negative impact on project outcomes.

Table A2. Estimation Results of the DID Models on Project Completion

Model	(1)		(2)		(3)		(4)	
Dependent variable	Completion		Log_completion_day		Rating_score		Bad_rating_dummy	
Time_based	-0.003*	(0.002)	1.069***	(0.153)	-0.026	(0.055)	-0.005	(0.028)
Time_based*After	-0.003	(0.007)	-0.066	(0.131)	0.020	(0.053)	0.001	(0.024)
Log_budget_max	-0.001	(0.002)	0.296***	(0.042)	0.003	(0.015)	-0.006	(0.007)
Log_title_length	0.001	(0.002)	-0.018	(0.083)	0.018	(0.032)	-0.008	(0.016)
Log_desc_length	0.002	(0.001)	0.078**	(0.039)	-0.020	(0.014)	0.016**	(0.007)
Category dummies	Yes		Yes		Yes		Yes	
Month dummies	Yes		Yes		Yes		Yes	
Employer dummies	Yes		Yes		Yes		Yes	
Clusters(employers)	1,261		1,261		1,261		1,261	
Observations	2,976		2,976		2,976		2,976	
R-squared	0.009		0.068		0.009		0.016	

Notes: (a) Robust standard errors clustered on employers are reported in parentheses. (b) The results are consistent if we use 1th, 5th, or 10th percentile of the chi-squared distribution of the Mahalanobis distance from the iterative basic set as a threshold to separate outliers from nonoutliers. (c) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table A3. Estimation Results of the DID Models on Project Ratings

Model	(1)		(2)		(3)		(4)		(5)	
Dependent variable	Communication		Expertise		Hire_again		Quality		Professionalism	
Time_based	-0.043	(0.060)	-0.014	(0.054)	-0.000	(0.061)	0.006	(0.055)	0.002	(0.052)
Time_based*After	0.014	(0.050)	-0.032	(0.052)	0.006	(0.061)	-0.015	(0.050)	-0.026	(0.052)
Log_budget_max	0.004	(0.014)	0.006	(0.013)	0.020	(0.016)	0.013	(0.012)	0.008	(0.014)
Log_title_length	0.016	(0.031)	0.001	(0.032)	0.027	(0.037)	-0.007	(0.034)	0.019	(0.031)
Log_desc_length	-0.020	(0.013)	-0.020	(0.014)	-0.023	(0.016)	-0.031**	(0.014)	-0.018	(0.013)
Category dummies	Yes		Yes		Yes		Yes		Yes	
Month dummies	Yes		Yes		Yes		Yes		Yes	
Employer dummies	Yes		Yes		Yes		Yes		Yes	
Clusters(employers)	1,259		1,259		1,259		1,259		1,259	
Observations	2,970		2,970		2,970		2,970		2,970	
R-squared	0.013		0.014		0.013		0.018		0.013	

Notes: (a) Robust standard errors clustered on employers are reported in parentheses. (b) The results are consistent if we use 1th, 5th, or 10th percentile of the chi-squared distribution of the Mahalanobis distance from the iterative basic set as a threshold to separate outliers from nonoutliers. (c) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

APPENDIX D

ROBUSTNESS CHECK WITH DIFFERENT OBSERVATIONAL WINDOWS

To show the robustness our findings, we rerun the model with a shorter observational window (six months before and after) to see if we still find a similar treatment effect for the introduction of monitoring systems. As Table A4 shows, we find that employers are subject to home bias and that their home bias decreases significantly after the monitoring system introduction, based on both the full sample panel and the matched sample panel during a shorter observational window.

Table A4. Estimation Results Based on a Different Observational Windows

Sample	Full sample (Six months before and after)				Matched sample (Six months before and after)			
Model	Logit		LPM		Logit		LPM	
	DV: whether the bidder is awarded							
Homecountry	0.279***	(0.101)	0.021**	(0.009)	0.351*	(0.181)	0.051**	(0.023)
Time-based×Homecountry	0.773***	(0.237)	0.094***	(0.026)	0.647**	(0.300)	0.067*	(0.036)
After×Homecountry	-0.007	(0.121)	0.006	(0.011)	-0.098	(0.217)	-0.011	(0.030)
Time-based×After ×Homecountry	-0.857**	(0.357)	-0.112***	(0.041)	-0.922**	(0.451)	-0.123**	(0.055)
Same language	0.406***	(0.041)	0.016***	(0.002)	0.440***	(0.071)	0.028***	(0.005)
Same currency	0.057*	(0.033)	0.005***	(0.002)	0.067	(0.055)	0.007	(0.005)
Same time zone	0.249***	(0.088)	0.025***	(0.008)	0.523***	(0.145)	0.075***	(0.019)
Log bid price	-1.862***	(0.028)	-0.105***	(0.002)	-2.013***	(0.051)	-0.161***	(0.004)
Log milestone percentage	-0.058***	(0.022)	-0.003***	(0.001)	-0.136***	(0.037)	-0.011***	(0.003)
Log count rating	0.091***	(0.011)	0.005***	(0.001)	0.100***	(0.018)	0.008***	(0.002)
Log avg rating	0.077***	(0.014)	0.002***	(0.001)	0.049**	(0.023)	0.002	(0.002)
Log bid order rank	-0.350***	(0.020)	-0.020***	(0.001)	-0.328***	(0.037)	-0.028***	(0.003)
Preferred freelancer	0.508***	(0.027)	0.031***	(0.002)	0.469***	(0.047)	0.045***	(0.005)
Bidder country dummy	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	155,471		155,471		35,110		35,110	
R-squared	0.488		0.049		0.488		0.079	
LogLik	-20,815				-6,225			
AIC	41,704				12,524			
BIC	42,072				12,837			
Number of projects	10,925		10,925		4,003		4,003	

Notes: a) All bids which submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared in the logit model is calculated based on the maximum likelihood R-squared; f)* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

APPENDIX E

ROBUSTNESS CHECK REGARDING DUAL-TYPE WORKERS

Further, to reduce the potential difference between the labor supply for time-based and fixed-price projects, we reexamine our DID estimate by limiting our sample to those bids which are submitted by workers bidding on both fixed-price and time-based projects (dual-type workers). This sampling approach helps ensure that workers are comparable and similar between the treatment group and the control group. Again, the results of the restricted sample are still highly consistent with our main findings.

Table A5. Estimation of Employers' Home Bias Based on Dual-type Workers

Sample	Dual-type sample	
Model	Logit	LPM
	DV: whether the bidder is awarded	
Homecountry	0.387***(0.054)	0.036***(0.005)
Same language	0.407***(0.029)	0.019***(0.001)
Same currency	0.051** (0.023)	0.004***(0.001)
Same time zone	0.240***(0.065)	0.025***(0.006)
Log bid price	-1.779***(0.020)	-0.101***(0.001)
Log milestone percentage	-0.096***(0.017)	-0.006***(0.001)
Log count rating	0.085***(0.008)	0.005***(0.000)
Log avg rating	0.112***(0.010)	0.004***(0.001)
Log_bid_order_rank	-0.307***(0.015)	-0.017***(0.001)
Preferred freelancer	0.476***(0.019)	0.028***(0.001)
Country dummy	Yes	Yes
Project fixed effects	Yes	Yes
Observations	297,724	297,724
R-squared	0.472	0.047
LogLik	-40,433	
AIC	80,935	
BIC	81,295	
Number of projects	21,129	21,129

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included into our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared in the logit model is calculated based on the maximum likelihood R-squared; f) * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table A6. DID Estimation Results Based on Dual-type Workers

Sample	Full sample		Matched sample	
Model	Logit	LPM	Logit	LPM
	DV: whether the bidder is awarded			
Homecountry	0.254***(0.091)	0.024***(0.009)	0.446***(0.162)	0.071***(0.024)
Time-based× Homecountry	0.773***(0.185)	0.094***(0.023)	0.625***(0.241)	0.064* (0.034)
After×Homecountry	0.174* (0.101)	0.020* (0.010)	0.009 (0.183)	-0.001 (0.027)
Time-based ×After ×Homecountry	-1.212***(0.245)	-0.140***(0.028)	-1.080***(0.312)	-0.134***(0.042)
Same language	0.428***(0.029)	0.020***(0.001)	0.397***(0.048)	0.031***(0.004)
Same currency	0.046* (0.023)	0.004***(0.001)	0.072* (0.038)	0.007** (0.004)
Same time zone	0.239***(0.067)	0.026***(0.006)	0.386***(0.104)	0.054***(0.013)
Log bid price	-1.783***(0.020)	-0.105***(0.001)	-1.865***(0.035)	-0.158***(0.003)
Log milestone percentage	-0.094***(0.017)	-0.006***(0.001)	-0.238***(0.028)	-0.020***(0.002)
Log count rating	0.082***(0.008)	0.005***(0.000)	0.082***(0.013)	0.007***(0.001)
Log avg rating	0.063***(0.011)	0.002***(0.001)	0.037**(0.017)	0.001 (0.001)
Log_bid_order_rank	-0.305***(0.015)	-0.017***(0.001)	-0.312***(0.026)	-0.026***(0.002)
Preferred freelancer	0.473***(0.019)	0.028***(0.001)	0.439***(0.033)	0.042***(0.004)
Country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	285,187	285,187	67,553	67,553
R-squared	0.451	0.047	0.448	0.075
Number of projects	20,786	20,786	7,857	7,857

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is only limited to projects with only one winner. b) Log (bidder tenure) is not included into our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price to the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared in the logit model is calculated based on the maximum likelihood R-squared; f)* p<0.1, ** p<0.05, *** p<0.01.

APPENDIX F

ROBUSTNESS CHECK WITH ALTERNATIVE MATCHING METHOD

We further employ the PSM method to regenerate a comparable sample. We find that after the matching, the related covariates are not significantly different between the two groups (Table A7). The results of our DID estimate based on the matched sample are still highly consistent with our main findings.

Table A7. Balance Check for Propensity Score Matching[49]

Variable	Sample	Mean		%bias	% reduced |bias|	t-test	
		Treated	Control			t	p>|t|
Project desc length	Unmatched	66.322	89.535	-28.400	91.500	-13.710	0.000
	Matched	66.322	68.302	-2.400		-1.130	0.260
Paid amount	Unmatched	209.030	183.670	3.600	65.700	2.880	0.004
	Matched	209.030	217.750	-1.200		-0.390	0.697
Title length	Unmatched	5.921	5.852	2.200	27.000	1.180	0.238
	Matched	5.921	5.971	-1.600		-0.620	0.536
Employer tenure month	Unmatched	24.618	29.153	-15.300	94.200	-7.710	0.000
	Matched	24.618	24.356	0.900		0.370	0.710
Bid count	Unmatched	14.400	17.732	-21.500	98.400	-10.620	0.000
	Matched	14.400	14.455	-0.400		-0.160	0.875
Median bid ratio	Unmatched	0.772	0.804	-6.100	99.000	-2.590	0.010
	Matched	0.772	0.772	0.100		0.030	0.973
Employer overall rating	Unmatched	4.906	4.926	-4.700	80.900	-2.720	0.006
	Matched	4.906	4.902	0.900		0.320	0.750

Notes: a) Results of Nearest Neighbor (1) Matching Method are presented. b) Within the matched sample, the group means of all the month dummies are not significantly different between time-based projects and fixed-price projects. Balance checks of all the month dummies are omitted for brevity.

Table A8. Estimation Results of Linear Probability Model and Conditional Logit Model

Sample	Full sample		Matched sample	
Model	Logit	LPM	Logit	LPM
Homecountry	0.219**(0.076)	0.016***(0.006)	-0.138 (0.246)	-0.008 (0.019)
Time-based×Homecountry	0.759**(0.175)	0.089***(0.020)	1.144***(0.288)	0.116***(0.027)
After×Homecountry	0.223**(0.083)	0.020***(0.007)	0.705** (0.292)	0.064** (0.026)
Time-based×After ×Homecountry	-1.116**(0.232)	-0.119***(0.025)	-1.599***(0.374)	-0.162***(0.036)

[49] We match fixed-price projects with time-based projects by using the Nearest Neighbor (1) matching method. In order to reduce the potential effect of various auction types, we limit our sample to projects with the common public auction format and exclude projects which require NDA contracts, are featured or sealed, etc.

Same language	0.415** (0.026)	0.018*** (0.001)	0.530*** (0.070)	0.026*** (0.003)
Same currency	0.063** (0.021)	0.004*** (0.001)	-0.015 (0.054)	-0.000 (0.004)
Same time zone	0.250** (0.057)	0.025*** (0.005)	0.263* (0.134)	0.030** (0.013)
Log bid price	-1.736** (0.018)	-0.090*** (0.001)	-1.374*** (0.049)	-0.087*** (0.003)
Log milestone percentage	-0.066** (0.016)	-0.003*** (0.001)	-0.378*** (0.032)	-0.024*** (0.002)
Log review count	0.096** (0.007)	0.005*** (0.000)	0.123*** (0.018)	0.008*** (0.001)
Log avg rating	0.103** (0.009)	0.003*** (0.000)	0.082*** (0.023)	0.002 (0.001)
Log bid order rank	-0.313** (0.014)	-0.016*** (0.001)	-0.197*** (0.035)	-0.011*** (0.002)
Preferred freelancer	0.471** (0.018)	0.026*** (0.001)	0.636*** (0.046)	0.046*** (0.004)
Bidder country dummy	Yes	Yes	Yes	Yes
Project fixed effects	Yes	Yes	Yes	Yes
Observations	371,968	371,968	46,655	46,655
R-squared	0.489	0.047	0.431	0.075
LogLik	-47,545		-6,845	
AIC	95,163		13,764	
BIC	95,564		14,087	
Number of projects	23,943	23,943	3,563	3,563

Notes: a) The results are estimated based on the matched sample with the CEM approach. The results are highly consistent if we estimate the model based on the full sample. b) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. c) Log(bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log(bidder tenure) instead of Log(count rating). d) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. e) Robust standard errors clustered by projects are reported in parentheses. f) R-squared in the logit model is calculated based on the maximum likelihood R-squared. g) * p<0.1, ** p<0.05, *** p<0.01.

APPENDIX G

ROBUSTNESS CHECK WITH MORE CONTROLS

Next, we explore the robustness of our DID estimate by adding more time-varying or project-specific controls. Specifically, we control for potential time-varying variations in levels of competitiveness and "market tightness" from each worker's fellow countrymen (e.g. the number of bidders from the same country and the average rating of bidders from the same country who bid on the specific project). Second, we add the country-month two-way fixed effect as the control for the country-specific time trend on the supply side. Tables A9 and A10 summarize the results of these two robustness checks, respectively. Overall, the results are consistent with our main findings.

Table A9. Estimation Results of Linear Probability Model and Conditional Logit Model

Sample	Full sample				Matched sample			
Model	Logit		LPM		Logit		LPM	
Homecountry	0.226***	(0.076)	0.015**	(0.006)	0.314**	(0.138)	0.040**	(0.016)
Time-based×Homecountry	0.743***	(0.174)	0.088***	(0.020)	0.694***	(0.222)	0.080***	(0.027)
After×Homecountry	0.218***	(0.083)	0.019***	(0.007)	0.174	(0.151)	0.016	(0.018)
Time-based×After ×Homecountry	-1.103***	(0.230)	-0.120***	(0.025)	-1.071***	(0.286)	-0.130***	(0.034)
Same language	0.412***	(0.026)	0.017***	(0.001)	0.383***	(0.043)	0.026***	(0.003)
Same currency	0.065***	(0.021)	0.004***	(0.001)	0.086**	(0.035)	0.008***	(0.003)
Same time zone	0.250***	(0.057)	0.025***	(0.005)	0.402***	(0.090)	0.051***	(0.010)
Log bid price	-1.735***	(0.018)	-0.090***	(0.001)	-1.849***	(0.032)	-0.139***	(0.002)
Log milestone percentage	-0.067***	(0.016)	-0.003***	(0.001)	-0.200***	(0.026)	-0.016***	(0.002)
Log count rating	0.097***	(0.007)	0.005***	(0.000)	0.092***	(0.012)	0.007***	(0.001)
Log avg rating	0.102***	(0.010)	0.001***	(0.000)	0.065***	(0.015)	0.000	(0.001)
Log bid order rank	-0.312***	(0.013)	-0.016***	(0.001)	-0.330***	(0.024)	-0.025***	(0.002)
Preferred freelancer	0.474***	(0.018)	0.026***	(0.001)	0.456***	(0.032)	0.041***	(0.003)
Log avg country rating	0.001	(0.011)	0.009***	(0.001)	0.043**	(0.017)	0.012***	(0.001)
Log country bidder	-0.093***	(0.015)	-0.002	(0.001)	-0.129***	(0.027)	-0.008***	(0.002)
Bidder country dummies	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	371,968		371,968		86,840		86,840	
R-squared	0.490		0.045		0.487		0.072	
LogLik	-47,526				-14,809			
AIC	95,129				29,696			
BIC	95,551				30,061			
Number of projects	23,943		23,943		9,028		9,028	

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared in the logit model is calculated based on the maximum likelihood R-squared; f) * p<0.1, ** p<0.05, *** p<0.01.

Table A10. Estimation Results of Linear Probability Model and Conditional Logit Model

Sample	Full sample				Matched sample			
Model	Logit		LPM		Logit		LPM	
Homecountry	0.306***	(0.084)	0.021***	(0.007)	0.377**	(0.156)	0.047***	(0.017)
Time-based×Homecountry	0.732***	(0.180)	0.084***	(0.020)	0.731***	(0.235)	0.076***	(0.027)
After×Homecountry	0.087	(0.096)	0.011	(0.008)	0.049	(0.174)	0.005	(0.019)
Time-based×After ×Homecountry	-1.043***	(0.235)	-0.113***	(0.025)	-1.053***	(0.299)	-0.123***	(0.034)
Same language	0.418***	(0.027)	0.017***	(0.001)	0.394***	(0.045)	0.025***	(0.003)
Same currency	0.068***	(0.022)	0.004***	(0.001)	0.096***	(0.036)	0.009***	(0.003)
Same time zone	0.248***	(0.057)	0.024***	(0.005)	0.417***	(0.093)	0.051***	(0.010)
Log bid price	-1.744***	(0.018)	-0.090***	(0.001)	-1.870***	(0.033)	-0.138***	(0.002)
Log milestone percentage	-0.072***	(0.016)	-0.003***	(0.001)	-0.213***	(0.027)	-0.016***	(0.002)
Log count rating	0.098***	(0.007)	0.005***	(0.000)	0.093***	(0.012)	0.007***	(0.001)
Log avg rating	0.101***	(0.010)	0.001***	(0.000)	0.064***	(0.016)	0.000	(0.001)
Log bid order rank	-0.314***	(0.014)	-0.016***	(0.001)	-0.338***	(0.025)	-0.026***	(0.002)
Preferred freelancer	0.477***	(0.018)	0.026***	(0.001)	0.459***	(0.032)	0.040***	(0.003)
Log avg country rating	0.005	(0.012)	0.009***	(0.001)	0.052***	(0.017)	0.012***	(0.001)
Log country bidder	-0.079***	(0.015)	-0.002	(0.001)	-0.107***	(0.028)	-0.009***	(0.003)
Bidder country and month two-way fixed effects	Yes		Yes		Yes		Yes	
Project fixed effects	Yes		Yes		Yes		Yes	
Observations	371,968		371,968		86,840		86,840	
R-squared	0.270		0.048		0.522		0.082	
LogLik	-51,816				-14,495			
AIC	104,806				30,146			
BIC	111,161				35,563			
Number of projects	23,943		23,943		9,028		9,028	

Notes: a) All bids submitted by workers having prior collaboration experience with the employer are dropped. Moreover, our sample is limited to projects with only one winner. b) Log (bidder tenure) is not included in our model because it is highly correlated with log (count rating). The results are highly consistent if we include Log (bidder tenure) instead of Log (count rating). c) The results are highly consistent if we control for the country-month two-way fixed effect and the potential time-varying levels of competitiveness from a worker's fellow countrymen. The results are highly consistent if we limit our sample to bids submitted by workers who bid on both fixed-price and time-based projects (named as "dual-type workers") (Lin et al. 2016). The results are highly consistent if we include the original bid price instead of the log-transformed bid price in the model. d) Robust standard errors clustered by projects are reported in parentheses; e) R-squared in the logit model is calculated based on the maximum likelihood R-squared; f) * p<0.1, ** p<0.05, *** p<0.01.

BIOGRAPHICAL SKETCH

Chen Liang is a Ph.D. candidate in Information Systems at the W.P. Carey School of Business at Arizona State University. Her research interests focus on gig economy, platform economics, and gender wage gap. She holds a bachelor's degree in Management Information Systems and a master's degree in Information Science from Sun Yat-sen University. Her work has appeared in *Management Science* and *Information Systems Research*. She has received several academic awards as well as research grants, such as the WISE Best Paper Award, NSF dissertation improvement grant, NET Institute grant.