Gender Discrimination in the Gig Economy: Evidence from Online Auctions for Freelancing

Ga Young Ko*

December 11, 2021

For the most recent version, click here.

Abstract

I study gender discrimination in an online auction-based platform for freelance jobs. To this end, I build an equilibrium model of demand and supply for freelance jobs, in which workers bid prices for each job they are interested in and the employer (who posted the job ad) makes a discrete choice from the offers tendered. The demand for workers in my model nests both *taste-based* and *statistical* discrimination against a gender within a random utility framework. I use rich and novel data from an online platform for different kinds of freelancing jobs (e.g., cleaning, moving, and gardening), which enables me to quantify variation in discrimination across job categories. To distinguish the two sources of gender discrimination, I combine past, present, and future performance measures of a worker to estimate workers' true quality, which is not observed by the employer before hiring. I show that observing this measure is sufficient to separate the effect of taste-based discrimination from statistical discrimination in the hiring process. My estimates suggest that taste-based discrimination is the primary form of discrimination in most jobs. If the platform imposes a gender-blind hiring policy, I find that the welfare of the disfavored group increases by 2% to 18%, depending on the job category.

JEL Codes: J71, L13, L14, D44, D82

Keywords: Discrimination, Gender, Auctions, Gig Economy, Information Asymmetry

*Department of Economics, University of Virginia, gk9xs@virginia.edu. Many thanks to Simon Anderson, Gaurab Aryal, Federico Ciliberto, Maxim Engers, Peter Troyan, Andrew Kloosterman, Natasha Zhang Foutz, and Moonju Cho, as well as past and present participants in the UVa IO Graduate Meetings for their advice and support. I also thank Donghyuk Shin for data access. I am grateful for financial assistance from the Bankard Fund for Political Economy.

1 Introduction

The gig economy is characterized by short-term contracts for freelance work through online platforms. This new labor market environment has the potential to mitigate discrimination in the hiring process that has long been documented in traditional labor markets (Neumark et al., 1996; Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004). Employers in the gig economy hire workers without face-to-face interviews, and thus the hiring process can be conducted anonymously, without information that reveals the group to which an individual worker belongs. However, contrary to this expectation, recent research shows that discrimination in hiring is prevalent even in the gig economy, through worker identity being discovered by profile photos or names (Pope and Sydnor, 2011; Edelman et al., 2017; Chan and Wang, 2018).[1]

Given the enormous volume of transactions, the gig economy has been accused of increasing discrimination (Fisman and Luca, 2016). In response to growing criticism, some platforms have changed their policy for revealing user demographic information. In 2018, Airbnb, the online marketplace for short-term housing rentals, changed how it displays guest profile photos.[2] Removing demographic information could improve the welfare of disfavored groups; however, we know little about those gains and how the market reacts; it might even adversely affect other groups.

Thus it seems reasonable to say that before changing any policy, we should (a) determine whether there is evidence of gender discrimination and (b) if so, identify the cause of the discrimination — e.g., whether it is "taste-based" discrimination (Becker, 1957) or information-based, i.e., "statistical" discrimination (Phelps, 1972). The efficacy of a policy response to mitigate discrimination depends on knowing its source. Despite the urgency and need, there isn't any empirical framework to address these questions and evaluate the effects of different policy interventions, such as a gender-blind hiring policy. One reason for this lack of an empirical framework is that identifying these two forms of discrimination is difficult because they reinforce each other. Therefore, we

[1] Pope and Sydnor (2011) find evidence of racial disparities in a peer-to-peer lending market. They show that loan listings with blacks in the attached picture are 25% to 35% less likely to receive funding than those of whites with similar credit profiles. Edelman et al. (2017) find that applications from guests with African American names are 16% less likely to be accepted relative to identical guests with white names in Airbnb. Chan and Wang (2018) find evidence of gender discrimination in an online labor market.

[2] Under the new policy, hosts can see a guest's photo only after they have accepted the booking request. https://hbr.org/2020/03/research-to-reduce-gender-bias-anonymize-job-applications This reduces the chance that the identify of guests is revealed before the booking is accepted, and thus prevents hosts from making decisions based on gender or race.

need a new approach to address these issues, which is the objective of my paper.

In particular, I contribute to this line of research by studying gender discrimination in an online auction-based platform for freelance jobs. An employer posts a job on the platform with some description, and workers submit bids ("ask prices") for jobs they are interested in. Then the employer makes a discrete choice decision from the offers tendered, taking into account all bids and workers' characteristics, which includes their gender and work reviews from previous jobs on the platform.

To answer these questions, I develop an equilibrium model of demand and supply for freelance jobs. On the demand side, an employer either chooses a worker from those who submitted bids for the job he posted or opts out in favor of outside options (e.g., using an offline local market or another freelance site or doing the job himself). An employer does not necessarily choose the worker with the lowest bid, because he may also care about non-price attributes such as gender and quality. To capture this feature, I augment the standard discrete choice random utility specification to include taste or animus toward a particular gender and expected quality of the worker by following the specification of the canonical Phelps model (Phelps, 1972).

On the supply side, I model oligopolistic competition with incomplete information about rivals' costs. Workers submit bids that maximize their profits given their own attributes, their rivals' attributes, their costs for completing the task, the distributions of their rivals' costs, and the employer's perceived preferences. To the best of my knowledge, this paper is the first to nest both taste-based and statistical discrimination within a random utility framework and incorporate an oligopolistic supply side with asymmetric information about costs.

To estimate the model, I use rich and novel data from an online platform for freelancing jobs from October 2012 to October 2017. In my data, I observe 1.3 million jobs posted by 500,000 unique employers, and for each job I observe all workers who submit a bid, their gender, their reviews, and several performance measures. In total there are 137,000 unique freelancers, many of whom work repeatedly on the platform. The platform hosts jobs in various categories; for this study I focus on cleaning, moving, gardening, repairs, and deliveries, which are relatively easy to measure and require face-to-face interaction with employers.

An employer who dislikes females, for instance, would be observationally equivalent to someone who thinks females have lower job quality than males. This substitutability between taste-based and information-based discrimination is what renders the identification problem challenging. To overcome this hurdle, I track workers' performance (past and future), such as reviews, ratings, and completion rates, during their tenure in my sample and combine them to construct a measure of workers' true qualities. The novelty of my approach is that I observe the past, present, and future performance of a worker so that I can construct the quality measure. In contrast, employers only observe, at best, the worker's past performance. I show that this gap between the employer's expectation about the worker's qualities and their true qualities is sufficient to separate taste-based discrimination from statistical discrimination.

The observed bids reflect workers' strategic behavior in response to employer preferences. Workers have incentive to bid up if they are members of the favored group. They also react to the composition of rivals by bidding down as the number of bidders from the favored group increases. Not only are workers' strategic behavior reflected in the observed bids but also workers' private costs. After controlling for workers' strategic behavior, the variation in bids stems solely from variation in the private costs of workers.

To estimate the cost distributions of workers, I follow the methodology commonly used in empirical auctions (Guerre et al., 2000), but rely on a parametric assumption, following Krasnokutskaya and Seim (2011). Specifically, I estimate the winning probability by making a parametric assumption on the bid distribution. I assume that bids follow a log-normal distribution with parameters as a function of the worker and rivals' attributes. Combining random draws from the estimated bid distribution and the parameters of employers' preferences on the demand side, I recover pseudo values of costs using the first-order condition for optimal bidding. I then estimate the cost distributions nonparametrically using these values.

I find evidence of discrimination against female workers in three out of the five job categories: moving, gardening, and repairs. The magnitude of discrimination is large and economically significant. For instance, to win a moving job, a woman has to bid \$2.40 less than an (observationally) equivalent male. This "gender tax" amounts to approximately 9% of the median hourly wage. Out of this \$2.40, \$1.40 is due to statistical discrimination and \$1 to taste-based discrimination. Consistent with these breakdowns, I find that female workers have lower quality for moving than male workers on average, which leads to statistical discrimination in hiring a mover.

For gardening and repairs, the magnitude of discrimination is even larger. The "gender tax" for a female worker is \$6.10 and \$7.10, respectively, and is approximately 23% and 28% of the median hourly wage. For these jobs, taste-based discrimination explains more than 85% of the total discrimination. I find evidence of discrimination against male workers in cleaning jobs, with a "gender tax" for a male worker of \$3, which is 11% of the median hourly wage. For cleaning jobs, taste-based discrimination accounts for most of the total discrimination. Lastly, there is no evidence of discrimination in delivery jobs.

I find that workers strategically decide their optimal bids. Estimates from the bid distribution show that workers decrease their bids for cleaning jobs as the number of female workers (favored group) increases, while they increase their bids as the number of male workers (disfavored group) increases. However, on average, female workers bid lower than their male counterparts even if they are favored by employers, which implies that female workers have lower costs than male workers.

My estimates suggest that the average cost for female workers is \$14, whereas that for male workers is \$20 for cleaning jobs. Interestingly, female workers have lower costs than male workers for most jobs except repairs. It may be counterintuitive that female workers have lower costs for moving and gardening. However, considering opportunity costs, the results are plausible since male workers usually have better other options than female workers.

Using these estimates, I consider a counterfactual exercise in which the platform blinds the gender information of workers. This policy rules out both taste-based and statistical discrimination. My findings show that (a) workers from the disfavored group increase their bids and (b) the likelihood that they are chosen as a service provider goes up after removing gender information. My estimates suggest that welfare for these workers increases by 2% to 18% depending on the level of discrimination for each job category. These estimates are expected to be even larger if I account for a response in participation behavior.

Related Literature

This paper relates to three broad streams of the literature. First and foremost, this paper contributes to the literature on sources of discrimination. A vast literature has used either a taste-based or statistical discrimination model to examine whether the patterns observed in the data are consistent with the given model. The statistical discrimination model implies that more information on individual worker quality reduces the level of discrimination. Wozniak (2015) shows that adoption of drug-testing legislation increases black employment, and Agan and Starr (2017) find that "Ban the Box" policies encourage racial discrimination, which suggests that the observed discrimination is driven by statistical discrimination. Farber and Gibbons (1996); Altonji and Pierret (2001); and Arcidiacono et al. (2010) examine whether wages become more correlated with hard-to-observe worker quality (e.g., AFQT test scores) and less correlated with easily observed worker characteristics (e.g., gender, race, or education) as employers learn about worker quality.[3] My model is in line with these papers, in that I leverage the gap between true versus observed quality at the time of hiring. On the other hand, Becker's (1957) taste-based discrimination model implies that employers with prejudice forgo profits, and thus they are forced to leave the market. Based on this idea, Berkovec et al. (1998) and Weber and Zulehner (2014) investigate whether firms with strong preferences for discrimination are more prevalent in less competitive markets. While this approach is intuitive and easy to test, it does not allow me to quantify how much of the observed discrimination is attributable to taste-based or statistical discrimination. Very few studies develop models that nest both taste-based and statistical discrimination within the same framework to design a test to tease out the two sources of discrimination. Knowles et al. (2001) and Anwar and Fang (2006) propose a test for distinguishing between the two by developing a simple model of police and motorist behavior. The test is based on the idea that if the police are trying to maximize arrests, then the success rate of searches should be the same across groups even if the police are more likely to search the vehicles of a particular group. My model shares the idea of the test used in these papers — the so-called "outcome-based test"; however, my model applies to a more general setting, i.e., employers' hiring decisions.

Second, this paper also relates to the small but growing literature on multi-attribute (or beauty contest) auctions. Krasnokutskaya et al. (2020) propose a multi-attribute auction model that allows buyers to take into account both seller bid price and non-price characteristics. Yoganarasimhan (2013, 2016) also presents an empirical framework for beauty contest auctions. My setting is similar to their setting; however, I focus specifically on the demand side by incorporating discriminatory buyer behavior.

Third, more broadly, this paper adds to the growing literature on the gig economy. Studies have

[3] The employer learning framework is also useful for identifying the returns to education. Aryal et al. (2019) identify private returns to education and social returns to education using a variable that employers do not observe and a variable that they do, respectively.

investigated the impact of various aspects and components of online labor markets on hiring outcomes such as valuation and competition uncertainty (Hong et al., 2016); reputation transferability (Kokkodis and Ipeirotis, 2016); gender-based stereotypes (Chan and Wang, 2018); monopsony and labor elasticity (Dube et al., 2020); country development level (Kanat et al., 2018); network effects and geographic heterogeneity (Cullen and Farronato, 2020); and text-based messaging systems (Hong et al., 2021). In terms of discrimination, Fisman and Luca (2016) argue that online labor market platforms need to be mindful of the potential for discrimination and open to experimentation when making choices about automation, algorithms, and the use of identifying data. For example, Edelman et al. (2017) demonstrate that platform design choices regarding how buyer names are displayed can facilitate discrimination. Platforms may be constrained in their ability to address inequality (Cullen et al., 2018) and it has been argued that there is no reason to expect the gig economy to close the gender gap (Cook et al., 2021). The literature on the topic is mostly nonstructural, whereas I propose a structural model of workers and employers that enables me to conduct counterfactual simulations.

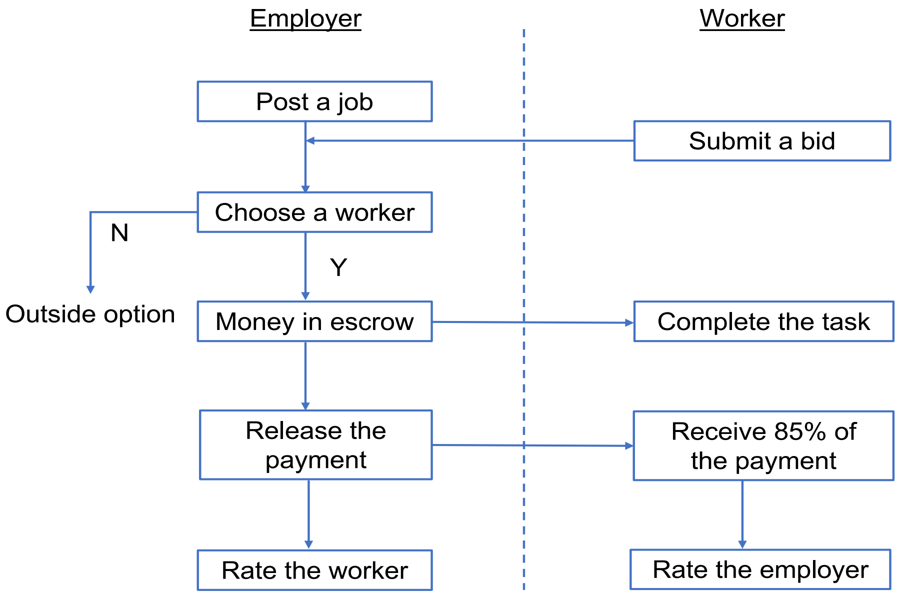
The paper proceeds as follows. Section 2 presents institutional details of the hiring process on the platform, describes the data, and provides descriptive statistics. Section 3 presents the model and Section 4 discusses identification. The estimation procedure and results are given in Section 5. Section 6 provides the counterfactual analysis, and Section 7 considers extensions. Section 8 concludes.

2 Background and Data

2.1 Hiring Process

The empirical context of the study is a major online freelance marketplace. The platform matches employers who post short-term tasks (the demand side) with workers who bid for these tasks (the supply side). The details of the hiring process are as follows. The auction begins as soon as an employer posts a task that includes a task name, task description, auction due date, and a price suggested by the employer (referred to hereafter as a *project value*). Some information about the employer such as average rating, first name, and profile photo are shown on the task page and more information is provided on the employer's profile page. Any worker can submit a bid with

Figure 1: Process flow



optional comments to provide additional information or ask questions. Workers can bid higher or lower than the project value. The auction format is a sealed bid auction, so the bids are visible only to the employer. Even though workers cannot observe the bids of their competitors, they can see who submitted bids on the task so far.

The employer can either choose a worker from those who participated in the auction, cancel the auction before the due date, or let the auction expire. The last two choices can be interpreted as an outside option including using offline local market or another freelance site, or doing it themselves. The employer makes a decision based on the bid prices and the workers' non-price attributes such as demographics and ratings from past employers. After the employer chooses a worker, the agreed upon amount is in escrow until the task is complete. Once the task is complete, the employer releases the payment. There are no fees for either posting auctions or for bidding, but, the platform charges a 15% commission on the transaction amount, and transfers the rest to the worker. After each transaction, employers and workers are allowed to rate each other on a five-star scale with optional written remarks, where a rating 1 stands for very bad and 5 for excellent. Figure 1 illustrates the process flow of a typical task on the platform.

2.2 Gender Inference

The platform does not collect demographic information from its users, and thus does not explicitly provide gender information of users. However, a user can infer the gender of other users by looking at their first names and/or profile photos. Users must verify their names (e.g., by credit card or cellphone) and upload a profile photo if they want to submit bids. Not all users post photos of their own faces. Gender is not identified from photos for those users who post photos of scenery, pets, etc.

To ascertain the gender of a user as perceived by other users from first names, I use the *gender* R package (Blevins and Mullen, 2015). The package computes the probability of being male or female given a first name based on the U.S. Census and Social Security data. I classify the gender of a user from first names as "unknown" if the predicted probability is less than 95%.

Compared to names, profile photos convey more direct and salient information about gender. Gender inference from photos consists of four stages: detection, alignment, representation, and prediction. The first step is to detect faces in the photos using computer vision and deep learning models. Photos commonly include not only faces but also background, clothing, and hair. I use the single shot multibox detector model (Liu et al., 2016) to obtain the face bounding boxes and ignore other areas of the photo. For cases where multiple people appear in the photo, I obtain multiple face bounding boxes. The second step is to align faces using an affine transformation (Jaderberg et al., 2015). The method relies on the facial landmarks (e.g., eyes, mouth, nose, jaw, etc.) to obtain a normalized rotation, translation, and scale of the face. The third step is to transform the normalized image into a common representation through deep learning approaches. I use a convolutional neural network (CNN) model (Krizhevsky et al., 2012; LeCun et al., 2015), which automatically discovers robust representations needed for accurate classification of images. Specifically, I use the VGG-Face model (Parkhi et al., 2015), which is an open-sourced pre-trained model for face recognition that achieves state-of-the-art performance on benchmark datasets. The output of the second-to-last layer of the model's architecture generates a robust representation of the input image (referred to as CNN codes). The last step is to predict the gender using the learned representation from the previous step via transfer learning. Since gender prediction is a classification task, the last layer of the VGG-Face model is replaced with a custom classification layer consisting of two nodes (i.e., one for male and the other for female prediction). Using the learned parameters of the base VGG-Face model as a starting point, the modified model is fine-tuned using facial images with ground-truth gender labels from publicly available datasets (Philip and Cuixian, 2018). If the photo includes more than one face, I treat the gender of a user from pictures as "unknown" unless all predicted labels are of the same gender.

Combining the approaches described above, I infer the gender for 82% of workers in my sample. To check the overall accuracy level of the final labels, 1,000 sellers are randomly selected as a test set for whom the ground-truth gender is determined manually. The total agreement between these labels and those obtained using deep learning models and name-gender probability is 92.4%.

2.3 Data

I use rich and novel data from an online platform for freelancing jobs that consists of records for all transactions between October 2014 to October 2017. This includes highly detailed information on 1,259,550 jobs listed and the 3,971,022 bids to these tasks placed by 137,483 workers. The data include information about the posted job and worker characteristics that include gender, bids, and performance.

Tasks are classified by the platform into 10 broad classes, and each task is then further divided into finer categories within these classes. The most popular finer task categories are cleaning (17.0%), moving (12.8%), gardening (7.8%), repairs (6.2%), and delivery (4.9%) The analysis focuses on these five task categories that are relatively easy to measure and require face-to-face interaction with employers.

2.4 Summary Statistics

Auction/Employer Level Statistics Table 1 describes some auction or employer-level statistics for the aforementioned five categories. Jobs in the sample data set are everyday tasks (e.g., cleaning, moving, and repairs) rather than professional tasks (e.g., programming and translation), and thus the majority of jobs are very small: the price suggested by employers (project value) tend to be below $100. An average auction has a deadline of 6 days because the default of the deadline is a 7 days while employers are allowed to change the auction deadline.[4] Employers can make hiring

[4]This is the deadline for auctions, not for jobs to be delivered.

decisions before the deadline. About 80% of auctions end with choosing a worker within 24 hours from when the auction was initiated.

The majority (72.7%) of employers in the platform are one-time users: they post only one job and exit the platform (or post a job in different category). This results in a sizable number of employers (73.7%) having no past ratings. Most ratings are a perfect score of 5-stars, and therefore, the mean of the average rating for employers who have been rated at least once in the past is 4.94.

Table 1: Summary Statistics of Auction and Employer Attributes

Employer Attributes	Mean	Std. Dev.	25th	50th	75th	(Min, Max)
Project value	102.16	151.40	45	63	100	(5,2000)
Auction deadline (days)	6.04	6.50	1.47	6.07	7.58	(0, 993)
Duration between posting and assignment (hrs)	19.67	59.80	0.65	3.28	20.14	(0, 18594)
Number of auctions per employer	1.95	2.52	1	1	2	(1, 295)
Number of employer ratings	1.31	5.04	0	0	1	(0, 273)
Average ratings	1.64	2.33	0	0	5	(0, 5)
Average ratings (if rated)	4.94	0.27	5	5	5	(1,5)

Worker Level Statistics Table 2 provides the summary statistics of worker and bid attributes. The average price submitted by workers is $141. The bid price normalized by project value is on average 1.56. While a sizable number of workers (37.6%) submit the same price as the project value of auctions, many workers also submit higher or lower prices than the project value. When workers submit higher prices than the project value, their prices on average exceed the project value by 103%, with the median being 50%. Workers tend to submit bids much earlier than the deadline. The majority of bids (93%) are submitted within 24 hours of the job being posted. The average number of bids a worker submits during their tenure (duration between their first and his last bidding) is 28. An average worker wins 3.7 auctions during their tenure.

In addition to the ratings which are platform-verified information, workers voluntarily provide additional information for employers' hiring decisions. Workers can upload personal information to their profiles such as a narrative description of skills and a personal picture. The number of words in the worker description is 60 on average. Workers are also allowed to leave comments when submitting bids, which could be a useful channel to provide more information not stated in profiles or ask a question to the employer. The volume of the information represented by number of words is on average 41.

As stated above, the majority of employers in the data are short-lived users. As a result, an employer very rarely works with the same worker repeatedly. About 0.8% of the workers have interacted with the employer in the past.

Table 2: Summary Statistics of Worker Attributes

Worker Attributes	Mean	Std. Dev.	25th	50th	75th	(Min, Max)
Bid price	141	156	65	100	155	(5, 2000)
Bid price (normalized by project value)	1.56	3.77	1	1.16	1.6	(0.0025,2000)
Duration between posting and bidding (hrs)	8.66	827	0.174	0.078	4	(0, 2355)
Number of bids ending in winning	3.71	21.11	0	0	1	(0, 1228)
Number of worker ratings	56.3	113.89	2	15	56	(0, 1381)
Average ratings	3.98	1.86	4.5	4.9	5	(0, 5)
Average ratings (if rated)	4.84	0.25	4.8	4.9	5	(1,5)
Number of words in seller description	60.06	143.16	1	31	67	(0, 4618)
Number of words in seller comments	40.68	31.44	19	34	56	(0, 1873)
Indicator for past interaction with employer	0 = 99.20%, 1 = 0.80%					

Employer Choice About 56.5% of the auctions end with no worker being allocated. 13.2% receive no offers, and the other 43.3% are canceled by the employer even if there are offers. These cases are interpreted as the employer choosing outside options.

Auctions attract 3 bids on average. Among auctions where a worker is chosen, about 37% of the jobs are allocated to a worker who submits a price above the lowest price submitted in the auction. When such a bid is chosen, the percentage difference between the chosen bid and the lowest bid is 53% on average. This suggests that factors other than price play a significant role in employers' hiring decisions.

Table 3: Summary Statistics of Employer Choices

	Mean	Std.Dev.	25th	50th	75th	(Min, Max)
Number of bids received	3.28	3.21	1	2	4	(0, 71)
Indicator for choosing the lowest bid	0 = 37.19%, 1 = 62.81%					
(Winning bid - lowest bid)/lowest bid	53.45	416.48	11.54	20	42.05	(0.10, 39900)

Table 4 provides model-free evidence of gender discrimination in hiring. For cleaning jobs, female workers are more likely to be chosen as a service provider. When they submit 100 bids, about 17 bids win auctions. In contrast, male workers have higher chance of winning for moving, gardening, and repairs. The results should be interpreted cautiously because I do not control other variables that may affect whether to win the auction.

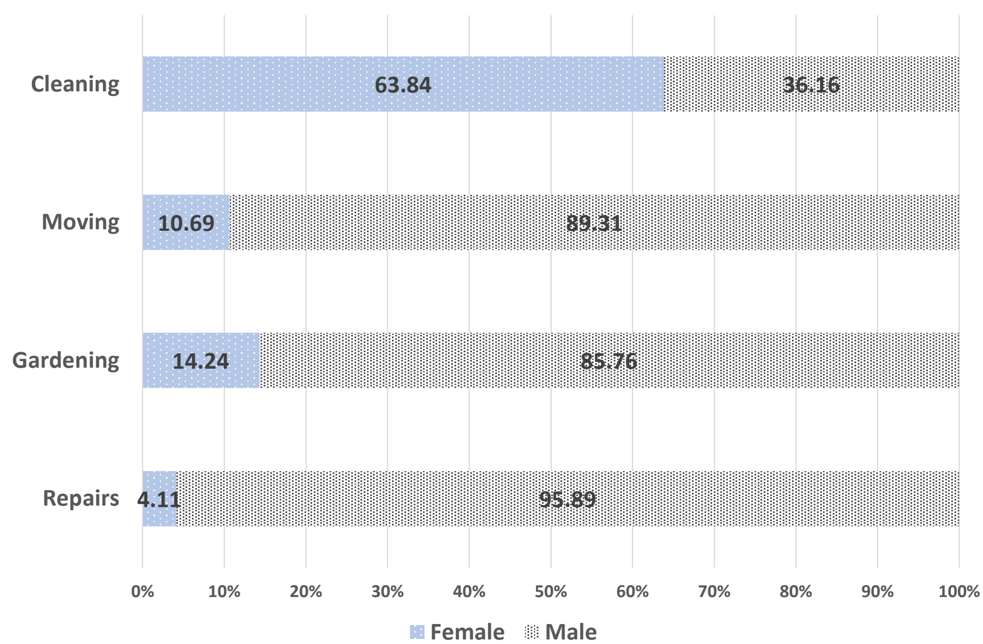
Figure 2 illustrates worker presence by gender for each job category. As the figure indicates, most bids (64%) for cleaning are submitted by female workers, while male workers account for more than 85% of bids for moving, gardening, and repairs. The results show the gender segregation across job categories, which may be consequences of gender discrimination. Workers may be more likely to bid for jobs in which they are favored, whereas they may be less likely to bid for jobs in which they are disfavored.

Table 4: Winning Probability by Gender

	Male	Female
Cleaning	12.7	16.6
Moving	19.1	16.8
Gardening	21.6	18.0
Repairs	20.2	18.7
Deliveries	19.9	22.3

Note: Winning probability (%) = total number of winners / total number of bidders × 100

Figure 2: Worker Composition by Gender



Note: The values in this figure represent the share of bids by gender and job category.

3 Model

In this section I first present the baseline equilibrium model of demand and supply for freelance jobs, and then discuss other specifications. There are two types of agents in the model: employers and workers. I outline the timing of the game and then detail the optimization problem for each agent. The timing of the game is as follows. At stage 0, employers post jobs on the platform. At stage 1, workers bid for jobs they are interested in. Each worker chooses the bid that gives the highest payoff. At stage 2, employers either choose a worker from those who submitted bids for their jobs or opt out in favor of outside options, whichever yields the highest utility.

Employers observe the submitted bid prices and worker characteristics for each offer. Employers do not necessarily choose the worker with the lowest bid because they also care about other worker characteristics that affect their utility, such as gender and quality. Thus, I use a random utility framework which is often used to model the individual's choice among a discrete set of alternatives. By engaging the canonical Phelps model of statistical discrimination this framework, the model nests both taste-based and statistical discrimination and thus enables me to generate an empirical test that distinguishes between the two.

When submitting the bids, workers know their own characteristics and private costs. They observe the characteristics of the other workers in the auction but do not observe the bid prices and costs of the other bidders. I set up a model of oligopolistic competition with incomplete information to capture these features.[5]

The setting of the model follows a static structural framework. I focus on the employer's decision to whom to award the job unlike Yoganarasimhan (2013) focused on the employer's dynamic decision whether to choose a bid from the current set of bids, cancel the auction, or wait for more bids.[6] I also assume a simultaneous sealed-bid auction rather than a sequential auction. Workers can observe who participated in the auctions so far, but cannot see the prices they have submitted. Furthermore, they can update their bids until the auction ends. I focus on the last bid they submitted.

[5]In auction terminology, this game is often called a multi-attribute auction or a beauty contest auction. Note that this setting differs from a scoring auction where seller characteristics are choice variables and the allocation rule is pre-announced.

[6]The data show that about 80% of auctions end with choosing a worker within 24 hours since the auction is initiated. Thus, I assume that auction deadlines are fixed and that employers only make hiring decisions among the bids.

3.1 Employer Hiring Decision

The employer's objective is to maximize his utility. I first define the employer's utility and incorporate the potential sources of discrimination.

Utility. Let N_ℓ denote the set of workers who submit the bids for job ℓ. Throughout the paper I use l to index an employer, the job he posted, and the corresponding auction synonymously. The employer either chooses a worker from A_ℓ, or opt for an outside option that gives him a payoff $U_{0\ell}$. The utility employer ℓ receives from choosing worker j is:

$$U_{j\ell} = \alpha f_j + \beta q_j - b_{j\ell} + \mathbf{x}_{j\ell}\boldsymbol{\gamma} + \epsilon_{j\ell} \qquad \forall j \in \{1, 2, ..., J_\ell\},\ \forall \ell \in \{1, 2, ..., L\} \tag{1}$$

where f_j is an indicator for whether the worker is female, q_j is the quality index for worker, $b_{j\ell}$ is the bid price for job, and $\mathbf{x}_{i\ell}$ is the set of the observable worker attributes other than gender, price, and quality. L represents the total number of employers (auctions) and J_ℓ represents the number of workers in the auction ℓ. The employer-worker match term, $\epsilon_{j\ell}$, is drawn from a Type I Extreme Value distribution. The match term includes individual employer idiosyncratic taste for the worker. The vector of coefficients (α, β, γ) measures the employer tastes for worker characteristics. In this baseline specification, I assume a fixed-coefficients model, which means all employers have the same preferences over worker characteristics. I extend the model to allow employers to have heterogeneous preferences for characteristics in Section 7.

Expected Quality. The utility described above assumes that the employer observes the true quality of the worker. However, it generally may not be directly observed at the time of hiring. Thus, the employer makes hiring decisions based on the expected quality rather than the true quality. By introducing the statistical model of Phelps (1972) and Aigner and Cain (1977), I formulate that the employer estimates the quality of a worker under limited information and how this belief results in discrimination.

The essential features of the statistical model follow. Consider a worker whose gender is $g_j \in \{F, M\}$ and unobservable quality is q_j, where $q_j \sim N(\mu_g, \sigma_g^2)$ with gender mean $\mu_g \in \mathbb{R}$ and gender variance $\sigma_g^2 > 0$. Quality is fixed across time and jobs. The worker generates a noisy signal $s_{j\ell} = q_j + \eta_{j\ell}$, where $\eta_{j\ell} \sim N(0, \tau_g^2)$ with gender variance $\tau_g^2 > 0$. Note that the signal is an unbiased estimator of the true quality by construction, $\mathbb{E}(s_{j\ell}|q_j) = q_j$.

Assume that the parameters μ_g, σ_g^2, and τ_g^2 are common knowledge and form the employer's prior belief about worker quality.[7] They combine their prior belief about the group average with the signal using Bayes' rule to form his posterior belief about quality. Given the normality assumption, the posterior belief of quality is normally distributed with mean $\frac{\mu_g \tau_g^2 + s_j \sigma_g^2}{\tau_g^2 + \sigma_g^2}$ and variance $\frac{\tau_g^2 \sigma_g^2}{\tau_g^2 + \sigma_g^2}$. That is, the expected value of quality given worker gender and signal is:

$$\begin{aligned} \mathbb{E}(q_j | s_{j\ell}, g_j) &= \frac{\mu_g \tau_g^2 + s_j \sigma_g^2}{\tau_g^2 + \sigma_g^2} \\ &= (1 - \delta_g)\mu_g + \delta_g s_j \end{aligned} \tag{2}$$

where $\delta_g = \sigma_g^2 / (\tau_g^2 + \sigma_g^2)$ and $1 - \delta_g$ are interpreted as the weight on signal versus group average quality. If the signal is less informative (high τ_g^2), then employers will put less weight on it and more weight on the group mean quality (low δ_g). On the other hand, if the group mean quality is less informative (high σ_g^2), then employers will put more weight on it and more weight on the signal (high δ_g).

To simplify notation and discussion of estimation, I rewrite (2) with respect to the female indicator for the two groups, F and M.

$$\mathbb{E}(q_j | s_{j\ell}, f_j) = \big[(1 - \delta_F)\mu_F + \delta_F s_{j\ell}\big] f_j + \big[(1 - \delta_M)\mu_M + \delta_M s_{j\ell}\big](1 - f_j) \tag{3}$$

Expected Utility. The Phelps model gives an expression for the expected quality. I then plug it into (1) to get the expected utility. The employer chooses the worker who gives him the highest

[7] This implies that employers share the same accurate beliefs about average quality, the dispersion of quality, and the reliability of signal for each group. If employers have inaccurate beliefs about these parameters, they will mistakenly overpay workers from a particular group. Such mistaken behavior will not persist in rational expectation equilibrium. A few papers discussed whether beliefs were accurate or inaccurate (Agan and Starr, 2017; Arnold et al., 2018; Bohren et al., 2019b,a).

expected utility:

$$
\begin{aligned}
\mathbb{E}(U_{j\ell}|s_{j\ell}, f_j) &= \alpha f_j + \beta \mathbb{E}(q_j|s_{j\ell}, f_j) - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell} \\
&= \alpha f_j + \beta \big\{ \big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] f_j + \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big](1-f_j) \big\} - b_{j\ell} + \gamma_\ell \mathbf{x}_{j\ell} + \epsilon_{j\ell} \\
&= \left[\underbrace{\alpha}_{\substack{\text{taste-based} \\ \text{discrimination}}} + \underbrace{\beta \big\{ \big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big] \big\}}_{\substack{\text{statistical} \\ \text{discrimination}}} \right] f_j + \\
&\quad \beta(1-\delta_M)\mu_M + \beta \delta_M s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell} \\
&= \nu + (\phi + \rho s_{j\ell}) f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell}
\end{aligned}
\tag{4}
$$

where $\nu = \beta(1-\delta_M)\mu_M$, $\phi = \alpha + \beta[(1-\delta_F)\mu_F - (1-\delta_M)\mu_M]$, $\rho = \beta(\delta_F - \delta_M)$, and $\theta = \beta\delta_M$.

Discrimination. Discrimination is defined as differential treatment in favor or (or against) a person based on the group to which that person belongs rather than on individual merit. In this context, discrimination is the difference between the hiring decision for male workers versus female workers, holding fixed the worker attributes such as the signal of quality. In (4), we say that the employer ℓ exhibits discrimination in favor of female workers if $\mathbb{E}(U_{j\ell}|s_{j\ell}, f_j = 1) - \mathbb{E}(U_{j\ell}|s_{j\ell}, f_j = 0) > 0$.

Employers may discriminate against a gender either because of their innate preference for gender or because of their expectation about gender differences in quality, or both. My model nests both channels within (4) and allows me to quantify the two sources of discrimination.

First, the employer may have a taste towards a gender, which is denoted by α in the model. If $\alpha > 0$, the employer receives additional utility from interacting with female workers, and consequently would hire female workers over male workers even if the employer expects that both groups have the same expected quality. We call this type of this discrimination taste-based discrimination.

Second, there might be significant differences between male and female workers in terms of average quality μ_g, the dispersion of quality σ_g^2, the signal reliability τ_g^2, or any combination thereof. These differences lead the employer to form different expectations about quality by gender and consequently differential hiring decisions even if they have the same signal of quality. We call this type of this discrimination statistical discrimination (Phelps, 1972; Aigner and Cain, 1977). In (4), the statistical discrimination is represented by $\beta\{[(1-\delta_F)\mu_F + \delta_F s_{j\ell}] - [(1-\delta_M)\mu_M + \delta_M s_{j\ell}]\}$, which is the difference in the expected quality for female workers versus male workers multiplied by the employer's preference over the expected worker quality. I describe how each primitive related to statistical discrimination generates discriminatory behavior below. For simplicity, suppose that there is no taste-based discrimination, $\alpha = 0$.

Case 1. unequal average quality: $\mu_F > \mu_M$

Suppose that the average quality of female workers is perceived to be higher than that of male workers. For simplicity, let each group have the same variance for quality and the same variance of signal errors ($\sigma_F = \sigma_M, \tau_F = \tau_M$ and thus $\delta_F = \delta_M = \delta$ where $\delta = \sigma^2/(\tau^2 + \sigma^2)$). Then, the difference of quality distributions between male and female workers reduces to:

$$\begin{aligned}\mathbb{E}(q_j|s_{j\ell}, f_j = 1) - \mathbb{E}(q_j|s_{j\ell}, f_j = 0) &= [(1-\delta_F)\mu_F + \delta_F s_{j\ell}] - [(1-\delta_M)\mu_M + \delta_M s_{j\ell}] \\ &= (1-\delta)(\mu_F - \mu_M) > 0\end{aligned} \tag{5}$$

The expected quality for a given signal is higher for female workers, leading to discrimination in favor of female workers, holding fixed the other attributes. This effect becomes greater as the reliability of the signal (δ) is lower (more weight on the group average of quality).

Case 2. unequal dispersion of quality: $\sigma_F > \sigma_M$

Suppose that the distribution of quality for female workers is more dispersed than that for male workers. It implies that the group average of quality for female workers is less informative, so the lower weight is imposed on the group average of quality. For simplicity, I assume that each group has the same average quality and the same variance of signal errors ($\mu_F = \mu_M = \mu, \tau_F = \tau_M = \tau$ and thus $\delta_F > \delta_M$). Then, I have the following expression for the expected quality difference:

$$\mathbb{E}(q_j|s_{j\ell}, f_j = 1) - \mathbb{E}(q_j|s_{j\ell}, f_j = 0) = \big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big]$$
$$= (\delta_F - \delta_M)(s_{j\ell} - \mu) \tag{6}$$

Since $\delta_F > \delta_M$, which group has the higher expected quality depends on whether the signal of a worker is above or below the mean quality. For a high signal the female worker is predicted to do the job better than a male worker with the same signal, while for a low signal the male worker is predicted to do the job better than a female worker with the same signal. This leads to differential hiring decisions for workers from the two gender groups with the same signals.

Case 3. unequal reliability of signal: $\tau_F < \tau_M$

Suppose that the signal is a more reliable measure of quality for female workers than male workers. Assuming the same quality distributions for both male and female workers ($\mu_F = \mu_M = \mu, \sigma_F = \sigma_M = \sigma$ and thus $\delta_F > \delta_M$), the expected quality difference is (6). Therefore, the results are the same as the case 2.

3.2 Worker Bidding Decision

The worker's objective is to maximize their payoff. To model workers, I make several assumptions. First, workers only know rivals' cost distributions, but do not know their actual costs. Second, private costs are independent across workers. Third, workers can observe rivals' characteristics but do not observe their bids. Lastly, workers are risk-neutral. Based on these assumptions, I can define the worker's payoff function.

Expected Payoff. Let $\Lambda_\ell = (\lambda_{j\ell}, \lambda_{-j\ell})$ denote the composition of the set of workers who participate in auction ℓ in terms of characteristics such as gender, signal, and other worker characteristics, $\boldsymbol{X} \equiv (f, s, \boldsymbol{x})$. Let $\boldsymbol{A}$ be a set of auction characteristics. Each worker in the auction ℓ draws a private cost $c_{j\ell}$ from a distribution $F_{C|\boldsymbol{X},\boldsymbol{A}}$. It means that a worker's cost of completing a job vary with her own attributes and the auction characteristics. For simplicity, I suppress auction characteristics, $\boldsymbol{A}$. The ex-ante expected payoff for a worker j can be written as:

$$\Pi_j(b_{j\ell}, c_{j\ell}, b_{-j\ell}; \Lambda_\ell) = [b_{j\ell}(1-r) - c_{j\ell}]P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) \tag{7}$$

where $r = 0.15$ is the commission fee paid by the winning bidder and c is the private cost for completing the task.

Winning Probability. The winning rule is not deterministic. First, the worker j does not know the costs of the other bidders in the same auction. Therefore, the probability of winning depends on the distribution of other worker is:

$$P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) = \int p_j(b_{j\ell}, b_{-j\ell}(c_{-j\ell}); \Lambda_\ell) dF(c_{-j\ell}) \tag{8}$$

where $F(c_{-j})$ is the joint distribution of rivals' costs. Second, the employer's choice is random given worker characteristics and prices by construction. The winning probability $p_j(b_{j\ell}, b_{-j\ell}(c_{-j\ell}); \Lambda_\ell)$ can be represented as below.

$$\Pr\Bigg(\max\Big(U_{0\ell}, \max_{i \neq j}\{(\phi+\rho s_{i\ell})f_i+\theta s_{i\ell}-b_{i\ell}+\gamma\mathbf{x}_{i\ell}+\epsilon_{i\ell}\}\Big) \leq (\phi+\rho s_{j\ell})f_j+\theta s_{j\ell}-b_{j\ell}(c_{-j\ell})+\gamma\mathbf{x}_{j\ell}+\epsilon_{i\ell}\Bigg) \tag{9}$$

Therefore, the worker decides which bid to submit considering her attributes, her rivals' attributes, her costs for completing the task, her rivals' costs, and her belief on the employer's preferences.

3.3 Equilibrium

I focus on type-specific where workers of the same type, $(f, s, \boldsymbol{x})$, use the same strategy. A *type-symmetric pure strategy Bayesian Nash equilibrium (BNE)* is a profile of bidding strategies b^* such that

$$b^*(c_{j\ell}, \Lambda_\ell) = \arg\max_{b_{j\ell}} [b_{j\ell}(1-r) - c_{j\ell}]P_j(b_{j\ell}, b^*_{-j\ell}; \Lambda_\ell) \tag{10}$$

The first order condition to this problem is:

$$[b_{j\ell}(1-r) - c_{j\ell}]\frac{\partial P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell)}{\partial b_{j\ell}} + (1-r)P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) = 0 \tag{11}$$

Existence of Equilibrium. By construction, the probability of winning is continuous and consequently worker's profit Π is continuous in b for any c. Also, Π satisfies the single-crossing property of incremental returns in (b, c) whenever other bidders use non-decreasing pure strategies. Hence, there exists a non-decreasing pure strategies following Corollary 2.1 in Athey (2001).

4 Identification

The model is characterized by two sets of model primitives. The first set of model primitives is on the demand side (i.e., employer hiring decisions) including employer preferences over worker characteristics, the distributions of worker quality and signal errors. The second is cost distributions on the supply side (i.e., worker bidding decisions). I show that all the underlying structural parameters of the model are identified.

4.1 Identification of Demand Primitives

Identification Problem. From Section 3.1, (4) is repeated below:

$$\begin{aligned}
\mathbb{E}(U_{j\ell}|s_{j\ell}, f_j) &= \alpha f_j + \beta \mathbb{E}(q_j|s_{j\ell}, f_j) - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell} \\
&= \left[\underbrace{\alpha}_{\substack{\text{taste-based} \\ \text{discrimination}}} + \underbrace{\beta \big\{ \big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big] \big\}}_{\substack{\text{statistical} \\ \text{discrimination}}} \right] f_j + \\
&\quad \beta(1-\delta_M)\mu_M + \beta\delta_M s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell} \\
&= \nu + (\phi + \rho s_{j\ell}) f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell}
\end{aligned}$$

where $\nu = \beta(1-\delta_M)\mu_M$, $\phi = \alpha + \beta[(1-\delta_F)\mu_F - (1-\delta_M)\mu_M]$, $\rho = \beta(\delta_F - \delta_M)$, and $\theta = \beta\delta_M$.

The ultimate goal is to distinguish taste-based discrimination from statistical discrimination. There is an identification problem in that different sets of taste-based and statistical discrimination give rise to the same level of discrimination. The level of total discrimination in hiring is directly identified $(\phi + \rho s_{j\ell})$ from data. I show that knowing true quality is sufficient to identy each source of discrimination. Therefore, I first define true quality of a worker and the signal of quality.

Quality and Reliability of Signal. I use performance variables for the jobs that a given worker has completed to proxy for the underlying quality of the worker. The variables are evaluations that the worker has received from past employers and job completion rate. I will detail the variable selection procedure in Section 5.1.

True quality has some parts that are invisible to employers. The signal is observable to employers but is an error-ridden measure of quality. Higher quality generates a higher signal. To capture these features, I combine past, present, and future performances of a worker to construct a single measure of true quality of the worker. The novelty of my approach is that I, as a researcher, observe all the performances, whereas employers only observe past performances at best. I rely on this gap between these two measures (observed and true) of quality to infer statistical discrimination. I define the signal as the observed measure of quality.

The true quality and signal are measured in the same units. Therefore, I can compute the signal errors by subtracting true quality from the signal for a given worker, $\eta_{j\ell} = s_{j\ell} - q_j$. Once I have the true quality and signal for every worker, the parameters (mean μ_g and variance σ_g) of quality distribution by gender and the parameter (variance τ_g) of signal errors distribution by gender are identified.

Preferences. The next step is to recover parameters relevant to total discrimination. I exploit variation in observed choice, prices, and worker characteristics (e.g., gender and signal) across auctions to examine whether there exists a difference in hiring between male and female workers. The coefficients of female indicator, ϕ, signal, θ, and the interaction term between the two, ρ, and the other preference parameters, γ, are identified in this step.

Source of Discrimination. The final step is to determine how much of the observed discrimination is attributable to taste-based versus statistical discrimination. From (4), ϕ, θ, ρ are functions of the parameters of quality distributions (μ_g and σ_g), the parameter of errors of signal distributions (τ_g), employer taste for gender (α), and employer preference for expected quality (β). In the previous steps, ϕ, θ, ρ, μ_g, σ_g, and τ_g are identified. From the expression for $\theta (= \beta\delta_M)$, β is identified, where β represents how much employers value the expected quality of a worker. I am now able to back out the magnitude of statistical discrimination, which is represented as $\beta\big\{\big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big]\big\}$ for a given s_{ij}. The taste-based discrimination parameter, α, is recovered by subtracting statistical discrimination from total discrimination.

$$
\begin{aligned}
&\underbrace{(\phi+\rho s_{j\ell})}_{\substack{\text{total}\\\text{discrimination}}} - \underbrace{\beta\big\{\big[(1-\delta_F)\mu_F+\delta_F s_{j\ell}\big]-\big[(1-\delta_M)\mu_M+\delta_M s_{j\ell}\big]\big\}}_{\substack{\text{statistical}\\\text{discrimination}}}\\
&=\Big(\underbrace{\alpha+\beta[(1-\delta_F)\mu_F-(1-\delta_M)\mu_M]}_{\phi}+\underbrace{\beta(\delta_F-\delta_M)}_{\rho}s_{j\ell}\Big)\\
&-\underbrace{\beta\big\{\big[(1-\delta_F)\mu_F+\delta_F s_{j\ell}\big]-\big[(1-\delta_M)\mu_M+\delta_M s_{j\ell}\big]\big\}}_{\substack{\text{statistical}\\\text{discrimination}}}\\
&=\underbrace{\alpha}_{\substack{\text{taste-based}\\\text{discrimination}}}
\end{aligned}
\tag{12}
$$

Discussion. The method used in this paper is in line with the typical outcome-based test used in discrimination literature (Knowles et al., 2001; Anwar and Fang, 2006; Antonovics and Knight, 2009). The outcome-based test compares differences between groups in evaluators' relevant decision to differences between groups in the true underlying distribution of the factor most relevant for the decisions in order to measure the relative role of statistical discrimination compared to taste-based discrimination (Bohren et al., 2019a). For example, Knowles et al. (2001) tests racial profiling by comparing disparities between racial groups in police traffic searches to disparities between racial groups in the probability of finding contraband. The empirical test of this paper derives from incorporating the Phelps model into a random utility framework. It shares the idea of the outcome-based test, but the methodology applies to more general settings of hiring decisions.

4.2 Identification of Supply Primitives

The identification of the cost distribution of workers follows from the identification results in empirical auctions (Guerre et al., 2000; Krasnokutskaya and Seim, 2011; Athey et al., 2011). The winning probability given the submitted bid and observable attributes is directly identified from data, and so is its derivative. Then using the first-order condition for optimal bidding the private costs are recovered via:

$$c_{j\ell} = (1-r)\left[b_{j\ell} + \frac{\Pr(\text{j wins} \mid b_{j\ell},; \Lambda_\ell)}{\partial \Pr(\text{j wins} \mid b_{j\ell},; \Lambda_\ell)/\partial b_{j\ell}}\right] \tag{13}$$

5 Estimation and Results

I now turn to the estimation strategy and results for the source of discrimination and cost distributions. The estimation argument closely follows the identification argument. I first outline the order of the estimation and then detail each step separately.

Step 1 Construct indices for true quality and signal using principal component analysis.

Step 2 Estimate employers' preferences for worker characteristics using conditional logit regression.

Step 3 Quantify taste-based vs. statistical discrimination, given estimates from Steps 1 and 2.

Step 4 Fit a log-normal distribution to the observed bids.

Step 5 Estimate the winning probability, given the simulated draws from Step 4 and parameters from Step 2.

Step 6 Recover cost distributions using the first-order condition for optimal bidding strategy.

5.1 Step 1: Indices for Quality and Signal

In Section 4, I define quality as a single measure that combines past, present, and future performance variables of a worker. The signal is defined in the same way as quality but it only includes past performances at the time of hiring. The remaining question is how to combine multiple variables into a single measure.

I adapt the commonly used methodology for constructing composite indicators in various national and international agencies. The construction of typical composite indicators broadly follow three steps: data selection, multivariate analysis, and sensitivity analysis.[8] In this section, I explain

[8] Nardo et al. (2008) recommends following the ten steps including theoretical framework, data selection, imputation of missing data, multivariate analysis, normalization, weighting and aggregation, uncertainty and sensitivity analysis, back to the data, links to other indicators, and visualization of the results. I cluster some steps and omit the steps that are not relevant to my analysis.

data selection and multivariate analysis in detail, and leave sensitivity analysis for future work.[9]

Data Selection. Variables should be selected on the basis of their relevance, measurability, analytical soundness, etc (Nardo et al., 2008). I use six variables that are relevant to worker performance: simple average rating, Bayesian adjusted average rating (Yang and Zhang, 2013), number of reviews, average sentiment score of textual reviews, average sentiment score adjusted by the number of words in the textual reviews, and completion rate. I list the definitions of the selected variables in Appendix A.1 and present their summary statistics in Table A.1. To calibrate rating scores by the variation in the number of reviews, I compute Bayesian adjusted average rating, which is a weighted measure between overall average rating of the same job category (prior) and individual average rating (posterior). To obtain sentiment scores of review text, I use the Amazon Comprehend text analysis API service, which provides deep learning models capable of accurately determining the sentiment of a given document.[10] Examples of sentiment scores and the distributions of adjusted sentiment scores by star ratings are shown in Appendix A.1. As expected, sentiment scores exhibit an increasing pattern with respect to star ratings.

Multivariate Analysis. The next step is to combine these six variables into a single index. I use principal component analysis (PCA), which is a widely used technique for variable reduction. The idea of PCA is to transform a set of correlated variables into a smaller set of uncorrelated variables called principal components. The principal components are linear combinations of the original variables. Their coefficients called loadings are determined in a way that the largest loading is assigned to the variable that has the largest variation across workers. Thus, they represent the statistical importance of the individual variables. The first principal component captures most of the variation of the original variable set, the second principal component captures most of the remaining variation, and so on. Appendix Figure A.2 depicts the proportion of variance explained by each of the six principal components. I use the first principal component as an index of quality and signal, respectively. It accounts for 46% of variance in the original data set. The first component is represented as:

$$q_j = \phi_{11} X_{1j} + \phi_{21} X_{2j} + ... + \phi_{61} X_{6j} \tag{14}$$

[9] The robustness section will be added in the updated version.

[10] https://docs.aws.amazon.com/comprehend/latest/dg/how-it-works.html

where q_j is a quality index, X_{kj} is the variable k, and ϕ_{k1} is the factor loading of the first principal component for factor k.

To examine the validity of PCA, I first see correlations among the six variables. If the correlation between individual variables is too small, PCA is not an appropriate method to combine a set of variables. I also see correlations between individual variables and the first principal component. As shown in Appendix Table A.4, the individual variables show high degrees of correlation with each other and with the first principal component.

I also use the second principal component to increase the explained variance in the original data for robustness check. As shown in Table A.3, the first component can be interpreted as quality measure that represents how a worker is good at a given job because relatively high weights are imposed on ratings and sentiment scores. Similarly, the second component can be interpreted as experience or reliability because relatively high weights are imposed on the number of reviews and completion rate. I compute the first and second component respectively and then combine them to construct a single quality measure using the explained variance associated with each component as weights.[11] Then, about 63% of variance of the total variance is explained by a new measure of quality. The estimation results using this measure are consistent with the main results.

Results. Once I have a quality and signal index for each worker, I calculate signal errors for each worker using $\eta_{j\ell} = s_{j\ell} - q_j$. Then, I fit a normal distribution to the data by gender to get parameters of quality distributions and errors of signal distributions by gender. Workers may have different qualities and signals for different kinds of jobs. For this reason, I construct quality and signal distributions for each job category.

Figure 3 presents quality distributions by gender for five kinds of jobs: cleaning, moving, gardening, repairs, and delivery. It shows that female workers for moving have lower mean of quality compared to their male counterparts. It also shows that female workers for repairs have large variance of quality compared to their male counterparts. I do not find any significant differences in quality between male and female for cleaning and delivery jobs.

Figure 4 displays the distributions of signal errors by gender for the five kinds of job. It shows that female workers for repairs have larger signal error variance compared to their male

[11] The weight for the first component is $0.46/(0.46+0.17) = 0.72$ and that for the second component is $0.17/(0.46+0.17) = 0.28$.

counterparts, which means that the signal is a less reliable measure of quality for female workers than male workers. Large variance of quality leads employers to put less weight on group average quality and put more weight on signal, while the opposite would be the case with large variance of signal errors. Combining these two effects, there is no difference between male and female workers in the relative weight on signal for repair jobs. Indeed, $\delta_M = \delta_F = 0.6$ where $\delta_g = \sigma_g^2/(\tau_g^2 + \sigma_g^2)$, σ_g^2 is the variance of quality for group g, and τ_g^2 is the variance of signal errors for group g. The parameter estimates for distributions are shown in Appendix Table C.1.

5.2 Step 2: Conditional Logit Regression

From Section 4.1, the equation of interest is repeated below:

$$U_{j\ell}^* = \mathbb{E}(U_{j\ell}|s_{j\ell}, f_j) = \eta + \phi f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell} + \epsilon_{j\ell} \tag{4 revisited}$$

The employer chooses worker j, $Y_{j\ell} = 1$, if and only if $U_{j\ell}^* \geq U_{0\ell}^*$ and $U_{j\ell}^* \geq U_{i\ell}^*$ for all $i \neq j$ who are present in the auction; $Y_{j\ell} = 0$ otherwise. Since $\epsilon_{j\ell} \sim$ EV, we estimate the parameters of the probability below using maximum likelihood estimation.

$$P(Y_{j\ell} = 1) = \frac{\exp[\phi f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell}]}{1 + \sum_{i=1}^{J_\ell} \exp[\phi f_i + \rho s_{i\ell} f_i + \theta s_{i\ell} - b_{i\ell} + \gamma \mathbf{x}_{i\ell}]} \tag{15}$$

The estimates from this regression are shown in Appendix Table C.2. I find evidence of discrimination in favor of women for cleaning jobs, while there is discrimination in favor of men unless the signal of quality is below the 12$^{\text{th}}$ percentile for moving jobs. Discrimination in favor of men is observed regardless of signal for gardening and repairing jobs. For any given signal, I find no evidence of discrimination for delivery jobs for any given signal.

5.3 Step 3: Taste-based vs. Statistical Discrimination

Given the estimated parameters in Steps 5.1 and 5.2, I quantify taste-based and statistical discrimination using expressions for each source of discrimination. The first column of Table 5 displays employer willingness to pay to hire a female worker rather than a male worker. Each auction has its project value. To control for auction heterogeneity, I conduct my analyses using bids normalized by

Figure 3: Quality Distributions by Gender

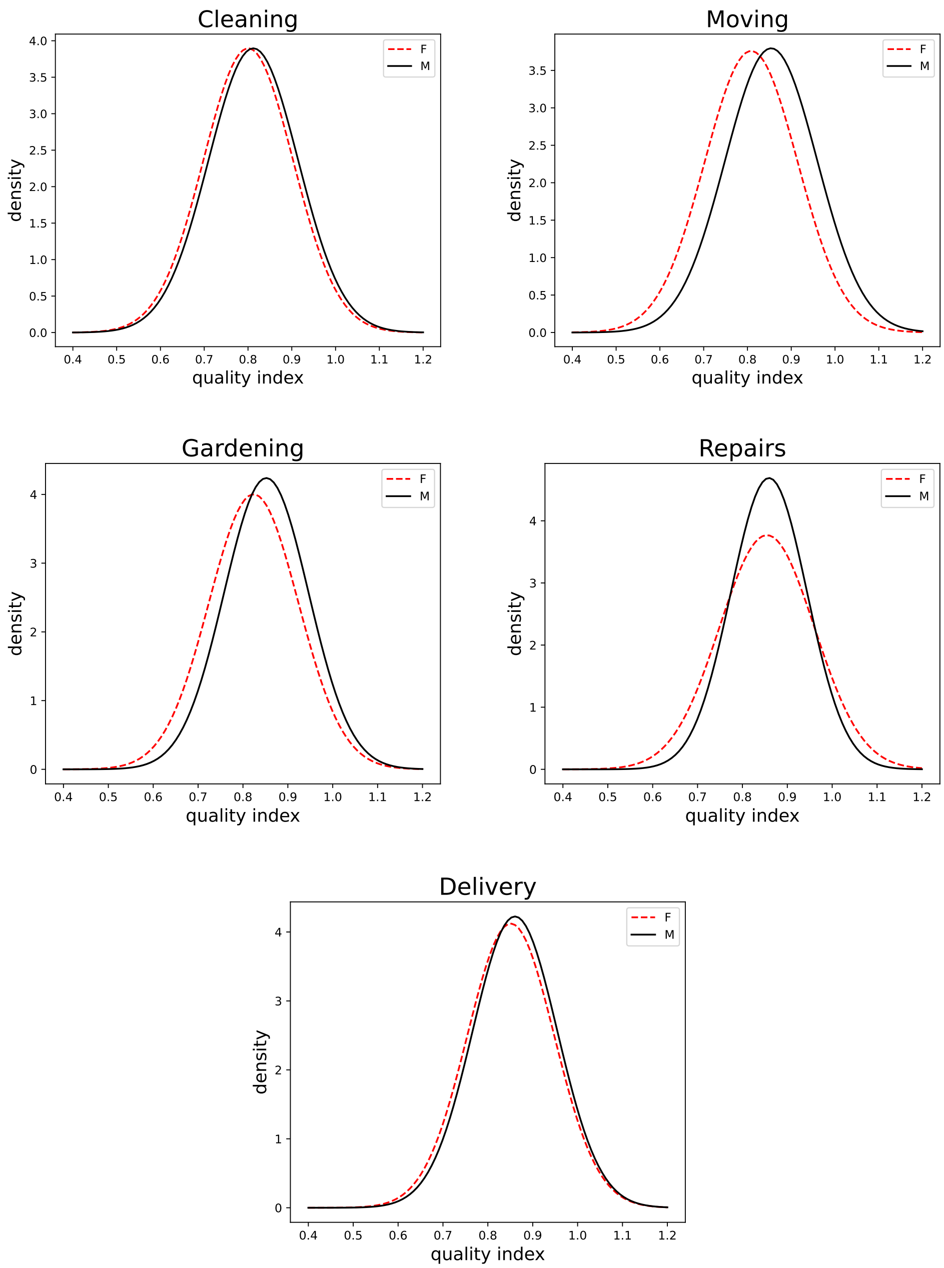
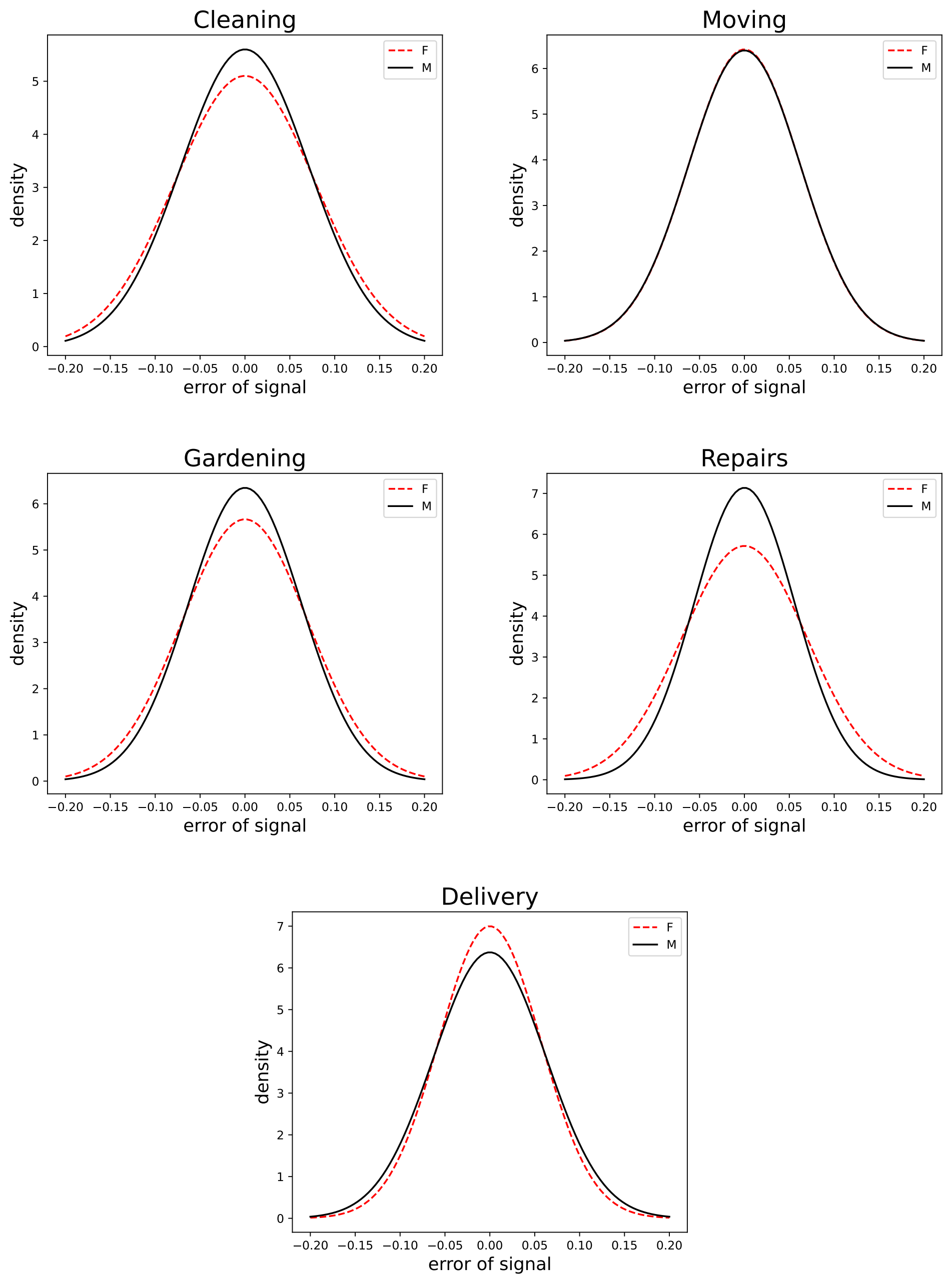


Figure 4: Errors of Signal Distributions by Gender



their project values. Thus, the willingness pay is also measured as a fraction of the project value. To get a sense of how much the willingness to pay is in US dollars, I assume that the project value is $37 and recover the unnormalized price.[12] Jobs with a project value of $37 are most common across job categories. Furthermore, these jobs are most likely to be 1-hour tasks, so it allows me to compute the relative magnitude of the willingness pay compared with the median hourly wage.

I find evidence of discrimination against female workers in three out of the five job categories: moving, gardening, and repairs. The magnitude of discrimination is large and economically significant.

For instance, to win a moving job, a woman has to bid $2.5 less than an (observationally) equivalent male. This "gender tax" amounts to approximately 9% of the median hourly wage. Out of this $2.50, $1.40 is due to statistical discrimination and $1 to taste-based discrimination. Consistent with these breakdowns, I find that female workers have lower quality for moving than male workers on average, which leads to statistical discrimination in hiring a mover.

The employer who looks for a cleaner is willing to pay $3 more to hire a female worker. The employer willingness to pay can be interpreted as a "gender premium" from the worker's perspective. A female worker can bid $3 more than an otherwise identical male worker. This amounts to 11% of the median hourly wage. In Figure 3, there are low disparities between male and female workers doing cleaning jobs in both quality and signal errors. Therefore, taste-based discrimination is 13% of the median hourly wage, while statistical discrimination is limited to 1% of the median hourly wage.

In contrast to cleaning jobs, an employer who wants to hire a mover is willing to receive $7 to compensate for hiring a female. From the worker's perspective, a woman has to bid $2.50 less than an (observationally) equivalent male to win a moving job. This gender tax amounts to approximately 9% of the median hourly wage. Out of this $2.50, $1.40 is due to statistical discrimination and $1 to taste-based discrimination. Consistent with these breakdowns, I find that female workers have lower quality for moving than male workers on average, which leads to statistical discrimination in hiring a mover.

For gardening and repairs, the magnitude of discrimination is even larger. The gender tax for a female worker is $6.10 and $7.10, respectively, and is approximately 23% and 28% of the

[12] Project values are generally multiples of 10. $37 is calculated accounting for the exchange rate.

median hourly wage. For these jobs, taste-based discrimination explains more than 85% of the total discrimination. Lastly, there is no evidence of discrimination in delivery jobs.

To summarize, I find that taste-based discrimination is the primary form of discrimination in most jobs, except for moving. There is taste-based discrimination against male workers for female-type jobs (e.g., cleaning) and taste-based discrimination against females for male-typed jobs (e.g., gardening and repairs). Taste-based discrimination is interpreted such that employers simply may not like men doing cleaning and not like women doing gardening or repairs, even if there is no difference in quality between men and women.

Table 5: Taste-based versus Statistical Discrimination

	Total		Taste-based		Statistical	
	WTP	% of MHW	WTP	% of MHW	WTP	% of MHW
Cleaning	2.97	11.28	3.35	12.74	-0.38	1.46
Moving	-2.44	9.27	-1.03	3.92	-1.41	5.35
Gardening	-6.05	23.03	5.16	19.65	-0.89	3.38
Repairs	-7.05	27.6	-6.99	26.58	-0.27	1.02
Delivery	0.67	2.53	0.88	3.36	-0.21	0.83

Note: WTP represents employer willingness to pay to hire a female worker compared with a male worker. Amounts are reported in US dollars, assuming that the project value is \$37. MHW represents the median hourly wage.

5.4 Step 4: Bid Distribution

To recover cost distributions, I first need to estimate the winning probability and its derivative. The winning probability can be estimated non-parametrically following the two-step method of Guerre et al. (2000). This approach is valid with few homogeneous bidders. If auctions or bidders are heterogeneous, we have to condition on all covariates to avoid misspecification error (Krasnokutskaya and Seim, 2011); however, conditioning large covariates with finite samples is not feasible to purely non-parametric methods, leading the "curse of dimensionality." In this paper, I estimate the winning probability making a parametric assumption on bid distributions due to the high dimensionality of attributes. I should note that there is also a semi-paramteric method (Aryal et al., 2021).

Recall that the winning probability takes the following form given that demands are logit and costs are private.

$$P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) = \int p_j(b_{j\ell}, b_{-j\ell}(c_{-j\ell}); \Lambda_\ell) dF(c_{-j\ell})$$
$$= \int \frac{\exp[\phi f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell}]}{1 + \sum_{i=1}^{J_\ell} \exp[\phi f_i + \rho s_{i\ell} f_i + \theta s_{i\ell} - b_{i\ell}(c_{i\ell}; \Lambda_\ell) + \gamma \mathbf{x}_{i\ell}]} dF(c_{-j\ell}) \tag{16}$$

where $F(c_{-j})$ is the joint distribution of rivals' costs and $\Lambda_\ell = (\lambda_{j\ell}, \lambda_{-j\ell})$ is the composition of the set of workers in the auction ℓ in terms of characteristics. Due to the uncertainty about rivals' costs, the worker expects that rivals' equilibrium bids are stochastic. Each rival's bid is drawn from a distribution, $G(b|\Lambda_\ell)$. Thus, I can rewrite 16 as:

$$P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) = \int \frac{\exp[\phi f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell}]}{1 + \sum_{i=1}^{J_\ell} \exp[\phi f_i + \rho s_{i\ell} f_i + \theta s_{i\ell} - b_{i\ell} + \gamma \mathbf{x}_{i\ell}]} dG(b_{-j\ell}|\Lambda_\ell) \tag{17}$$

where $G(b_{-j})$ is the joint distribution of rivals' bids. I assume that bids follow a log-normal distribution with parameters depending on worker and rivals' attributes:

$$G(b_{i\ell}|\Lambda_\ell) = \Phi\Big(\frac{\ln b_{i\ell} - \mu(\Lambda_\ell)}{\sigma(\Lambda_\ell)}\Big) \tag{18}$$

where Φ is the cumulative distribution of the standard distribution. Here, μ is the mean and σ is the variance of the distribution, parameterized as $\mu = \alpha_0 + \alpha_1 V_\ell + \alpha_2 f_i + \alpha_3 F_\ell + \alpha_4 M_\ell + \alpha_5 MS_\ell + \alpha_6 HS_\ell$ and $\sigma = \exp(\beta_0 + \beta_1 V_\ell + \beta_2 f_i)$ where V_ℓ is the project value of the auction ℓ, f is an indicator whether the worker i is female, F is the total number of female workers in the auction ℓ, M is the total number of male workers, MS is the total number of workers whose signal is in the middle level (from 33$^{\text{th}}$ to 66$^{\text{th}}$ percentile), and HS is the total number of workers whose signal is in the high level (above 66$^{\text{th}}$ percentile).

Appendix Table D.1 reports the estimated coefficients of the bid distribution for cleaning jobs. The estimated coefficients are of the expected sign. Since female workers are preferred by employers for cleaning, the bid price decreases as the number of female bidders increases, while it increases as the number of male bidders increases. Similarly, bidders with higher signals are preferred by employers, and thus bids decrease as the number of bidders with a high or mid level signal increases. Female workers are more likely to submit lower bids compared to their male counterparts. We would expect that female workers tend to have lower costs than male workers for cleaning.

I assume that private costs are independent across workers, and thus rivals' bids are also independently drawn from the bid distribution. Therefore, the joint bid distribution can be written as the product of individual rival bid distributions.

$$G(b_{-j\ell}|\Lambda_\ell) = \prod_{i\neq j} G(b_{i\ell}|\Lambda_\ell) \tag{19}$$

5.5 Step 5: Winning Probability

Given the estimated bid distribution, I obtain numerical estimates of the probability that the worker j wins, $P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell)$, and its derivative, $\partial P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell)/\partial b_{j\ell}$ where $\Lambda_\ell = (\lambda_{j\ell}, \lambda_{-j\ell})$. The procedure is:

1. For each rival $i(\neq j)$ in auction ℓ, draw b_i from $G(b|\lambda_{i\ell}, \lambda_{-i\ell})$.

2. Calculate the probability that worker j wins given her own bid $b_{j\ell}$, her attributes, her rivals' attributes, and the bids drawn in step 1. Calculate its derivative. This constitutes one simulation.

$$\begin{aligned} p_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) = \mathbb{P}_j &= \frac{\exp[\phi f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} - b_{j\ell} + \gamma \mathbf{x}_{j\ell}]}{1 + \sum_{i=1}^{J_\ell} \exp[\phi f_i + \rho s_{i\ell} f_i + \theta s_{i\ell} - b_{i\ell} + \gamma \mathbf{x}_{i\ell}]} \\ \frac{\partial p_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell)}{\partial b_{j\ell}} &= -\mathbb{P}_j(1 - \mathbb{P}_j) \end{aligned} \tag{20}$$

3. Repeat step 1 and 2 N times (say 1,000) and take averages to get the winning probability and its derivative, respectively.

$$\begin{aligned} P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell) &= \frac{1}{N}\sum_{n=1}^{N} p_j(b_{j\ell}, b^n_{-j\ell}; \Lambda_\ell) \\ \frac{\partial P_j(b_{j\ell}, b_{-j\ell}; \Lambda_\ell)}{\partial b_{j\ell}} &= \frac{1}{N}\sum_{n=1}^{N} \frac{\partial p_j(b_{j\ell}, b^n_{-j\ell}; \Lambda_\ell)}{\partial b_{j\ell}} \end{aligned} \tag{21}$$

5.6 Step 6: Cost Distribution

Given the winning probability, I recover a sample of pseudo costs using the first-order condition for optimal bidding (13). I then use the pseudo costs to estimate the distribution of workers' costs nonparametrically. Figure 5 depicts the estimated distributions of costs by gender for each job category. The cost distribution of female workers is first-order stochastic dominated by that of male workers for most jobs except repairs. Table 6 presents the median costs by gender for each job category. The median cost for female workers is \$14, whereas that for male workers is \$20 for cleaning jobs. Interestingly, female workers have lower costs than male workers for most jobs except repairs. It may be counterintuitive that female workers have lower costs for moving and gardening. However, considering opportunity costs, the results are plausible, since male workers usually have better other options than female workers.

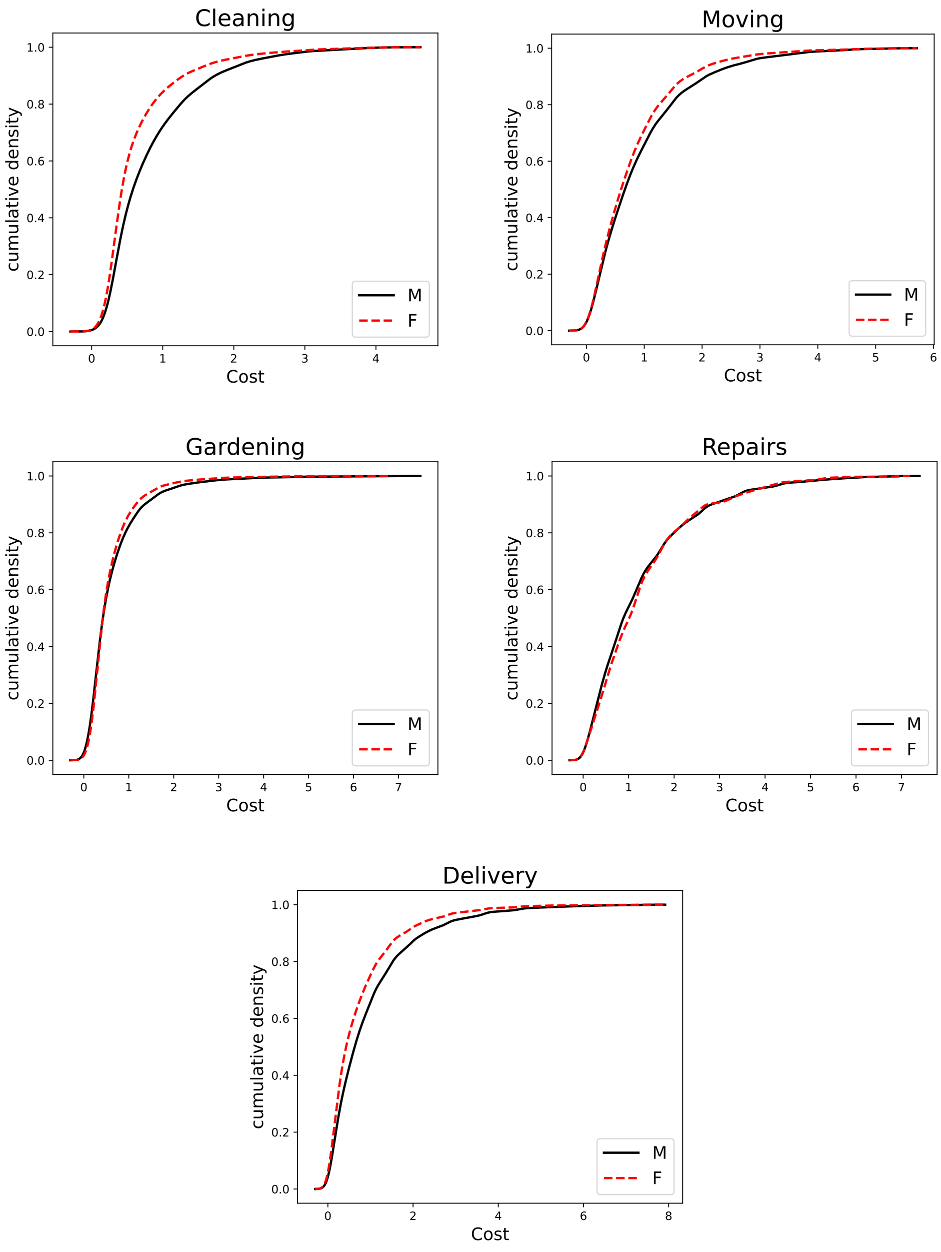
On the other hand, worker market power (measured by margins) by gender differs across job categories. Two factors affect workers' market power: comparative advantage in terms of costs and gender discrimination in the market. Figure 6 presents the estimated distributions of margins by gender for each job category. From Table 6, for cleaning jobs, I find that the margins of female workers are \$20 and \$18 for males, which means that female workers have more market power. This is as expected, because female workers both have lower cost and are preferred by employers, and thus they can charge a gender premium. I do not find stochastic dominance of market power for moving or gardening jobs. There is discrimination against female workers, even though female workers have a cost advantage. The two effects work in opposite directions, and on net they do not have significantly different market power. In contrast, male workers have more market power than female workers for repairs. This is because there is gender discrimination against female workers while there is no cost difference between males and females. For deliveries, female workers are more competitive because they have lower costs on average in the market without discrimination. Table 7 summarizes the directions of the two effects for each category and their consequences for market power.

Table 6: Bids, Costs, and Margins

	Bids (USD)			Costs (USD)			Margins (USD)		
	M	F	Total	M	F	Total	M	F	Total
Cleaning	45.63	39.54	42.58	20.46	13.98	16.38	18.32	19.63	19.82
Moving	56.81	53.84	56.15	24.08	21.88	23.88	24.22	23.88	23.85
Gardening	43.80	43.07	43.80	14.93	14.71	14.90	22.30	21.90	22.33
Repairs	73.00	80.30	73.00	31.05	37.64	31.24	31.00	30.61	30.81
Delivery	57.36	50.19	54.75	22.74	14.92	21.35	26.02	27.74	25.19

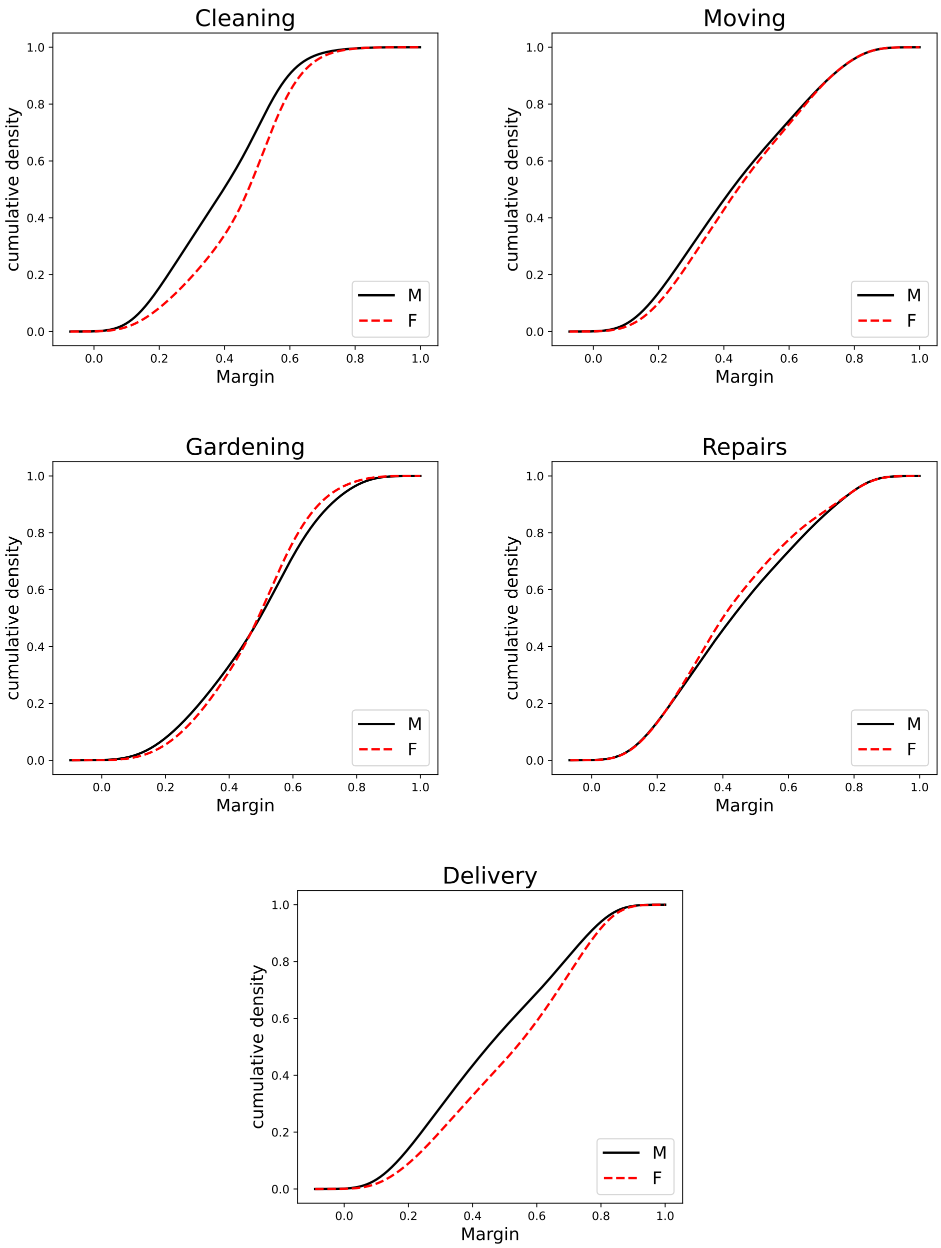
Note: The values in this table are median values measured in US dollars, assuming that the project value is \$37. margin = 0.85×bid - cost.

Figure 5: Cost Distributions by Gender



Note: Costs are normalized by their project values.

Figure 6: Margin Distributions by Gender



Note: Margins are normalized by their project values.

Table 7: Gender Premium, Cost Advantage, and Market Power

	Gender Premium	Cost Advantage	Market Power
Cleaning	F	F	F
Moving	M	F	-
Gardening	M	F	-
Repairs	M	-	M
Delivery	-	F	F

5.7 Discussion

Bias in Reviews. I constructed a quality measure using job completion rate and the ratings (or reviews) workers had received from employers. The completion rate, which is defined as the percentage of completed jobs out of all assigned jobs, is an objective measure of quality. On the other hand, ratings may not reflect a worker's true quality because evaluators make their own subjective assessment of the worker's performance. This can threaten the validity of the construction of the quality measure and thus the distinction between taste-based and statistical discrimination.

Three main channels cause bias in reviews on online platforms. First, employers may refrain from leaving negative feedback due to fear of retaliation, which can result in overwhelmingly positive ratings on the platform. Bolton et al. (2013); Nosko and Tadelis (2015); and Filippas et al. (2019) argue that leaving a negative rating is more costly than leaving a positive rating because of the possibility of retaliation and harassment. This problem may be more severe on the platform I study because jobs such as cleaning and moving require face-to-face interaction with employers. Indeed, the distribution of simple average ratings is highly skewed toward perfect scores in my sample. Appendix Figure A.3 shows that about 90% of workers' average ratings are greater than 4.5 out of 5.[13] For this reason, I use not only the simple average rating but also adjusted rating measures such as a Bayesian adjusted rating, the sentiment score in textual reviews, and so on. This leads to more variations in ratings across workers, which enables me to make fine distinctions among workers with good ratings.

Another channel through which bias in reviews may occur is fake reviews. Sellers may buy good

[13]It is worth noting that this inflation is not wholly explained by retaliation. It is also explained by user survival patterns on online platforms. Users with low ratings are unable to get hired again, so they exit the market. Thus, the majority of platform users end up having high ratings, which is common in online marketplaces (Hu et al., 2009). In such a case, the overwhelmingly positive ratings may not be a critical issue.

reviews they do not deserve, which is a common issue across many in online marketplaces such as Amazon, eBay, and Yelp (Tadelis, 2016). However, the platform in this study only allows employers to leave reviews of workers they have actually hired through the platform. In such a case, the cost of a fake review on this platform is much higher than the cost on a platform on which anyone can leave a review. Therefore, I assume that bias in reviews due to fake reviews is negligible in my setting.

Last, and importantly, there may be gender bias in reviews. Previous research finds evidence for gender bias in customer evaluations of service quality, especially in the context of student evaluations of teaching (Boring, 2017; Mengel et al., 2018). These studies show that students systematically give lower teaching evaluations to female teachers than male teachers. In the context of the gig economy, Bohren et al. (2019b) test the existence of gender bias in evaluations using a field experiment on a large online Q&A forum. They find evidence for gender bias in evaluations of questions (employer evaluations rated by worker) but no evidence of answers (worker evaluations rated by employer). Thus, they use the evaluations of answers to proxy for underlying quality, similar to my study.

The ideal setting to test the existence of gender bias in reviews is a field experiment or a correspondence study, in which we can exogenously vary the gender and quality associated with a worker. Unfortunately, in this observational setting, I cannot determine whether a gender gap in ratings is due to quality differences or gender bias. I use reviews to construct worker quality under the implicit assumption that reviews are not biased. It is worth noting that measurement errors due to bias in reviews can be calculated from the model. According to (4), taste-based discrimination is represented as α and statistical discrimination is represented as $\beta\big\{\big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big]\big\}$, where μ_g is average quality by gender and δ_g is a function of quality dispersion by gender and signal errors dispersion by gender. Suppose that employers systematically give lower ratings to female workers than male workers, regardless of quality. Assume that the amount of bias in ratings is represented by x, and thus average quality based on ratings is now $\mu'_F = \mu_F - x$ and $\mu'_M = \mu_M$, where μ_g is actual average quality and μ'_g is biased average quality (based on ratings). Replacing μ_F and μ_M with μ'_F and μ'_M, the level of taste-based discrimination is $\alpha - \beta(1-\delta_F)x$. Thus, my results may overestimate (or underestimate) taste-based discrimination by $\beta(1-\delta_F)x$ in the presence of bias in reviews.

Evolution of Worker Quality. The canonical Phelps model specifies the expected quality for a given worker with the assumption that the worker's latent quality is static. As with the Phelps model, most discrimination literature assumes that quality is fixed over time.[14] However, worker quality may be dynamic and evolve with time or gained experience. If worker quality is dynamic, combining past, present, and future reviews is not valid for constructing the quality of a given worker at the time of hiring. Given this, I run a regression of ratings on worker experience. I use worker fixed effects to control for time-invariant worker quality. I also control for auction or job characteristics such as project value and year and month in which the auction was held. As shown in Appendix Table B.1, I find no evidence for performance dynamics. Worker experience, which is measured by the number of reviews a worker has received so far, has no significant effect on star rating, sentiment score of textual reviews, or sentiment score adjusted by the number of words in textual reviews. Therefore, the assumption of static quality is valid.

Retention and Quality. If the quality measure I constructed accurately reflects a worker's true quality, it will be positively correlated with employer retention. The idea behind this argument is that an employer who has had a better experience on the platform is more likely to continue to hire workers via the platform in the future. Therefore, employers who were interacted with high-quality workers are more likely to remain on the platform. Moreover, employer retention should not be affected by the worker's signal of quality after controlling for the worker's true quality. Signal is an error-ridden measure of true quality and affects employer retention only through the quality channel.

In line with this idea, I run a Probit regression for two specifications. In the first specification, I use a signal index as an explanatory variable along with worker gender. In the second, I add a quality index as another explanatory variable. The dependent variable is an indicator that takes a value of 1 if an employer makes another transaction later and 0 otherwise. Appendix Table B.2 presents my estimates of the effect of worker quality on employer retention. Column (a) shows that employers who hired workers with high signal are more likely to remain on the platform. However, as shown in Column (b), this effect becomes insignificant after taking true quality into account. The results support my construction of quality and signal measures. Signal and quality measures

[14] For example, Farber and Gibbons (1996), Altonji and Pierret (2001), and Arcidiacono et al. (2010) use AFQT scores as a proxy for worker quality.

are highly correlated with each other (correlation = 0.72), however, the quality measure includes components that are hard to observe from the signal and is a more accurate measure of true quality that directly affects employers' satisfaction.[15]

The results also mitigate concerns about review bias. Given that the quality measure based on reviews is correlated with employer retention, reviews are based on the worker's true qualities and are not contaminated by fake reviews or gender bias in ratings.

6 Counterfactual Analyses

What if the platform removed information that identifies the gender group to which an individual worker belongs? This policy has been implemented in some hiring processes, especially in online settings. For instance, when employers hire a worker without conducting face-to-face interviews, they may take a name and/or a profile photo from a resume. Even in the offline world, we may be able to blind gender information. For example, Goldin and Rouse (2000) show that the probability a woman will be hired increases if audition procedures for symphony orchestras adopt blind auditions with a screen.

Some experimental research has shown that the acceptance rate of members of an underrepresented group increased after they blinded information about group identity. However, little is known about how much the market outcomes and welfare of each agent change in markets.

In the previous section, I find that taste-based is the main driver of discrimination on the platform. In this case, removing gender information may be the best solution to curb employers' discriminatory behavior. Providing more information about an individual worker's quality will not help to reduce differential hiring outcomes if taste-based discrimination is dominant. Not only does it rule out taste-based discrimination, but removing gender information ensures that all types of discrimination is prevented, including statistical discrimination.

Given the primitives estimated in the previous section, I conduct a counterfactual simulation in which the gender of a worker is blinded. Under the counterfactual scenario, the employer's expected

[15] The results should be interpreted with some caution. First, there can be potential selection problems because employers are not randomly matched with workers. Heavy users may be more likely to choose high-quality workers and to remain on the platform. Second, the analysis does not consider disintermediation. Once an employer finds a high-quality worker, he may prefer to transact with the worker outside the platform to avoid commission fees. Since it is difficult to observe and measure disintermediated transactions, I am unable to consider this possibility in my analysis.

utility becomes

$$\begin{aligned}\mathbb{E}(U_{j\ell}|s_{j\ell},g_j) &= \alpha f_j + \beta\mathbb{E}(q_j|s_{j\ell},g_j) - b_{j\ell} + \gamma\boldsymbol{x}_{j\ell} + \epsilon_{j\ell} \\ &= \Pr(g_j = F)\alpha + \beta[\Pr(g_j = F)\{(1-\delta_F)\mu_F + \delta_F s_{j\ell}\} \\ &+ (1-\Pr(g_j = F)\{(1-\delta_M)\mu_M + \delta_M s_{j\ell}\}] - b_{j\ell} + \gamma\boldsymbol{x}_{j\ell} + \epsilon_{j\ell},\end{aligned} \tag{22}$$

where $\Pr(g_j = F)$ is the proportion of female workers for a specific job category.

Since the employer does not observe the gender of a worker, he relies on the unconditional expectation about the worker's gender when making hiring decisions. For example, he expects to receive additional utility from choosing a worker, $\Pr(g_j = F) \times \alpha + (1 - \Pr(g_j = F)) \times 0$, for his innate preference for female workers. With a $\Pr(g_j = F)$ chance, he can work with a female worker and get additional utility of α.

I recovered employers' preferences over worker characteristics and workers' costs in the previous section. Since primitives of the model are unlikely to change under the counterfactual scenario, I continue to use them. Given the estimates and the new form of utility function, I compute the new optimal bid for each worker. Then, I determine who wins the auction and examine changes in welfare for each agent: employer, worker, and platform.

Table 8 describes the changes in market outcomes after blinding gender information. I focus on the four job categories in which evidence of discrimination was found in the previous section. The results show that workers from the disfavored group increase their prices, and the likelihood that they are chosen as a service provider goes up after removing gender information. The changes become larger as the worker composition by gender is highly skewed and the level of discrimination is higher.[16]

Table 9 provides the changes in welfare for each agent type in the market. For jobs in which one gender is highly dominant (e.g., moving, gardening, repairs), the surplus of the disfavored group

[16]Consider the auctions for gardening with two bidders: one female and one male. Before removing gender information, the male bidder (a member of the favored group) could charge a premium for his gender, while the female bidder (a member of the disfavored group) discounted her bid. After removing gender information, the female bidder would increase her bid substantially relative to her original bid because the employer would expect that she is female with only 14% chance. On the other hand, the male bidder would decrease his bid a little because the employer believes that the worker is male with 86% chance. The change of bid relative to original bid is smaller if the workers with the same gender competed in the auction because they did not set a price high even before removing gender information.

increases by 2-18%, while the surplus of the favored group does not change much. The platform revenue decreases by 4.5% for cleaning jobs after removing gender information, because employers who post these jobs are more likely to opt out for outside options rather than choosing a worker on the platform.

The results must be interpreted with some caution, as there are a number of caveats. First, I assume that the set of employers who post jobs is fixed, which may not be true in the real world. Rather, employers adjust their entry decisions in response to the policy change. Employers with strong animus toward a specific gender would choose not to post jobs at all if the platform removes gender information. In such a case, my counterfactual results my overestimate the gains of the disfavored group. I will discuss a model of employer entry in Section 7.2. Second, I further assume that the set of workers who bid for a given task is fixed, which may not be realistic. After the gender-blind policy is implemented, workers from the disfavored group are more likely to enter auctions while those from the favored group are less likely to enter. Thus, the total gains to the disfavored group are expected to be even larger if the model allows for endogenous worker entry. The idea of a model of worker entry will be discussed in Section 7.3. Lastly, workers may signal their gender groups via other methods such as conversations between workers and employers or bid prices. In such a case, my estimates would provide a upper bound on the gains to the disfavored group.

7 Extensions

In this section, I extend the baseline model to allow for employers' heterogeneous preferences and endogenous employer and worker entry. I then briefly discuss how my counterfactual estimates would change under the extended models.

7.1 Heterogeneous Preferences

So far, I have assumed that all employers have the same preferences over worker characteristics. However, they are more likely to have heterogeneous preferences, especially for gender. In reality, some — but not all — employers like men, some dislike men, and others are indifferent between

Table 8: Changes in Market Outcomes under the Counterfactual Scenario

	Optimal bid		Margin (%)		Winning Price	Allocation (%)		
	M	F	M	F		M	F	O
Before:								
Cleaning	1.554	1.347	38.87	44.84	1.140	14.55	33.77	51.68
Moving	1.853	1.705	43.97	45.56	1.355			
Gardening	1.438	1.350	48.12	47.74	1.162			
Repairs	2.528	2.541	44.48	42.95	1.880	39.78	1.64	58.59
After:								
Cleaning	1.561	1.343	39.40	44.49	1.138	14.66	31.28	54.06
Moving	1.852	1.712	43.90	46.00	1.354	47.38	5.59	47.04
Gardening	1.433	1.376	47.78	49.68	1.162	54.85	9.05	36.1
Repairs	2.527	2.565	44.43	44.12	1.879	39.81	1.7	58.49
Change (%):								
Cleaning	0.45	-0.34	0.53	-0.36	-0.17	0.11	-2.49	2.38
Moving	-0.06	0.37	-0.07	0.44	-0.04	-0.1	0.15	-0.04
Gardening	-0.31	1.92	-0.33	1.94	-0.03	-0.2	1.03	-0.82
Repairs	-0.04	0.94	-0.05	1.17	-0.05	0.03	0.06	-0.1

Note: Optimal bids and winning prices are normalized by their project values.

Table 9: Welfare Changes

	Platform Revenue	Worker (per capita)		
		M	F	Total
Cleaning	-4.59	2.29 (1.52)	-8.07 (-0.74)	-4.67
Moving	0.27	-0.39 (-0.17)	6.63 (3.73)	0.30
Gardening	1.35	-1.09 (-0.73)	18.35 (4.88)	1.29
Repairs	0.35	-0.05 (-0.13)	14.03 (9.88)	0.44

Note: The results in the table show the percentage change. Per capita parentheses.

men and women. Moreover, employers differ in how strongly they like or dislike a given gender.

I consider that employers have heterogeneous preferences for taste-based discrimination. Since I assume that information about group-specific quality distributions and signal error distributions are common knowledge, it is reasonable to say that every employer is homogeneous when it comes to statistical discrimination.

For simplicity, suppose that employers are homogeneous for worker other characteristics but not for taste-based discrimination only. The expected utility for employer ℓ is now represented as

$$\begin{aligned}\mathbb{E}(U_{j\ell}|s_{j\ell}, f_j) &= \alpha_\ell f_j + \beta\mathbb{E}(q_j|s_{j\ell}, f_j) - b_{j\ell} + \boldsymbol{\gamma}\mathbf{x}_{j\ell} + \epsilon_{j\ell} \\ \alpha_\ell &= \alpha_0 + \alpha_1 f_\ell^e + \alpha_2 u_\ell^e + \alpha_3 na_\ell + \alpha_4 \ln(pv_\ell).\end{aligned} \tag{23}$$

Employer ℓ's value of worker gender, α_ℓ, is a function of his characteristics, including demographic groups and auction characteristics. Specifically, f_ℓ^e is an indicator for whether the employer is female, u_ℓ^e is an indicator for whether the gender of the employer is not identified, na_ℓ is the number of jobs the employer has assigned in the past, and pv_ℓ is the project value of the job.[17]

To simplify estimation, I rewrite the utility function:

$$\begin{aligned}\mathbb{E}(U_{j\ell}|s_{j\ell}, f_j) &= \big[\alpha_0 + \alpha_1 f_\ell^e + \alpha_2 u_\ell^e + \alpha_3 na_\ell + \alpha_4 \ln(pv_\ell)\big] f_j + \beta\mathbb{E}(q_j|s_{j\ell}, f_j) - b_{j\ell} + \gamma\mathbf{x}_{j\ell} + \epsilon_{j\ell} \\ &= \Big[\underbrace{\alpha_0 + \alpha_1 f_\ell^e + \alpha_2 u_\ell^e + \alpha_3 na_\ell + \alpha_4 \ln(pv_\ell)}_{\substack{\text{taste-based} \\ \text{discrimination}}} \\ &\quad + \underbrace{\beta\big\{\big[(1-\delta_F)\mu_F + \delta_F s_{j\ell}\big] - \big[(1-\delta_M)\mu_M + \delta_M s_{j\ell}\big]\big\}}_{\substack{\text{statistical} \\ \text{discrimination}}}\Big] f_j + \\ &\quad \beta(1-\delta_M)\mu_M + \beta\delta_M s_{j\ell} - b_{j\ell} + \gamma\mathbf{x}_{j\ell} + \epsilon_{j\ell} \\ &= \phi f_j + \alpha_1 f_\ell^e f_j + \alpha_2 u_\ell^e f_j + \alpha_3 na_\ell f_j + \alpha_4 \ln(pv_\ell) f_j + \rho s_{j\ell} f_j + \theta s_{j\ell} + \nu - b_{j\ell} + \gamma\mathbf{x}_{j\ell} + \epsilon_{j\ell}\end{aligned} \tag{24}$$

where $\nu = \beta(1-\delta_M)\mu_M$, $\phi = \alpha_0 + \beta[(1-\delta_F)\mu_F - (1-\delta_M)\mu_M]$, $\rho = \beta(\delta_F - \delta_M)$, and $\theta = \beta\delta_M$.

Appendix Table E.1 provides the conditional logit estimates for cleaning jobs. I find no evidence of in-group favoritism. Rather, male employers are more likely to hire female workers compared with female employers ($\alpha_1 < 0$). Experienced employers may have weak preferences for a particular gender as they interact with the overall pool of workers on the platform. In contrast to my expectation, employers tend to reinforce their gender prejudice as they gain more experience. Employers who posted jobs with higher project value are less likely to rely on their taste for gender.

If I allow for heterogeneous preferences, aggregate discrimination should be carefully defined. I say that there is aggregate taste-based discrimination in favor of female workers if $E[\alpha_\ell] > 0$ where

[17] Employers are less likely to post photos of their own faces compared with workers. In my sample, the gender of about 16.6% of employers is not identified.

the expectation is taken over the population of employers on the platform. My estimates show that about 89.1% of employers have animus toward male workers while only 10.9% of employers have animus toward female workers, which leads to aggregate taste-based discrimination against male workers for cleaning jobs.

7.2 Endogenous Employer Entry

In the counterfactual analysis, I treated employers' entry (or posting) decisions as given and focused on changes in employers' hiring decisions. This approach, however, ignores the fact that employers may not even enter the platform under different counterfactual scenarios. To obtain more realistic estimates for any policy changes, I need to develop a model of employer entry and solve employers' entry and hiring decisions simultaneously.

Before posting jobs, employers have uncertainty regarding both the number of bids they expect to receive and the attributes of bidders. Thus, I model employers' expectations about the number of bids and bidder attributes. I then characterize employers' entry decisions and discuss expected outcomes after imposing the gender-blind hiring policy.

I assume that the number of bidders in auction ℓ follows a Poisson distribution, with rate λ_ℓ, which depends on auction and employer attributes $\boldsymbol{A}_\ell$. The Poisson distribution is appropriate in this setting for two reasons. First, it models count variables. Second, it is the limiting distribution of binomial distribution with small probability of success; Since there are a large number of workers on online platforms, only a few decide to bid in a given auction (Bajari and Hortacsu, 2003). Thus, the Poisson arrival process is justified in my case. The expected conditional probability of receiving n bids is

$$f(n;\boldsymbol{\zeta}_b) = \Pr(N = n|\boldsymbol{A}_\ell, \boldsymbol{\zeta}_b) = \frac{\lambda_\ell^n e^{-\lambda_\ell}}{n!}, \tag{25}$$

where $\ln \lambda_\ell = \boldsymbol{A}_\ell' \boldsymbol{\zeta}_b$ and $\boldsymbol{\zeta}_b$ is a parameter vector to be estimated. The set of auction and employer attributes is listed in Appendix Table F.1. The results show that female employers are likely to get more bids compared with male employers. It also shows that auctions with detailed job descriptions or higher project value attract more bids. Employers get fewer bids if they write long self-descriptions on their profile pages. I find no significant effect of employer experience or review scores on the number of bids except for cleaning jobs, possibly due to lack of variation in employer experience.[18]

Employers are uncertain attributes for a given bidder, as well as the number of bidders. Note that I considered two primary bidder attributes in the baseline model: worker gender and signal.[19] Employers expect that worker attributes and bids are drawn from joint distributions. I model each of these attributes and then estimate the joint distributions of equilibrium bids and worker attributes. First, I specify a logit model of worker's gender as a function of employer and auction attributes $\boldsymbol{A}_\ell$ that were used in the Poisson model. The probability that bidder j in auction ℓ belongs to gender group g is

$$\begin{aligned} \Pr(g_{j\ell} = female|\boldsymbol{A}_\ell, \boldsymbol{\zeta}_g) &= \frac{\exp(\boldsymbol{A}_\ell' \boldsymbol{\zeta}_g)}{1 + \exp(\boldsymbol{A}_\ell' \boldsymbol{\zeta}_g)} \\ \Pr(g_{j\ell} = male|\boldsymbol{A}_\ell, \boldsymbol{\zeta}_g) &= \frac{1}{1 + \exp(\boldsymbol{A}_\ell' \boldsymbol{\zeta}_g)}. \end{aligned} \tag{26}$$

As shown in Appendix Table F.2, female workers are less likely to bid for jobs with a lower project value. Therefore, employers would expect relatively more male bidders when they post jobs with a high project value. Second, worker's signal is modeled using a normal distribution with parameters as a function of auction and employer attributes $\boldsymbol{A}_\ell$, and worker's gender $g_{j\ell}$. Third, I draw an equilibrium bid for each worker from a log-normal distribution with parameters as a function of worker and rivals' attributes. The distribution has already been estimated in section 5.4.

The expected utility of an employer ℓ from posting a job can be represented as

$$\mathbb{E}U_\ell = \int\limits_{n_\ell} \int\limits_{(g_{j\ell}, s_{j\ell}, b_{j\ell})|_{j=1}^{n_\ell}} \max U_\ell \cdot dH\Big((g_{j\ell}, s_{j\ell}, b_{j\ell})|_{j=1}^{n_\ell} | \boldsymbol{A}_\ell, n_\ell\Big) \cdot dF(n_\ell | \boldsymbol{A}_\ell, \boldsymbol{\zeta}_b), \tag{27}$$

where $\max U_\ell$ is the expected maximum utility from posting job ℓ and receiving n_ℓ bids with worker attributes $\{g_{j\ell}, s_{j\ell}, b_{j\ell}\}_{j=1}^{n_\ell}$. Assuming that employer-worker match terms follow a Type I Extreme Value distribution, $\max U_\ell$ can be expressed by the following logsum formula:

$$\max U_\ell = \ln\Big(1 + \sum_{j=1}^{n_\ell} \exp\left[(g_{j\ell}, s_{j\ell}, b_{j\ell}) \cdot \boldsymbol{\zeta}_u\right]\Big), \tag{28}$$

[18] In my sample, the majority of employers (73.7%) have no past ratings.

[19] Employers also care about workers' other attributes such as race, the number of words in the worker self-description, and the number of words in the message sent by the worker. For simplicity, I focus on gender and signal in the model of employer entry.

where $\boldsymbol{\zeta}_u$ is the set of employer preferences over worker attributes that has been estimated in the baseline model. The expected utility is integrated over their beliefs about the number of bids and worker attributes.

The platform does not charge any fees for posting or canceling auctions. However, there are indirect costs such as time costs, which lead employers to post jobs only if their expected utility from posting the job is greater than that from not posting. Suppose that the utility from not entering the market is U_ℓ^{ne}. Together with the above expression for expected utility, employers' entry decisions are characterized as $\mathbb{1}\{\mathbb{E}U_\ell > U_\ell^{ne}\}$. The entry probability is directly identifiable from the data after defining potential employers.[20]

If the platform blinds gender information, the expected utility from choosing a worker becomes (22). The policy change harms employers with strong preferences for a specific gender, and thus they would not enter the market if the loss of utility is not sufficiently offset by the decrease in equilibrium bid prices. Thus, after taking endogenous employer entry into account, the increase in the surplus of the disfavored group under the counterfactual scenario may shrink because some auctions would not be held.

7.3 Endogenous Worker Entry

As shown in Figure 2, occupational segregation by gender is severe on the platform. Gender segregation across job categories may be a consequence of gender discrimination. If it is costly to submit a bid, workers with unfavorable attributes (e.g., gender) or those with very high costs may choose not to enter the auction because their expected payoffs are lower than the entry cost. This self-reinforcing behavior perpetuates gender imbalances on the platform.

I may model worker entry as a dynamic across-auction optimization problem or a static within-auction problem. In the first, workers compare jobs posted on the platform and choose the job that gives them the highest utility. In the second, workers enter a given auction if the expected utility from entering is greater than the entry costs. In this section, I sketch ideas for the static within-auction model and explain data availability.

Unlike traditional procurement auctions, bidding is not costly on online auction platforms.

[20] I define potential employers as all employers who post similar jobs in the past (broad definition) or those who started to write postings including not released to the public.

However, workers do not bid for every job, which implies that the cost of bidding is not actually zero. There is no direct (explicit) cost for placing a bid, but there might be an indirect (implicit) cost. The platform imposes a penalty if workers win an auction but do not complete the task. Thus, the entry cost may reflect their reputation cost.

If the model allows worker entry decisions, a type-symmetric pure strategy Bayesian Nash equilibrium is now a profile of two strategies, bidding and entry. Borrowing notation from Krasnokutskaya et al. (2020), for each private cost for completing job c and entry cost e, workers' optimal bidding, σ^*, and entry strategy, τ^*, are defined as

$$\begin{aligned}\sigma^*(c_{j\ell}, \Lambda_\ell) &= \arg\max_{b_{j\ell}}[b_{j\ell}(1-r) - c_{j\ell}]P_j(b_{j\ell}, \sigma^*_{-j\ell}, \tau^*_{-j\ell}, \Lambda_\ell) \\ \tau^*(e_{j\ell}, \Lambda_\ell) &= \mathbb{1}\{e \leq \mathbb{E}[\Pi_{j\ell}(\sigma(C_{j\ell}, \Lambda_\ell), C_{j\ell}, \sigma^*_{-j\ell}, \tau^*_{-j\ell}, \Lambda_\ell).\end{aligned} \tag{29}$$

In the extended model, Λ_ℓ is a composition of "potential" bidders in terms of worker attributes. I define the set of potential bidders as all workers who (a) submit at least one bid for jobs in the same category as the job posted and (b) live within driving distance (e.g., 20 miles), because jobs such as cleaning and moving require face-to-face interaction with employers. The former condition ensures that each worker is aware of potential rivals' presence, so they strategically decide their bidding and entry strategy given the composition of potential bidders.

With respect to endogenous participation, the gender-blind hiring policy would increase entry of workers from low cost and disfavored groups. Thus, the total surplus of the disfavored group (e.g., females for moving jobs) would be larger than that under fixed participants. Per capita surplus would decrease due to the increase in competition among disfavored groups.

8 Conclusion and Future Work

Due to its online nature, the gig economy can serve as an effective and scalable experimental platform to test various policies for addressing discrimination, which can be quite difficult in traditional labor markets. Gig economy platforms have tested and implemented various policies, such as blinding gender or racial information or "de-stereotyping" by displaying females in construction outfits.

In this paper, I examine whether there is gender discrimination and investigate the source of discrimination: taste-based or statistical. To answer these questions, I develop a model of demand and supply for freelance jobs. Identifying the two sources is challenging, because different combinations of taste-based and statistical discrimination generate the same level of observed discrimination. I combine the canonical Phelps model into a random utility framework and propose a test to identify the main driver of discrimination. I show that knowing true quality is sufficient to identify each source of discrimination.

My estimation suggests that the magnitude of discrimination is large and economically significant. For instance, to win a repair job, a female has to bid \$7.10 less than an otherwise equivalent male, and this amounts to approximately 28% of the median hourly wage. The primary form of discrimination is taste-based, except for moving jobs. The results indicate that removing gender information could be an effective policy. Indeed, I find that the welfare of the disfavored group increases by 2% to 18%, depending on the job category. The effect is expected to be larger if workers adjust their entry decisions in response to the new policy.

There are at least two avenues for future research. One is to consider endogenous entry in auctions. In this paper, I simulate a gender-blind hiring policy by taking the number of bidders as given. In practice, the change in policy attracts workers in the disfavored group to enter the auction, and thus the number of bidders or the composition of bidders would change. Therefore, counterfactual estimates with fixed entry may underestimate the effect of the new policy. Another avenue of research is to examine auction and employer heterogeneity. I currently focus on five job categories such as cleaning, moving, and gardening, which require face-to-face interaction with employers. Differential outcomes would be expected if I consider online jobs such as programming, web design, and translation.

References

Agan, Amanda and Sonja Starr (2017) "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *The Quarterly Journal of Economics*, 133 (1), 191–235.

Aigner, Dennis J. and Glen Cain (1977) "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30 (2), 175–187.

Altonji, Joseph G. and Charles R. Pierret (2001) "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116 (1), 313–350.

Antonovics, Kate and Brian G. Knight (2009) "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics*, 91 (1), 163–177.

Anwar, Shamena and Hanming Fang (2006) "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review*, 96 (1), 127–151.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010) "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2 (4), 76–104.

Arnold, David, Will Dobbie, and Crystal S Yang (2018) "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, 133 (4), 1885–1932.

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2019) "Signaling and Employer Learning with Instruments," NBER Working Paper Series 25885, National Bureau of Economic Research.

Aryal, Gaurab, Maria F. Gabrielli, and Quang Vuong (2021) "Semiparametric Estimation of First-Price Auction Models," *Journal of Business & Economic Statistics*, 39 (2), 373–385.

Athey, Susan (2001) "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica*, 69 (4), 861–889.

Athey, Susan, Jonathan Levin, and Enrique Seira (2011) "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions," *The Quarterly Journal of Economics*, 126 (1), 207–257.

Bajari, Patrick and Ali Hortacsu (2003) "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *RAND Journal of Economics*, 34 (2), 329–55.

Becker, Gary S. (1957) *The Economics of Discrimination*: The University of Chicago Press.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan (1998) "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending," *The Review of Economics and Statistics*, 80 (2), 241–250.

Bertrand, Marianne and Sendhil Mullainathan (2004) "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94 (4), 991–1013.

Blevins, Cameron and Lincoln Mullen (2015) "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction," *Digital Humanaties Quarterly*, 9 (3).

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope (2019a) "Inaccurate Statistical Discrimination: An Identification Problem," NBER Working Paper Series 25935, National Bureau of Economic Research.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg (2019b) "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 109 (10), 3395–3436.

Bolton, Gary, Ben Greiner, and Axel Ockenfels (2013) "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management Science*, 59 (2), 265–285.

Boring, Anne (2017) "Gender Biases in Student Evaluations of Teaching," *Journal of Public Economics*, 145, 27–41.

Chan, Jason and Jing Wang (2018) "Hiring Preferences in Online Labor Markets: Evidence of a Female Hiring Bias," *Management Science*, 64 (7), 2973–2994.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer (2021) "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers," *The Review of Economic Studies*, 88 (5), 2210–2238.

Cullen, Zoë B, John Eric Humphries, and Bobak Pakzad-Hurson (2018) "Gender and Sorting in the On-demand Economy," in *ASSA Annual Meeting*.

Cullen, Zoë and Chiara Farronato (2020) "Outsourcing Tasks Online: Matching Supply and Demand on Peer-to-Peer Internet Platforms," *Management Science*, 67 (7), 3985–4003.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri (2020) "Monopsony in Online Labor Markets," *American Economic Review: Insights*, 2 (1), 33–46.

Edelman, Benjamin, Michael Luca, and Dan Svirsky (2017) "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, 9 (2), 1–22.

Farber, Henry S. and Robert Gibbons (1996) "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111 (4), 1007–1047.

Filippas, Apostolos, John J Horton, and Joseph M Golden (2019) "Reputation Inflation," NBER Working Paper Series 25857, National Bureau of Economic Research.

Fisman, Raymond J. and Michael Luca (2016) "Fixing Discrimination in Online Marketplaces," *Harvard Business Review*, 94, 88–95.

Goldin, Claudia and Cecilia Rouse (2000) "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90 (4), 715–741.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong (2000) "Optimal Nonparametric Estimation of First-price Auctions," *Econometrica*, 68 (3), 525–574.

Hong, Yili, Jing Peng, Gordon Burtch, and Ni Huang (2021) "Just DM Me (Politely): Direct Messaging, Politeness, and Hiring Outcomes in Online Labor Markets," *Information Systems Research*, 32 (3), 786–800.

Hong, Yili, Chong (Alex) Wang, and Paul A. Pavlou (2016) "Comparing Open and Sealed Bid Auctions: Evidence from Online Labor Markets," *Information Systems Research*, 27 (1), 49–69.

Hu, Nan, Paul Pavlou, and Jie Zhang (2009) "Overcoming the J-Shaped Distribution of Product Reviews," *Communications of the ACM*, 52, 144–147.

Jaderberg, Max, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu (2015) "Spatial Transformer Networks," in *Proceedings of the Advances in Neural Information Processing Systems*, 2017–2025.

Kanat, Irfan, Yili Hong, and T. S. Raghu (2018) "Surviving in Global Online Labor Markets for IT Services: A Geo-Economic Analysis," *Information Systems Research*, 29 (4), 893–909.

Knowles, John, Nicola Persico, and Petra Todd (2001) "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 109 (1), 203–232.

Kokkodis, Marios and Panagiotis G. Ipeirotis (2016) "Reputation Transferability in Online Labor Markets," *Management Science*, 62 (6), 1687–1706.

Krasnokutskaya, Elena and Katja Seim (2011) "Bid Preference Programs and Participation in Highway Procurement Auctions," *American Economic Review*, 101 (6), 2653–2686.

Krasnokutskaya, Elena, Kyungchul Song, and Xun Tang (2020) "The Role of Quality in Internet Service Markets," *Journal of Political Economy*, 128 (1), 75–117.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton (2012) "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the Advances in Neural Information Processing Systems*, 1097–1105.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015) "Deep Learning," *Nature*, 521 (7553), 436–444.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg (2016) "SSD: Single Shot MultiBox Detector," in *Proceedings of the European Conference on Computer Vision*, 21–37.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz (2018) "Gender Bias in Teaching Evaluations," *Journal of the European Economic Association*, 17 (2), 535–566.

Nardo, Michela, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffman, and Enrico Giovannini (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*: OECD publishing.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort (1996) "Sex Discrimination in Restaurant Hiring: An Audit Study," *The Quarterly Journal of Economics*, 111 (3), 915–941.

Nosko, Chris and Steven Tadelis (2015) "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment," NBER Working Paper Series 20830, National Bureau of Economic Research.

Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman (2015) "Deep Face Recognition," in *Proceedings of the British Machine Vision Conference (BMVC)*, 41.1–41.12.

Phelps, Edmund S. (1972) "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62 (4), 659–661.

Philip, Smith and Chen Cuixian (2018) "Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation," in *Proceedings of the IEEE International Conference on Big Data*, 2564–2571.

Pope, Devin G. and Justin R. Sydnor (2011) "What's in a Picture? Evidence of Discrimination from Prosper.com," *Journal of Human Resources*, 46 (1), 53–92.

Tadelis, Steven (2016) "Reputation and Feedback Systems in Online Platform Markets," *Annual Review of Economics*, 8 (1), 321–340.

Weber, Andrea and Christine Zulehner (2014) "Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail?" *Journal of the European Economic Association*, 12 (2), 492–521.

Wozniak, Abigail (2015) "Discrimination and the Effects of Drug Testing on Black Employment," *The Review of Economics and Statistics*, 97 (3), 548–566.

Yang, Xiao and Zhaoxin Zhang (2013) "Combining Prestige and Relevance Ranking for Personalized Recommendation," in *Proceedings of the ACM International Conference on Information & Knowledge Management*, 1877–1880.

Yoganarasimhan, Hema (2013) "The Value of Reputation in an Online Freelance Marketplace," *Marketing Science*, 32 (6), 860–891.

——— (2016) "Estimation of Beauty Contest Auctions," *Marketing Science*, 35 (1), 27–54.

Appendices

A Indices for Quality and Signal

A.1 Selected Variables

Definition of Variables (All variables are at the user-level)

1. Number of reviews: the total number of reviews user i received, denoted as n_i.
2. Average rating: $b_i = \sum_j r_{ij}/n_i$, where r_{ij} is the rating user i received for job j.
3. Bayesian adjusted average rating: weighted average rating based on the relative number of reviews:

$$\text{Bayesian Adjusted Average Rating} = \frac{m_c \times a_c + n_i \times b_i}{m_c + n_i},$$

   where m_c and a_c is the average number of reviews and average rating for job category c, respectively.
4. Average sentiment score: $\sum_j s_{ij}/n_i$, where s_{ij} is the sentiment score of review text that user i received for job j. Sentiment score is defined as the likelihood that the text has a positive sentiment. If it is below a certain threshold (which is optimally determined in the model), the text is classified to have a negative sentiment. I scale the scores such that the sentiment score is equal to 0 at the threshold. Then, the sentiment score is negative for text classified as having a negative sentiment.
5. Adjusted average sentiment score: $\sum_j s_{ij} \log(w_{ij})/n_i$, where w_{ij} is the number of words in review text that user i received for job j.
6. Completion rate: percentage of completed jobs for user i among assigned jobs to him/her.

Table A.1: Summary Statistics of Selected Variables

	Mean	Std.Dev.	25th	50th	75th	(Min, Max)
Average ratings	4.87	0.33	4.9	5	5	(1, 5)
Bayesian ratings	4.79	0.04	4.8	4.8	4.8	(4.3, 5)
Number of reviews	10.25	26.9	1	3	9	(1, 718)
Average sentiment score	0.64	0.10	0.66	0.67	0.68	(-0.32,0.68)
Adjusted average sentiment score	1.66	0.46	1.44	1.68	1.91	(-1.28,3.34)
Completion rate	0.79	0.22	0.67	0.85	1	(0,1)

Note: The results in this table are based on a sample of cleaning jobs.

Figure A.1: Distribution of Sentiment Scores by Rating

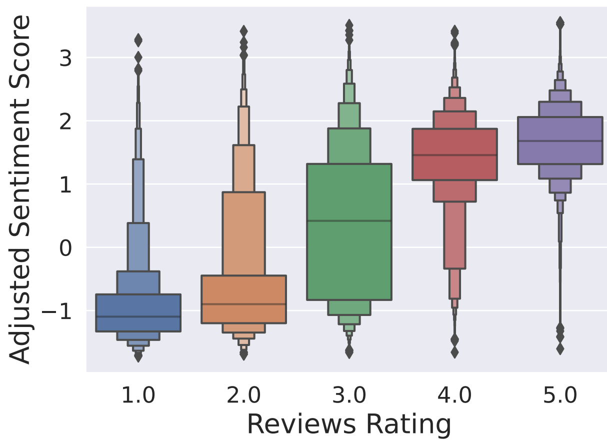


Table A.2: Examples of Sentiment Scores

Review Text	Sentiment Score	Adjusted Sentiment Score
Very impressed with the service and high quality of work. Very friendly guy and would not hesitate to recommend. We will actually be getting Harry back for some more work. Thank you so much Harry.	0.68	2.38
Nice guy. Very punctual. Would happily hire again.	0.67	1.54
Job was not completed & difficulty in communicating.	-0.32	-0.69

A.2 Principal Component Analysis

Table A.3: Factor Loadings (Weights) of the Corresponding Principal Component

	PC1	PC2
Average rating	0.53	-0.06
Bayes adjusted average rating	0.45	0.15
Number of reviews	0.05	0.74
Sentiment score	0.53	-0.15
Adjusted sentiment score	0.47	-0.14
Completion rate	0.12	0.62

Note: PC1 denotes the loadings of the first principal component, and PC2 denotes that of the second principal component. The results in this table are based on a sample of cleaning jobs.

Figure A.2: Scree Plot

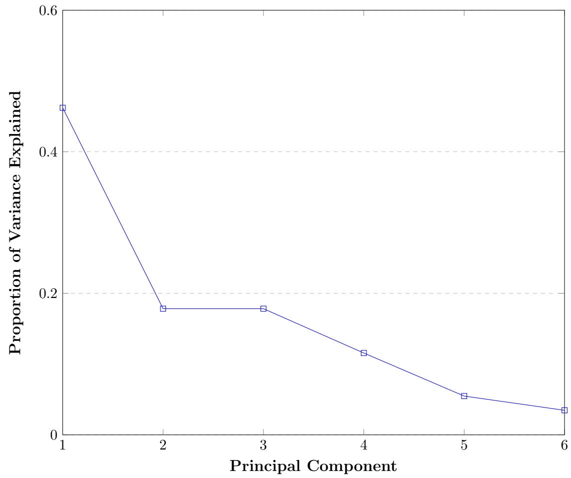
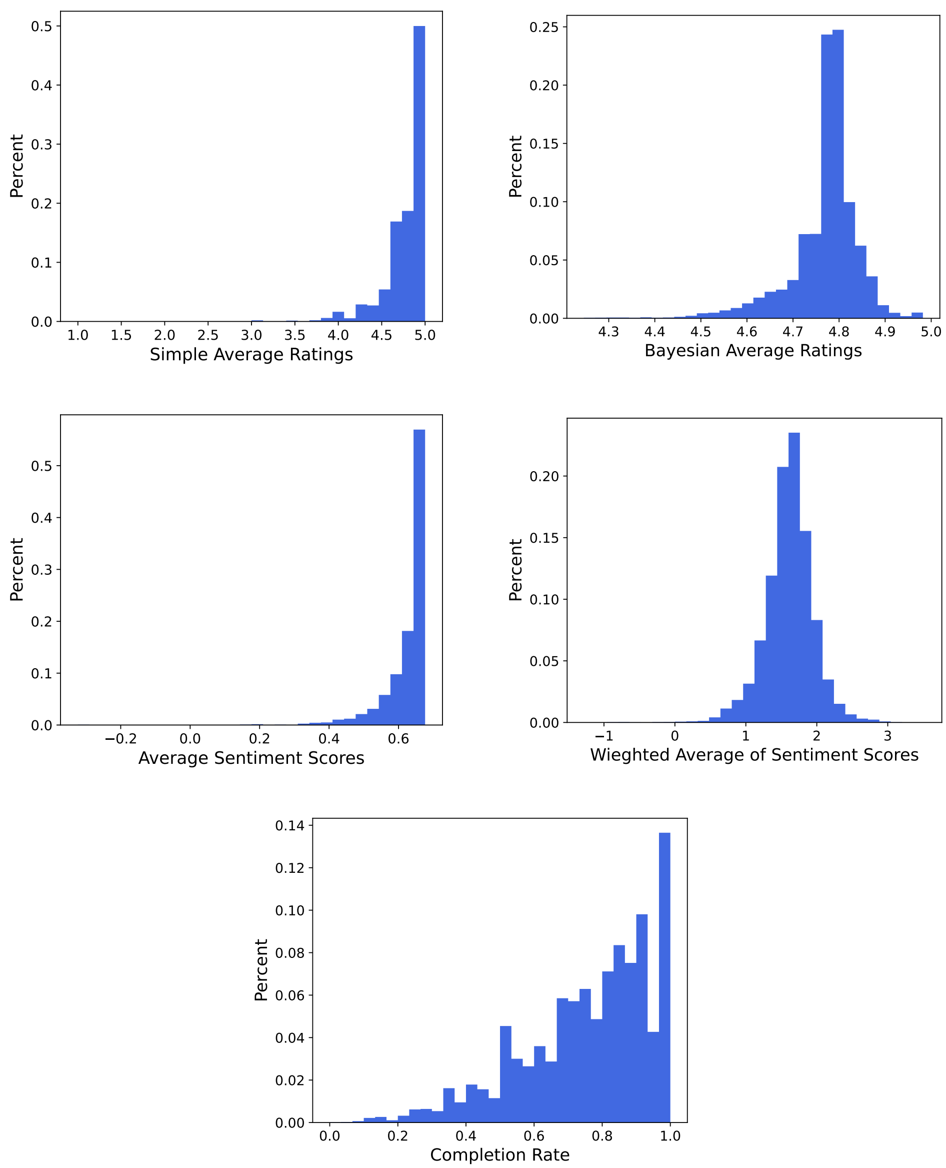


Table A.4: Correlation Matrix

	1	2	3	4	5	6	pc1
1. Average ratings	1	0.67	0.03	0.71	0.53	0.09	0.88
2. Bayesian ratings		1	0.10	0.47	0.38	0.16	0.75
3. Number of reviews			1	0.02	0.01	0.07	0.08
4. Average sentiment score				1	0.71	0.07	0.88
5. Adjusted average sentiment score					1	0.10	0.79
6. Completion rate						1	0.20

Note: The results in this table are based on a sample of cleaning jobs.

Figure A.3: Distributions of Performance Measures



B Discussion

Table B.1: Effects of Worker Experience on Performance

	Average Rating	Sent. Score	Adj. Sent. Score
Employer experience	0.0002	0.00004	0.0001
	(0.0001)	(0.00005)	(0.0002)
ln(Project value)	-0.0468***	-0.0146***	0.0173**
	(0.0063)	(0.0020)	(0.0079)
Worker fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The results in this table are based on a sample of cleaning jobs.

Table B.2: Effects of Quality on Employer Retention

	(1)	(2)
Signal	0.0131***	-0.0008
	(0.0048)	(0.0069)
Quality		0.0160***
		(0.0057)
Female indicator	0.0209	0.0214
	(0.0161)	(0.0161)

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is an indicator that take on a value of 1 if an employer makes another transaction later. The results in this table are based on a sample of cleaning jobs.

C Demand-Side Parameter Estimates

Table C.1: Parameters of Quality / Signal Errors Distribution

	Mean Quality		Variance of Quality		Variance of Signal Errors		Weight on Signal	
	M	F	M	F	M	F	M	F
Cleaning	0.812	0.800	0.102	0.102	0.071	0.078	0.59	0.57
Moving	0.854	0.808	0.105	0.106	0.062	0.062	0.63	0.63
Gardening	0.852	0.823	0.094	0.100	0.063	0.070	0.60	0.59
Repairs	0.859	0.855	0.085	0.106	0.056	0.070	0.60	0.60
Delivery	0.860	0.851	0.094	0.097	0.063	0.057	0.60	0.63

Table C.2: Conditional Logit Estimates

	female (ϕ)	female x signal (ρ)	signal (θ)
Cleaning	0.159***	0.036***	0.122***
Moving	-0.105***	-0.080***	0.122***
Gardening	-0.300***	-0.003	0.122***
Repairs	-0.223***	0.017	0.123***
Delivery	0.027	-0.005	0.085***

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Supply-Side Parameter Estimates

Table D.1: Estimated Parameters of Log-Normal Distribution of Bids

	Coefficient
Mean	
Constant	1.1485
ln(project value)	0.8173
Female	-0.0731
Number of female workers	-0.0149
Number of male workers	0.0124
Number of workers with mid-level signal	-0.0118
Number of workers with high-level signal	-0.0044
Standard deviation of log-bids	
Constant	-1.0378
ln(project value)	0.0125
Female	-0.1024

Note: The results in this table are based on a sample of cleaning jobs.

E Heterogeneous Preferences

Table E.1: Conditional Logit Estimates with Heterogeneous Preferences

	Coefficient	Std. Error
Female indicator for worker (f_j)	1.165***	0.124
Interaction between female indicator for worker and:		
female indicator for employer ($f^e_\ell \times f_j$)	-0.140***	0.030
unknown indicator for employer ($u^e_\ell \times f_j$)	-0.276***	0.041
number of jobs that the employer has assigned ($na_\ell \times f_j$)	0.057***	0.004
log of project value ($\ln(pv_\ell) \times f_j$)	-0.226***	0.028
signal ($s_\ell j \times f_j$)	0.039***	0.013
Signal	0.121***	0.011

Note: The results in this table are based on a sample of cleaning jobs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

F Endogenous Employer Entry

Table F.1: Estimates of Poisson bid arriaval model

	(1) Cleaning	(2) Moving	(3) Gardening	(4) Repairs	(5) Deliveries
Employer gender = female	0.003	0.026***	0.032***	0.055***	0.022
	(0.007)	(0.008)	(0.009)	(0.013)	(0.014)
Employer gender = unknown	-0.015	0.000	0.020*	0.024	0.027
	(0.010)	(0.011)	(0.012)	(0.017)	(0.019)
ln(Number of words in job description)	0.046***	0.015***	0.014**	0.028***	0.021**
	(0.004)	(0.005)	(0.006)	(0.008)	(0.010)
ln(Number of words in employer description)	-0.014***	-0.009***	-0.005	-0.013**	-0.003
	(0.004)	(0.003)	(0.005)	(0.007)	(0.005)
ln(Project value)	0.102***	0.026***	0.061***	0.009	0.050***
	(0.006)	(0.005)	(0.006)	(0.006)	(0.008)
ln(Number of employer reviews)	-0.027***	-0.015*	-0.018	-0.002	-0.004
	(0.009)	(0.009)	(0.011)	(0.014)	(0.010)
ln(Average employer ratings)	0.009***	0.001	0.002	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Constant	0.533***	0.933***	0.795***	0.824***	0.780***
	(0.029)	(0.026)	(0.029)	(0.038)	(0.043)
Observations	31396	26960	20202	11356	8322

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F.2: Estimates of Logit model of worker gender

	(1) Cleaning	(2) Moving	(3) Gardening	(4) Repairs	(5) Deliveries
Employer gender = female	-0.004	-0.047*	0.090***	0.092	0.036
	(0.015)	(0.025)	(0.026)	(0.061)	(0.036)
Employer gender = unknown	-0.001	-0.027	0.001	0.078	0.008
	(0.020)	(0.035)	(0.036)	(0.082)	(0.049)
ln(Number of words in job description)	0.177***	-0.052***	-0.027*	0.033	0.045*
	(0.008)	(0.017)	(0.016)	(0.037)	(0.025)
ln(Number of words in employer description)	-0.047***	-0.011	0.010	-0.003	0.025*
	(0.008)	(0.012)	(0.013)	(0.032)	(0.013)
ln(Project value)	-0.263***	0.044**	0.027	-0.106***	-0.197***
	(0.013)	(0.017)	(0.017)	(0.032)	(0.022)
ln(Number of employer reviews)	0.121***	-0.029	0.004	-0.073	0.044*
	(0.020)	(0.031)	(0.031)	(0.070)	(0.026)
ln(Average employer ratings)	-0.008	-0.003	0.004	-0.008	-0.034***
	(0.006)	(0.009)	(0.009)	(0.021)	(0.011)
Constant	0.975***	-2.083***	-1.907***	-2.676***	-0.804***
	(0.062)	(0.085)	(0.084)	(0.185)	(0.113)
Observations	100262	80355	61171	30533	23916

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$