Submitted to *Management Science*
manuscript MS-INS-19-00733.R3

Learning to Successfully Hire in Online Labor Markets

Marios Kokkodis, Sam Ransbotham
Boston College, kokkodis@bc.edu, sam.ransbotham@bc.edu,

(Management Science, Forthcoming)

Hiring in online labor markets involves considerable uncertainty: which hiring choices are more likely to yield successful outcomes, and how do employers adjust their hiring behaviors to make such choices? We argue that employers will initially explore the value of available information. When employers observe successful outcomes, they will keep reinforcing their hiring strategies; but when the outcomes are unsuccessful, employers will adjust their hiring behaviors. To investigate these dynamics, we propose a two-component framework that links hiring choices with task outcomes. The framework's first component, a Hidden Markov Model, captures how employers transition from unsuccessful to successful hiring decisions. The framework's second component, a conditional logit model, estimates employer hiring choices. Analysis of 238,364 hiring decisions from a large online labor market shows that, often, employers initially explore cheaper contractors with a lower reputation. When these options result in unsuccessful outcomes, employers learn and adjust their hiring behaviors to rely more on reputable contractors who are not as cheap. Such hirings tend to be successful, guiding employers to reinforce their hiring processes. As a result, the market observes employers transition from cheaper, lower-reputation options with poorer performance to more expensive reputable options with better performance. We attribute part of this behavior to employer confidence and risk attitude, which can change over time. This work is the first to investigate how employers learn to make successful hiring choices in online labor markets. As a result, it provides platform managers with new knowledge and analytics tools to target employer interventions.

---

1. Introduction

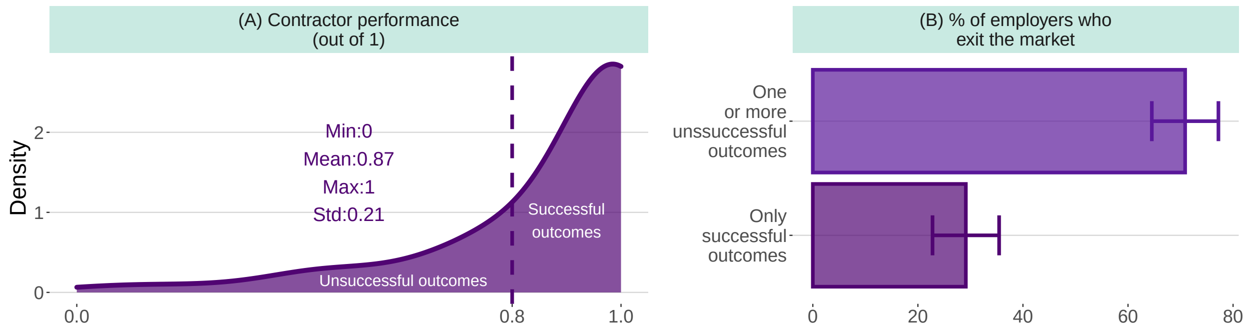
Online platforms overcome many market inefficiencies by facilitating transactions between consumers and producers throughout the world. Online labor markets, such as Upwork and Freelancer, are prime examples; they connect independent employers with contractors who accomplish diverse tasks. These platforms boomed over the past decade: Upwork, for example, now has more than fourteen million contractors and more than five million employers; these contractors complete three million tasks annually with a total transaction volume of over one billion U.S. dollars (Lauren 2017, Brier and Pearson 2018). This growth will likely continue (if not accelerate) in the future as the sharing economy continues to structure the future of work (Sundararajan 2016, Jain et al. 2018).

Uncertainty in these markets makes hiring difficult. While online markets are similar to their offline counterparts, experience using offline markets does not entirely translate to these growing platforms. Available information, for example, differs between offline and online markets. To identify the best candidate, employers in both markets must assess both observed and latent contractor characteristics. Observed characteristics include a contractor's skillset, work history, and certifications. Latent characteristics include a contractor's true knowledge and abilities (Geva and Saar-Tsechansky 2016). But employers have additional information available in online markets, particularly for signals of latent characteristics (e.g., worker reputation; Kokkodis and Ipeirotis 2016). The existence of latent characteristics and heterogeneity in the observed ones (Kokkodis and Ipeirotis 2014) creates an uncertain environment where hiring decisions rely on idiosyncratic assessments of fit between task and contractor characteristics.

This uncertainty in hiring decisions often results in unsuccessful hirings that are costly for the employer—and the platform. Figure 1A illustrates the distribution of task outcomes (i.e., the normalized contractor performance) in an online labor market. On average, contractors perform well (mean performance is 0.87 out of 1). However, a significant portion (27%) of hiring choices results in outcomes where contractor performance is well below 0.8. These unsuccessful outcomes are costly for the marketplace, as they discourage employers from continuing to participate on the platform (Tripp and Grégoire 2011). In fact, 71% of employers who exited the market in our data (i.e., did not post a new task for more than three consecutive months) had one or more unsuccessful outcomes; only 29% of those employers had only successful hiring choices (Figure 1B).

Given how costly unsuccessful hiring outcomes are, it is crucial for employers and platform managers to understand what leads to successful decisions and whether (and how) employers learn

Figure 1 Uncertainty in hiring decisions yields unsuccessful and costly outcomes



Employers often make hiring choices that result in unsuccessful outcomes (Figure A). Such outcomes are costly, as they discourage employers from participating in the market (Figure B).

and improve their hiring behaviors. Despite extensive research on the drivers of hiring decisions in online labor markets (Yoganarasimhan 2013, Pallais 2014, Moreno and Terwiesch 2014, Kokkodis and Ipeirotis 2016, Lin et al. 2016, Gefen et al. 2016), we know very little about how these hiring decisions connect to task outcomes. For example, employers tend to hire reputable contractors (Lin et al. 2016, Yoganarasimhan 2013, Hong and Pavlou 2017) and are willing to trade off price and reputation (Moreno and Terwiesch 2014). Yet, what is not known, is how these hiring preferences affect hiring outcomes and, more importantly, how employers learn to adjust their behaviors to achieve better outcomes. Hence, we ask:

Which hiring choices are more likely to yield successful outcomes, and how do employers adjust their hiring behaviors to make such choices?

To investigate, we focus on how employers adjust their assessment of characteristics that affect hiring outcomes in online labor markets. We argue that market participants will initially explore the value of platform information (Busenitz and Barney 1997, Dequech 1999, Peterson and Pitz 1988). While exploring, some employers might be prone to try lower-cost options that, on average, have lower reputation (Moreno and Terwiesch 2014). Sometimes, such hires will be successful; but, on average, they will result in poorer outcomes (Kokkodis and Ipeirotis 2016). Research on individual exploitation and exploration suggests that when employers observe successful outcomes, they will keep exploiting knowledge they have developed and reinforce their hiring strategies; but when employers observe unsuccessful outcomes, they will adjust their hiring behaviors and keep exploring market information (Mom et al. 2007, Lee and Meyer-Doyle 2017, Lee 2019). Combined, these dynamics suggest that the market will observe employers moving from lower-reputation and cheaper options to higher-reputation and more expensive ones that will result in more successful outcomes.

To empirically test these mechanisms we propose a new two-component framework that facilitates the investigation of both hiring behaviors and outcomes. In particular, the first component of the framework, a Hidden Markov Model (HMM), captures how employers transition from *unsuccessful* to *successful hiring decisions*. The second component, a conditional logit model, estimates employer hiring choices. This structure allows for a comparison of hiring behaviors as employers move between successful and unsuccessful decisions, and hence it links *hiring choices with task outcomes* for the first time.

Analysis of 238,364 hiring decisions from a large online labor market provides empirical evidence supporting the relationships we theorize. Specifically, we show that (1) employers indeed initially explore available signals and they often select lower cost and lower reputation alternatives that are more likely to yield unsuccessful outcomes, and (2) employers then learn from exploration and start to exploit by reinforcing successful prior hiring behaviors. Finally, the two-component framework shows that employers who learn to rely more on contractor reputation and to not chase cheap contractors become more successful.

Why do some employers choose initially to hire lower-cost contractors with relatively worse reputation? We argue that confidence, defined as "belief in success" might be a plausible mechanism. Employers with higher confidence will likely have higher risk attitudes and hence try to beat the market by hiring cheaper contractors that they consider as good deals. Indeed, some of these hires will end up being successful, but most, will likely underperform (Kokkodis and Ipeirotis 2016, Danescu-Niculescu-Mizil et al. 2009, Liu et al. 2008, Lu et al. 2010, Kokkodis 2021). Our learning framework suggests that employers whose choices underperform will keep exploring and adjust their hiring behaviors, and as a result, re-evaluate their belief in success and potentially adjust their confidence and risk attitudes. On the other hand, employers with lower confidence will have lower risk attitudes and will hire conservatively by selecting contractors who are not as cheap but have good reputation. These choices will tend to be more successful, and will guide many employers to reinforce (exploit) their hiring behaviors. As a result of these dynamics, some higher-confidence employers who experience unsuccessful outcomes will adjust their hiring preferences and transition to lower-risk choices that are, on average, more successful.

This research extends our understanding of how users learn in online markets, particularly in online labor markets. It is the first study to explain how hiring choices affect outcomes and how employers learn to adjust their hiring behaviors to achieve better outcomes. Through a new theoretical framework, our work argues that employers prone to making decisions that are more likely to fail will keep exploring and adjust their hiring behaviors. On the other hand, employers who hire successfully, will keep reinforcing their hiring behaviors that exploit already acquired knowledge. Empirical evidence confirms this framework, showing that as employers transition from unsuccessful to successful hiring choices, they rely more on contractors' feedback scores and become less price-sensitive. Additionally, we provide some evidence that struggles on the platform could stem from employers beliefs in their ability to take advantage of perceived pricing anomalies, but subsequently learn that the market prices and contractor reputation tend to correlate with probability of successful outcomes.

Methodologically, our work provides a new two-component framework that links hiring choices with task outcomes. This framework provides future research a way to study transactions in these markets while incorporating both employer evolving preferences and task outcomes. Even further, our analysis helps platform managers understand how employers make successful hiring decisions. Such understanding is fundamentally important as both (1) employers who hire successfully and (2) contractors who receive positive feedback are more likely to keep participating (Tripp and

Grégoire 2011, Jerath et al. 2011). In addition, our framework allows managers to dynamically measure and track the likelihood of employers making successful hiring choices. Knowledge of such employer-evolving trends gives platform managers opportunities to target interventions and incentivize employers to adjust their hiring behaviors.

2. Theory

Online platforms in general—and online labor markets in particular (Agrawal et al. 2015, Lauren 2017, Brier and Pearson 2018)—have grown substantially over the past decades supporting the transition of many business functions from offline to online. In adopting online platforms, individuals use cognitive structures from their prior experience "that allow them to form perceptions about uncertainty" (Barr et al. 1992) to combine multiple sources of information and assess incomplete signal in new contexts (Busenitz and Barney 1997, Dequech 1999, Peterson and Pitz 1988). New online platform users encounter a mix of more familiar and less familiar information that they must assess to use online platforms successfully. For more familiar information, cognitive structures they formed outside the market can help their decisions in the market, but may be less relevant in the new context than they realize. For less familiar information, new users may be unsure how to adapt prior cognitive structures that lack platform-specific context. As a result, new users must learn to assess market information in the new context to reduce the likelihood of unsuccessful outcomes.

2.1. Learning in online labor markets

We use online labor markets as a context to study learning in such new platforms for three reasons. First, online labor markets are growing in importance (Agrawal et al. 2015, Lauren 2017, Brier and Pearson 2018). Second, users must combine observable and latent information to make a hiring decision. Third, these markets have offline parallels from which make some, but not all, platform information familiar from prior experience. Indeed, the theoretical importance of online labor markets motivated multidisciplinary research in areas such as algorithmic contractor management (Horton and Chilton 2010, Mason and Watts 2010, Sheng et al. 2008, Ipeirotis et al. 2010), design choices that increase market efficacy (Horton 2010, Arora and Forman 2007, Allon et al. 2012, Gopal et al. 2003, Dey et al. 2010, Chen and Horton 2016, Liang et al. 2017, Kokkodis and Ipeirotis 2016), and, particularly relevant for our context, employer hiring choices (Yoganarasimhan 2013, Pallais 2014, Moreno and Terwiesch 2014, Hong et al. 2015, Lin et al. 2016).

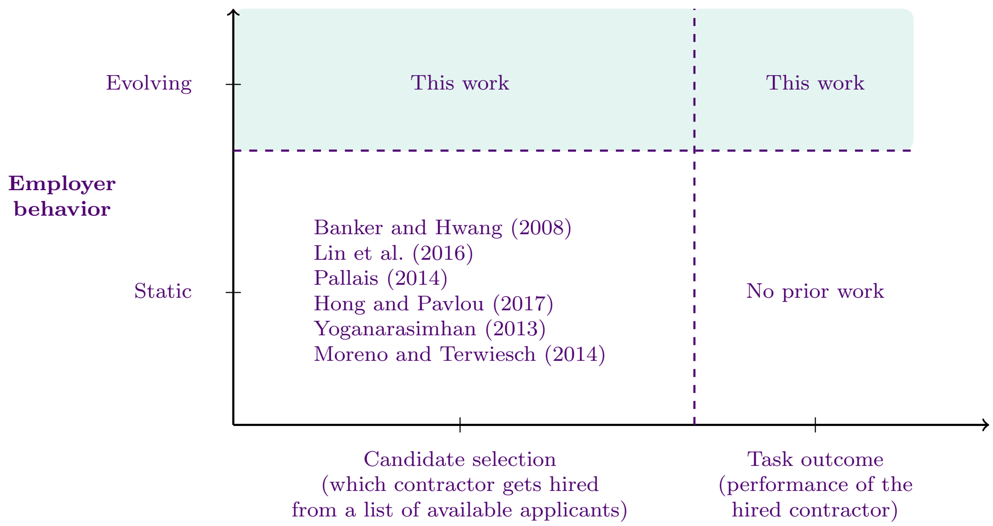
In online labor markets, potential employers post the specifications for work they need, mostly contract work of fixed time or scope. Once posted, jobs typically attract numerous applications. Employers then base hiring decisions on expectations of contractor performance. Once hired, the contractors complete the work and the employer observes actual performance and can assess the success of the hire.

As such, labor is an "experience" good (Nelson 1970): employers do not know about a contractor's performance on a task before the contractor completes the task. When hiring contractors, employers must consider observable (e.g., listed skills, certifications, price) and latent (actual abilities) characteristics to estimate contractor performance. The existence of latent characteristics and the heterogeneity in the observable ones (Kokkodis and Ipeirotis 2014) increases the uncertainty in hiring choices: employers must use imperfect information to make noisy estimates about contractors' abilities.

But processing signals from such imperfect sources is subjective and correlates with an individual's ability to acquire and process information (Grinold and Kahn 2000). Individuals can potentially learn from observing the outcomes of their decisions and adjust how they process market signals accordingly. And while prior research explains how employers choose contractors in online labor markets (Moreno and Terwiesch 2014, Chan and Wang 2017, Kokkodis 2018), it does not yet explain how (and what) employers learn to make successful hiring choices.

Specifically, prior relevant works are limited in two important ways. First, studies of hiring choices alone do not provide information about the *result* of these choices—we know a considerable amount about what contractors employers choose, but we know less about how well these choices work out. Second, static analysis of hiring choices does not reflect how (or whether) employers adjust their hiring behaviors over time. While it is helpful to understand employer hiring characteristics at a

Figure 2 **Learning from hiring decisions in labor markets**



This research extends prior research by incorporating two important aspects: (1) Employers and platforms care about outcomes, not just hiring choices. (2) Employers evolve and potentially learn to make better hiring choices.

given time, task outcomes (successful or unsuccessful completion of a task) provide considerable valuable information to employers about the relevance of various market signals, hence helping them to re-assess and learn from their prior behaviors. As a result, a study that investigates how employer learning occurs—and if occurs—from prior outcomes helps us better understand how online labor markets work.

Figure 2 compares this work with the most relevant papers: no prior work investigates how employer hiring preferences evolve and whether they learn to make better hiring choices over time.

2.2. Employer learning through exploration and exploitation

As employers use an online labor market, their diverse hiring choices result in different observed outcomes. Following a learning curve (Yelle 1979, Schilling et al. 2003), employers can learn from exploration to estimate better the fit between a task's requirements and an applicant's abilities. Exploratory learning can happen in two ways: when employers observe a successful outcome, they can reinforce (i.e., use applicant information similarly in subsequent hirings), *exploiting* those successful hiring behaviors (Mom et al. 2007, Lee 2019) that maximized their expected return (Erev and Roth 2014). But when they observe an unsuccessful outcome, they can *explore* by changing how they use information in subsequent hirings in hopes of achieving better outcomes.

In particular, exploitation allows employers to refine existing certainties (Mom et al. 2007) and increase decision reliability (Levinthal and March 1993, Good and Michel 2013) by refining what they already know (Good and Michel 2013, Holmqvist 2004), for example, in a new context. On the other hand, unsuccessful hiring choices will lead employers to keep exploring and augment what they already know (Gibson and Birkinshaw 2004, Mom et al. 2009, Rogan and Mors 2014, Rosing and Zacher 2017, Lee and Meyer-Doyle 2017, Lee 2019, Skinner 1990) through a trial-and-error process (Hull 1930, Jueptner et al. 1997, Evans et al. 2000). Because exploration yields uncertain and often negative returns (March 1991), and because exploitation usually has "positive, proximate, and predictable" outcomes, we observe that:

Proposition 1 *Employers will initially explore the new market environment; over time, they will learn from observed outcomes by adjusting hiring behaviors that lead to unsuccessful outcomes and by reinforcing (exploiting) hiring behaviors that lead to successful outcomes. As a result, the initial exploration choices will be, on average, less successful than the later exploitation ones.*

2.3. Learning to use market information

What attributes do employers use when making hiring choices in online labor markets?

Like new users of other online platforms, new employers in online labor markets confront an assortment of more familiar and less familiar information when making hiring decisions. We focus on two important attributes in hiring choices (Yoganarasimhan 2013, Moreno and Terwiesch 2014, Kokkodis and Ipeirotis 2016)—price and reputation.

As in any market, expense for the employer and revenue for the contractor is crucial. Research in consumer psychology (Shiv et al. 2005) and the neural representations of the brain (Plassmann et al. 2008) shows that consumers associate quality with higher prices. Higher prices signal higher quality in multiple product categories, including frequently purchased convenience goods (Caves and Greene 1996) and relatively cheap products (Rao and Monroe 1989).

In online labor markets, price is a fundamental attribute of a contractor's bid and is clearly observable (Grewal et al. 1994). Expensive contractors are likely of higher quality and perform better than cheap contractors (Svveeney 1999). As such, pricing has attracted considerable research attention. For example, tasks of higher value (Snir and Hitt 2003) and longer duration (Gefen et al. 2016) attract higher bids but of lower average quality. Sealed bid auctions attract more contractors, but open bid auctions create higher surpluses for the employers (Hong et al. 2015). As bid price dispersion increases, so does employer indecision and contractor regret, hence hurting matching (Zheng et al. 2016). Even further, employers tend to be more price-sensitive in fixed-price contracts than in hourly contracts (Lin et al. 2016).

New online labor market employers can bring considerable experience using price information for transactions in general and perhaps even for hiring in offline settings. Price, however, is noisy and imperfectly correlates with product (or service) quality (Gerstner 1985, Erdem et al. 2008), especially when individuals experience a new context. As a result, some might initially try to *explore* lower-cost options (Mom et al. 2007). (Section 6 discusses why some employers might be more prone to exploring such lower-cost options.) For instance, some new employers may select lower-than-average price contractors to avoid overpaying or to try to "beat the market" (Grinold and Kahn 2000). Sometimes, these lower-price options do result in a good deal for the employer but, more often, they result in poorer performance (Kokkodis and Ipeirotis 2016). Hence, compared with more experienced employers who have developed increased familiarity with the value of price in this new context and are reinforcing prior successful behavior (Proposition 1), relatively newer employers who are still exploring the value of price in this new context might often make decisions that result in less successful outcomes:

Hypothesis 1 *Employers who explore cheaper contractors will likely have less successful than average outcomes.*

Besides price, online labor markets recognize employers' need for information about a contractor's past performance and include some measure of contractor reputation. Current reputation systems allow contractors to receive feedback for the tasks they complete (Filippas et al. 2018). As contractors complete tasks, feedback scores typically accumulate to generate a contractor's reputation on the platform (Rahman 2018). These contractor reputation scores usually predict future performance (Kokkodis and Ipeirotis 2016, Danescu-Niculescu-Mizil et al. 2009, Liu et al. 2008, Lu et al. 2010, Kokkodis 2021), even in markets with inflated scores (such us online labor markets, Filippas et al. 2018, Rahman 2018). Hence, it is not a surprise that reputation drives the choice of contractors in online labor markets (Yoganarasimhan 2013, Kokkodis 2021). In fact, even having a reputation (compared to being new in the market) significantly improves a contractor's current (Lin et al. 2016) and subsequent hiring chances (Pallais 2014).

Unlike price, reputation is a platform-specific signal that new employers might initially be less familiar with as these reputation scores do not have exact equivalents outside online labor markets. In fact, while practically all online labor markets include some reputation scores, the details underlying how they work differ idiosyncratically by platform. Hence, some employers might initially explore by discounting the value of reputation, and, as mentioned above, try to hire cheaper contractors who usually have lower (or no) reputation (Moreno and Terwiesch 2014, Pallais 2014). But since contractors with lower reputation scores on average underperform those with higher scores (Kokkodis and Ipeirotis 2016), such lower-reputation exploratory hires will, on average, result in less successful outcomes:

Hypothesis 2 *Employers who explore hiring lower-reputation contractors will likely have less successful than average outcomes.*

Thus price and reputation signals contain information and the two signals correlate as contractors with higher reputation charge higher premiums (Yoganarasimhan 2013, Pallais 2014, Moreno and Terwiesch 2014). *But how exactly do employers trade off price and reputation over time?*

Consider a group of employers who choose to ignore poor reputation and focus only on lower prices. Stochastically, Hypotheses 1 and 2 suggest that such contractors are more likely to perform poorly. Following Proposition 1, some of those employers whose hired choices underperformed will learn from their bad experience and re-adjust how they assess market information (e.g., price and reputation). A different group of employers might choose to incorporate reputation even at the cost of slightly higher prices. Such reputable workers are more likely to perform well (Kokkodis and Ipeirotis 2016, Danescu-Niculescu-Mizil et al. 2009, Liu et al. 2008, Lu et al. 2010, Kokkodis 2021). As a result, these employers will be satisfied with their hired choices, and, as per Proposition 1, reinforce their hiring behavior. Combined, these dynamics suggest that the first group of employers will move from lower-to-higher reputation and from cheaper to more expensive contractors, while the second group will continue hiring highly reputable contractors who might not be as cheap. Hence we argue:

Hypothesis 3 *Over time, employers learn to increase (or reinforce) their reliance on contractor reputation and decrease (or reinforce) their price sensitivity. This behavior results in more successful than average outcomes.*

3. Modeling Evolving Employer Hiring Choices and Outcomes

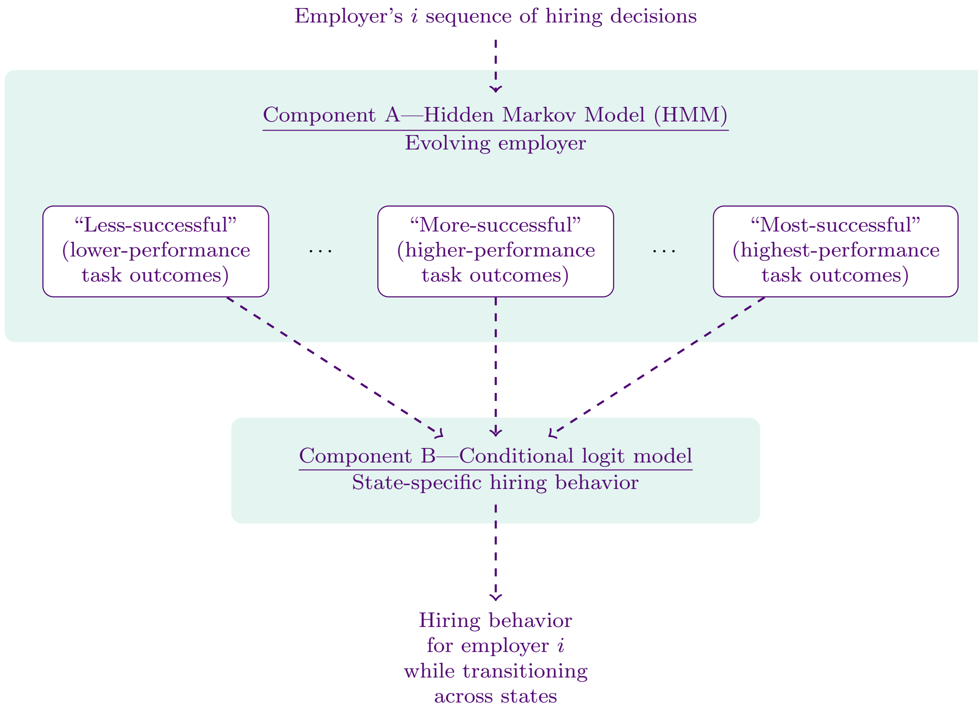
Testing these hypotheses requires a framework that models employer evolution and, at the same time, tracks hiring outcomes.

3.1. The challenges of simultaneously modeling hiring choices and outcomes

While important, incorporating both employer evolution and hiring outcomes is difficult—no prior research study includes both (Figure 2). One possible reason for this is that hiring choices and task outcomes occur at different times and in different populations (i.e., task applicants vs. hired contractors). Employers first choose a contractor to hire, and then they observe the outcome of their collaboration with the chosen contractor. As a result, a model needs to capture first whether or not a given contractor gets hired (i.e., a binary outcome "hired" vs. "no-hired" across task applicants), and second, the performance score of the hired contractor (continuous outcome, across hired contractors).

We propose a two-component framework that directly models these steps (Figure 3). In the Component A, a Hidden Markov Model (HMM) captures employer evolution across various hiring-behavior states. As a result, this component provides up-to-date estimates of employers' propensities to hire contractors successfully. Component B uses state-specific conditional logit models to

Figure 3 A two-component framework of employer learning



Component A models the evolution of employers across various hiring-behavior states that represent different propensities of observing successful outcomes. Component B captures the hiring behavior of each employer within each state. The result allows for a comparison of hiring behaviors as employers transition from a less successful to a more successful state.

explain how successful and unsuccessful employers choose contractors. With these two components, the framework *compares employers' hiring behaviors as they evolve.*[1]

[1] The two-step process in our context has both similarities and differences with traditional two-stage selection models (Heckman 1979). Like these models, our framework models two stochastic events: the probability of an employer having a given level of hiring ability (HMM) and the employer's subsequent probability of choosing an applicant (conditional logit). Yet, traditional selection models cannot directly model the first step of our approach because this step does not rely on an observed outcome (e.g., the choice to participate), but instead, it estimates *unobserved* ability states. On the other hand, two-stage selection models can capture the choice to participate or exit the market; we implement such models to test alternative explanations in Section A.1.2.

3.2. Component A: modeling an employer's evolving hiring behavior

An online labor market transaction starts with an employer creating a new opening. The opening describes task requirements and contractor characteristics (such as skills and experience) that the employer seeks. Contractors who are searching for opportunities and see a fit between their interests and the task opening can submit an application. For each contractor j that applies for a task o the employer i makes a hiring decision. Once the contractor completes the task, the employer rates the performance of the *hired* contractor (y_{io}).

Each employer who joins the platform has an unobserved process of evaluating available market information to hire contractors. By hiring contractors and observing task outcomes some employers might learn and adjust their hiring behaviors (Section 2.3). This evolution suggests that employers transition across a series of states, such as "Less-successful," "More-successful," "Most-successful." Even though we do not observe these states, we observe the outcome of each completed task (i.e., the performance rating y_{io}), which correlates with the underlying employer's hiring behavior.

An HMM naturally models such an employer evolution across a set of states $\mathcal{S} = \{s_1, ..., s_K\}$. At any given point in time t, each employer i operates from a state $S_t \in \mathcal{S}$ that represents i's propensity to hire successfully. Upon completion of a new task, the employer emits an observation y_{it} (i.e., the performance rating of the hired contractor at time t—we consider that the completion of a new task defines a unit of time and, as a result, we omit the task subscript o because the time subscript t subsumes it). Each state models these observations through continuous probability distributions, $Y_{it} \sim f(S_{it}), Y_{it} \in [0,1], S_{it} \in \mathcal{S}$. ($Y_{it}$ is a random variable that describes observations y_{it}.)

Model structure: A complete definition of an HMM requires (1) a matrix T of the transition probabilities between states, (2) a matrix E of the state-specific emission probability distributions of the observed outcomes, and (3) a vector of initial state probabilities $\boldsymbol{\pi}$.

We assume an employer- and state-specific transition matrix T. We allow the history of each employer i up to time $t-1$ (captured by a vector of covariates $\boldsymbol{Z}_{it-1}$ discussed in Appendix B.2) to affect the employer's propensity to transition to a new state. These employer-history effects are also state-specific: given $\boldsymbol{Z}_{it-1}$, the transition probability for employer i to move from a state S_{it-1} to a state S_{it} at time t is:

$$\lambda^{s_k,s_v}_{\boldsymbol{\gamma}_k,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} = \Pr(S_{it}=s_v|S_{it-1}=s_k;\boldsymbol{\gamma}_k,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}) \tag{1}$$

where:

$$\lambda^{s_k,s_v}_{\boldsymbol{\gamma}_k,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} = \begin{cases} \Pr(\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}+\varepsilon_i \le \alpha^1) = \Lambda(\alpha^1-\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}), & \text{if } v=1 \\ \Pr(\alpha^{v-1} < \boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}+\varepsilon_i \le \alpha^v) = \Lambda(\alpha^v-\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}) - \Lambda(\alpha^{v-1}-\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}), & \text{if } 1<v<K \\ \Pr(\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}+\varepsilon_i > \alpha^{K-1}) = 1-\Lambda(\alpha^{K-1}-\boldsymbol{\gamma}_k\boldsymbol{Z}_{it-1}) & \text{if } v=K \end{cases} \cdot$$

In Equation 1, $\boldsymbol{\gamma}_k$ is the vector of coefficients of state s_k that define the weights of $\boldsymbol{Z}_{it-1}$ in estimating the transition probability to state s_v. We estimate this probability through an ordered logit formulation, where ε_i is the unobserved error term that we assume to follow a logistic distribution (i.e., $\varepsilon_i|\boldsymbol{Z}_{it-1} \sim Logistic(0,1)$), Λ is the logit function, and $\boldsymbol{\alpha} = [\alpha^1,\alpha^2,...,\alpha^{K-1}]'$ are state-specific thresholds (Wooldridge 2010). As a result, the transition matrix at time t for employer i is

$$T(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}) = \begin{bmatrix} \lambda^{s_1,s_1}_{\boldsymbol{\gamma}_1,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} & \lambda^{s_1,s_2}_{\boldsymbol{\gamma}_1,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} & \dots & \lambda^{s_1,s_K}_{\boldsymbol{\gamma}_1,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{s_K,s_1}_{\boldsymbol{\gamma}_K,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} & \lambda^{s_K,s_2}_{\boldsymbol{\gamma}_K,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} & \dots & \lambda^{s_K,s_K}_{\boldsymbol{\gamma}_K,\boldsymbol{\alpha},\boldsymbol{Z}_{it-1}} \end{bmatrix}, \tag{2}$$

where $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1,...,\boldsymbol{\gamma}_K]'$.

The emission matrix E describes the conditional probabilities of observed outcomes given the current state of a user i:

$$\Pr(Y_{it}|S_t=s_k;\boldsymbol{\theta}_k) = f(\boldsymbol{\theta}_k)\,, \tag{3}$$

where $f(.)$ is a truncated continuous probability distribution (e.g., truncated Gaussian) and $\boldsymbol{\theta}_k$ is the parameter vector of this distribution. (Appendix B.3 discusses alternative distribution choices for f.) The emission matrix E is then as follows:

$$E(\boldsymbol{\theta}) = \left[f(\boldsymbol{\theta}_1), f(\boldsymbol{\theta}_2),\dots,f(\boldsymbol{\theta}_K)\right]\,, \tag{4}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_K]'$.

Finally, the framework estimates the probability vector $\boldsymbol{\pi} = [\pi_1, ..., \pi_K]'$ and the parameter vectors $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ through maximum likelihood estimation. Appendix B presents the derivation of the global likelihood and the parameter estimation details.

3.3. Component B: Within state employer hiring behavior

The utility that an employer i in state $s_k \in \mathcal{S}$ derives by hiring a contractor j for a task o is

$$u_{io}^j(s_k) = \boldsymbol{\beta}_{s_k} \boldsymbol{X}_{jo} + \kappa_o + \varepsilon_{io}^{js_k}\,, \tag{5}$$

where $\boldsymbol{X}_{jo}$ is the vector of characteristics of the contractor j for task o, κ_o captures static task-specific effects, and $\varepsilon_{io}^{js_k}$ is the unobserved error term. Similar to prior literature on hiring decisions (Moreno and Terwiesch 2014, Chan and Wang 2017), we describe this utility through a conditional logit model (McFadden 1973). The conditional logit model assumes that consumers make choices among a set of products in a way that maximizes their utility functions (Guadagni and Little 1983). In the context of this work, employers choose the candidate that maximizes their state-specific utility function (Equation 5).

The conditional logit model assumes that the error term $\varepsilon_{io}^{js_k}$ follows a Type I extreme value distribution, and hence the probability of choosing contractor j for task o is:

$$\Pr(H_{io}^j = 1 | \boldsymbol{X}_{1o}, \boldsymbol{X}_{2o}, \ldots, \boldsymbol{X}_{Jo}, s_k) = \frac{\exp\left(\sum_{s_k \in \mathcal{S}} \mathbb{1}_{s_k} \boldsymbol{\beta}_{s_k} \boldsymbol{X}_{jo}\right)}{\sum_{d=1}^{J} \exp\left(\boldsymbol{\beta}_{s_k} \boldsymbol{X}_{do}\right)}\,, \tag{6}$$

where H_{io}^j is an indicator function that is one when contractor j is considered to get hired for task o, and $\mathbb{1}_{s_k}$ is an indicator function that is 1 when the current state of the employer is s_k. The vector of utility weights $\boldsymbol{\beta}_{s_k}$ identifies the importance of each characteristic in $\boldsymbol{X}_{jo}$ when employer i is at state $s_k \in \mathcal{S}$.

4. Empirical Setup and Results

To empirically test our theoretical framework, we analyze a dataset of hiring decisions from a major online labor market. The focal dataset includes 238,364 hiring decisions made by 2,932 employers on 10,385 tasks. For all employers, we observe their complete history (i.e., from the time they joined the platform) and how they evolve over consecutive hiring decisions and completed tasks. The data covers diverse tasks from various categories, including administrative, sales, marketing, writing, data science, and web development (Table 15 in Appendix E).

In this market, employers and contractors participate globally. Employers begin a new transaction by posting task openings with specific skill requirements and a respective task description. Available contractors then can apply without cost. On average, each task opening in our dataset receives 52 applications (median 35, Table 14 in Appendix E). Employers then hire from the available applicants. Once the hired contractor completes the task, the employer evaluates the contractor's performance.

The two-component framework (Figure 3) trains on two different datasets. The HMM component trains only on hired contractors, as it requires task-outcome information. We measure task outcomes through privately assigned feedback from employers. (Employer ratings are only visible to the platform.) Appendix B describes the details and the results of this process. On the other hand, the conditional logit model trains on all applications. Hence, it allows us to study how employers choose contractors as they evolve across different ability states. The remainder of this section focuses on the results of the conditional logit model.

4.1. Focal and control variables

Several measures (vector $\boldsymbol{X}_{jo}$) may affect employer hiring choices.

Focal variables: The contractor's profile lists the feedback score at the time of application. This score reflects the accumulated public feedback that prior employers have given the contractor for all previously completed tasks on the platform ("Feedback score"). Similarly, contractors reveal their premiums at the time of application. Because online labor markets are highly heterogeneous in terms of tasks and task requirements, bid prices are only comparable within tasks (Kokkodis et al. 2015). Hence, to study how employers respond to price, we create a relative, task-specific measure of the available bids: we consider whether each bid price is above the median *task-specific* bid price ("Bid price").

Table 1 Descriptive statistics of the dependent, focal, and control variables

Variable	Units	Mean	Median	StD	Min	Max
Dependent variable						
Hired (H_{io}^{j})	Hired = 1	0.04	0	0.2	0	1
Focal variables						
Feedback score	Index	0.94	0.98	0.1	0	1
Bid price	Above median = 1	0.37	0	0.48	0	1
Task-varying variables						
Assigned tasks	Count	3.4	2	5	0	90
Skills IP	Count	1	1	1.2	0	16
Certifications PMI	Index	2.4	2.3	1.7	−0.71	7
Hours of work	Hours	551	44	1,401	0	27,662
Time of application	Ordinal	25	15	31	0	263
Gender	Male = 1	0.81	1	0.39	0	1
Contractor completed tasks	Count	5	0	14	0	356
Contractor experience	Years	4.5	4	4.1	0	30
Invited	Invited = 1	0.12	0	0.33	0	1
Countries PMI	Index	−0.5	−0.46	0.43	−4.9	5

StD is the standard deviation. PMI is point-wise mutual information; IP is inner product (see Appendix E.1 for additional details).

Task-varying variables: To better isolate the effect of the focal variables, we control for the observed history of each contractor on the platform through the number of completed tasks ("Contractor completed tasks") and the number of completed hours of work ("Hours of work," Chan and Wang 2017). Furthermore, we control for the number of tasks that a contractor is concurrently working on at the time of application ("Assigned tasks") and the self-reported years of contractor experience ("Contractor experience").

Besides platform history, we need to control for other contractor characteristics that often affect employer decisions. Specifically, the models control for a contractor's gender ("Gender"; Chan and Wang 2017), location, skills, and qualifications ("Countries PMI," "Skills IP," "Certifications PMI"; Appendix E.1). Finally, we also control for timing effects by including the order in which a contractor's application arrives ("Time of application"), and for employer preference effects by annotating contractors that the employer invited to apply ("Invited"; Chan and Wang 2017).

Table 1 summarizes the dependent, focal, and control variables. Appendix E presents additional data statistics, including a short discussion on feature engineering (Section E.1) and correlations between these variables (Figure 11). We log-transform variables with long tails. We standardize all non-binary variables for faster convergence and easier interpretability. All control variables vary across applicants within each task; otherwise, task fixed effects would absorb these.

4.2. Results

Which hiring choices are more likely to yield successful outcomes, and how do employers adjust their hiring behaviors to make such choices?

Component B's (Figure 3) conditional logit model estimates each employer's behavior within each state. The main specification (Equation 6) controls for observed and unobserved confounding factors that could endogenize the estimates of the focal variables. Specifically, task-varying variables control for the observed heterogeneity across the set of available candidates for each task (Table 1). Opening fixed effects account for private information that employers have when making hiring decisions, and hence they eliminate the time-invariant portion of the unobserved error. Opening fixed effects further eliminate any employer-specific, category-specific, and time-specific static unobserved effects as these quantities do not vary within tasks.

To identify the parameters of Equation 6, we first need to estimate each employer's most likely state at the time of the hiring decision (i.e., $s_k \in \mathcal{S}$). Appendix B details the HMM implementation and estimation process. Through this analysis, the three latent states that best describe the employers' hiring behavior in our data (Figure 7 in Appendix B) are as follows:

- "Less-successful" state: employers' decisions result in unsuccessful task outcomes. (Employers hire contractors that, on average, perform 0.53 out of 1.)
- "More-successful" state: employers' decisions result in more successful task outcomes. (Employers hire contractors that, on average, perform 0.82 out of 1.)
- "Most-successful" state: employers' decisions result in successful task outcomes. (Employers hire contractors that, on average, perform 0.99 out of 1.)

For these estimated latent states $\mathcal{S} = \{\text{Less-successful, More-successful, Most-successful}\}$, we identify the coefficients of Equation 6. Table 2 shows the results for each of the state-specific focal

Table 2 State-specific effects of feedback score and bid price on hiring choices

	All employers		Evolving employers
	(Without HMM)	(With HMM)	(With HMM)
DV: Hire	(A1)	(A2)	(A3)
Bid price	-0.131^{***}		
	(0.023)		
Feedback score	0.288^{***}		
	(0.018)		
Bid price × Less-successful		-0.350^{***}	-0.374^{***}
		(0.053)	(0.058)
Bid price × More-successful		-0.146^{**}	-0.173^{***}
		(0.048)	(0.051)
Bid price × Most-successful		$-0.053.$	-0.038
		(0.030)	(0.036)
Feedback score × Less-successful		0.142^{***}	0.142^{***}
		(0.033)	(0.037)
Feedback score × More-successful		0.248^{***}	0.266^{***}
		(0.037)	(0.040)
Feedback score × Most-successful		0.379^{***}	0.391^{***}
		(0.027)	(0.033)
Task-varying variables	Yes		
Task-varying variables × Less-successful		Yes	Yes
Task-varying variables × More-successful		Yes	Yes
Task-varying variables × Most-successful		Yes	Yes
Opening FE	Yes	Yes	Yes
Observations	238,364	238,364	174,873
AIC	52,833	52,763	39,113

Results of conditional logit models. The first two columns show results of the complete dataset (all employers). The last column focuses only on employers who transition across two or more states. Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. Tables 16 and 17 show the state-specific coefficients for all task-varying variables. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

variables. (Table 16 in Appendix F shows the coefficients of the complete vector $\boldsymbol{\beta}_{s_k} \forall s_k \in \mathcal{S}$.) Column A1 in Table 2 shows the average effect of the two focal variables across states. Column A2 splits this effect into state-specific effects. First, the decrease in AIC score from Column A1 to Column A2 suggests that the state-specific estimates better explain the observed variance in hiring choices—a log-likelihood ratio test (Wooldridge 2010) rejects the null hypothesis that the static model (Column A1) fits the data better ($p < 0.001$), providing further statistical evidence that modeling employer evolution through an HMM better explains the observed variance in hiring choices.

Second, as employers move from less successful to more successful states, they increase their reliance on the observed feedback score and tend to hire relatively more expensive contractors (Hypothesis 3). The focal coefficients differ both practically and statistically ($p < 0.001$) across states. Specifically, in line with Hypothesis 1, the coefficient of bid price is significantly smaller for employers in more successful states: an employer in the "Most-successful" state is 85% less price-sensitive than an employer in the "Less-successful" state. In addition, in line with Hypothesis 2, the coefficient of feedback score is significantly larger for employers in more successful states: an employer in the "Most-successful" state values reputation 267% more than an employer in the "Less-successful" one.

The results in Column A2 include all employers, even those who never transition to different states. As a result, they might represent disjoint sets of employers with different hiring preferences and not employers who evolve across states. To investigate, we consider only *evolving employers* who transition across two or more states. This subsample compares how *the same employers* hire in "Less-successful," versus "More-successful," versus "Most-successful," states. Hence, it identifies whether the changes in employers' hiring behaviors drive the observed results. The results (Column A3 of Table 2) are consistent with the entire sample results, and, because they focus on evolving employers, they provide stronger support for Hypothesis 3: as employers move from less successful to more successful hiring choices, they learn to adjust their hiring behaviors by relying more on feedback score and becoming less price-sensitive.

4.3. Additional evidence in support of the theoretical framework

The previous analysis supports Hypotheses 1, 2, and 3. To strengthen this empirical support, we investigate whether data evidence also aligns with the theoretical arguments that derived these hypotheses (Section 2.3). Specifically, our conceptualization argues that (1) lower prices and lower reputation scores yield, on average, less successful results, and (2) employers learn and adjust their hiring behavior (by exploring) after unsuccessful outcomes, while they reinforce their hiring behaviors (by exploiting) after successful ones. Next, we provide empirical evidence in support of these arguments.

4.3.1. Further support for Hypotheses 1 and 2: *Are employers who choose cheaper contractors with lower reputation more likely to yield unsuccessful outcomes?*

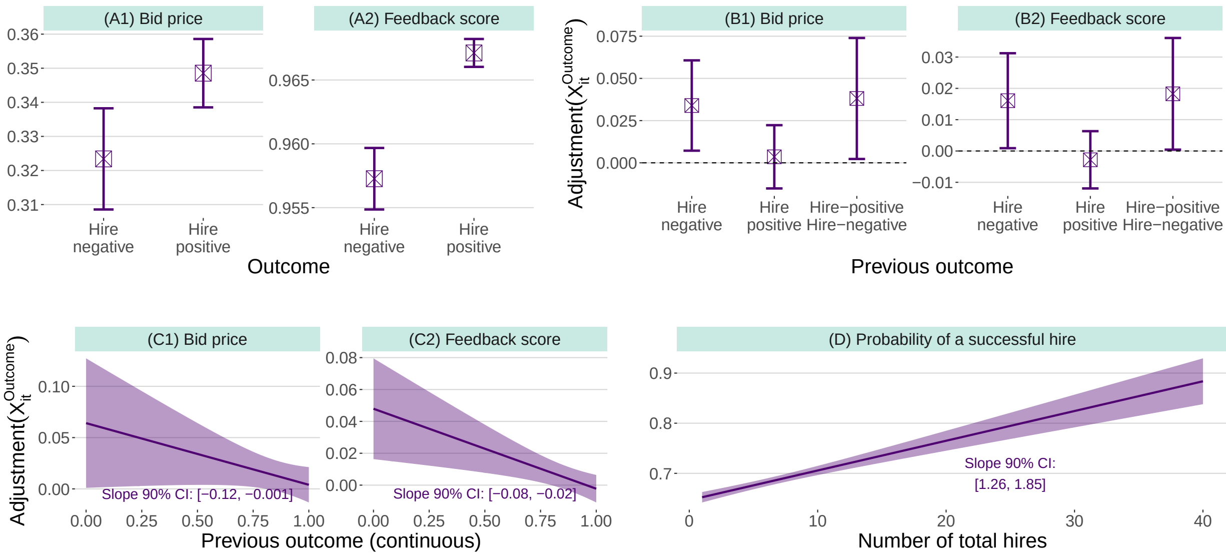
The main analysis in Table 2 shows that employers who are more likely to make successful hiring choices choose, on overage, more expensive contractors with higher reputation (supporting Hypotheses 1 and 2). An alternative way to show that indeed such choices correlate with more successful outcomes is to estimate the average bid price and feedback score for various outcomes. First, we separate successful and unsuccessful outcomes by adapting the internal annotation of the focal platform: if the employer privately rates the performance of the contractor with a score greater or equal to 80%, the platform labels that outcome as successful; otherwise, the platform labels the outcome as unsuccessful. For simplicity, we call successful outcomes "Hire-positive" and unsuccessful ones "Hire-negative." (Figure 1A shows how this threshold separates "Hire-positive" from "Hire-negative" outcomes.) Figures 4A1 and 4A2 show that successful outcomes correlate with significantly higher feedback score ($p < 0.01$) and bid price choices ($p < 0.05$), hence providing additional support for Hypotheses 1 and 2.

4.3.2. Further support for Hypothesis 3: *Do employers learn to adjust their hiring preferences?*

Column A3 in Table 2 shows that indeed, many employers transition across states, and when they move from less to more successful states they adjust their hiring criteria in accordance with Hypothesis 3. An alternative way to test our arguments, is to track how employers react after successful and unsuccessful hiring decisions. Given an employer i and a hiring outcome at time t, we define employer's i hiring adjustment of feature X as follows:

$$\text{Adjustment}(X_{it}^{Outcome}) = X_{it} - X_{it-1}^{Outcome}\,. \tag{7}$$

Figure 4 Additional support for Hypotheses 1, 2, and 3



Higher bid price and feedback score correlate with more successful outcomes (A1, A2). After unsuccessful outcomes, employers explore by choosing contractors with higher bid price and feedback score (B1, B2; the $y-$axis shows the adjustment between consecutive hires of the same employer). The more negative the outcome, the larger is the adjustment (C1, C2). Overall, the more hiring choices an employer makes, the higher is the likelihood of a successful outcome (D). Error bars show 95% bootstrapped confidence intervals. CI: confidence interval.

We estimate the average adjustment for $X \in$\{Bid price, Feedback score\} and $Outcome \in$ \{Hire-positive, Hire-negative\}. Figures 4B1 and 4B2 show that indeed, after unsuccessful outcomes, employers explore and hire contractors with significantly ($p < 0.05$) better feedback scores who are more expensive. On the other hand, after successful outcomes, employers exploit and reinforce their behavior (insignificant adjustment, $p > 0.1$). The figures also show that a similar adjustment occurs when employers observe a less successful (Hire-negative) outcome after previously observing a more successful (Hire-positive) one. Combined, this evidence provides empirical support that, on average, employers indeed learn and adjust their behaviors in accordance to Hypothesis 3.

Figures 4C1 and 4C2 show the linear relationship between outcome adjustment (Equation 7) and the continuous version of each employer's previous hiring outcome. The figures show that for both bid price and feedback score, the slope is negative and significant (at least $p < 0.1$), showing that larger adjustments correlate with a worse performance of the previous hire. As a result, these results align with those of figures Figures 4B1 and 4B2, and they further support our core arguments about employer learning.

Finally, Figure 4D shows employer learning through a simpler regression: the probability of a successful outcome increases ($p < 0.05$) as employers make more hiring choices. This increase suggests that over time employers move from less to more successful hiring choices, further corroborating our findings in support of Hypothesis 3.

In summary, our analyses in this section support our conceptualization and show that many employers do indeed learn to adjust their preferences and hire more successfully over time. On the other hand, support for the proposed mechanisms does not necessarily exclude alternative mechanisms that could also be driving the results. Section 5 and Appendix A test ten potential alternative mechanisms.

5. Alternative Explanations

To increase confidence in the hypothesized conceptualization, we:

- ◇ Incorporate the choice of an employer to participate or exit the market (using a customized exit-state HMM and Heckman two-stage selection models, Appendix A.1)
- ◇ Focus only on repeat employers (using subsample analyses, Appendix A.2);
- ◇ Control for employer hiring ability (using instrumental variables, Appendix A.3);
- ◇ Examine whether more experienced or better employers attract better contractors, and as a result learning does not actually occur (normalization of reputation within each opening, distribution of applicants, Appendix A.4);
- ◇ Consider alternative measures of outcome success (such as rehires, Appendix A.5);
- ◇ Control for task characteristics (using matched samples in conditional logit models, Appendix A.6);
- ◇ Explore how hiring abilities differ between technical and non-technical tasks (using subsample analyses, Appendix A.7);
- ◇ Investigate how focused or diverse employers drive the results (using subsample analyses, Appendix A.8);

◇ Model employer evolution and hiring choices with alternative approaches (using clustering algorithms and various fixed effects specifications, Appendix A.9); and,

◇ Consider different specifications that use employer fixed effects and control for time-varying employer attributes (e.g., experience) that opening fixed effects absorb in our main specification (Appendix A.10).

Table 3 summarizes these tests, which corroborate our main findings and provide additional evidence that employer learning occurs, as, on average, employers transition from choosing cheaper contractors with lower reputation to choosing more expensive contractors with higher reputation.

6. Potential Mechanism: Employer Confidence

The empirical analysis clarifies the dynamics of learning in online labor markets. However, our conceptualization in Section 2 does not explain why some platform participants might be more likely to explore lower-cost options.

One possible explanation is confidence (i.e., belief in success; Snyder and Lopez 2009). An individual's confidence structures their attitude towards information (Kreye et al. 2012) as it governs the perception of loss associated with specific actions (Forlani and Mullins 2000). Higher-confidence lowers perceptions of risk (Busenitz and Barney 1994, Palich and Bagby 1995, Robinson and Marino 2015) and results in "optimistic judgments of risk" (Kahneman and Lovallo 1993).

6.1. Confidence in online labor markets

In an online labor market, an employer's confidence and the resulting risk assessment may influence how an employer assess available market information. Employers with higher confidence will be more likely than employers with lower confidence to risk and try to beat the market (Grinold and Kahn 2000) by looking for cheap contractors (with likely lower reputation scores—Section 2.3). But our theory and empirical analysis finds that cheaper contractors on average yield worse outcomes (Section 2.3, Hypotheses 1 and 2, and Figures 4A1 and 4A2). As a result, some of these higher-confidence employers who hire unsuccessfully will re-adjust their confidence and risk attitudes, and hence learn to better assess the available information sources (Proposition 1, Hypothesis 3).

Table 3 Investigation of alternative explanations

Analysis	Bid price (support for H1)	Feedback score (support for H2)	Discussion
Alternative explanation: Self-selection of employers to keep participating drives the results			
Solution: Modeling the choice to participate:			
Exit state HMM	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.1,
Heckman two-stage selection model	✓($p < 0.001$)	✓($p < 0.001$)	Table 4
Alternative explanation: Repeat employers always had the observed behavior (i.e., no learning over time)			
Solution: Focus only on repeat employers who hire:			
Two or more times	✓($p < 0.001$)	✓($p < 0.001$)	
Three or more times	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.2,
Four or more times	✓($p < 0.001$)	✓($p < 0.001$)	Table 5
Five or more times	✓($p < 0.001$)	✓($p < 0.001$)	
Alternative explanation: The ability of employers who end up staying on the platform drives the observed results			
Solution: Instrumental variables capture employer state:			
State of every other employer at hire o	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.3,
% of employers who survive at hire o	✓($p < 0.001$)	✓($p < 0.001$)	Table 6
% of employers who survive at hire o after outcome	✓($p < 0.001$)	✓($p < 0.001$)	
Alternative explanation: More experienced employers attract better contractors			
Solution: Normalize reputation within openings	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.4,
Distributions of applicants across employers	L	L	Table 8, Fig. 6
Alternative explanation: The results are an artifact of our measurement of success			
Solution: Rehires as measure of success	✓($p < 0.001$)	✓($p < 0.001$)	AppendixA.5, Table 7
Alternative explanation: Systematically different types of tasks end up being in different states			
Solution: Propensity score matching on task characteristics	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.6, Table 9
Alternative explanation: Employers might have different abilities across different types of tasks			
Solution: Separate analysis of technical and non-technical tasks	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.7, Table 10
Alternative explanation: Focused or diverse employers drive the results			
Solution: Separate analysis of focused and diverse employers	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.8, Table 10
Alternative explanation: Results are an artifact of the HMM modeling			
Solution: Alternative modeling of Component A (K-means)	✓($p < 0.001$)	✓($p < 0.001$)	Appendix A.9, Table 11
Alternative explanation: Results are an artifact of the conditional logit model approach			
Solution: More experienced employers attract better contractors / Alternative modeling of Component B:			
Linear models with employer FE	✓($p < 0.001$)	✓($p < 0.001$)	Section A.10,
Logit models with employer FE	✓($p < 0.001$)	✓($p < 0.001$)	Table 12

✓: Hypothesis supported. (p) the level of significance that the respective coefficient of the "Most-successful" state is different than the respective coefficient of the "Less-successful" state. L suggests support of our core employer learning argument.

Conversely, employers with lower confidence may minimize risk by hiring contractors with higher reputation albeit at a higher price. Their choices will likely yield, on average, more successful outcomes (Section 2.3, Hypotheses 1 and 2, and Figures 4A1 and 4A2), which in turn might reinforce their confidence and risk attitudes (exploitation strategies; March 1991, Denrell and March 2001, Lee and Meyer-Doyle 2017, Menkhoff et al. 2006, Goodman-Delahunty et al. 2010). Some of these contractors might then explore more risky choices, until they learn to appropriately assess available information (Hypothesis 3, Figures 4B1 and 4B2).

These dynamics suggest that over a period of time, the market will observe employers with higher confidence to move from higher-risk (i.e., cheaper contractors with lower reputation) to lower-risk (i.e., more expensive contractors with higher reputation) choices, while employers with lower confidence to reinforce their behavior or explore higher-risk choices that, when unsuccessful, will push them back to lower-risk choices. As a result, the net-effect of these dynamics will be a transition from lower to higher contractor reputation and premium, and, on average, from less to more successful hiring outcomes—the invisible hand at work.

6.2. Empirical evidence of the confidence mechanism

The market does not directly observe employers' confidence levels. However, text analysis can detect concepts related to confidence. Specifically, we use the Linguistic Inquiry and Word Count (LIWC) package to analyze the unstructured text of employers' task descriptions (LIWC 2018). LIWC measures dimensions such as social processes (Sridhar and Srinivasan 2012), emotion (Hong et al. 2016), popularity (Goes et al. 2014), or anxiousness (Yin et al. 2014) using validated dictionaries. Like other research (Smith-Keiling and Hyun 2019), we measure confidence using the LIWC *clout* metric, defined as "the relative social status, confidence, or leadership that people display through their writing or talking" (LIWC 2018).

Alternatively, the task description's *emotional tone* may also correlate with confidence. A higher emotional tone conveys a more positive attitude, while a lower emotional tone reveals greater anxiety (LIWC 2018). Given that higher confidence associates with a more positive attitude (Snyder and Lopez 2009), emotional tone might serve as an additional signal of employers' confidence.

Do higher-confidence hiring choices result in less successful outcomes? Figure 5A suggests that the answer is probably yes: higher clout ($p < 0.05$) and emotional tone ($p < 0.1$) positively correlate with the likelihood that the hiring outcome will be unsuccessful (Figure 5A). This observation aligns with the argument that higher-confidence employers make higher-risk choices of cheaper contractors with lower reputation scores, who, on average, result in less successful outcomes (Section 2.3, Hypotheses 1 and 2, and Figures 4A1 and 4A2).

Do employers adjust their confidence and risk attitudes over time? To investigate, we compute the employer-specific confidence adjustment at time t as follows:

$$\text{Normalized } C_{it} = C_{it} - C_{i1} , \tag{8}$$

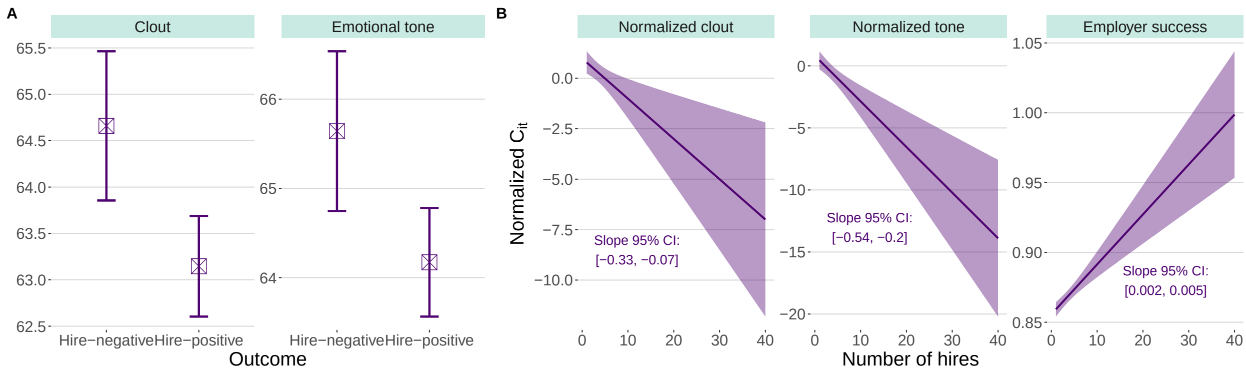
where i is the focal employer and $C \in \{\text{Clout, Emotional tone}\}$. This normalized score proxies how employers adjust their confidence level over time compared with their own initial confidence (comparison level).

As employers hire more contractors, they generally adjust their confidence downward (negative relationship, $p < 0.001$, in the first two panels in Figure 5B). Since employers mostly adjust their behaviors after unsuccessful outcomes (Figures 4B1, 4B2, 4C1, and 4C2), the observed effect likely stems from employers of higher-confidence and higher-risk attitudes. Furthermore, confidence re-adjustment suggests that the transitions from higher- to lower-risk choices should result (on average) in more successful outcomes. Indeed, over time, employers hire contractors who perform better ($p < 0.001$, last panel in Figure 5). (Appendix B.4 provides additional empirical evidence in support of the confidence explanation.)

7. Discussion

How do employers adjust their behaviors as they learn to make successful hiring decisions in online labor markets? We theorised that employers initially explore in their assessment of available market information; some—potentially because of higher confidence—will try to beat the market and hire contractors who are cheaper and have lower reputation. But these contractors are more likely to

Figure 5 Over time, higher-confidence (proxied by clout and emotional tone) employers adjust their hiring preferences



Higher-confidence employers (as measured by higher clout and emotional tone) have on average less successful outcomes (A). Over consecutive hiring choices ($x-$axis), employers adjust their confidence (as measured by clout and emotional tone), and this adjustment correlates with more successful hiring choices (B). CI: confidence interval.

have poor performance. Hence, employers who observe such unsuccessful outcomes will learn and adjust their hiring behavior by further exploring the value of available market information. On the other hand, employers who hire successfully will reinforce (exploit) their hiring strategies. We empirically test these mechanisms by building a two-component framework (HMM, conditional logit models) that links hiring behaviors with task outcomes. Hiring decisions from a large online labor market provide empirical evidence in support of this theory: employers learn and adjust their hiring behaviors by transitioning from hiring cheaper contractors with lower reputation to hiring more expensive contractors with higher reputation. The latter yield significantly more successful results.

7.1. Contributions to research

This research extends current literature in online labor markets and hiring decisions by investigating a new question (Figure 2): which hiring choices are more likely to yield successful outcomes, and how do employers learn to adjust their hiring behaviors to achieve better outcomes? Through a new theoretical framework, our work argues that over time, many employers learn to make successful hiring choices. This theoretical framework deepens our understanding of how employers' hiring preferences evolve and guides the empirical identification of the studied effects. Through a comprehensive empirical analysis, our study is the first to show that as employers transition from unsuccessful to successful hiring decisions, they rely more on contractors' feedback scores and become less price-sensitive.

This paper is the first to link task outcomes with employer hiring behaviors and show that successful outcomes often associate with higher contractor feedback scores and higher premiums. The two-component framework provides a guideline for both studying hiring decisions in a dynamic context and linking task outcomes with hiring choices. As a result, future research can study matching in digital workplaces through evolving employer and contractor interactions while incorporating task outcomes.

Additionally, this methodological contribution (i.e., the two-component framework) spans beyond the specific context of online labor markets. For instance, the proposed framework can investigate employer evolving behavior in the offline setting (e.g., through platforms such as LinkedIn or BurningGlass). Task outcomes, in that case, could be promotions and other aspects of career trajectories. Overall, organizations stand to benefit by applying the two-component framework to study hiring behaviors and outcomes formally.

7.2. Implications for platform managers

Given the high potential for online work growth in the coming years (Agile-1 2016, Sundararajan 2016), understanding how employers can improve their likelihood of hiring successfully can provide platform managers with an additional tool to facilitate employer success. In particular, our work provides platforms with an explanation of why some hiring choices are more likely to be unsuccessful. Understanding what makes a successful hiring decision and how employers learn to make successful candidate choices is fundamentally important for an online labor market—both (1) employers who hire successfully and (2) contractors who receive positive feedback are more likely to keep participating in the platform (Tripp and Grégoire 2011, Jerath et al. 2011).

Furthermore, our study provides online labor markets with a data-driven approach (HMM framework) that models the evolving hiring abilities of the market's clients (employers). Through the proposed framework, managers can better understand the market's employer population: at any point in time, managers can identify employers who are struggling to make successful hiring choices. Managers can then intervene and provide these employers with information and guidance on how to minimize uncertainty and hire candidates that are likely to perform well. Even further, and depending on the market's policies, managers can experiment with subsidizing such employers (e.g., by reducing the platform's commission) to further increase their chances of success and subsequent use of the platform.

7.3. Limitations and future research

Hiring decisions in online labor markets depend on factors other than price and reputation. For example, employers base hiring decisions on contractor demographic attributes, as they prefer contractors from their own country (Gefen and Carmel 2008, Ghani et al. 2014, Lin and Viswanathan 2015) because communication is easier and trust and familiarity increases when interacting with local contractors (Arora and Forman 2007, Kossinets and Watts 2009). Furthermore, contractor gender (Chan and Wang 2017), skills, and certifications also affect employer choices (Goes and Lin 2013, Kokkodis et al. 2015). In this work, we focus on price and reputation because they can sufficiently illustrate how users learn in online labor markets. Our empirical analysis however, uses all these additional characteristics as control variables (Table 1), and in Appendix F and Table 16 we briefly mention how their estimated coefficients fit within the context of this framework. Future research can use this framework to further theorize and explore how employers adjust their hiring behavior across these additional characteristics.

A limitation of our study is that it relies on observational data to investigate how employers learn. The empirical evidence in Section 4 and Appendix A supports that at least some employers become successful through the learning path discussed in Section 2.3. Yet, we cannot claim that learning is the only causal path that drives these results. Hypothetically, a field experiment could randomly assign contractors (by drawing from a pre-determined pool of contractors with certain characteristics) of varying reputation and bid prices to be chosen by employers. This way, the experiment would enforce choice sets that are almost identical (same number of applicants, with similar characteristics) across employers, and hence eliminate any employer ability to attract better (or worse) pools of applicants. Next, the platform would observe how employers react after different types of outcomes and whether they adjust their behaviors after observing hiring outcomes. By ensuring that no employer exits the market, the experiment could show whether employer learning is indeed the sole causal path that drives these results.

The above hypothetical experiment has practical and ethical limitations (e.g., random assignment of applicants). Our study, even though observational, takes significant steps to investigate whether employers attract systematically better (or worse) pools of applicants (Appendices A.3, A.4, Figure 6), and whether employers who exit the market affect the results (Appendices C, A.1, A.2, A.3). However, we acknowledge that our observational analysis cannot fully eliminate these mechanisms, which could also explain part of the observed results.

Section 6 suggests that as employers explore market information and encounter unsuccessful outcomes, they adjust their behavior by reducing risk and achieve more successful outcomes. However, some users who take on additional risk are successful; future research could help better identify conditions in which seemingly risky selections of lower reputation contractors, for example, may be worthwhile.

Additionally, we expect that new users similarly adapt in other platforms and future research can help better understand how users learn and adjust their decisions in such different contexts. For example, platform designers could make prior outcomes more or less salient to users. Or platform managers could clarify or obscure how user decisions deviate from other users' decisions. Importantly, user adaptation likely depends on many context attributes, such as frequency of use, subjectivity of outcomes, process steps, composition and adaptation of other users, and consequence of unsuccessful outcomes. Future research can clarify these issues and many others.

7.4. Conclusion

Business is increasingly digital with online platforms underlying an ever-growing portion of modern commerce. These digital platforms overcome many market inefficiencies by facilitating transactions between consumers and producers throughout the world. But as new digital markets and platforms emerge, new platform users come in with diverse backgrounds that lead to varied ability to navigate the new markets. As a result, many online platform users struggle to adapt their offline experiences to digital platforms. These struggles are costly for both the user and the platform.

Yet, despite the importance of learning on digital platforms, we know little about how new users adjust their behavior based on experiences on the platform. Our work on employer learning in online labor markets opens a discussion on how platform users can learn and adjust their behaviors as they gain experience. Beyond hiring choices, in digital platforms such as ride-matching[2] or hotels,[3] many users might need to learn new managerial skills. Our framework provides a guideline on platform learning that other platforms can customize to understand how their users adjust their behavior over time in order to succeed.

[2] https://www.uber.com/newsroom/first-trip-tips-riders/

[3] https://money.usnews.com/money/blogs/my-money/2015/06/23/6-things-to-know-when-using-airbnb-for-the-first-time

References

Agile-1. 2016. Gig economy. http://www.hrotoday.com/wp-content/uploads/2016/07/Whitepaper_Agile2016-single.pdf. [Online; accessed 28-November-2020].

Agrawal, Ajay, John J. Horton, Nicola Lacetera, Elizabeth Lyons. 2015. Digitization and the contract labor market. *Economic Analysis of the Digital Economy* 219.

Agrawal, Ajay, Nicola Lacetera, Elizabeth Lyons. 2013. Does information help or hinder job applicants from less developed countries in online markets? Tech. rep., National Bureau of Economic Research.

Allon, Gad, Achal Bassamboo, Eren B. Çil. 2012. Large-scale service marketplaces: The role of the moderating firm. *Management Science* **58** 1854–1872.

Aral, Sinan, Lev Muchnik, Arun Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106** 21544–21549.

Arora, Ashish, Chris Forman. 2007. Proximity and information technology outsourcing: How local are it services markets? *Journal of Management Information Systems* **24** 73–102.

Austin, Peter C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* **46** 399–424.

Banker, Rajiv D., Iny Hwang. 2008. Importance of measures of past performance: Empirical evidence on quality of e-service providers. *Contemporary Accounting Research* **25** 307–337.

Barr, Pamela S., John L. Stimpert, Anne S. Huff. 1992. Cognitive change, strategic action, and organizational renewal. *Strategic Management Journal* **13** 15–36.

Baum, Christopher, Schaffer Mark, Stillman Steven. 2002. Ivreg2: Stata module for extended instrumental variables/2sls and gmm estimation. http://www.repec.org/bocode/i/ivreg2.html. [Online; accessed 28-November-2020].

Benson, Alan, Aaron Sojourner, Akhmed Umyarov. 2020. Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Science* **66** 1802–1825.

Brier, Elisabeth, Rich Pearson. 2018. Upwork's SVP of marketing explains what it takes to perfect an offering that relies on people. https://www.prnewswire.com/news-releases/snagajob-appoints-former-upwork-ceo-to-board-of-directors-300417689.html. [Online; accessed 28-November-2020].

Broyden, Charles G. 1970. The convergence of a class of double-rank minimization algorithms. General considerations. *Journal of Applied Mathematics* **6** 76–90.

Busenitz, Lowell W., Jay B. Barney. 1994. Biases and heuristics in strategic decision making: Differences between entrepreneurs and managers in large organizations. *Academy of Management Proceedings* **1994** 85–89.

Busenitz, Lowell W., Jay B. Barney. 1997. Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing* **12** 9–30.

Caves, Richard E., David P. Greene. 1996. Brands' quality levels, prices, and advertising outlays: Empirical evidence on signals and information costs. *International Journal of Industrial Organization* **14** 29–52.

Chan, Jason, Jing Wang. 2017. Hiring preferences in online labor markets: Evidence of a female hiring bias. *Management Science* **64** 2973–3468.

Chen, Daniel L., John J. Horton. 2016. Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research* **27** 403–423.

Church, Kenneth W., Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* **16** 22–29.

Danescu-Niculescu-Mizil, C., G. Kossinets, J. Kleinberg, L. Lee. 2009. How opinions are received by online communities: a case study on Amazon.com helpfulness votes. *International Conference on World Wide Web*. ACM, 141–150.

Denrell, Jerker, James G. March. 2001. Adaptation as information restriction: The hot stove effect. *Organization Science* **12** 523–538.

Dequech, David. 1999. Expectations and confidence under uncertainty. *Journal of Post Keynesian Economics* **21** 415–430.

Dey, Debabrata, Ming Fan, Conglei Zhang. 2010. Design and analysis of contracts for software outsourcing. *Information Systems Research* **21** 93–114.

Erdem, Tülin, Michael P Keane, Baohong Sun. 2008. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science* **27** 1111–1125.

Erev, Ido, Alvin E Roth. 2014. Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences* **111** 10818–10825.

Evans, Jonathan J., Barbara A. Wilson, Uwe Schuri, Jackie Andrade, Alan Baddeley, Olga Bruna, Tony Canavan, Sergiola Del Sala, Robin Green, Ritva Laaksonen, et al. 2000. A comparison of "errorless" and "trial-and-error" learning methods for teaching individuals with acquired memory deficits. *Neuropsychological Rehabilitation* **10** 67–101.

Filippas, Apostolos, John J. Horton, Joseph Golden. 2018. Reputation inflation. *Conference on Economics and Computation*. ACM, 483–484.

Forlani, David, John W. Mullins. 2000. Perceived risks and choices in entrepreneurs' new venture decisions. *Journal of Business Venturing* **15** 305–322.

Forney, David G. 1973. The viterbi algorithm. *IEEE* **61** 268–278.

Gefen, David, Erran Carmel. 2008. Is the world really flat? A look at offshoring at an online programming marketplace. *MIS Quarterly* 367–384.

Gefen, David, Gavriel Gefen, Erran Carmel. 2016. How project description length and expected duration affect bidding and project success in crowdsourcing software development. *Journal of Systems and Software* **116** 75–84.

Gerstner, Eitan. 1985. Do higher prices signal higher quality? *Journal of Marketing Research* 209–215.

Geva, Tomer, Maytal Saar-Tsechansky. 2016. Who's a good decision maker? Data-driven expert worker ranking under unobservable quality. *International Conference on Information Systems*.

Ghani, Ejaz, William R. Kerr, Christopher Stanton. 2014. Diasporas and outsourcing: Evidence from odesk and india. *Management Science* **60** 1677–1697.

Ghose, Anindya, Vilma Todri. 2016. Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly* **40** 889–910.

Gibson, Cristina B., Julian Birkinshaw. 2004. The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of Management Journal* **47** 209–226.

Goes, Paulo, Mingfeng Lin. 2013. Does information really "unravel" understanding factors that motivate sellers to seek third-party certifications in an online labor market. (Working paper).

Goes, Paulo B, Mingfeng Lin, Ching-man Au Yeung. 2014. "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research* **25** 222–238.

Good, Darren, Eric J. Michel. 2013. Individual ambidexterity: Exploring and exploiting in dynamic contexts. *The Journal of Psychology* **147** 435–453.

Goodman-Delahunty, Jane, Pär Anders Granhag, Maria Hartwig, Elizabeth F. Loftus. 2010. Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy, and Law* **16** 133.

Gopal, Anandasivam, Konduru Sivaramakrishnan, Mayuram S. Krishnan, Tridas Mukhopadhyay. 2003. Contracts in offshore software development: An empirical analysis. *Management Science* **49** 1671–1683.

Grewal, Dhruv, Jerry Gotlieb, Howard Marmorstein. 1994. The moderating effects of message framing and source credibility on the price-perceived risk relationship. *Journal of Consumer Research* **21** 145–153.

Grinold, Richard, Ronald Kahn. 2000. *Active portfolio management*. McGraw Hill New York, NY;.

Guadagni, Peter M., John D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* **2** 203–238.

Heckman, James J. 1979. Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society* 153–161.

Holmqvist, Mikael. 2004. Experiential learning processes of exploitation and exploration within and between organizations: An empirical study of product development. *Organization Science* **15** 70–81.

Hong, Yili, Ni Huang, Gordon Burtch, Chunxiao Li. 2016. Culture, conformity, and emotional suppression in online reviews. *Journal of the Association for Information Systems* **17** 2.

Hong, Yili, Paul A. Pavlou. 2017. On buyer selection of service providers in online outsourcing platforms for IT services. *Information Systems Research* **28** 547–562.

Hong, Yili, Chong Wang, Paul A. Pavlou. 2015. Comparing open and sealed bid auctions: Evidence from online labor markets. *Information Systems Research* **27** 49–69.

Horton, John J. 2010. *Online labor markets*. Springer.

Horton, John J., Lydia B. Chilton. 2010. The labor economics of paid crowdsourcing. *Conference on Electronic Commerce* 209–218.

Hull, Clark Leonard. 1930. Simple trial and error learning: A study in psychological theory. *Psychological Review* **37** 241.

Ipeirotis, Panagiotis G., Foster Provost, Jing Wang. 2010. Quality management on amazon mechanical turk. *Workshop on Human Computation* 64–67.

Jain, Hemant, Balaji Padmanabhan, Paul A. Pavlou, Raghu T. Santanam. 2018. Call for papers–special issue of information systems research–humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Information Systems Research* **29** 250–251.

Jerath, Kinshuk, Peter S. Fader, G.S. Bruce Hardie. 2011. New perspectives on customer "death" using a generalization of the pareto/nbd model. *Marketing Science* **30** 866–880.

Jueptner, Markus, Chris D. Frith, David J. Brooks, Richard Frackowiak, Richard Passingham. 1997. Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *Journal of Neurophysiology* **77** 1325–1337.

Kahneman, Daniel, Dan Lovallo. 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science* **39** 17–31.

Kokkodis, Marios. 2018. Dynamic recommendations for sequential hiring decisions in online labor markets. *International Conference on Knowledge Discovery & Data Mining*. ACM, 453–461.

Kokkodis, Marios. 2021. Dynamic, multi-dimensional, and skillset-specific reputation systems for online work. *Information Systems Research* (Articles In Advance).

Kokkodis, Marios, Panagiotis G. Ipeirotis. 2014. The utility of skills in online labor markets. *International Conference on Information Systems*.

Kokkodis, Marios, Panagiotis G. Ipeirotis. 2016. Reputation transferability in online labor markets. *Management Science* **62** 1687–1706.

Kokkodis, Marios, Panagiotis G. Ipeirotis. 2021. Demand-aware career path recommendations: A reinforcement learning approach. *Management Science* (Articles In Advance).

Kokkodis, Marios, Panagiotis Papadimitriou, Panagiotis G. Ipeirotis. 2015. Hiring behavior models for online labor markets. *International Conference on Web Search and Data Mining*. 223–232.

Kossinets, Gueorgi, Duncan J Watts. 2009. Origins of homophily in an evolving social network. *American Journal of Sociology* **115** 405–450.

Kreye, Melanie E, Yee M Goh, Linda B Newnes, P Goodwin. 2012. Approaches to displaying information to assist decisions under uncertainty. *Omega* **40** 682–692.

Lauren, Dyke. 2017. Snagajob appoints former Upwork CEO to board of directors. https://www.prnewswire.com/news-releases/snagajob-appoints-former-upwork-ceo-to-board-of-directors-300417689.html. [Online; accessed 28-November-2020].

Le, Quoc, Tomas Mikolov. 2014. Distributed representations of sentences and documents. *International Conference on Machine Learning*. 1188–1196.

Lee, Sunkee. 2019. Learning-by-moving: Can reconfiguring spatial proximity between organizational members promote individual-level exploration? *Organization Science* **30** 467–488.

Lee, Sunkee, Philipp Meyer-Doyle. 2017. How performance incentives shape individual exploration and exploitation: Evidence from microdata. *Organization Science* **28** 19–38.

Levinthal, Daniel A., James G. March. 1993. The myopia of learning. *Strategic Management Journal* **14** 95–112.

Liang, Chen, Yili Hong, Bin Gu. 2017. Moral hazards and effects of IT-enabled monitoring systems in online labor markets. *International Conference on Information Systems*. AIS.

Lin, Mingfeng, Yong Liu, Siva Viswanathan. 2016. Effectiveness of reputation in contracting for customized production: Evidence from online labor markets. *Management Science* **64** 345–359.

Lin, Mingfeng, Siva Viswanathan. 2015. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* **62** 1393–1414.

Liu, Yang, Xiangji Huang, Aijun An, Xiaohui Yu. 2008. Modeling and predicting the helpfulness of online reviews. *International Conference on Data Mining,* 443–452.

LIWC. 2018. Linguistic Inquiry and Word Count. [Online; accessed: 28-November-2020].

Lu, Yue, Panayiotis Tsaparas, Alexandros Ntoulas, Livia Polanyi. 2010. Exploiting social context for review quality prediction. *International Conference on World Wide Web,* 691–700.

Mannes, Albert E, Richard P Larrick, Jack B Soll. 2012. *The social psychology of the wisdom of crowds*. Psychology Press.

March, James G. 1991. Exploration and exploitation in organizational learning. *Organization Science* **2** 71–87.

Mason, Winter, Duncan J Watts. 2010. Financial incentives and the performance of crowds. *SIGKDD Explorations Newsletter* **11** 100–108.

McFadden, Daniel. 1973. *Conditional logit analysis of qualitative choice behavior*. University of California Berkley.

Menkhoff, Lukas, Ulrich Schmidt, Torsten Brozynski. 2006. The impact of experience on risk taking, overconfidence, and herding of fund managers: Complementary survey evidence. *European Economic Review* **50** 1753–1766.

Mom, Tom J.M., Frans A.J. Van Den Bosch, Henk W. Volberda. 2007. Investigating managers' exploration and exploitation activities: The influence of top-down, bottom-up, and horizontal knowledge inflows. *Journal of Management Studies* **44** 910–931.

Mom, Tom J.M., Frans A.J. Van Den Bosch, Henk W. Volberda. 2009. Understanding variation in managers' ambidexterity: Investigating direct and interaction effects of formal structural and personal coordination mechanisms. *Organization Science* **20** 812–828.

Moreno, Antonio, Christian Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research* **25** 865–886.

Murphy, Kevin P. 2012. *Machine learning: A probabilistic perspective*. The MIT Press.

Nelson, P. 1970. Information and consumer behavior. *The Journal of Political Economy* **78** 311–329.

Palich, Leslie E, D Ray Bagby. 1995. Using cognitive theory to explain entrepreneurial risk-taking: Challenging conventional wisdom. *Journal of Business Venturing* **10** 425–438.

Pallais, Amanda. 2014. Inefficient hiring in entry-level labor markets. *American Economic Review* **104** 3565–3599.

Peterson, Dane K, Gordon F Pitz. 1988. Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14** 85.

Plassmann, Hilke, John O'Doherty, Baba Shiv, Antonio Rangel. 2008. Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences* **105** 1050–1054.

Rahman, Hatim. 2018. Don't worship the stars: Ratings inflation in online labor markets. *International Conference on Information Systems*.

Rao, Akshay R, Kent B Monroe. 1989. The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research* **26** 351–357.

Robinson, Anthony T., Louis D. Marino. 2015. Overconfidence and risk perceptions: Do they really matter for venture creation decisions? *International Entrepreneurship and Management Journal* **11** 149–168.

Rogan, Michelle, Marie Louise Mors. 2014. A network perspective on individual-level ambidexterity in organizations. *Organization Science* **25** 1860–1877.

Rosing, Kathrin, Hannes Zacher. 2017. Individual ambidexterity: The duality of exploration and exploitation and its relationship with innovative performance. *European Journal of Work and Organizational Psychology* **26** 694–709.

Schilling, Melissa A, Patricia Vidal, Robert E Ployhart, Alexandre Marangoni. 2003. Learning by doing something else: Variation, relatedness, and the learning curve. *Management Science* **49** 39–56.

Sheng, Victor S, Foster Provost, Panagiotis G. Ipeirotis. 2008. Get another label? Improving data quality and data mining using multiple, noisy labelers. *International Conference on Knowledge Discovery and Data Mining*. ACM, 614–622.

Shiv, Baba, Ziv Carmon, Dan Ariely. 2005. Placebo effects of marketing actions: Consumers may get what they pay for. *Journal of Marketing Research* **42** 383–393.

Sinchaisri, Park, Gad Allon, Maxime Cohen. 2019. The impact of behavioral and economic drivers on gig economy workers. *Academy of Management Proceedings*, vol. 1. 10216.

Skinner, Burrhus Frederic. 1990. *The behavior of organisms: An experimental analysis*. BF Skinner Foundation.

Smith-Keiling, Beverly L, Hye In F Hyun. 2019. Applying a Computer-Assisted Tool for Semantic Analysis of Writing: Uses for STEM and ELL. *Journal of Microbiology & Biology Education* **20**.

Snir, Eli M, Lorin M Hitt. 2003. Costly bidding in online markets for it services. *Management Science* **49** 1504–1520.

Snyder, Charles R, Shane J Lopez. 2009. *Oxford Handbook of Positive Psychology*. Oxford Library of Psychology.

Sridhar, Shrihari, Raji Srinivasan. 2012. Social influence effects in online product ratings. *Journal of Marketing* **76** 70–88.

Stock, James H, Motohiro Yogo. 2002. Testing for weak instruments in linear IV regression. Tech. rep., National Bureau of Economic Research.

Sundararajan, Arun. 2016. *The sharing economy: The end of employment and the rise of crowd-based capitalism*. Mit Press.

Svveeney, Llian C. 1999. The role of perceived risk in the ouality-value relationship: A study in a retail environment. *Journal of Retailing* **75**.

Tripp, M. Thomas, Yany Grégoire. 2011. When unhappy customers strike back on the internet. *MIT Sloan Management Review* **52** 37–44.

Wooldridge, M. Jeffrey. 2010. *Econometric analysis of cross section and panel data*. MIT Press.

Xu, Yuqian, Baile Lu, Anindya Ghose, Hongyan Dai, Weihua Zhou. 2020. How do ratings and penalties moderate earnings on crowdsourced delivery platforms? (Working paper).

Yelle, Louis E. 1979. The learning curve: Historical review and comprehensive survey. *Decision Sciences* **10** 302–328.

Yin, Dezhi, Samuel Bond, Han Zhang. 2014. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly* **38** 539–560.

Yoganarasimhan, Hema. 2013. The value of reputation in an online freelance marketplace. *Marketing Science* **32** 860–891.

Zheng, Alvin, Yili Hong, Paul A. Pavlou. 2016. Matching in two-sided platforms for it services: Evidence from online labor markets. (Working paper).

Zucchini, Walter, Iain L. MacDonald, Roland Langrock. 2017. *Hidden Markov models for time series: An introduction using R*. CRC press.

Appendix A Further Examination of Alternative Explanations

The next sections examine alternative explanations and test the robustness of the observed results across different models and specifications.

A.1 Modeling the choice to participate

The main analysis does not directly model an employer's choice to keep participating or exit the market. The self-selection of employers to keep participating might affect the observed results.

A.1.1 An exit-state HMM: To test this alternative explanation, we first directly model an employer's choice to exit the market through an adjusted HMM. In particular, we assume an HMM that emits the following set of discrete observations (Section 4.3.1):

$$Y_{it} = \{\text{Hire-positive, Hire-negative, Idle}\}\,, \tag{9}$$

where Idle captures the choice of the employer to not participate in the marketplace during the next 30 days. Compared with our main HMM, we model these three discrete outcomes through a multinomial logit distribution. Appendix C details the estimation, along with information about the resulting hidden states. (Discrete outcomes are necessary for the HMM to model an exit state).

The exit-state HMM estimates the current state of each employer at the time of a hiring decision, while it also estimates contractors who have likely exited the market. The results (Table 4, Column A2) align with our main results (repeated for convenience in Column A1). Over time, employers adjust their hiring behaviors by assigning greater weights on feedback scores and becoming less price-sensitive.

A.1.2 Two-stage selection model: Two-stage selection models (Heckman 1979) can also model employers' choice to participate in the marketplace. Specifically, a first stage probit model estimates the likelihood of an employer to post an opening and hire a contractor. To model this first stage, we follow the same assumption as in the HMM-exit that a month of inactivity signals that the employer is not currently actively recruiting. This assumption generates a binary outcome dataset that allows the estimation of the likelihood of each employer to participate and hire:

$$\Pr(\text{Participate}_{it} = 1 | \mathbf{Z}_{it-1}) = \Phi(\alpha \mathbf{Z}_{it-1})\,, \tag{10}$$

Table 4 The employer choice to participate does not drive the main results

	Conditional Logit		Heckman	
	Main HMM	Exit-state HMM	Two-step	
DV: Hire	(A1)	(A2)	(B1)	(B2)
Feedback score × Less-successful	0.142^{***}	0.174^{***}	0.002^{*}	0.003^{***}
	(0.033)	(0.032)	(0.001)	(0.001)
Feedback score × More-successful	0.248^{***}	0.288^{***}	0.004^{***}	0.005^{***}
	(0.037)	(0.039)	(0.001)	(0.001)
Feedback score × Most-successful	0.379^{***}	0.356^{***}	0.006^{***}	0.006^{***}
	(0.027)	(0.027)	(0.001)	(0.000)
Bid price × Less-successful	-0.350^{***}	-0.287^{***}	-0.011^{***}	-0.012^{***}
	(0.053)	(0.047)	(0.002)	(0.002)
Bid price × More-successful	-0.146^{**}	-0.122^{*}	-0.006^{**}	-0.006^{***}
	(0.048)	(0.048)	(0.002)	(0.002)
Bid price × Most-successful	$-0.053.$	-0.069^{*}	-0.002	-0.001
	(0.030)	(0.031)	(0.001)	(0.001)
Inverse Mills ratio			0.043^{***}	0.035^{***}
			(0.007)	(0.007)
Task-varying variables × Less-successful	Yes	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes	Yes
Opening FE	Yes	Yes	No	No
Employer FE	NA	NA	No	Yes
Time FE	NA	NA	No	Yes
Task-category FE	NA	NA	No	Yes
Observations	238,364	238,364	245,438	238,364
AIC	52,763	52,822		
R-squared			0.212	0.044

The first two columns show results of conditional logit models that use latent states of an exit-state HMM (Appendix C) that directly models the choice of each employer to participate or exit the market. The last two columns show results of a Heckman two-stage model. Clustered standard errors in parentheses (robust standard errors for Column B1). The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

where $\mathbf{Z}_{it-1}$ is the same vector of covariates we used to model state transitions in the HMM (Appendix B.2) and it is disjoint from vector $\mathbf{X}_{it}$. Through Equation 10, we estimate the inverse Mills ratio, which captures the predicted probability of each employer to participate in the market. We then include this ratio in a linear model to control for the self-selection of employers to participate. (A conditional logit or a linear model with opening fixed effects is not applicable in this case as the opening fixed effects would eliminate the inverse mills ratio.) Table 4 shows the results without including any fixed effects (Column B1) and with employer, time, and task-category fixed effects (Column B2). Both results align with our main analysis and provide further support for Hypotheses 1, 2 and 3.

A.2 Analysis of repeat employers

Another alternative explanation for the observed results is that employers who stay on the platform for a long time have always been looking for certain-type contractors. This explanation argues that successful employers who keep participating do not change their behavior, and hence the observed effects stem from comparisons between successful and unsuccessful employers who eventually exit the market. The previous analysis that models the probability to participate in the market controls for part of this selection issue. Alternatively, we investigate this problem by focusing only on repeat employers and dropping employers who exit the market early. The results for subsamples of employers who hire 2, 3, 4, 5 or more contractors (Table 5) align with our main results, hence further supporting our theory that employers adjust their hiring behavior over repeated hires.

A.3 Instrumenting employer ability

Relatedly, unobserved reasons could influence some employers to exit and others to continue to participate in the market. Our previous analyses provide empirical evidence that our observations are not an artifact of comparisons between employers who exited with those who did not exit but, instead, result from our theorized mechanism of employer learning. Alternatively, we could use instrumental variables to capture the state of each employer and their choice to continue participating. Inspired by prior research (Sinchaisri et al. 2019, Xu et al. 2020), we consider the following instrumental variables for the current employer state:

- ◇ State of every other employer at o: We first consider a Hausman-based instrument that predicts the state of each focal employer by getting the most frequent state of every other employer in the dataset who have completed o hires. To do so, we use both the mean and the median state of every other employer.
- ◇ Percent of employers who continue at o: Next, we consider the percentage of employers (excluding the focal employer) who keep participating after completing $o-1$ hires.

Table 5 **Comparisons between repeat employers and employers who exit the market do not drive the results**

	All employers	$O \geq 2$	$O \geq 3$	$O \geq 4$	$O \geq 5$
DV: Hire	(A1)	(A2)	(A3)	(A4)	(A5)
Feedback score × Less-successful	0.142^{***}	0.150^{***}	0.139^{***}	0.128^{**}	0.141^{*}
	(0.033)	(0.035)	(0.037)	(0.046)	(0.057)
Feedback score × More-successful	0.248^{***}	0.253^{***}	0.255^{***}	0.306^{***}	0.297^{***}
	(0.037)	(0.038)	(0.041)	(0.053)	(0.067)
Feedback score × Most-successful	0.379^{***}	0.376^{***}	0.378^{***}	0.369^{***}	0.356^{***}
	(0.027)	(0.027)	(0.029)	(0.033)	(0.038)
Bid price × Less-successful	-0.350^{***}	-0.346^{***}	-0.361^{***}	-0.372^{***}	-0.378^{***}
	(0.053)	(0.055)	(0.058)	(0.073)	(0.089)
Bid price × More-successful	-0.146^{**}	-0.165^{***}	-0.176^{***}	-0.238^{***}	-0.309^{***}
	(0.048)	(0.049)	(0.052)	(0.065)	(0.081)
Bid price × Most-successful	$-0.053.$	$-0.058.$	-0.041	$-0.070.$	-0.067
	(0.030)	(0.030)	(0.032)	(0.037)	(0.044)
Task-varying variables × Less-successful	Yes	Yes	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes	Yes	Yes
Opening FE	Yes	Yes	Yes	Yes	Yes
Observations	238,364	224,758	206,832	144,172	99,397
AIC	52,763	49,889	45,906	32,313	22,635

Results of conditional logit models. The first column shows results of the complete dataset (all employers). The next four columns show results for employers who make O or more hiring choices. Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

◇ Percent of employers who continue at o after outcome: Finally, we consider the percentage of employers who continue participating after completing $o-1$ hires (excluding the focal employer) and whose the hiring outcome of their $o-1$ hire was the same as the outcome of the $o-1$ hire of the focal employer.

Table 6 shows the two-stage least squares (2SLS) results along with two tests regarding the statistical validity of the instrumental variables: a test that examines whether the focal specification is underidentified and a weak identification test that examines whether the instrumental variables are only weakly correlated with latent employer states (Baum et al. 2002). Each column shows the results of a different instrumental variable. Based on the tests, the instrumental variables do not underidentify (reject the null at least with $p < 0.001$) and are not weakly correlated with the HMM states ($F-$statistic ranges between 7,176 and 22,018). Hence, we can interpret the observed results with increased statistical comfort. Across all instrumental variables, the results remain

Table 6 The main results remain unchanged after instrumental variables control for employers' latent states

	Two-Stage Least Squares			
DV: Hire	(A1)	(A2)	(A3)	(A4)
Feedback score × Less-successful	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Feedback score × More-successful	0.004^{***}	0.005^{***}	0.005^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Feedback score × Most-successful	0.006^{***}	0.006^{***}	0.006^{***}	0.007^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
Bid price × Less-successful	-0.012^{***}	-0.012^{***}	-0.012^{***}	-0.012^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Bid price × More-successful	-0.006^{**}	-0.006^{***}	-0.007^{***}	-0.007^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Bid price × Most-successful	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Instruments for Experience/state				
State of everyone else at o (mean)	Yes	No	No	No
State of everyone else at o (median)	No	Yes	No	No
Percent of employers who continue at o	No	No	Yes	No
Percent of employers who continue at o after outcome	No	No	No	Yes
Instrument validity				
Underidentification ($p-$value of K-P rk LM statistic)	0.000	0.000	0.000	0.000
Weak identification (K-P rk $F-$statistic)	22,018	1,381,561	19,494	7,176
Task-varying variables × Less-successful	Yes	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes	Yes
Employer FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Task-category FE	Yes	Yes	Yes	Yes
Observations	238,364	238,364	238,364	238,364

Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. K-P stands for Kleibergen-Paap. The null hypothesis of the underidentification test is that the equation is underidentified. The null hypothesis of the weak identification test is that the instruments are only weakly correlated with the endogenous regressors (Baum et al. 2002); for one instrument, we reject the null hypothesis at $p < 0.05$ if the $F-$statistic is greater than 6.66 (Stock and Yogo 2002). ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

robust: employers tend to rely more on feedback score and be less price-sensitive as they move from Less-successful to Most-successful states.

Table 7 The main results hold for alternative measures of success such as rehires

	All employers		Evolving employers
DV: Hire	(A1)	(A2)	(A3)
Feedback score	0.289^{***}		
	(0.018)		
Bid price	-0.133^{***}		
	(0.023)		
Feedback score × Less-successful		0.185^{***}	0.163^{***}
		(0.030)	(0.034)
Feedback score × Most-successful		0.366^{***}	0.360^{***}
		(0.026)	(0.034)
Feedback score × Rehire-prone		0.242^{***}	0.304^{**}
		(0.053)	(0.114)
Bid price × Less-successful		-0.269^{***}	-0.320^{***}
		(0.043)	(0.050)
Bid price × Most-successful		-0.099^{***}	-0.114^{**}
		(0.029)	(0.039)
Bid price × Rehire-prone		0.034	0.028
		(0.068)	(0.146)
Task-varying variables	Yes		
Task-varying variables × Less-successful		Yes	Yes
Task-varying variables × Most-successful		Yes	Yes
Task-varying variables × Rehire-prone		Yes	Yes
Opening FE	Yes	Yes	Yes
Observations	238,364	238,364	135,010
AIC	52,595	52,563	29,861

Results of conditional logit models. The first two columns show results of the complete dataset (all employers). The last column focuses only on employers who transition across two or more states. Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

A.4 Do more experienced employers attract better contractors?

Our main analysis does not rank contractors according to their reputation within openings; we chose to only normalize biding prices since they vary significantly between openings. On the other hand, contractor reputation is universal on the platform and is not opening-dependant. Yet, as we discussed in the previous section, different types of employers might attract different type of contractors in terms of reputation, and that might be interfering with our results. Table 8 shows that this is likely not the case: our findings remain qualitatively the same once we normalize

Table 8 State-specific effects of normalized feedback score and bid price on hiring choices

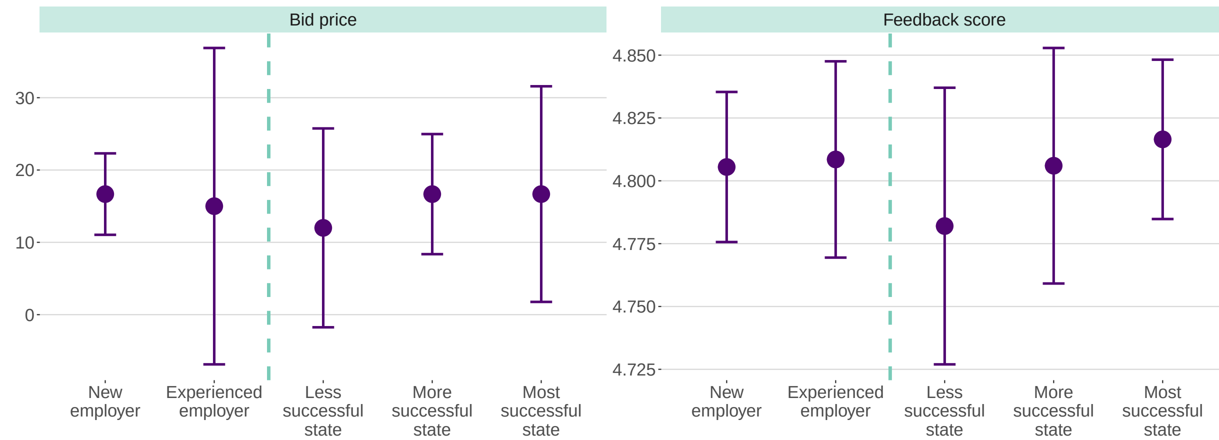
	All employers		Evolving employers
	(Without HMM)	(With HMM)	(With HMM)
DV: Hire	(A1)	(A2)	(A3)
Bid price	-0.127^{***}		
	(0.023)		
Feedback score	0.354^{***}		
	(0.022)		
Bid price × Less-successful		-0.348^{***}	-0.373^{***}
		(0.053)	(0.058)
Bid price × More-successful		-0.140^{**}	-0.165^{**}
		(0.048)	(0.051)
Bid price × Most-successful		-0.048	-0.032
		(0.030)	(0.036)
Feedback score (Binary) × Less-successful		0.180^{***}	0.161^{**}
		(0.050)	(0.055)
Feedback score (Binary) × More-successful		0.305^{***}	0.321^{***}
		(0.045)	(0.049)
Feedback score (Binary) × Most-successful		0.428^{***}	0.464^{***}
		(0.028)	(0.035)
Task-varying variables	Yes		
Task-varying variables × Less-successful		Yes	Yes
Task-varying variables × More-successful		Yes	Yes
Task-varying variables × Most-successful		Yes	Yes
Opening FE	Yes	Yes	Yes
Observations	238,364	238,364	174,873
AIC	52,882	52,822	39,145

Feedback score is binarized such that contractors who have reputation above the median contractor reputation take the value 1, while the rest of the contractors take the value 0. Results of conditional logit models. The first two columns show results of the complete dataset (all employers). The last column focuses only on employers who transition across two or more states. Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. Tables 16 and 17 show the state-specific coefficients for all task-varying variables. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

reputation scores within openings (hence making the choices depend on the contractor ranking within a given opening).

Furthermore, Figure 6 shows the median bid price and feedback scores of applicants that job postings by new, experienced, "less-successful," "more-successful," and "most-successful" employers attract. We observe no significant difference ($p > 0.1$) between these distributions, hence corroborating the results of Table 8 and showing that more experienced and more successful employers do not attract systematically better contractors.

Figure 6 Correlogram of the focal and task-varying variables



A.5 Alternative measures of success

For the main HMM, we use the private feedback that employers assign to contractors (Appendix B, Table 13). However, this feedback is employer-dependent. For instance, some employers might be more lenient than others. The focal dataset allows us to evaluate an alternative measure of success: rehires. With rehires, we can be more confident that the prior collaboration was successful.

Specifically, we train an adjusted HMM that emits the following set of discrete observations:

$$Y_{it} = \{\text{Hire-positive, Hire-negative, Rehire}\}. \tag{11}$$

Appendix D describes the estimation details along with information about the resulting hidden states of this HMM. Table 7 shows the results for all employers (Columns A1 and A2) and for evolving employers (Column A3). The results show a new Rehire-prone state that includes employers with the highest likelihood of rehiring; the Most-successful state still represents employers with a higher likelihood of hiring successfully; the Less-successful state consolidates employers who hire unsuccessfully (Figure 10). The results align with our main analysis: as employers move from Less-successful to Most-successful states, they become less price-sensitive, and they rely more on feedback score. In the Rehire-prone state, employers put smaller weights on feedback scores ($p < 0.001$) than in the Most-successful state and are even less price-sensitive (positive but insignificant coefficients of bid price). The reduced importance of feedback score for rehires is consistent with our theoretical development about learning—for rehires, employers can rely on even more direct information about the contractor (i.e., their personal experience) rather than the available information from the platform.

A.6 Analysis of matched tasks

Selection bias might exists because systematically, different types of tasks end up being in the same states. For instance, it might be that cheap tasks end up in the "Most-successful" state, while expensive tasks end up in the "Less-successful" state. To test whether such a selection bias affects the observed results, we match tasks according to their observed characteristics. Specifically, to match tasks, we use "Task category" dummies, the "Number of bids," and the "Mean task bid." Like previous studies that matched tasks in online labor markets (Lin et al. 2016), we use propensity score matching (Austin 2011) to estimate each task's likelihood to be in each state. Since we consider three states, we use one-hot encoding to define treatment and match without replacement treated and untreated tasks with predicted probabilities within a caliper size equal to 0.001. We drop instances that do not meet this criterion (Aral et al. 2009). Table 9 shows the results, which align with our main results and support Hypotheses 1 and 2. Hence, we conclude that task types and characteristics likely do not drive our findings.

A.7 Analysis of technical and non-technical tasks

An additional potential alternative explanation is that employers might have different abilities for estimating the success of different task characteristics. For instance, an employer might have a higher ability to identify good programmers but a lower ability to identify good administrative assistants. As a result, our findings might be driven by comparing employers' varying abilities across different types of tasks.

Table 9 Results remain robust across samples of matched tasks

	Less-successful vs. All	More-successful vs. All	Most-successful vs. All
DV: Hire	(A1)	(A2)	(A3)
Feedback score × Less-successful	0.139^{***}	$0.139.$	0.146^{***}
	(0.033)	(0.081)	(0.034)
Feedback score × More-successful	0.295^{**}	0.248^{***}	0.248^{***}
	(0.099)	(0.037)	(0.037)
Feedback score × Most-successful	0.376^{***}	0.315^{***}	0.405^{***}
	(0.071)	(0.054)	(0.032)
Bid price × Less-successful	-0.351^{***}	-0.262^{*}	-0.348^{***}
	(0.053)	(0.120)	(0.053)
Bid price × More-successful	-0.107	-0.145^{**}	-0.144^{**}
	(0.124)	(0.048)	(0.048)
Bid price × Most-successful	-0.073	0.011	-0.032
	(0.076)	(0.064)	(0.035)
Task-varying variables × Less-successful	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes
Opening FE	Yes	Yes	Yes
Observations	80,931	93,181	221,770
AIC	16,620	20,701	45,893

The table shows conditional logit estimates. The first four column matches tasks from Less-successful state and any other state; the second column matches tasks from More-successful state and any other state; the last column matches tasks from Most-successful state and any other state. Caliper size is equal to 0.001. Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

One way to test this explanation is to split the dataset into different types of tasks. Specifically, we group tasks that require technical skills (e.g., software development, engineering, and accounting skills) and tasks that require non-technical skills (e.g., administrative support, data entry, and customer service skills). Columns A1 and A2 in Table 10 show the results, which align with our main results and support Hypotheses 1 and 2. Hence, this analysis shows that it is unlikely for employers' task-specific abilities to drive our findings.

A.8 Analysis of diverse and focused employers

Similarly, another alternative explanation could argue that the results are either due to employers who hire on focused tasks or employers who choose to hire on diverse tasks. To test this explanation, we split our data into a set that includes employers who hire only in one task category (focused) and into a set that includes employers who hire in more than one task category (diverse). Columns

Table 10 Results remain robust across non-technical and technical tasks, and across focused and diverse employers

DV: Hire	Non-technical tasks (A1)	Technical tasks (A2)	Focused employers (B1)	Diverse employers (B2)
Feedback score × Less-successful	0.135^{**}	0.151^{**}	0.099^{*}	0.174^{***}
	(0.042)	(0.054)	(0.049)	(0.045)
Feedback score × More-successful	0.271^{***}	0.199^{***}	0.244^{***}	0.252^{***}
	(0.047)	(0.060)	(0.059)	(0.048)
Feedback score × Most-successful	0.391^{***}	0.362^{***}	0.334^{***}	0.401^{***}
	(0.035)	(0.042)	(0.046)	(0.033)
Bid price × Less-successful	-0.316^{***}	-0.411^{***}	-0.407^{***}	-0.320^{***}
	(0.066)	(0.089)	(0.087)	(0.067)
Bid price × More-successful	-0.118^{*}	-0.206^{*}	$-0.145.$	-0.156^{**}
	(0.058)	(0.085)	(0.078)	(0.060)
Bid price × Most-successful	-0.012	-0.129^{**}	0.019	-0.087^{*}
	(0.037)	(0.050)	(0.053)	(0.036)
Task-varying variables × Less-successful	Yes	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes	Yes
Opening FE	Yes	Yes	Yes	Yes
Observations	152, 392	85, 972	73, 247	165, 117
AIC	34, 313	18, 447	17, 595	35, 183

The table shows conditional logit estimates. The first two columns estimate how employers adjust their hiring preferences across non-technical and technical tasks. The last two columns estimate how focused and diverse employers adjust their hiring preferences over time. Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

B1 and B2 in Table 10 show the results. These results align with our main findings and show that independent of whether employers are focused or diverse, they value feedback scores more and are less price-sensitive as they move from low to Most-successful states. Hence, an employer's level of diversification is unlikely to drive the main findings.

A.9 Alternative modeling of component A

What if the results are an artifact of our modeling choices? Clustering approaches could also group tasks according to their success level and replace the HMM in the framework's component A. For instance, a $K-$means clustering approach could replace the proposed HMM. However, such an approach would rely on simple thresholds (e.g., contractor performance higher or lower than a given threshold) and would completely omit the dependence between observations of the same employer. These are critical benefits of the HMM approach, as they protect employers' state annotations from incidental fluctuations—for example, an unsuccessful instance of an employer who mostly makes successful hiring choices. More importantly, such a clustering technique cannot model employer evolution, which is one of this work's side goals with significant managerial implications (Section 7). (The HMM resolves these shortcomings as it (1) naturally models employer evolution, (2) it encodes the dependence between observations of the same employer, and (3) it is robust across incidental observations as it stochastically estimates an employer's success level through the Viterbi algorithm.)

Despite these drawbacks of a clustering approach compared with our HMM, it is important to test whether our results are susceptible to different modeling choices. Table 11 shows the results of a $K-$means approach (with $K=3$). These results further support Hypotheses 1 and 2, indicating that our main observations are not an artifact of our HMM approach. Furthermore, note the HMM states explain a higher amount of variance in the dataset than the Kmeans states, as captured by the lower AIC score of Column A2 of Table 2. Table 2.

A.10 Alternative fixed effects models

The task fixed effects of the conditional logit specification eliminate the static unobserved error per task. Because the time, the task category, the employer, and the total number of received bids do not vary within tasks, we drop them from the main specification. Alternatively, we could include these variables into different FE specifications. In particular, we consider the following:

$$\Pr(H_{io}^{j}=1|\boldsymbol{X}_{jo}) \sim \boldsymbol{\beta}'\boldsymbol{X}_{jo} + \boldsymbol{\delta}[\text{Number of bids, Employer reputation, Skillset D2V}]'_o \\ + T_o + E_o + TC_o + \varepsilon \,, \tag{12}$$

where T_o are the time fixed effects, E_o the employer fixed effects and TC_o the task-category fixed effects. Note that these fixed effects have the subscript o, as they are task-specific.

The previous specification also controls for (1) the popularity of each task opening through the variable "Numbers of bids," the employer's accumulated reputation on the platform [4](Benson et al.

[4] In our focal platform, at the end of every collaboration, the contractor gets the chance to privately rate each employer. The platform uses this feedback internally. However, employer reputation is not visible to contractors when

Table 11 Alternative modeling for component A does not affect the main results ($K-$means instead of HMM)

	All employers		Evolving employers
	(Without $K-$means)	(With $K-$means)	(With $K-$means)
DV: Hire	(A1)	(A2)	(A3)
Feedback score	0.288^{***}		
	(0.018)		
Bid price	-0.131^{***}		
	(0.023)		
Feedback score × Less-successful		0.102^{*}	0.106^{*}
		(0.046)	(0.050)
Feedback score × More-successful		0.199^{***}	0.199^{***}
		(0.035)	(0.038)
Feedback score × Most-successful		0.360^{***}	0.347^{***}
		(0.024)	(0.031)
Bid price × Less-successful		-0.346^{***}	-0.327^{***}
		(0.077)	(0.082)
Bid price × More-successful		-0.265^{***}	-0.267^{***}
		(0.050)	(0.054)
Bid price × Most-successful		-0.063^{*}	-0.052
		(0.027)	(0.036)
Task-varying variables	Yes		
Task-varying variables × Less-successful		Yes	Yes
Task-varying variables × More-successful		Yes	Yes
Task-varying variables × Most-successful		Yes	Yes
Opening FE	Yes	Yes	Yes
Observations	238,364	238,364	154,020
AIC	52,833	52,799	34,111

The table shows conditional logit estimates. The first two columns show the results of the complete dataset (all employers). The last column focuses only on employers who transition across two or more states. Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

2020), the employer experience, and the specific skills that the task is requesting ("Skillset D2V"). In particular, to control for any skillset-specific effects, we use document embeddings (D2V, Le and Mikolov 2014): we assume that each skillset is a document, and then we decompose them to a vector of real numbers (for this decomposition, we use 20 dimensions; Kokkodis and Ipeirotis 2021).

they apply. Contractors typically can only observe the experience (number of jobs) of each employer at the time of application. Despite the fact that employer reputation is invisible to the applying contractors, we would like to explore whether other employer characteristics can proxy employer reputation and hence attract systematically better candidates.

As a result, this analysis controls for task popularity, employer reputation, and skillset variation, characteristics that our main analysis captures through the opening fixed effects.

We estimate Equation 12 through a linear probability and a logit model. Table 12 presents the results. Columns A1 to A5 show models that include different combinations of the three task-specific variables. Column B1 shows the coefficients of a complete logit specification. All models provide additional support for Hypotheses 1 and 2 ($p < 0.01$). Overall, neither the model choice nor the additional task-specific variables are likely to drive our main results.

Table 12 Alternative fixed effects models further support our main results

	Linear models					Logit
DV: Hire	(A1)	(A2)	(A3)	(A4)	(A5)	(B1)
Feedback score × Less-successful	0.003***	0.003***	0.003***	0.003***	0.003***	0.118***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.031)
Feedback score × More-successful	0.005***	0.005***	0.005***	0.005***	0.005***	0.206***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.034)
Feedback score × Most-successful	0.007***	0.006***	0.006***	0.006***	0.006***	0.331***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.024)
Bid price × Less-successful	−0.012***	−0.012***	−0.012***	−0.012***	−0.012***	−0.321***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.051)
Bid price × More-successful	−0.006**	−0.006**	−0.006**	−0.006**	−0.006**	−0.140**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.046)
Bid price × Most-successful	−0.002	−0.001	−0.001	−0.001	−0.001	−0.029
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.029)
Number of bids		−0.001***	−0.001***	−0.001***	−0.001***	−0.030***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Employer reputation				0.000	0.000	−0.004
				(0.000)	(0.000)	(0.007)
Employer experience				−0.000	−0.000	−0.002
				(0.000)	(0.000)	(0.004)
Task-varying variables × Less-successful	Yes	Yes	Yes	Yes	Yes	Yes
Task-varying variables × More-successful	Yes	Yes	Yes	Yes	Yes	Yes
Task-varying variables × Most-successful	Yes	Yes	Yes	Yes	Yes	Yes
Skillset D2V (20 variables)	NA	No	Yes	Yes	Yes	Yes
Opening FE	Yes	No	No	No	No	No
Employer FE	NA	Yes	Yes	Yes	Yes	Yes
Time FE	NA	Yes	Yes	Yes	Yes	Yes
Task-category FE	NA	Yes	Yes	Yes	Yes	Yes
R-squared	0.080	0.048	0.048	0.048	0.048	
Observations	238,364	238,364	238,364	238,364	238,364	238,364
AIC						66,306

Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

Appendix B Implementation Details of the Main HMM

The next paragraphs describe the estimation process of the HMM along with the resulting hidden states that feed Component B in Figure 3.

B.1 HMM Estimation

To estimate the parameter vectors $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ and the initial probabilities $\boldsymbol{\pi}$ the framework maximizes the conditional probability of the observations given the model's structure. Assume that the sequence of M observations for a given employer i is $\boldsymbol{Y}_i = Y_{i1}, Y_{i2}, \ldots, Y_{it}, \ldots Y_{iM}$. Also, assume that $\boldsymbol{Y}_i$ is the result of a sequence of latent states, $\boldsymbol{S}_i = S_{i1}, S_{i2}, \ldots, S_{it}, \ldots S_{iM}$, where $S_{it} \in \mathcal{S}$. This sequence of states is affected by the employer's history, i.e., vectors $\boldsymbol{Z}_{i1:M-1} = \boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}, \ldots, \boldsymbol{Z}_{it}, \ldots, \boldsymbol{Z}_{iM-1}$.

Based on the structure of the HMM, the conditional likelihood of observing $\boldsymbol{Y}_i$ is

$$\Pr(\boldsymbol{Y}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) = \prod_{t=1}^{M} \Pr(Y_{it}|S_{it};\boldsymbol{\theta})\,. \tag{13}$$

The conditional probability of observing the sequence $\boldsymbol{S}_i$ is

$$\Pr(\boldsymbol{S}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\pi}) = \boldsymbol{\pi}(S_{i1}) \prod_{t=2}^{M} \Pr(S_{it}|S_{it-1}\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha})\,. \tag{14}$$

Hence, the likelihood of this sequence of observations for employer i is

$$\begin{aligned}
l(\boldsymbol{Y}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi}) &= \Pr(\boldsymbol{Y}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi}) \\
&= \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i,\boldsymbol{S}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi}) \\
&\overset{\text{HMM structure}}{=} \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) \Pr(\boldsymbol{S}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\pi}) \\
&\overset{\text{Equations 13,14}}{=} \Pr(Y_{i1}|S_{i1};\boldsymbol{\theta}) \times \boldsymbol{\pi}(S_{i1}) \times \qquad (15) \\
&\qquad \sum_{\forall \boldsymbol{S}_i} \prod_{t=2}^{M} \Pr(Y_{it}|S_{it};\boldsymbol{\theta}) \Pr(S_{it}|S_{it-1}\boldsymbol{Z}_{it-1};\boldsymbol{\gamma},\boldsymbol{\alpha})\,, \qquad (16)
\end{aligned}$$

where the structure of the HMM allows decomposition of the joint probability of $Pr(\boldsymbol{Y}_i,\boldsymbol{S}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi})$ (Murphy 2012). The total likelihood for N employers is as follows:

$$L(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi}) = \prod_{i=1}^{N} l(\boldsymbol{Y}_i|\boldsymbol{Z}_{i1:M-1};\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\pi})\,. \tag{17}$$

Table 13 Descriptive statistics of the outcome and transition variables of the HMM

	Mean	Median	StD	Min	Max
Main HMM outcome variable					
Contractor performance (y_i^j)	0.87	1	0.21	0	1
Transition variables ($\boldsymbol{Z}_{it-1}$)					
Employer prior success	0.68	0.8	0.36	0	1
Employer money spent	331	85	872	0	21669
Employer experience	3.3	2	5	0	88
Employer total rehires	0.03	0	0.24	0	6

We estimate the parameters $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\pi}$ through maximizing this likelihood function (minimize its negative log-likelihood) numerically by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden 1970). Once we estimate the parameters, the Viterbi algorithm (Forney 1973) calculates each employer's most likely state sequence. For the states to make economic sense, we impose an order constraint, such that a state s_k represents a greater likelihood of successful outcomes than a state s_v when $k > v$ (Ghose and Todri 2016, Zucchini et al. 2017).

B.2 HMM outcome and transition variables

The HMM includes an outcome variable Y_{it} and a historical vector $\mathbf{Z}_{it-1}$ that customizes the transition probabilities of each employer to a new state.

HMM outcome variable: Each HMM state emits continuous observations of contractor performance (Equation 3). To learn the emission distributions, we use the scores that employers rate contractors at the completion of each task ("Contractor performance"). These performance scores are private and only available to the platform.

Transition variables: The focal dataset provides several measures that drive employer evolution (i.e., variables that form vectors $\boldsymbol{Z}_{it-1}$). The HMM considers a vector of covariates $\boldsymbol{Z}_{it-1}$ that affect employers' transition probabilities to higher (or lower) success states. This vector encodes the observed history of each employer in the marketplace *up to time* $t-1$. To capture historical events that affect employers' learning and calibration processes, vector $\boldsymbol{Z}_{it-1}$ includes the ratio of the tasks that an employer has previously completed successfully ("Employer prior success"), the total money spent on previous tasks ("Employer money spent"), the total number of completed tasks ("Employer experience"), and the total number of repeat collaborations ("Employer total rehires"). Table 13 shows the descriptive statistics of these variables.

B.3 Description of the best-fitted HMM

The HMM framework requires two choices: the number of states (K) and the emission function f. To make the best possible choices, we compare various configurations and calculate their Bayesian Information Criterion (BIC) scores (Murphy 2012). In particular, the framework considers the following continuous probability distributions for modeling function f:

$$f \in \{\text{ Normal, Gamma, Lognormal}\}\,, \tag{18}$$

and the following set of number of states:

$$K \in \{2, 3, 4, 5\}\,. \tag{19}$$

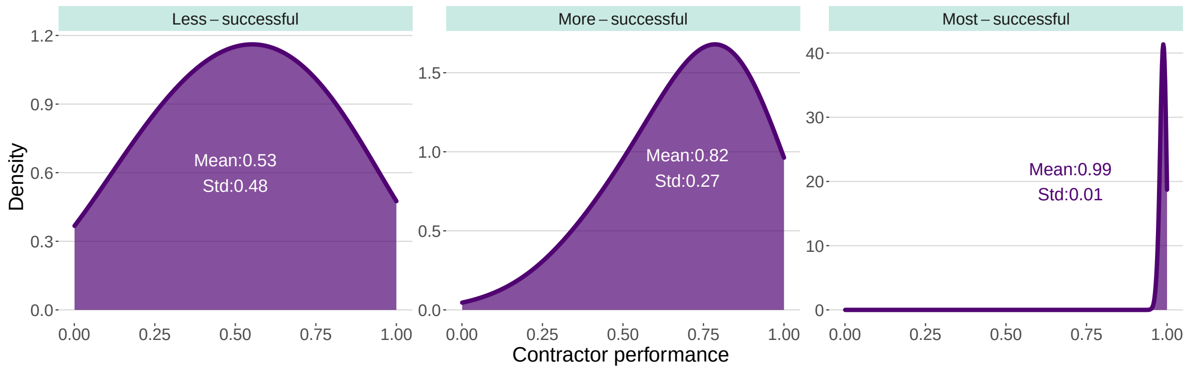
For each combination in $\{K \times f\}$, we estimate the set of parameters $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\pi}$ that maximize the conditional likelihood of the observed behavior (Equation 17). Maximization processes such as these depend on the parameter initialization and are thus prone to finding local maxima rather than the global maximum. Hence, to increase the likelihood of selecting the best number of states K and the most appropriate emission function f, we conduct a search of 1,000 randomly generated initial configurations for each combination in $\{K \times f\}$. Based on this analysis, we choose:

- f to follow a Normal distribution.
- $K = 3$.

Figure 7 shows the emission distributions of each one of the three identified HMM states. For readability, we label the three states as:

◇ "<u>Less-successful</u>" state: employers make higher-variance decisions that result in unsuccessful task outcomes (mean contractor performance of 0.53 out of 1).

Figure 7 Performance distributions of the best-fitted HMM



The HMM identifies three different states. The "Less-successful" state where employers hire contractors who underperform (Mean 0.53), the "More-successful" state where employers hire contractors with better performance (Mean 0.82), and the "Most-successful" state where employers hire contractors who perform well (Mean 0.99).

- ◇ "More-successful" state: employers make lower-variance decisions that result in more successful task outcomes (mean contractor performance of 0.82 out of 1).
- ◇ "Most-successful" state: employers make lower-variance decisions that result in successful task outcomes (mean contractor performance of 0.99 out of 1).

B.4 HMM-based evidence in support of the confidence mechanism

Adding to the empirical evidence presented in Section 6.2, Figure 8 shows how employer confidence and contractor performance change when employers transition from less-to-more and more-to-less successful HMM states. The figure measures confidence through clout and emotional tone extracted from each opening's task description text (Section 6.2). On the $y-$axis, the figure shows the transition adjustment of employer i at time t, defined as:

$$\text{Transition adjustment}_{it} = X_{it} - X_{it-1} \,, \tag{20}$$

where $X \in \{\text{Clout, Emotional tone, Contractor performance}\}$. Figures 8B and 8C show that as employers move from more-to-less successful states, their confidence increases (clout at $p < 0.01$, emotional tone at $p < 0.1$). The opposite is true when employers move from less-to-more successful states. Furthermore, Figure 8C shows that as employers move to more successful states, their hired contractor performance increases ($p < 0.001$). This analysis adds extra support to our arguments presented in Section 6.

Figure 8 Adjustment of confidence across state transitions

Employers appear to post tasks with increased confidence as they move to less successful states, and decreased confidence as they moved to more successful states (A,B). The opposite is true in terms of outcomes (C). Error bars show 95% bootstrapped confidence intervals. The figure ignores transitions to the same state, and only considers transitions to different states.

Appendix C An Exit-State HMM

The main HMM presented in Appendix B does not model an employer's likelihood to participate in the market. As we discuss in Section A.1, this source of selection bias could be driving the observed results.

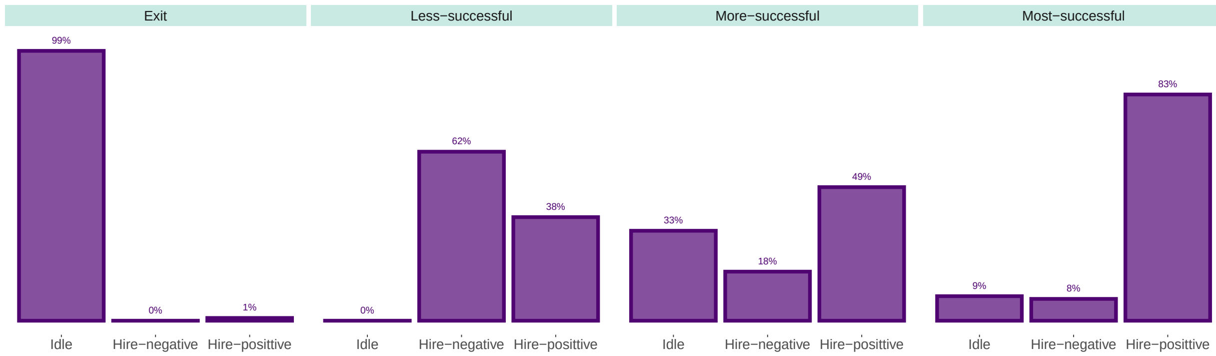
One way to evaluate this alternative explanation is to explicitly model each employer's likelihood to not participate in the market. Specifically, we adjust our HMM to have the following set of discrete outcomes (Section A.1.1):

$$Y_{it} = \{\text{Hire-positive, Hire-negative, Idle}\}\,. \tag{21}$$

The estimation process of this HMM is the same as the one described in Appendix B.1, with the only difference that we model emission probabilities through a multinomial distribution. By using the transition variables presented in Table 13, we follow a similar tuning process as the one described in Appendix B.3, and we identify that the best-fitted HMM includes $K = 4$ states.

Figure 9 shows the discrete emission distributions of each one of the four identified HMM states. For readability, we label these states as:

Figure 9 Performance distributions of the best-fitted exit-state HMM



The HMM identifies four different states. An "Exit" state where employers choose to not participate in the marketplace, a "Less-successful" state where employers are more likely to hire contractors who perform poorly, a "More-successful" state where employers hire contractors who perform on average better than those in the "Less-successful" state, and a "Most-successful" state where employers hire contractors that frequently perform well.

- ◇ "Exit" state: employers emit mostly "Idle" observations and, as a result, choose to not participate in the market.
- ◇ "Less-successful" state: employers make higher-variance decisions that often result in unsuccessful task outcomes.
- ◇ "More-successful" state: employers make lower-variance decisions that result in more successful task outcomes.
- ◇ "Most-successful" state: employers make lower-variance decisions that result in successful task outcomes.

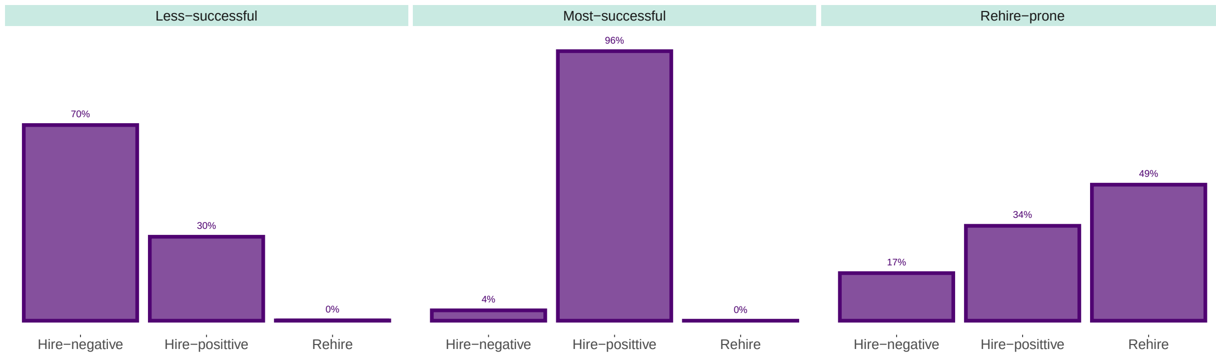
Appendix D A Rehire-State HMM

The focal dataset allows us to evaluate an alternative measure of success: rehires. A rehire decision signals that a prior hiring choice was successful. As a result, rehires can reveal information about an employer's hiring-behavior state. Specifically, we train an adjusted HMM that emits the following set of discrete observations:

$$Y_{it} = \{\text{Hire-positive, Hire-negative, Rehire}\}, \tag{22}$$

where a "Rehire" outcome signals the employer's choice to rehire a contractor.

Figure 10 Performance distributions of the best-fitted rehire-state HMM



The HMM identifies three different states. A “Less-successful” state where employers are more likely to hire contractors who perform poorly, a “Most-successful” state where employers hire contractors who perform well, and a “Rehire-prone” state where employers tend to hire contractors that have hired before.

The estimation process of this HMM is the same as the one described in Appendix B.1, with the only difference that we now model emission probabilities through a multinomial distribution. By using the transition variables presented in Table 13, we follow a similar tuning process as the one described in Appendix B.3, and we identify that the best-fitted HMM includes $K = 3$ states.

Figure 10 shows the discrete emission distributions of each one of the three identified HMM states. For readability, we label the three states as:

- ◇ “Less-successful” state: employers make higher-variance decisions that often result in unsuccessful task outcomes.
- ◇ “Most-successful” state: employers make lower-variance decisions that result in successful task outcomes.
- ◇ “Rehire-prone” state: employers are more likely to rehire contractors that they have previously hired.

Appendix E Additional Data Statistics and Variable Analyses

Next, we present details regarding some of the control variables we used in our conditional logit models, along with variable correlations and additional data statistics.

E.1 Feature engineering

Not every variable enters our model as is. Some variables require feature engineering. Contractor location affects employer choices (Gefen and Carmel 2008, Ghani et al. 2014, Lin and Viswanathan 2015, Hong and Pavlou 2017, Agrawal et al. 2013, Kokkodis et al. 2015). To include location in our models, we calculate the point-wise mutual information (PMI) of the employer's and the contractor's countries ("Countries PMI," Kokkodis et al. 2015). PMI measures how the probability of a particular co-occurrence between two events differs from what we would expect to see if the two events were independent. Formally (Church and Hanks 1990, Kokkodis et al. 2015):

$$\text{Countries-PMI}(C_i, C_j) = \log \frac{\Pr(C_i, C_j)}{\Pr(C_i)\Pr(C_j)}, \tag{23}$$

where C_i is the country of the employer and C_j the country of the hired contractor. PMI captures complicated location preferences that a binary variable (e.g., local vs. non-local) misses (Kokkodis et al. 2015).

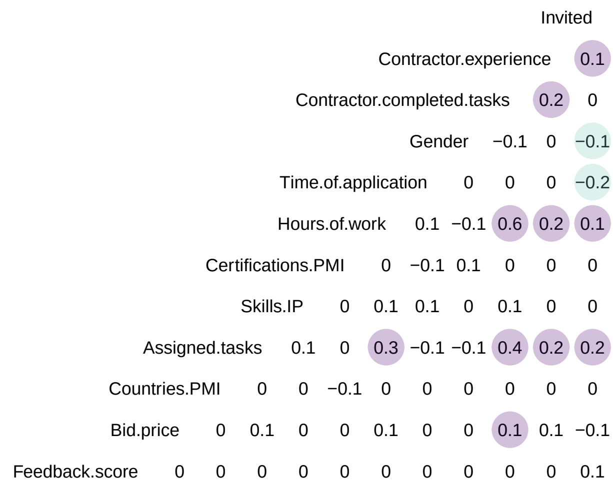
We follow the same process for certifications. Some certifications may reduce the likelihood of getting hired (Goes and Lin 2013), while other certifications might be necessary for specific tasks but irrelevant for others. To reflect this task-specific effect, we estimate the PMI between the set of required skills by the task and the contractor's set of certifications ("Certifications PMI").

Finally, we introduce a control variable for each contractor's skills by estimating the inner product between the set of the contractor's self-reported skills and the set of skills the task requires ("Skills IP," Kokkodis et al. 2015).

E.2 Additional data statistics

Figure 11 shows the correlogram of the focal and task-varying variables. Each number represents a correlation point between the row and the column variables. Most of the studied variables are not significantly correlated.

Figure 11 Correlogram of the focal and task-varying variables



Variables explain different parts of the observed variance. Shaded ellipses show absolute correlations greater than 0.1.

Table 14 shows descriptive statistics of the employers in our data. On average, each employer makes 52 decisions (hire or no-hire) per task (median is 35). On average, a Hire-positive occurs 2.5 times more often than a Hire-negative. In addition, employers seem to focus on specific task categories (average 1.8).

Finally, Table 15 shows the number of decisions per task category (number of applications employers choose to hire from). Software development and design tasks are the most prevalent in our data, even though a fair amount comes from Accounting, Engineering, Information Technology, and Customer Service. Overall, our data is quite diverse in terms of task categories.

Table 14 Descriptive statistics of employers in the focal dataset

	Mean	Median	StD	Min	Max
Number of bids	52	35	48	2	264
Employer reputation	0.91	1	0.18	0	1
“Hire-negative” per employer	1	1	1.4	0	28
“Hire-positive” per employer	2.5	2	2.4	0	32
“No hire” per employer	78	54	89	1	2368
Task categories per employer	1.8	2	0.92	1	7

Table 15 **Number of hiring decisions per task category**

Category	Number of hiring decisions
Software Development	77,002
Design	50,418
Administrative	41,644
Sales & Marketing	26,128
Writing	24,515
Translation	8,978
Accounting	3,414
Engineering	2,928
Information Technology	1,999
Customer Service	709

Categories ranked in decreasing order of number of hiring decisions.

Appendix F Additional Tables

Table 16 shows the complete set of coefficients β_{s_k}; Table 17 shows the same results if we build per-state conditional logit models. The evolution of the task-varying variables further supports our theoretical framework. For instance, the weight of the "Countries PMI" coefficient increases as employers move to more successful states; this suggests that as employers adjust their hiring criteria, they learn how to better assess country-specific information and regress to the average location preferences (i.e., "wisdom of crowds" location preferences). Such "wisdom of crowds" preferences are usually less uncertain and result in better outcomes (Mannes et al. 2012).

Table 16 State-specific effects on hiring choices (complete set of coefficients)

	All employers	Evolving employers
DV: Hire	(A1)	(A2)
Assigned tasks × Less-successful	0.396*** (0.035)	0.386*** (0.038)
Assigned tasks × More-successful	0.485*** (0.031)	0.492*** (0.033)
Assigned tasks × Most-successful	0.404*** (0.019)	0.417*** (0.024)
Skills IP × Less-successful	0.182*** (0.036)	0.215*** (0.039)
Skills IP × More-successful	0.156*** (0.034)	0.142*** (0.037)
Skills IP × Most-successful	0.189*** (0.022)	0.200*** (0.027)
Certifications PMI × Less-successful	0.041 (0.038)	0.053 (0.042)
Certifications PMI × More-successful	−0.009 (0.035)	0.015 (0.037)
Certifications PMI × Most-successful	−0.021 (0.022)	−0.028 (0.026)
Hours of work × Less-successful	−0.062 (0.043)	−0.081. (0.046)
Hours of work × More-successful	−0.096* (0.039)	−0.090* (0.041)
Hours of work × Most-successful	−0.139*** (0.024)	−0.154*** (0.029)
Time of application × Less-successful	−0.132. (0.068)	−0.119 (0.075)
Time of application × More-successful	0.034 (0.069)	0.040 (0.074)
Time of application × Most-successful	−0.050 (0.042)	0.025 (0.054)
Gender × Less-successful	−0.052 (0.067)	−0.079 (0.073)
Gender × More-successful	0.024 (0.061)	0.021 (0.065)
Gender × Most-successful	−0.145*** (0.038)	−0.188*** (0.047)
Contractor completed tasks × Less-successful	0.186*** (0.045)	0.204*** (0.050)
Contractor completed tasks × More-successful	0.071. (0.042)	0.073 (0.045)
Contractor completed tasks × Most-successful	0.127*** (0.027)	0.138*** (0.033)
Contractor experience × Less-successful	−0.026 (0.026)	−0.030 (0.028)
Contractor experience × More-successful	−0.000 (0.022)	−0.004 (0.023)
Contractor experience × Most-successful	0.022. (0.013)	0.015 (0.016)
Invited × Less-successful	0.664*** (0.095)	0.587*** (0.104)
Invited × More-successful	0.376*** (0.089)	0.399*** (0.096)
Invited × Most-successful	0.884*** (0.053)	0.952*** (0.066)
Countries PMI × Less-successful	0.228*** (0.041)	0.193*** (0.045)
Countries PMI × More-successful	0.276*** (0.034)	0.268*** (0.037)
Countries PMI × Most-successful	0.306*** (0.022)	0.286*** (0.028)
Feedback score × Less-successful	0.142*** (0.033)	0.142*** (0.037)
Feedback score × More-successful	0.248*** (0.037)	0.266*** (0.040)
Feedback score × Most-successful	0.379*** (0.027)	0.391*** (0.033)
Bid price × Less-successful	−0.350*** (0.053)	−0.374*** (0.058)
Bid price × More-successful	−0.146** (0.048)	−0.173*** (0.051)
Bid price × Most-successful	−0.053. (0.030)	−0.038 (0.036)
Opening FE	Yes	Yes
Observations	238,364	174,873
AIC	52,763	39,113

The table shows conditional logit estimates. The first column shows the results of the complete dataset (all employers). The last column focuses only on employers who transition across two or more states. Clustered standard errors in parentheses. Constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).

Table 17 State-specific effects on hiring choices (broken down per state)

	All employers			Evolving employers		
	Less-successful	More-successful	Most-successful	Less-successful	More-successful	Most-successful
DV: Hire	(A1)	(A2)	(A3)	(B1)	(B2)	(B3)
Assigned tasks	0.396^{***}	0.485^{***}	0.404^{***}	0.386^{***}	0.492^{***}	0.417^{***}
	(0.035)	(0.031)	(0.019)	(0.038)	(0.033)	(0.024)
Skills IP	0.182^{***}	0.156^{***}	0.189^{***}	0.215^{***}	0.142^{***}	0.200^{***}
	(0.036)	(0.034)	(0.022)	(0.039)	(0.037)	(0.027)
Certifications PMI	0.041	−0.009	−0.021	0.053	0.015	−0.028
	(0.038)	(0.035)	(0.022)	(0.042)	(0.037)	(0.026)
Hours of work	−0.062	-0.096^{*}	-0.139^{***}	−0.081.	-0.090^{*}	-0.154^{***}
	(0.043)	(0.039)	(0.024)	(0.046)	(0.041)	(0.029)
Time of application	−0.132.	0.034	−0.050	−0.119	0.040	0.025
	(0.068)	(0.069)	(0.042)	(0.075)	(0.074)	(0.054)
Gender	−0.052	0.024	-0.145^{***}	−0.079	0.021	-0.188^{***}
	(0.067)	(0.061)	(0.038)	(0.073)	(0.065)	(0.047)
Contractor completed tasks	0.186^{***}	0.071.	0.127^{***}	0.204^{***}	0.073	0.138^{***}
	(0.045)	(0.042)	(0.027)	(0.050)	(0.045)	(0.033)
Contractor experience	−0.026	−0.000	0.022.	−0.030	−0.004	0.015
	(0.026)	(0.022)	(0.013)	(0.028)	(0.023)	(0.016)
Invited	0.664^{***}	0.376^{***}	0.884^{***}	0.587^{***}	0.399^{***}	0.952^{***}
	(0.095)	(0.089)	(0.053)	(0.104)	(0.096)	(0.066)
Countries PMI	0.228^{***}	0.276^{***}	0.306^{***}	0.193^{***}	0.268^{***}	0.286^{***}
	(0.041)	(0.034)	(0.022)	(0.045)	(0.037)	(0.028)
Feedback score	0.142^{***}	0.248^{***}	0.379^{***}	0.142^{***}	0.266^{***}	0.391^{***}
	(0.033)	(0.037)	(0.027)	(0.037)	(0.040)	(0.033)
Bid price	-0.350^{***}	-0.146^{**}	−0.053.	-0.374^{***}	-0.173^{***}	−0.038
	(0.053)	(0.048)	(0.030)	(0.058)	(0.051)	(0.036)
Opening FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,958	52,026	138,380	39,952	45,408	89,513
AIC	10,041	11,785	30,936	8,421	10,315	20,377

The table shows conditional logit estimates. The first three columns show results of the complete dataset (all employers). The last three columns focus only on employers who transition across two or more states. Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05, . $p-$value < 0.1).