

How Does Collective Reputation Affect Hiring? Selection and Sorting in an Online Labour Market*

Guo Xu[†]

October 15, 2015

Abstract

Using data from an online labour market where the country of residence is the salient group characteristic, we document a mechanism through which collective reputation perpetuates group inequality. Using an IV strategy, we identify reputational externalities between an employer's first hire and the propensity to contract more workers from the same country. Employers, contingent on their first worker's performance, continue to almost exclusively hire from the same country. This coincides with a positive sorting response: Observing their predecessor's success, workers from the same country disproportionately apply and are of higher quality. Employers, facing better applicants, in turn provide higher ratings.

JEL codes: J7, J16, M51

*I am indebted to Oriana Bandiera and Robin Burgess for their supervision and continued guidance. I thank Esteban Aucejo, Michael Boehm, Jon DeQuidt, Thiemo Fetzer, Maitreesh Ghatak, Rocco Macchiavello, Dennis Novy, Heather Sarsons, Pedro Souza, Chris Stanton and Sandra Sequeira for their valuable comments. All errors remain my own. This paper supersedes a previous version under the title "The Value of Reputation: Evidence from Outsourcing".

[†]London School of Economics, Department of Economics and STICERD, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: g.xu2@lse.ac.uk)

1 Introduction

Technological advances like broadband and cloud storage have greatly facilitated remote work and outsourcing. In recent years, online labour markets such as oDesk, Guru, Elance, Rentacoder, Freelancer or Amazon MTurk have enabled businesses to conveniently tap into a global pool of IT workers. The potential gains from trade in these global markets are huge, as employers and workers can match across great distance in a near “zero gravity” environment, taking advantage of large wage differentials across regions and countries.

While varied in design, a common characteristic shared across all virtual markets is incomplete information: Participants on these markets never physically meet and often use aliases. Skills stated are hard to verify. In these settings, reputation mechanisms have proven successful: By allowing participants to publicly rate each other and provide feedback, dynamic incentives are introduced to alleviate moral hazard that is otherwise pervasive in environments of imperfect information (Stanton and Thomas, 2014; Cabral, 2012; Ghani et al., 2014).

In addition to measures of individual reputation, online markets typically provide group-specific information such as country of residence or gender. These collective traits can provide added information to employers (Phelps, 1972; Altonji and Pierret, 2001), but also enable workers to sort along a dimension which, ex-ante, may be unrelated to worker productivity. The possibility for workers to coordinate on a collective trait can, in theory, give rise to inefficient “self-fulfilling stereotypes” (Coate and Loury, 1993; Tirole, 1996; Moro and Norman, 2004). While the role of individual reputation in mitigating moral hazard is well documented, much less is empirically known about the role of collective reputation¹ in facilitating or inhibiting market transactions.

How does collective reputation affect hiring and selection into jobs? This paper delves into the role of collective reputation in a global online labour market, where the country of residence is the salient group characteristic. Using detailed hiring data on all public transactions, we first examine reputational externalities between an employer’s very first hire and the propensity to contract more workers from the same country (the “first hire” country). In absence of collective reputation, the individual rating given to the first worker should not extend to others. After empirically establishing the presence of reputational externalities, we examine changes in the applicant composition and the employers’ final choices to shed light on the underlying mechanisms at play.

¹Following the definition of individual reputation (Mailath and Samuelson, 2006; Cabral, 2005), collective reputation is “the situation when agents believe another particular agent to be something”, depending on an observable group-specific (collective) trait. Collective reputation is hence typically discussed within as (dynamic) statistical discrimination framework (Blume, 2006; Kim and Loury, 2014).

A recurring challenge in any hiring setting is the lack of experimental sources of variation that enables us to rule out alternative, observationally equivalent explanations. To obtain causal estimates, we introduce a novel instrumental variable (IV) strategy, where we exploit plausibly exogenous variation in vertical traits within the first applicant pool to predict the actual hire. The availability of rich data on the applicant pool level for over 25,000 employers is key to the empirical strategy, allowing us to take an unusually close look at the mechanisms through which collective reputation emerges and translates into market outcomes.

We present three key findings: First, we document the persistence of the “first hire” country. Our instrumental variable estimate suggests that employers are 3.1% points more likely to continue to hire from their “first hire” country than from other countries in their first applicant pool (mean: 10%). This observed persistence is primarily driven by the positive first rating given, consistent with employer learning and a reputational externality. Second, we document a strong and positive supply-side sorting response: Workers from the “first hire” country disproportionately apply after observing the first hire’s successful outcome: Following an earlier positive rating, workers from the same country are 5.1% points more likely to apply than in response to a negative rating (mean: 27.3%). These workers also tend to be of higher quality, as measured by their previous rating and experience. Finally, the sorting response amplifies the positive first hiring: Employers, faced with more high quality workers from the successful “first hire” country, are in turn 11.1% points more likely to continue providing top ratings for their later hires (mean: 90%). Overall, the results provide empirical evidence for the role of coordination in creating “self-fulfilling stereotypes”: An employer’s first rating given to a worker with an observable collective trait serves as a group-specific signal upon which later workers coordinate. Depending on the first experience, good workers attract more good workers from the same country and vice versa. The provision of collective traits hence perpetuates initial group inequalities by creating a “herding effect”. Most importantly, the results suggest that collective traits such as country of residence - even if uncorrelated to *any* economic fundamentals ex-ante - can persistently shape the way workers sort and apply to jobs.

We rule out competing explanations that appear observationally equivalent to collective reputation. To ensure that the observed persistence is indeed coming from a reputational externality across workers, we excluded rehires throughout the analysis. Other than collective reputation, the sorting response could however also reflect learning about the country-specific match productivity, revealed through the first hire’s country and rating. In this case, we expect the effect to decline once we control for bilateral confounds or remove language related tasks (e.g. translation). To ensure

that the persistence is not driven by same-country referral networks, we also explicitly omit private transactions that are exclusive for invited workers, focusing only on public transactions which are competitive and open to all. A remaining concern is that employers “signal” their preference for certain groups by providing a high rating in their first hire. If workers indeed sort in response to the employer’s revealed taste for the collective trait, we expect the sorting effect to be even stronger when the first worker appears particularly unattractive along vertical traits, such as the individual rating or experience. We find no evidence for these alternative explanations and provide further robustness checks to support our main findings.

1.1 Related literature and implications

The results contribute to several strands of literature: First, our paper adds to the emerging literature on online labour markets (Thomas, 2012; Cabral, 2012). While a large body of literature has documented the importance of individual reputation, as measured by online ratings and feedback (Resnick et al., 2006; Cabral and Hortacsu, 2006; Hortasu et al., 2009; Moreno and Terwiesch, 2013; Agrawal et al., 2013; Pallais, 2014), few papers have examined collective reputation. The closest papers to ours are Ghani et al. (2014) and Nosko and Tadelis (2015). Ghani et al. (2014) provide evidence from oDesk that ethnic diaspora Indians are more likely to initially outsource to India, and more likely to continue hiring from the first country of hire when the experience was positive. We confirm the correlations and provide causal evidence for a sorting response through which the country of hire may appear persistent. Nosko and Tadelis (2015) provide experimental evidence from eBay that buyers overly rely on the first transaction to learn about platform quality, creating a reputational externality. While the persistence of the first transaction appears observationally equivalent, we provide evidence for another mechanism on a two-sided market. In the labour market studied, the persistence is driven by the supply-side response of the workers (or sellers, in their terminology) who sort based on the rating and country of an employer’s first hire.

Second, this paper relates to the personnel economics literature on recruitment and selection. While a large body of literature has documented moral hazard and policies to motivate workers on the job, relatively little is known about how to attract the “right” workers in the first place (Lazear and Oyer, 2007; Paul and Scott, 2011). In contrast to the literature that examines selection by varying financial incentives (Bo et al., 2013; Deserranno, 2014; Ashraf et al., 2014), we focus on the impact of reputational externalities in attracting or deterring applicants.

More broadly, the contribution adds value by documenting statistical discrimination on the labour market (Altonji and Pierret, 2001; Bertrand and Mullainathan, 2004; Riach and Rich, 2006; Mill and Stein, 2015). Our results, in particular, suggest

that the provision of collective traits such as gender, race or nationality in applications (e.g. on CVs) may further amplify existing group inequalities as workers, anticipating discrimination, refrain from applying altogether. The sorting response documented, in particular, sheds light on a mechanism that could partly explain occupational sorting and persistent labour market differences along collective traits such as gender or ethnicities (Glover et al., 2015; Bertrand, 2011; Botticini and Eckstein, 2013).

The remainder of this paper is organized as follows: In Section 2, we introduce the empirical context and the data. Section 3 discusses the identification strategy. Section 4 presents the main results on the persistence of the first hire. Section 5 discusses the mechanisms by turning the focus to the applicant pool composition and final choices. Section 6 provides robustness checks and Section 7 concludes.

2 Context and market structure

2.1 Empirical context

The empirical evidence is from rentacoder.com, one of the largest first-generation virtual markets for outsourcing. In recent years, virtual markets have become increasingly attractive for scholars to study, mainly due to the availability of large datasets and the presence of information asymmetries (Cabral, 2012; Thomas, 2012). The availability of the applicant pools in this context is key for the implementation of our instrumental variables strategy.

In the market studied, employers contract out service jobs to workers who compete by bidding a fixed wage and revealing observable measures of quality.² The market chosen is appealing for several reasons: As one of the largest markets, it is representative of a range of competing outsourcing markets. Up to its acquisition by a competitor in 2012, the market under study was one of the five largest virtual markets for outsourcing, with 1.3 million tasks posted by employers and cumulative worker earnings of \$139 million. Typical tasks on the platform are data entry tasks, small programming tasks (e.g. creating websites) or simple design tasks.

The online platform offers two types of hiring for task assignment: Public market transactions, which are competitive and open to all workers, and private transactions, which restrict the set of potential workers to those invited by the employer. These private transactions almost exclusively comprise rehires or referrals of previous workers, where individual reputation is likely to be more important (Stanton and Thomas, 2014).

²This type of mechanism is often also referred to as multi-attribute auctions, as sellers do not only compete on price but several dimensions (e.g. rating and experience). In contrast to a *scoring auction* (Asker and Cantillon, 2008, 2010), the scoring rule of the employer is unknown to the bidding workers.

Since we focus on collective reputation, however, our main focus is on transactions made on the large and competitive public market.

With a public online market where entry is free and unregulated, information asymmetries are particularly pervasive (Agrawal et al., 2013; Moreno and Terwiesch, 2013): Anyone can sign up and bid as a worker and there is no minimum skill requirement, rendering types unobservable to employers. To sustain transactions given these information asymmetries, the market runs a reputation management system where participants rate each other after completion of tasks.

A typical public market transaction can be described as follows: Employers post jobs requests on the market along detailed specifications about the deliverables, contract type (e.g. fixed or hourly paid) and time frame. The task is then reviewed by the market operator. When approved, it is listed on the market and workers can bid for the job by submitting a wage for which they would be willing to deliver it. Workers do not observe other bidding workers, so there is limited scope for strategic interaction among them, a critical feature for our empirical strategy. The employer observes a list of all applicant workers along a narrow set of variables and chooses his preferred worker (the applicant pool). The set of well-defined variables comprise the bid wage, a rating for past quality and the number of tasks completed. The salient collective trait shown is the country of residence, as indicated by a country flag and the location of the worker.³ The employer can obtain additional information by viewing the complete profile of each worker. Figure 1 presents a typical list of bids.

[Figure 1 here]

Once an employer chooses a worker, the payment agreed upon is transferred from the employer to an escrow account to mitigate moral hazard on the employer's side. The worker then begins with the job, which can end in two ways: Once the worker reports the task complete within the time agreed upon, the worker is asked to upload the deliverables which will then be checked by the employer. If the employer is satisfied with the results, the task is reported complete in which case the payment (minus a commission to the market provider) is transferred to the worker. If unsatisfied, the employer may ask the worker to revise the work or cancel the task in which case the money is returned to the employer and the worker receives nothing. Similarly, when the worker failed to deliver on time, the employer may either extend the time frame or cancel the task. After completion or cancellation of the task, employers and workers are given the opportunity to rate each other on an integer scale from 1 (lowest) to 10 (highest). The rating is only revealed once both have submitted their rating or if two

³While the display of sub-national locations may suggest a localized role of collective reputation (e.g. city-level reputation for outsourcing hubs Gurgaon vs. Bangalore in India), the within country variation of locations is too small to allow for a conclusive test for sub-national collective reputation.

weeks have passed.⁴ Once revealed, the rating is visible to anyone on the market. The employers' hiring history is hence visible to potential applicant workers in later hires.

While entry is free and unregulated, re-entry is relatively costly. Employers and workers are required to provide their full contact and bank details to ensure payment. The contacts are verified by phone calls and double or fake accounts, when detected, are suspended. Finally, off-site communication is discouraged as they are not legally binding and will not be taken into account when tasks fail and employers report the case to the market provider for arbitration.

2.2 Data and descriptive statistics

We collected data for all public market transactions between 2001 and 2012, covering the entire period of market operation. The resulting core dataset contains 271,783 bids made by 60,083 workers for the respective first job of each of the 25,652 employers.

The online market is international, with the main direction of contracting from high income to low and medium income countries. Almost half of the employers are based in the United States, followed by the United Kingdom, Canada, Austria and Germany. Workers are primarily based in India, the United States, Romania and Pakistan.

The main measure of individual reputation is the average rating assigned to a worker. While ratings range between 1 (lowest) to 10 (highest), the majority of workers receive the highest ratings. This highly skewed distribution of ratings is common across virtual markets, as documented by Nosko and Tadelis (2015) and Dellarocas and Wood (2008). In order to obtain informative variation, we only focus on the top margin, examining the share of workers who have obtained a top rating of 10/10. This corresponds to about 70% of the cases.⁵

Using the binary measure for top ratings, we break down the average distribution of ratings by countries to examine cross-country differences in the individual reputation score. In Figure 2, we plot the average share of top rated workers for the largest countries, ranking them in descending order. The plot indicates substantial variation in average ratings across country: More than 85% of workers from Argentina, Germany and Bulgaria obtained top ratings, while the share of top ratings is almost 10% points lower for Bangladesh, Sri Lanka and India. These differences are jointly significant: The country fixed effects are jointly significant, even when controlling for employer and all observable individual worker characteristics (See Appendix A1)

[Figure 2 here]

⁴This is to avoid that poor workers, for example, do not strategically hide their anticipated bad rating by not submitting their own bid.

⁵In later robustness checks, we confirm the robustness of the results to the continuous measure as well as alternative dummies, e.g. based on above median rating (See Section 6).

The collective trait, hence, appears to contain added information to predict the performance of workers. The presence of differences in group-reputation alone, however, need not indicate collective reputation: For one, these differences could simply be taste-based. Some countries are more popular than others. Alternatively, the group differences could be an endogenous equilibrium outcome as workers sort across employers and tasks. Put differently, these average group differences do not inform us to whether these reflect actual (ex-ante) differences or are themselves a result, for example of self-fulfilling stereotypes.

To make progress, our empirical strategy explicitly focuses on the employers' very first hire on the market. This allows us to examine how the first exposure endogenously shapes subsequent hiring outcomes. More specifically, we first exploit a quasi-random source of initial assignment of employers to workers from various countries. We then use this first country-specific exposure to see how shocks to individual reputation extend to others to subsequently generate differences in collective hiring outcomes. We discuss the empirical strategy in the next section but before provide additional descriptive statistics about the first applicant pool.

The data allows us to examine composition changes among applicants. For each of the jobs, we observe the applicant pools with the characteristics of the competing applicant workers. These characteristics range from the wage bid submitted, to a rating (1 low to 10 high), the number of jobs completed, a measure of experience, to the country of residence. Table 1 summarizes the characteristics of the first applicant pools: For the pooled sample of all employers, the average number of bids in the first applicant pool is about 10.6, with bidding workers from an average of 5.7 countries. Furthermore, employers appear to be responsive to price and measures of reputation: The average bid price is \$323.3, but the chosen bid is only \$170.7 on average. Employers are also more likely to choose highly rated workers (an average share of 24% vs. 31% chosen), and workers with more experience on the market (average of 36.3 previous jobs vs. 59.4 chosen). These patterns are in line with studies of hiring determinants on online markets (Ghani et al., 2014; Moreno and Terwiesch, 2013).

[Table 1 here]

Only about a third of the employers continue to hire beyond the first hire. While unusual for physical labour markets, the high numbers of entry and exit are typical features of online markets (Nosko and Tadelis, 2015). The unconditional means are correspondingly low, with buyers on average hiring only 1.34 times beyond the first hire. The unconditional mean for the average subsequent amount paid per job is \$29.15 and the total volume is \$93.08.

3 Empirical model and identification

3.1 Regression model

We first test for reputational externalities between an employer’s very first hire and the propensity to contract workers from the same country in the later hires. In absence of collective reputation, the first hiring decision does not extend to workers sharing the same collective trait.

With the worker’s country as the salient collective trait, we collapse the data to the employer-worker country level. This allows us to compare employer-specific differences in hiring outcomes between countries. Since the majority of employers only hire once or twice, the simplification comes at no major empirical loss. For each of the $i = 1 \dots N$ employers, we compute the bilateral hiring intensity with respect to the $j = 1 \dots M_i$ worker countries from the first applicant pool. We then compare if an employer’s likelihood to hire from one of the M_i countries depends on the first country choice. Employer i ’s overall hiring intensity vis-à-vis worker country j then is:

$$y_{ij} = \beta \cdot \widehat{first_hire}_{ij}(z_{ij}) + c_i + d_j + \mathbf{x}'_{ij}\gamma + \epsilon_{ij} \quad (1)$$

where y_{ij} is a measure of the frequency of hiring between employer i and country j . To capture the extensive margin in hiring, we first use a dummy where $y_{ij} = 1$ if employer i hires any workers from country j for his second or later jobs. We also use the number of hires and average wages paid to workers to examine potential impacts on the intensive margin. The independent variable of interest is $first_hire_{ij}$, a bilateral dummy that is 1 if the employer’s first hire was from country j . In presence of a reputational externality, we would expect $H_0 : \beta \neq 0$.

The main challenge in this non-experimental setting is to rule out confounds that may bias the estimation. In particular, there are three main sources of bias that the identification strategy must address: First, there may be sorting among workers and employers. If only workers from j select into applicant pools of large employers (who are also more likely to remain on the market), for example, employers are mechanically more likely to hire from country j both in the first and later hires. We address selectivity by comparing only the set of countries that select into the same employer’s first applicant pool, implemented using the employer fixed effects c_i .

Second, there may exist time-invariant differences across worker country j that could drive the differential probability of hiring across all hires: The likelihood of a worker from country j submitting the lowest bid, for example, increases mechanically with the size (in terms of workers on the market) of the country. Similarly, if workers from country j consistently submit more competitive applications, for example due to

cross-country wage differences, a spurious persistence may be created by unobserved cross-country heterogeneity. As before, we address this econometric concern by introducing worker country fixed effects d_j that absorb cross-country level differences. In robustness checks, we also allow these country fixed effects to vary over time by using country-year and country-year-month fixed effects.⁶

Finally, a spurious effect may also appear in presence of bilateral employer-worker country confounders, inducing a correlation between $first_hire_{ij}$ and the error term ϵ_{ij} . If an employer simply has a preference for country j , the taste-based persistence in hiring will be observationally equivalent to a reputational externality. We address this issue by proposing a novel IV.

3.2 Predicting the first hire: IV

We use an instrumental variable strategy to generate exogenous variation in the country of first exposure. The intuition for the instrumental variable strategy can be described in three steps: First, since employers are *ceteris paribus* more willing to select workers with cheaper bids, we can predict the first hire country using the variation in wage bids among workers in the first applicant pool (relevance). Second, given the institutional features, the variation in wage bids within applicant pools is noisy, especially once absorbing cross-country heterogeneity through the worker fixed effects. Third, this noisy variation in the first applicant pool is unlikely to be correlated with the variation in wage bids in future applicant pools, other than through the first hire (exclusion).

More specifically, we exploit the residual variation in the realized distribution of wage bids to instrument for the first hire country. To see how the instrument can be implemented, let w_{ijk} denote the wage bids of workers $k = 1 \dots K_i$ from country j in employer i 's first applicant pool. We simply create a bilateral dummy that is 1 if a worker from country j submitted the cheapest bid in employer i 's first applicant pool:

$$z_{ij} = \mathbf{1}[w_{ij.} = \min(w_{i..})] \quad (2)$$

More generally, our instrumental variable strategy predicts the probability of a worker from country j being hired as a function of the first applicant pool's *realized* distribution of bids, with cross-country differences partialled out. In robustness checks, we also create IVs using other functions of the distribution of bids (e.g. the average deviation from the mean wage bid) but focus on the order statistic, the simplest instrument for the purpose of exposition. Figure 3 summarizes the final source of variation

⁶The results are robust to country-year and country-year-month fixed effects. The results also remain robust when omitting the first five nascent years of the online market where large changes (e.g. in terms of country composition) may have occurred (Table A3).

used to construct the instrument. The figure shows the distribution of (standardized) bids after partialling out both cross-applicant pool and cross-worker country differences. The residual variation in bids remains large. We exploit this arguably idiosyncratic source of variation to predict the country of first hire.

[Figure 3 here]

As described, the exclusion restriction in this context is that future (residual) realizations of the distribution of bids are uncorrelated with the first realization. In other words, the exclusion restriction maintains that the fact an employer first hired a worker from country j *just* because workers from other countries were more expensive is uncorrelated with the relative competitiveness among bid countries in later hirings, other than through the first hire. We argue that the exclusion restriction is reasonable in this empirical context. The market structure creates a high degree of uncertainty: Workers neither know the number, bids and type of other workers when applying for a job. Similarly, the nature of tasks varies across hiring pools, so there exists uncertainty about the actual costs required to deliver the project. Since the instrument uses the relative variation in wage bids *within* a given pool, we argue that this variation is very likely to be uncorrelated with the subsequent relative variation created by a very different set of workers, employers and application pool.

To complete the discussion of the empirical specification, we also include a vector \mathbf{x}_{ij} that controls for bilateral country-level confounders which may determine both first and subsequent hiring. In our context, the main measures are shared common languages, time zone differences and the geographic distance between the employer and worker country.⁷ Finally, the standard errors are clustered at the employer-level.

4 Results

4.1 Reputational Externalities

The main results are reported in Table 2. Column 1 and 2 report the OLS estimates. As expected, the inclusion of worker country fixed effects accounts for the upward bias driven by cross-country level differences. Column 3 reports the instrumental variable estimate: Compared to countries from which workers applied but were not chosen from, employers are 3.2% points more likely to continue hiring another worker from the first hire country. The point estimate is nearly unchanged when adding the common language, time zone differences and (log) distance as bilateral controls, adding support

⁷The bilateral data is drawn from the CEPII's Gravity Dataset, see Mayer and Zignago (2011).

to the validity of the instrumental variable strategy (Column 4). In Column 5 and 6, we report the results by replacing the binary outcome with the subsequent number of hires (Column 5) and the total wage payments (Column 6). Employers are more likely to hire and pay higher wages to countries of first hire, compared to countries from which workers applied initially but were not chosen from. The first stage of the two-stage estimation is strong: Employers are 27.2% points more likely to hire from a given country if one of the country’s worker submitted the lowest (residual) wage bid. The first-stage conveniently passes conventional tests of weak instruments, with a (Kleibergen-Paap) F-statistic of 5214.54.

[Table 2 here]

The estimates are economically significant: With the mean of hiring continuation at 10%, the increase by 3.1% points reflects a sizeable increase of 31% when evaluated at the mean of the dependent variable. For the total amount paid, the increase is even larger and at 56% when evaluated at the mean of \$16.35. Since the sample also includes a large number of employers who do not continue after the first hire, the coefficients are even larger when conditioning on the set of employers who at least hired twice (Table A3, Column 8). Since continuation itself is likely to be endogenous, however, our interpretation relies on the unconditional, lower estimates. Notice also that the effect does not include the rehiring of the same workers (as in first hire), hence capturing solely the externality generated by the shared group trait, the country of residence.

There are two potential reasons for why the instrumental variable estimates are larger than the OLS estimates. First, the results may be driven by measurement error, for example due to misreporting in the country of residence. Second, if the treatment effect is heterogeneous, the instrumental variable estimate will reflect a local average treatment effect (LATE) of those employers who are particularly responsive to variations in prices, especially to the cheapest price. In Section 6.2, we examine the complier population and the robustness of the estimates upon alternative construction of the instrument further (Table A9). For the main result, however, it is assuring that the OLS point estimate - which we suspect to be upward biased - is even smaller than the point estimate of the instrumental variable procedure.

To explore how persistent the reputational externality is, we focus on a subset of employers who continue to hire beyond the sixth hire (5% of all employers) and estimate the probability of hiring from the first hire country at each hire up to the sixth. We use the same specification as for the main result (4.1). The estimates are reported in Figure 4. The country of hire in the first job has a large impact on the probability of hiring from the same country in the second job. Conditioning on the set of employers who continue to hire, the point estimate is about twice the size of the overall unconditional first country of hire effect in Table 2. While the coefficient magnitude drops in the third

hire, we observe an uptick and statistically significant effects for the fourth and fifth hire, despite the imprecise estimates due to a substantially smaller sample. The results provide evidence for a persistent effect of the reputational externality that tapers off at the sixth hire. Although the duration of persistence appears short, the number of rehires is generally low on the market. About 90% of all employers hire only up to four times, which is in the range where the persistence remains large.

[Figure 4 here]

4.2 Treatment heterogeneity

The persistence identified in Section 4.1 suggests that the country-specific first exposure predicts later hiring patterns. In presence of a reputational externality, we also expect employers to learn and update their beliefs about the first worker’s country depending on his or her performance (Altonji and Pierret, 2001). We therefore expect the persistence to be even stronger when differentiating by the first job’s outcome.

To investigate this further, we extend the main regression model (1) by allowing for treatment heterogeneity. Employer i ’s hiring intensity vis-à-vis country j then is:

$$y_{ij} = \beta_0 \cdot \widehat{first}_{ij}(z_{ij}) + \beta_1 \cdot \widehat{first}_{ij}(z_{ij}) \cdot s_{ij} + c_i + d_j + \mathbf{x}'_{ij}\gamma + \epsilon_{ij} \quad (3)$$

We interact the first hire dummy \widehat{first}_{ij} with another dummy s_{ij} denoting the success of the first job. We define success ($s_{ij} = 1$) as having obtained a top rating of 10/10, which corresponds to 70% of the cases. With most of the jobs rated as a success, our margin of comparison here is hence to compare top rated workers to the rest.⁸ While the success of the first job s_{ij} is potentially endogenous, we argue that the interaction $\widehat{first}_{ij} \times s_{ij}$ is more likely to be quasi-random. The specification can also be interpreted as a difference-in-differences: intuitively, we now compare the subsequent hiring outcome of workers from two first hire countries, with the added difference that one worker received a high rating and the other a low rating.

Table 3 summarizes the results, now only reporting the instrumental variable estimates. In line with a reputational externality, the results suggest that the first hire effect is solely driven by the job’s outcome, as measured by the rating given. In Column 2, the interaction between the first hire and the top rating is positive and significant: As expected, the point estimate increases in contrast to the previous first hire effect

⁸In robustness checks, we also use different definitions of success, such as stricter definitions (e.g. rating of at least 2/10...7/10, 8/10... 10/10) or a more continuous measure (See Table A5)

reported in Section 4.1. Only conditional on success does the first hire translate into future hires.

[Table 3 here]

One potential concern is that the interaction is not capturing the actual rating given for the first job but the individual quality of the hired worker: If chosen workers with a high rating are more likely to succeed (and hence rated highly), the key interaction could be spuriously driven by the individual quality of the first worker. To address this concern, we control for the observable measures of individual reputation: In Column 3, we interact the first hire country with the (ex-ante) individual rating of the chosen worker at time of hire. Reassuringly, the interaction term is insignificant with a substantially smaller point estimate. In Column 4, we use experience as the second measure of individual quality but find a similarly insignificant result.⁹ In Column 5, we add all three interactions: The results confirm that it is indeed the experience with the first worker, as measured by the first rating given, that is driving the persistence.

In Column 6 and 7, we replace the dummy dependent variable with continuous variables. The results remain comparable: Conditional on success in the first hire, the times traded with a given country are significantly higher (Column 6). We also find the same result for the cumulative payments (Column 7).

To further explore how the persistence varies with the first rating given, we allow for a finer breakdown by ratings. If the heterogenous effect indeed reflects a reputational externality and employer learning, we expect the updating to be stronger the clearer the first signal, as captured by the rating, is. We therefore break down the rating into five groups: Worst rating (Below 4), bad rating (4-6), neutral rating (7-8), good rating (9) and best rating (10).¹⁰ We estimate the effects relative to the neutral group.

[Figure 5 here]

The results are summarized in Figure 5, where the effects are plotted by the constructed bins. The effect is symmetric: Employers are more (less) likely to continue hiring from the first country of hire if the first rating was positive (negative) relative to a neutral rating. This is consistent with learning, as employers react accordingly to both positive and negative signals. While the estimates for the negative ratings are somewhat noisier, this is due to the small number of bad ratings observed.

⁹Experience is a dummy than is 1 if the worker's number of jobs completed is above median (17).

¹⁰The bins are chosen to reflect the classifications provided by the market and to ensure sufficient numbers of observations by bin. As discussed, rating distributions are typically left-skewed, with most of the mass concentrated among the highest ratings. The results, however, do not change when we use alternative groupings (See Table A5).

5 Discussion and mechanisms

In the previous section, we found causal evidence consistent with a reputational externality: Shocks to the past performance of a worker extend over to the entire group. We now explore the underlying mechanisms of this persistence by exploring the role of coordination and worker sorting.

For a positive shock to collective reputation to generate the strong observed persistence, we require the shock to endogenously shift the composition of applicants so as to “rationalize” the employer’s belief about a given country’s average worker quality. In theory, we require a sufficiently strong strategic complementarity between the employer’s positive hiring response and the incentives of higher ability workers from the first hire country to apply. We test this condition by studying if the employer’s first exposure is indeed accompanied by a positive supply-side sorting response: More and better applicants apply in response to a positive shock to collective reputation. We conclude this section by switching to the employer’s side, comparing the final choices employers make. This allows us to assess how shifts in the applicant composition coincide with changes in the final choices.

5.1 Sorting and applicant composition

We use the same specification as before in (4), but now turning to the worker’s side by examining the applicant pool. The results are reported in Table 4.

We first examine if the propensity of workers to apply in later jobs relates to the employer’s first exposure (Column 1-2). The results show that workers are more likely to apply to employers who initially hired from their country. Comparing among countries in the first hire, workers from the same country as the employer’s first hire are 3% points more likely to apply (Column 1). Again, the effect increases when we allow for treatment heterogeneity (Column 2), suggesting that the effect is primarily driven by the first positive rating given. In Column 3 and 4, we replace the binary measure with the number of applicants.¹¹ The results remain similar and confirm the effect also in the intensive margin.

[Table 4]

In Column 5-8, we explore whether the composition shift among countries is also accompanied by a change in the quality of the applicants, as measured by the individual rating and past number of jobs completed. The results suggest that the sorting response is positive: Following a positive first exposure to a country, workers from the same

¹¹Since we hold constant the numbers of subsequent hires using the employer fixed effects, there is no need to deflate these numbers by the overall numbers of applicants or jobs.

country do not only increasingly apply but are also of higher quality. Compared to a negative first exposure, the number of top rated applicants per country is 0.63 higher after a positive first exposure. Evaluated against the mean number of 0.6 per country, this increase is large.

Interestingly, the results for experience also show that workers tend to sort along the (ex-ante) experience level of the first hired worker. An experienced worker in the first hire increases the average number of experienced workers in later applicant pools by 0.3. This suggests that the employer’s first choice may also signal the preferences for certain type of workers. We discuss this alternative mechanism in Section 6. Once again, the point estimates are larger when we condition on the set of employers who do not exit. As exit itself is endogenous, however, we prefer the unconditional specification.

Overall, the results are consistent with the role of the first hire in solving a co-ordination problem: As shocks to collective reputation induce positive sorting, the employer’s first impression is likely to be subsequently confirmed.

5.2 Comparing chosen workers and ratings

How does the composition change in the applicant pool affect the final choice? In order to make that comparison, and in contrast to previous specifications, we now condition on the endogenous set of final hires. For each of the i employers, we enumerate the t subsequent hires separately. The hires correspond to workers, indexed j and from country c . We estimate:

$$y_{ijct} = c_i + \beta \cdot \widehat{first}_{ij}(z_{ij}) + \beta_1 \cdot \widehat{first}_{ij}(z_{ij}) \times s_{ij} + d_c + \epsilon_{ijct} \quad (4)$$

As before, we include employer fixed effects c_i and worker country fixed effects d_c . The identifying variation is therefore comparable to the previous specification, but now only among the final hires (See Section 3.1). We also use the same instrumental variable strategy. The standard errors remain clustered at the employer-level.

The results are reported in Table 5. Among the set of hired workers, those from a first hire country with a positive first rating are more likely to receive a top rating themselves. Compared to a negative first rating, a positive first rating increases the probability of receiving a top rating by 11.1% (Column 2). This effect appears to be driven by the higher quality of selected workers from the first hire country (Column 6). Followed by a positive first rating, later hires from the country are 12.5% points more likely to have a top rating. In terms of wage and experience, however, the final choices are comparable (Column 4 and 8).

[Table 5]

The results confirm the presence of strong strategic complementarities that give rise to a “herding effect”: Higher ability workers, observing an employer’s first positive experience with workers from their country, disproportionately apply as they anticipate a higher probability of being chosen. The increased entry of workers from the first hire’s country, however, drives down their “country premium” so that workers from the first hire country do not receive higher average payments.¹² Faced with a larger pool of better applicants from successful first hire countries, however, employers are more likely to select a higher rated worker. These high reputation workers in turn perform better, further amplifying the initial public group-specific signal, thus endogenously creating persistent differences between countries. This closes the loop that gives rise to self-reinforcing differences in application rates across groups.

6 Alternative mechanisms and Robustness

6.1 Alternative mechanisms

We consider several competing mechanisms that may appear observationally equivalent to collective reputation. We first explore an alternative channel through which the applicant composition change could be driven by. While employers may not change their selection rule, they could strategically delay their job postings to local hours where workers from the first hire country are most likely to be awake. Akin to many online markets, workers respond timely to job posts as these are displayed chronologically. We explore whether employers change their time of job posting in order to increase the number of applications from the first worker country. The test is implemented by constructing a dummy for the hours when workers from the given country are most likely to be awake. The window of active local hours is defined to lie between 9am to 5pm local time in the worker’s country. For countries with several time-zones, we average across all zones as an approximation. The results are reported in Appendix Table A2 and suggest that employers are not more likely to strategically delay their job postings in response to the first exposure. This suggests that the applicant pool shift is indeed driven by the increased selection of first hire country workers into later applications.

Other than collective reputation, the sorting response could reflect learning about the country-specific match productivity, revealed through the first hire’s country and rating. First of all, however, it is unclear why there should be country-specific match heterogeneity in the first place. Most of the jobs on the platform are small tasks such

¹²In Appendix A6, we use an alternative specification by comparing the characteristics of the final choice relative to the applicant pool average. The results confirm the absence of the “country premium”.

as data entry or programming. In presence of country-specific match heterogeneity, however, we expect the effect to decline once we control for bilateral confounds or remove language related tasks (e.g. translation). Indeed, the share of direct language-related tasks like translation is less than 2% and the removal of these do not affect our results. Furthermore, we include a range of bilateral country measures and interact them with the first hire country to account for observable bilateral heterogeneity. The results, again, remain unchanged (Table A7).

The results may also be driven by referrals networks. If employers are more likely to hire referrals from the first hire country, the estimates would recover same-country propensities in friendship and referral networks. While we cannot definitely rule out unobserved referral networks, we argue that this mechanism is unlikely to be important in our context:¹³ In the market under study, referral hirings typically take place in private transactions, where employers can invite selected workers to bid. We explicitly exclude these private transactions throughout our analysis and focus only on public market transactions that are competitive and open to all workers.

Another final concern is that employers “signal” their preference for certain groups by providing a high rating in their first hire. If workers indeed sort in response to the employer’s revealed taste for some countries, however, we expect the sorting effect to be even stronger when the first worker appears particularly unattractive along vertical traits, such as the individual rating or experience. In Table 3 and Table 4, we allowed for the first hire effect to vary along the first worker’s (ex-ante) individual rating and experience. If employer’s signal their preference for workers from a certain country, we expect the sorting response to increase when the first hire was particularly low ranked in reputation and experience. At most we find the opposite: Experienced workers are more likely to select in if the first hire was particularly experienced.

6.2 Robustness

We conclude the results with a wide range of robustness checks. First, we show that the main results are robust upon alternative measures of hiring intensity, such as measures for wage payments above the 9th decile, or the ranking of overall payment flows (Table A3, Column 2-4). We split the sample by the median year and confine the sample to only US employers (about half of the sample) and employers who hire at least twice (Column 5-8). For transparency, we report the simple OLS and the reduced form regressions (Table A4). We also report results using alternative definitions of high/low rating (Table A5).

As another exercise, we experiment with alternative instruments and characterize

¹³Other markets, e.g. oDesk, explicitly allow for referrals by enabling workers to “affiliate” with established intermediaries (Stanton and Thomas, 2014). This feature however does not apply to the market under study.

the complier population in order to facilitate the interpretation of the LATE. Instead of using the order statistic (cheapest worker) to construct the cheapest wage as an instrument, we re-estimate the regression using a dummy for when at least one worker from the given country submitted a bid below the average wage bid (Table A8). We also create a continuous measure that captures the number of workers from a country with bids below average.

The different instruments provide similar results. In terms of magnitude of the first hire effect, the alternative wage instruments provide comparable point estimates ranging between 3% to 5% points. These differences in estimates are likely to reflect different complier groups: Compared to the first hire (treated) and untreated populations, the size of the complier population varies across the instruments (See Table A9). Note that our preferred instrument constitutes the lower bound of our range of estimates.

7 Conclusion

In this paper, we provide causal evidence for collective reputation as a coordination device. An employer’s first rating given to a worker with an observable collective trait serves as a group-specific signal upon which later workers coordinate. Depending on the first experience, good workers attract more good workers from the same country and vice versa. The provision of collective traits hence perpetuates initial group inequalities by creating an (unintended) “herding effect”. As the provision of collective traits like country of origin or gender is pervasive across online markets, our results are likely to extend beyond the online labour market under study.

More broadly, our results trace out a channel through which collective traits - even if uncorrelated to *any* economic fundamentals ex-ante - can persistently shape the way workers sort and apply to jobs. While the online labour market setting limits the external validity of our findings, the sorting mechanism documented is likely to also apply to physical labour markets. Our results, for example, suggest that the requirement to disclose collective traits such as gender, race or nationality in applications (e.g. on a CV) may create inefficiencies as workers from some groups, anticipating discrimination, refrain from applying to certain employers altogether.

References

- AGRAWAL, A. K., N. LACETERA, AND E. LYONS (2013): “Does Information Help or Hinder Job Applicants from Less Developed Countries in Online Markets?” Working Paper 18720, National Bureau of Economic Research.
- ALTONJI, J. G. AND C. R. PIERRET (2001): “Employer Learning And Statistical Discrimination,” *The Quarterly Journal of Economics*, 116, 313–350.
- ASHRAF, N., O. BANDIERA, AND S. LEE (2014): “Do-gooders and Go-getters: Career incentives, Selection, and Performance in Public Service Delivery,” STICERD - EOPP Discussion Papers Series 54, STICERD, LSE.
- ASKER, J. AND E. CANTILLON (2008): “Properties of Scoring Auctions,” *The RAND Journal of Economics*, 39, 69–85.
- (2010): “Procurement when price and quality matter,” *RAND Journal of Economics*, 41, 1–34.
- BERTRAND, M. (2011): *New Perspectives on Gender*, Elsevier, vol. 4 of *Handbook of Labor Economics*, chap. 17, 1543–1590.
- BERTRAND, M. AND S. MULLAINATHAN (2004): “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *American Economic Review*, 94, 991–1013.
- BLUME, L. (2006): “The Dynamics of Statistical Discrimination,” *Economic Journal*, 116, F480–F498.
- BO, E. D., F. FINAN, AND M. A. ROSSI (2013): “Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service,” *The Quarterly Journal of Economics*, 128, 1169–1218.
- BOTTICINI, M. AND Z. ECKSTEIN (2013): *The Chosen Few: How Education Shaped Jewish History, 70-1492*, vol. 1, Princeton University Press, 1 ed.
- CABRAL, L. (2005): “The Economics of Trust and Reputation: A Primer,” *mimeo*.
- (2012): “Reputation on the Internet,” *The Oxford Handbook of the Digital Economy*.
- CABRAL, L. AND A. HORTACSU (2006): “The Dynamics of Seller Reputation: Evidence from eBay,” *mimeo*.

- COATE, S. AND G. C. LOURY (1993): “Will Affirmative-Action Policies Eliminate Negative Stereotypes?” *The American Economic Review*, 83, pp. 1220–1240.
- DELLAROCAS, C. AND C. A. WOOD (2008): “The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias,” *Management Science*, 54, 460–476.
- DESERRANNO, E. (2014): “Financial incentives as signals: Experimental evidence from the recruitment of Health Workers,” *mimeo*.
- GHANI, E., W. R. KERR, AND C. STANTON (2014): “Diasporas and outsourcing: evidence from oDesk and India,” *Management Science*, 60, 1677–1697.
- GLOVER, D., A. PALLAIS, AND W. PARIENTE (2015): “Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores,” *mimeo*.
- HORTASU, A., F. A. MARTNEZ-JEREZ, AND J. DOUGLAS (2009): “The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre,” *American Economic Journal: Microeconomics*, 1, 53–74.
- KIM, Y.-C. AND G. LOURY (2014): “Collective Reputation and the Dynamics of Statistical Discrimination,” Tech. rep., *mimeo*.
- LAZEAR, E. P. AND P. OYER (2007): “Personnel Economics,” Working Paper 13480, National Bureau of Economic Research.
- MAILATH, G. AND L. SAMUELSON (2006): *Repeated Games and Reputations: Long-run Relationships*, Oxford scholarship online, Oxford University Press.
- MAYER, T. AND S. ZIGNAGO (2011): “Notes on CEPIIs distances measures: The GeoDist database,” Working Papers 2011-25, CEPII Research Center.
- MILL, R. AND L. STEIN (2015): “Race, skin color, and economic outcomes in early twentieth-century America,” *mimeo*.
- MORENO, A. AND C. TERWIESCH (2013): “Doing Business with Strangers: Reputation in Online Service Marketplaces,” *mimeo*.
- MORO, A. AND P. NORMAN (2004): “A General Equilibrium Model of Statistical Discrimination,” *Journal of Economic Theory*, 114, 1–30.
- NOSKO, C. AND S. TADELIS (2015): “The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment,” Working Paper 20830, National Bureau of Economic Research.

- PALLAIS, A. (2014): “Inefficient Hiring in Entry-Level Labor Markets,” *American Economic Review*, 104, 3565–99.
- PAUL, O. AND S. SCOTT (2011): “Chapter 20 - Personnel Economics: Hiring and Incentives,” Elsevier, vol. 4, Part B of *Handbook of Labor Economics*, 1769 – 1823.
- PHELPS, E. S. (1972): “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 62, 659–61.
- RESNICK, P., R. ZECKHAUSER, J. SWANSON, AND K. LOCKWOOD (2006): “The value of reputation on eBay: A controlled experiment,” *Experimental Economics*, 9, 79–101.
- RIACH, P. A. AND J. RICH (2006): “An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market,” *The B.E. Journal of Economic Analysis & Policy*, 5, 1–22.
- STANTON, C. AND C. THOMAS (2014): “Landing the First Job: The Value of Intermediaries in Online Hiring,” *CEP Discussion Paper 1316*.
- THOMAS, C. (2012): *The Economics of Online Recruitment*, Palgrave Macmillan, vol. 6 of *The New Palgrave Dictionary of Economics*, chap. 1.
- TIROLE, J. (1996): “A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality),” *Review of Economic Studies*, 63, 1–22.

Figure 1: Example of a typical applicant pool

Project Title: [REDACTED]
 Project Id: [REDACTED]

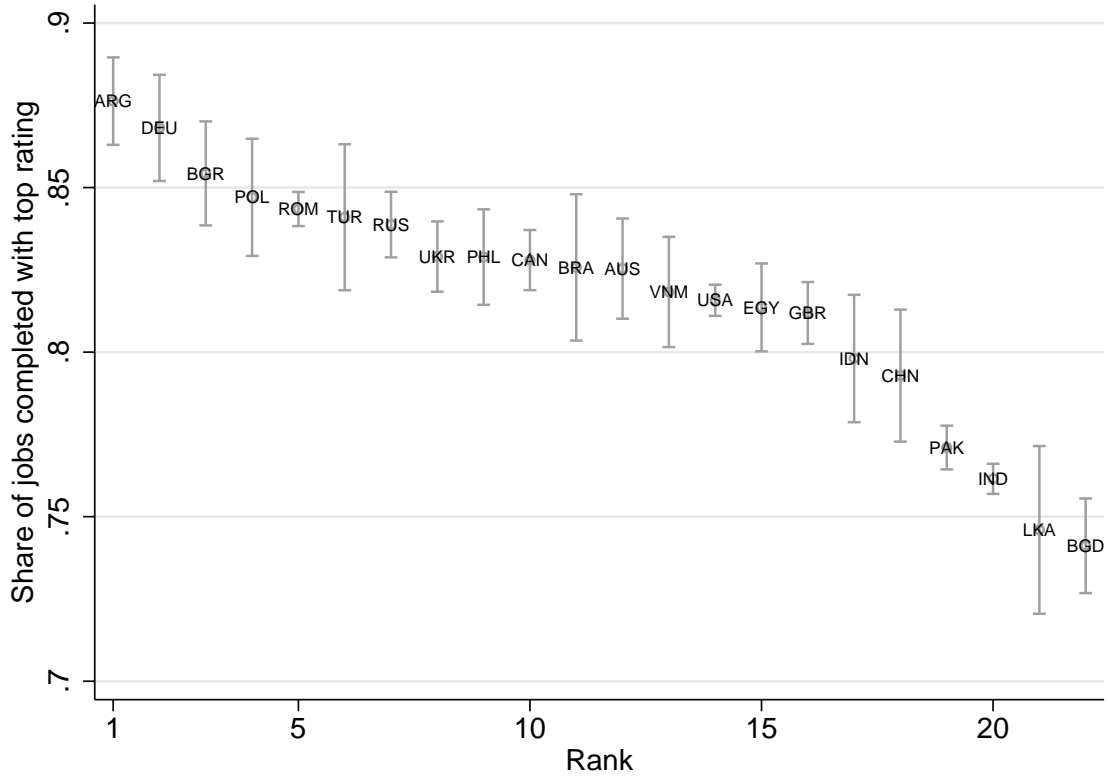
Bidding / message summary

Bids posted: 6

Date	Name	City / Country	Rating Avg. / # of Ratings	Worker Competition Ranking	Last Bid	Posts
Note: Click here to see only the final selected bidder.						
Jun 10, 2009 2:06:00 PM	[REDACTED]	In Santo Domingo, Dominican Republic	Very Good: 7.66 out of 10 from 99 ratings	Ranked #42,296 out of 390,128 (better than 89.15 %).	\$330.00 USD formerly shown as \$330.00 (why?)	2 messages
Jun 10, 2009 2:08:06 PM	[REDACTED]	In Gurgaon , India	Excellent: 9.98 out of 10 from 27 ratings	Ranked #526 out of 390,128 (better than 99.86 %).	None yet	2 messages
Jun 10, 2009 2:27:37 PM	[REDACTED]	In Khimki, Russian Federation	Superb: 9.4 out of 10 from 30 ratings	Ranked #2,517 out of 390,128 (better than 99.09 %).	\$400.00 USD formerly shown as \$400.00 (why?)	2 messages
Jun 10, 2009 2:45:33 PM	[REDACTED]	In Saint-Petersburg, Russian Federation	Excellent: 9.98 out of 10 from 112 ratings	Ranked #13 out of 390,128 (better than 99.99 %).	\$425.00 USD formerly shown as \$500.00 (why?) was accepted	35 messages
Jun 10, 2009 2:56:32 PM	[REDACTED]	In Shanghai, China	Excellent: 10 out of 10 from 8 ratings	Ranked #10,677 out of 390,128 (better than 97.26 %).	\$500.00 USD formerly shown as \$500.00 (why?)	4 messages
Jun 10, 2009 3:07:57 PM	[REDACTED]	In PUNE, India	Superb: 8.71 out of 10 from 7 ratings	Not yet ranked	\$450.00 USD formerly shown as \$450.00 (why?)	1 message
Jun 10, 2009 3:31:11 PM	[REDACTED]	In Striy, Ukraine	Superb: 9.41 out of 10 from 75 ratings	Ranked #258 out of 390,128 (better than 99.93 %).	\$500.00 USD formerly shown as \$500.00 (why?)	1 message

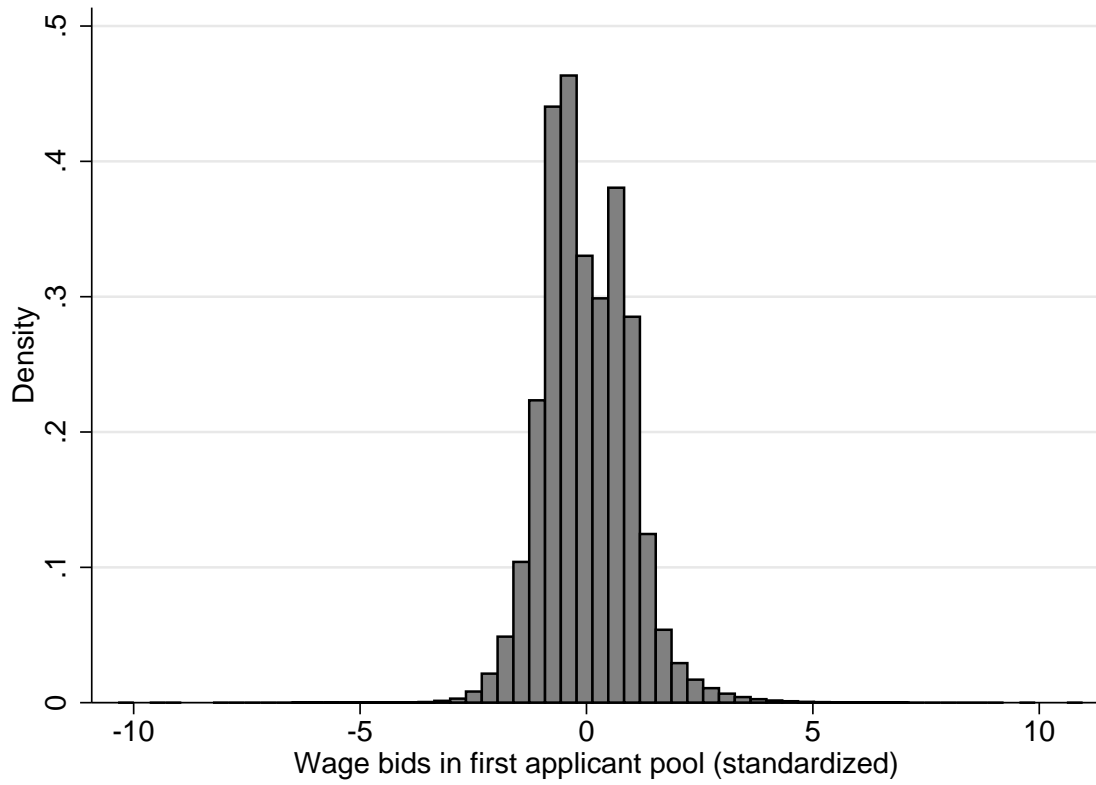
Names are anonymized but typically are aliases that do not enable employers to directly infer to underlying worker quality. The salient collective trait in this context is the country of residence, as indicated by the flag and the country name.

Figure 2: Average share of top rated jobs by worker country



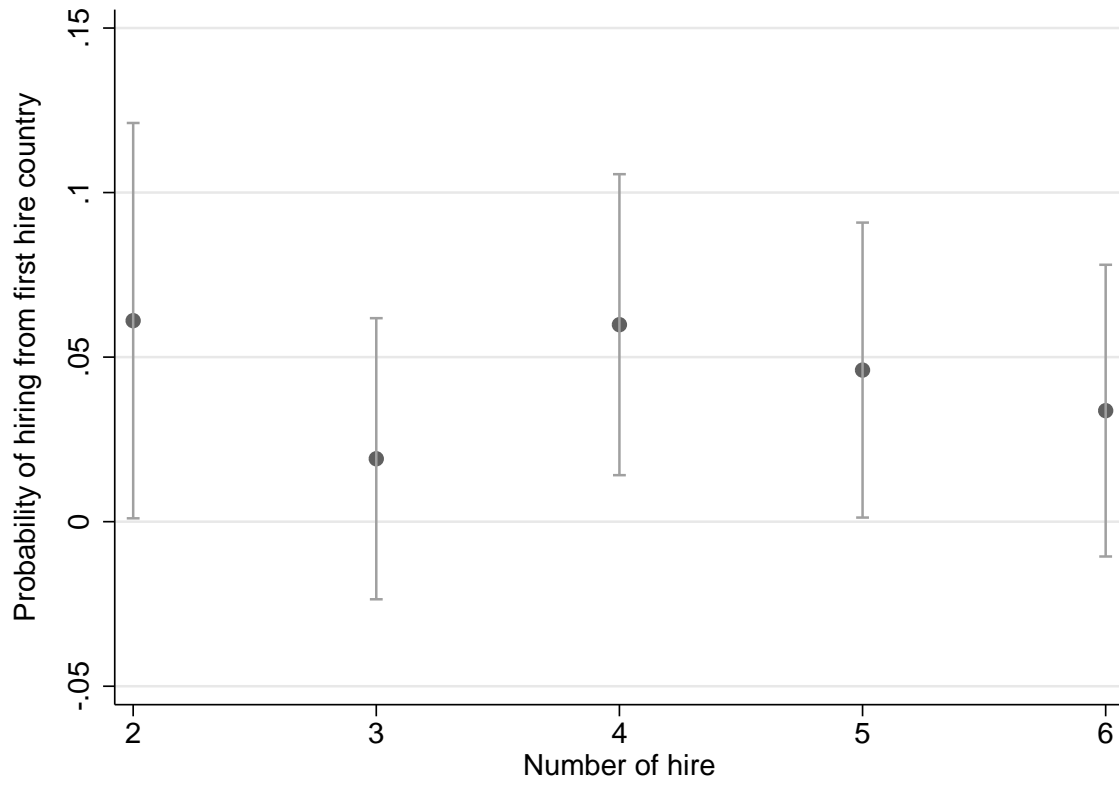
Average share of top rated jobs over the full sample period 2001-2012, by the worker's country (ISO-3 code) in descending order. 95% confidence intervals.

Figure 3: Residual *within* applicant pool distribution of wage bids



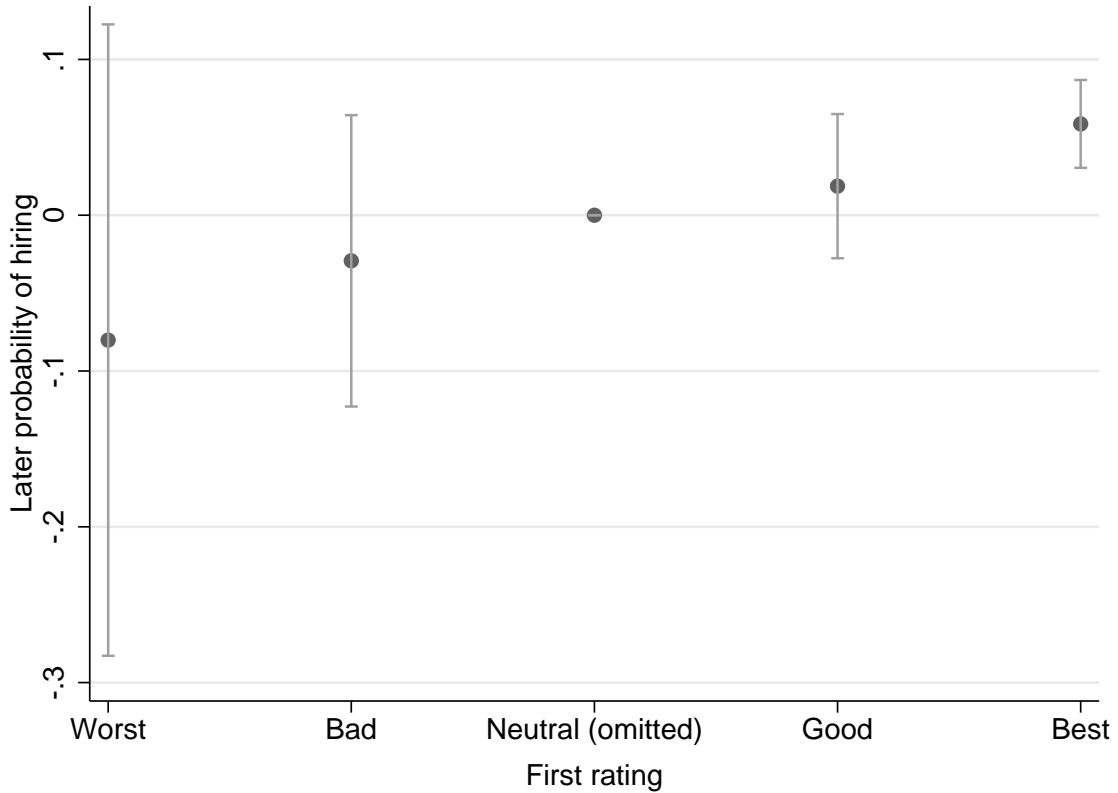
Residual *within* applicant pool variation in standardized (log) wage bids, with worker country fixed effects partialled out. Full sample period 2001-2012.

Figure 4: Times hired: Persistence of the first hire effect



Probability of hiring from the first hire country at the 2nd, 3rd, 4th, 5th and 6th hire among employers who hire beyond the 6th hire (5% of all employers). $N = 12498$.

Figure 5: Treatment heterogeneity: Probability of hiring again from the country of first hire as a function of the first rating given



Point estimates of the first hire effect as a function of the first rating given. Estimates are reported relative to the neutral rating (7-8 points). Worst rating is defined as a rating below 4 points. A bad rating is defined as a rating between 4-6 points. A good (best) rating is equivalent to a rating of 9 (10) points. 95% confidence intervals.

Table 1: First applicant pool characteristics and descriptive statistics of later hiring patterns

Panel A: Employer level	Applicant pool characteristics of the first job									
	All employers (N=25805)					Non-exit (N=8219)				
	Mean	SD	10%	50%	90%	Mean	SD	10%	50%	90%
Number of applicants	10.60	10.78	2	7	24	11.99	11.25	2	8	26
Number of countries	5.66	3.60	2	5	11	6.23	3.70	2	5	11
Wage bid (\$)	323.24	5925.94	23.12	96.52	591.65	228.43	1154.68	21.10	81.87	466.93
- Chosen worker	171.01	469.45	12.75	68	424.15	132.09	327.07	10.2	51	296.65
Top rating	0.24	0.19	0	0.22	0.50	0.24	0.18	0	0.22	0.50
- Chosen worker	0.31	0.46	0	0	1	0.32	0.46	0	0	1
Number of completed jobs	36.34	51.86	2	18.54	90.13	33.78	49.36	2	17.03	84.06
- Chosen worker	59.44	134.82	0	14	156	55.56	132.58	0	13	142
Panel B: Employer-country level	Employers hiring from countries of their first applicant pool, after the first hire									
	All employers (N=146273)					Non-exit (N=51545)				
	Mean	SD	10%	50%	90%	Mean	SD	10%	50%	90%
Hires from country after first job	0.100	0.301	0	0	1	0.286	0.452	0	0	1
- First hire country	0.141	0.348	0	0	1	0.440	0.496	0	0	1
Times hired	0.236	1.495	0	0	1	0.671	2.460	0	0	2
- First hire country	0.356	1.717	0	0	1	1.111	2.893	0	0	3
Total pay (\$)	16.35	163.34	0	0	0	46.41	272.61	0	0	85
- First hire country	24.92	198.82	0	0	12.75	77.79	345.39	0	0	170

Panel A reports the descriptive statistics for the employer’s first applicant pool. *Number of applicants* is the number of applicant workers in the employers’ first job applicant pool. *Number of countries* is the number of distinct countries from which workers applied in the first applicant pool. *Wage bid* is the fixed wage bid for completing the job in USD (\$). *Top rating* is a dummy that is 1 if the worker has a top rating and 0 otherwise (workers with no previous rating are hence coded 0). *Number of completed jobs* is the number of previous jobs completed on the platform. *Chosen worker* shows the summary statistics for the hired workers among the applicants. **Panel B** reports descriptive statistics for the employer’s subsequent transactions with countries from the applicant pool of the first job. *Hires from country after first job* is a dummy that is 1 if the employer ever hired from the country after the first hire. *Times hired* is the number of times the employer hired from the country after the first job. *Total pay* is the cumulative payments an employer made to workers from the given country, in USD (\$). *First hire country* shows the summary statistics for the country the employer first hired from. Reporting mean, standard deviation (SD), the 1st decile (10%), median (50%) and 9th decile (90%).

Table 2: Reputational externalities: First country hire and later hiring

Panel A: OLS and IV	(1)	(2)	(3)	(4)	(5)	(6)
	Dummy = 1 if employer ever hires from country after the first job				Times hired	Total pay
Mean dep. var.	0.100	0.100	0.100	0.100	0.236	10.026
First hire country	0.037*** (0.00)	0.011*** (0.00)	0.032*** (0.01)	0.031*** (0.01)	0.081** (0.03)	9.270** (4.21)
Estimation	OLS	OLS	IV	IV	IV	IV
Panel B: First stage	(8)	(9)	(10)	(11)	(12)	(13)
				First hire country		
Mean dep. var.	0.175			0.175	0.175	0.175
Cheapest	0.273*** (0.004)			0.272*** (0.004)	0.272*** (0.004)	0.272*** (0.004)
Employer FEs	X	X	X	X	X	X
Worker country FEs		X	X	X	X	X
Bilateral controls				X	X	X
Kleibergen-Paap F-statistic	5250.67			5214.54	5214.54	5214.54
Observations	146273	146273	146273	146273	146273	146273

The unit of observation is the employer-worker country pair. The dependent variable is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *Times hired* is the number of times the employer hired from the country after the first hire. *Total pay* is the cumulative wages paid by an employer to workers in the given country after the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. *Cheapest* is a dummy that is 1 if applicant workers from the country submitted the cheapest (residual) wage bid. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. The test for weak instruments is the Kleibergen-Paap Wald F-statistic. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Treatment heterogeneity: Later hiring contingent on the first hire's rating

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dummy = 1 if employer ever hires from country after the first job					Times hired	Total pay
Mean dep. var.	0.100	0.100	0.100	0.100	0.100	0.236	10.026
First hire country	0.031*** (0.01)	-0.006 (0.01)	0.026*** (0.01)	0.036*** (0.01)	-0.005 (0.01)	-0.040 (0.06)	1.101 (2.47)
First hire country \times Top rating given		0.055*** (0.01)			0.054*** (0.01)	0.136** (0.06)	4.229* (2.44)
First hire country \times Previous: Top rating			0.014 (0.01)		0.006 (0.01)	0.051 (0.07)	3.029 (2.77)
First hire country \times Previous: Experienced				-0.013 (0.01)	-0.007 (0.01)	0.033 (0.07)	-1.875 (2.73)
Estimation method	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X	X
First-stage Kleibergen-Paap F-statistic	5214.54	1771.78	2105.11	903.50	419.57	419.57	419.57
Observations	146273	146273	146273	146273	146273	146273	146273

The unit of observation is the employer-worker country pair. The dependent variable is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *Times hired* is the number of times the employer hired from the country after the first hire. *Total pay* is the cumulative wages paid by an employer to workers in the given country after the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. The test for weak instruments is the Kleibergen-Paap Wald F-statistic. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Mechanisms: Change in applicant pool composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Applied		No. applicants		Top rating		High exp.	
Mean dep. var.	0.273	0.273	0.155	0.155	0.573	0.573	0.196	0.196
First hire country	0.030*** (0.01)	-0.006 (0.01)	0.301 (0.22)	-0.621 (0.39)	0.205* (0.12)	-0.409** (0.21)	0.049 (0.06)	-0.245** (0.10)
First hire country × Top rating given		0.051*** (0.01)		1.092*** (0.41)		0.634*** (0.22)		0.256** (0.12)
First hire country × Previous: Top rating		0.009 (0.01)		0.214 (0.51)		0.194 (0.26)		0.038 (0.12)
First hire country × Previous: Experienced		-0.003 (0.01)		0.291 (0.48)		0.312 (0.25)		0.278** (0.12)
Estimation method	IV	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X	X	X
Observations	145281	145281	145281	145281	145281	145281	145281	145281

The unit of observation is the employer-country pair. *Applied* is a dummy that is 1 if workers from the country from which workers applied in the first applicant pool applied *after* the first hire. *No. of applicants* is the number of applicant workers from the country after the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. *Top rating* is the number of top rated applicant workers (10/10) from the country after the first hire. *High exp.* is the number of applicants with more than a median amount (17) of completed jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Mechanisms: Effect on later hiring outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Characteristics of hired worker							
	Top rating given		(ln) Wage		Top rating		Experienced	
Mean of dep. var.	0.900	0.900	3.677	3.677	0.772	0.772	0.314	0.314
First hire country	0.004 (0.02)	-0.072 (0.05)	0.037 (0.07)	0.132 (0.17)	-0.006 (0.04)	-0.136* (0.07)	0.003 (0.03)	0.026 (0.07)
First hire country \times Top rating given		0.111** (0.05)		-0.034 (0.18)		0.125* (0.08)		-0.057 (0.08)
First hire country \times Previous: Top rating		0.002 (0.04)		-0.148 (0.16)		0.045 (0.08)		-0.062 (0.08)
First hire country \times Previous: Experienced		-0.030 (0.04)		-0.038 (0.16)		0.039 (0.08)		0.112 (0.07)
Estimation method	IV	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X	X	X
Observations	20761	20761	20761	20761	20761	20761	20761	20761

The unit of observation is the completed job. *Top rating given* is a dummy that is 1 if the employer provided the highest rating (10/10) to the worker. *ln(Wage bid)* is the (log) wage paid to the hired worker. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. *Top rating* is a dummy that is 1 if the hired worker had a top rating at time of application (before receiving the rating for the current job). *Experienced* is a dummy that is 1 if the chosen worker completed more than a median amount (17) of completed jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

8 Appendix material

Table A1: Testing the joint significance of Worker country FEs (Full sample 2001-12)

	(1)	(2)	(3)
	Top rating received for job		
Mean dep. var.	0.804	0.804	0.804
Previous: Top rating			0.039*** (0.001)
Previous: Experienced			0.020*** (0.001)
Overall R^2	0.008	0.691	0.692
Observations	144516	144516	144516
Worker country FEs $F(21, 109319)$	52.958***	22.962***	19.423***
Employer FEs $F(35175, 109340)$		6.921***	6.729***
Individual background $F(2, 144513)$			223.59***

The unit of observation is the job. Reporting the F -test for the joint significance of the collective trait, the worker's country, in predicting job performance, as measured by whether a top rating was received. *Top rating* is a dummy that is 1 if the job was completed with a highest rating by the employer (10/10). *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Testing for changes in the employer’s job posting time

	(1)	(2)	(3)	(4)
	Local time at job posting			
	9am-5pm		8am-8pm	
Mean of dep. var.	0.345	0.345	0.495	0.495
First hire country	-0.051	0.008	-0.026	0.073
	(0.04)	(0.08)	(0.04)	(0.09)
First hire country × Top rating		-0.053		-0.039
		(0.08)		(0.09)
First hire country × Previous: Top rating		-0.018		-0.102
		(0.08)		(0.09)
First hire country × Previous: Experienced		-0.027		-0.080
		(0.08)		(0.09)
Estimation method	IV	IV	IV	IV
Employer FEs	X	X	X	X
Worker country FEs	X	X	X	X
Bilateral controls	X	X	X	X
Observations	20676	20676	20676	20676

The unit of observation is the job. The dependent variable *Local time at job posting* is a dummy that is 1 if the job was posted between 9am-5pm (8am-8pm) in the worker’s local country time. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Robustness: Alternative dependent variables, time periods and samples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: First hire effect	Hire after first job	Highest	Alt. dep. vars. \geq 9th decile	Rank	Time period		Employer sample	
					Pre 2006	Post 2006	US employers	$>$ 1 hires
Mean dep. var.	0.100	0.715	0.010	4.477	0.125	0.078	0.099	0.286
First hire country	0.031*** (0.006)	0.027*** (0.006)	0.004* (0.002)	0.130** (0.05)	0.031*** (0.01)	0.022** (0.01)	0.024*** (0.008)	0.073*** (0.02)
Estimation method	IV	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X	X
Observations	146273	146273	146273	146273	70989	75284	80474	51545
Panel B: Contingent on first outcome	Base Continue	Highest	Alt. dep. vars. \geq 9th decile	Rank	Time period		Employer sample	
					Pre 2006	Post 2006	US employers	$>$ 1 hires
Mean dep. var.	0.100	0.715	0.010	4.477	0.125	0.078	0.099	0.286
First hire country	-0.005 (0.01)	-0.004 (0.01)	-0.004 (0.00)	0.099 (0.11)	-0.001 (0.02)	-0.018 (0.02)	-0.018 (0.02)	0.037 (0.04)
First hire country \times Top rating	0.054*** (0.01)	0.051*** (0.01)	0.007 (0.00)	0.127 (0.12)	0.043** (0.02)	0.061*** (0.02)	0.059*** (0.02)	0.073* (0.04)
First hire country \times Previous: Top rating	0.006 (0.01)	0.000 (0.01)	0.008 (0.01)	0.015 (0.13)	0.010 (0.02)	-0.003 (0.02)	0.009 (0.02)	-0.002 (0.04)
First hire country \times Previous: Experienced	-0.007 (0.01)	-0.008 (0.01)	0.002 (0.01)	-0.057 (0.13)	-0.006 (0.02)	0.004 (0.02)	-0.004 (0.02)	-0.044 (0.04)
Estimation method	IV	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X	X
Observations	146273	146273	146273	146273	70989	75284	80474	51545

The unit of observation is the employer-country pair. **Panel A** re-estimates the main result in Section 4.1 with different measures of outcome (Column 2-4), time periods (5-6) and samples (7-8). *Highest* is a dummy that is 1 if an employer's largest cumulative wage payments went to the given country. \geq 9th decile is a dummy that is 1 if the cumulative wage payments were above the 9th decile among payments to all hired countries. *Rank* transforms the continuous payments into an ordinal ranking of flows. To facilitate comparison, the measure has been inverted so higher ranks indicate a higher payments made. Column 5-6 splits the sample by the median year (2006). Column 7 restricts the sample to US employers. Column 8 conditions only on employers who hire at least twice. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. **Panel B** repeats the same regressions for the secondary result in Section 4.2. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Robustness: OLS and reduced forms

Panel A: OLS	(1)	(2)	(3)	(4)	(5)	(6)
	Hire after first job		Times hired		Total pay	
Mean dep. var.	0.100	0.100	0.236	0.236	10.026	10.026
First hire country	0.010***	0.008	0.023**	0.025	1.399***	2.074**
	(0.00)	(0.01)	(0.01)	(0.02)	(0.43)	(0.99)
First hire country \times Top rating		0.011**		0.017		0.375
		(0.00)		(0.02)		(0.91)
First hire country \times Previous: Top rating		-0.005		-0.014		-0.414
		(0.01)		(0.03)		(1.00)
First hire country \times Previous: Experienced		-0.008*		-0.019		-1.594*
		(0.00)		(0.02)		(0.93)
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS
Employer FEs	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X
Observations	146273	146273	146273	146273	146273	146273
Panel B: Reduced form	(5)	(6)	(7)	(8)	(9)	(10)
	Continue		Times		Total pay	
Mean dep. var.	0.100	0.100	0.236	0.236	10.026	10.026
Cheapest	0.009***	0.001	0.023**	-0.005	1.177***	0.751
	(0.00)	(0.00)	(0.01)	(0.02)	(0.39)	(0.91)
Cheapest \times Top rating		0.015***		0.037*		1.128
		(0.00)		(0.02)		(0.80)
Cheapest \times Previous: Top rating		-0.000		0.011		0.644
		(0.00)		(0.02)		(0.92)
Cheapest \times Previous: Experienced		-0.007		-0.003		-1.121
		(0.00)		(0.02)		(0.85)
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS
Employer FEs	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X
Observations	146273	146273	146273	146273	146273	146273

The unit of observation is the employer-worker country pair. **Panel A** reports OLS estimates and **Panel B** reports reduced forms. The dependent variable is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *Times hired* is the number of times the employer hired from the country after the first hire. *Total pay* is the cumulative wages paid by an employer to workers in the given country after the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Robustness: Using different definitions of job rating

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Continue			
Mean dep. var.	0.100	0.100	0.100	0.100	0.100	0.100	0.100
First hire country	-0.106 (0.07)	-0.106** (0.05)	-0.048 (0.04)	-0.033 (0.03)	-0.013 (0.02)	-0.002 (0.01)	-0.181*** (0.06)
First hire country \times Rating ≥ 5	0.137* (0.07)						
First hire country \times Rating ≥ 6		0.138** (0.05)					
First hire country \times Rating ≥ 7			0.080* (0.04)				
First hire country \times Rating ≥ 8				0.066** (0.03)			
First hire country \times Rating ≥ 9					0.048** (0.02)		
First hire country \times Rating ≥ 10						0.039** (0.02)	
First hire country \times Rating [0,1]							0.224*** (0.07)
Estimation method	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X
Worker country FEs	X	X	X	X	X	X	X
Observations	146273	146273	146273	146273	146273	146273	146273

The unit of observation is the employer-worker country pair. The dependent variable is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Rating $\geq X$* is a dummy that is 1 if a rating is higher than $X/10$. *Rating [0,1]* is the normalized rating that ranges from 0 to 1, where 0 is the absence of rating. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Robustness: Final choice relative to applicant pool average

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Characteristics of hired worker compared to applicant pool							
	Highest rating		Highest experience		Cheapest		Dominant bid	
Mean of dep. var.	0.465	0.465	0.402	0.402	0.476	0.476	0.181	0.181
First hire country	-0.019	-0.090	0.021	-0.124*	-0.013	-0.042	0.015	-0.015
	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.06)	(0.03)	(0.06)
First hire country \times Top rating given		0.097		0.052		0.046		0.017
		(0.07)		(0.07)		(0.07)		(0.06)
First hire country \times Previous: Top rating		0.050		0.115		-0.033		0.054
		(0.07)		(0.07)		(0.07)		(0.06)
First hire country \times Previous: Experienced		-0.060		0.157**		0.011		-0.008
		(0.07)		(0.07)		(0.07)		(0.06)
Estimation	IV	IV	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X	X	X
Country FEs	X	X	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X	X	X
Observations	20761	20761	20761	20761	20761	20761	20761	20761

The unit of observation is the employer-worker country pair. The dependent variable captures the final choices of employer's relative to their applicant pools: In Column 1-2, the dependent variable is a dummy that is 1 if the employer hired the worker with the highest rating. In Column 3-4 (5-6), we capture whether the employer hired the most experienced (cheapest) worker. In Column 7-8, we summarize the previous three measures using a dummy whether the worker hired the dominant worker, based on the vertical traits of rating, experience and price. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Robustness: Controlling for observable country-specific match heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
	Employers			Workers		
	Hire	Times	Total pay	Applied	Top rated	High exp.
Mean of the dep. var.	0.100	0.236	16.357	0.273	0.573	0.196
First hire country	-0.016 (0.10)	0.154 (0.38)	0.618 (17.99)	-0.032 (0.10)	-2.867* (1.56)	-1.819** (0.80)
First hire country \times Top rating given	0.053*** (0.01)	0.132** (0.06)	4.347* (2.46)	0.051*** (0.01)	0.630*** (0.23)	0.259** (0.12)
First hire country \times Previous: Top rating	0.004 (0.01)	0.042 (0.07)	3.043 (2.80)	0.008 (0.01)	0.151 (0.26)	0.020 (0.12)
First hire country \times Previous: Experienced	-0.006 (0.01)	0.042 (0.07)	-1.658 (2.74)	-0.002 (0.01)	0.338 (0.25)	0.289** (0.12)
First hire country \times ln(Distance)	0.003 (0.01)	-0.014 (0.05)	0.334 (2.36)	0.005 (0.01)	0.368* (0.21)	0.228** (0.11)
First hire country \times Abs. time difference	-0.000 (0.00)	-0.004 (0.02)	-0.533 (0.57)	-0.001 (0.00)	-0.076 (0.06)	-0.049 (0.03)
First hire country \times Common language	-0.027* (0.02)	-0.111* (0.06)	1.821 (2.82)	-0.023 (0.01)	-0.680*** (0.25)	-0.271** (0.12)
Estimation	IV	IV	IV	IV	IV	IV
Employer FEs	X	X	X	X	X	X
Country FEs	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X
Kleibergen-Paap F-statistic	233.14	233.14	233.14	233.14	233.14	233.14
Observations	146273	146273	146273	145281	145281	145281

The unit of observation is the employer-worker country pair. The dependent variable *hire* is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *Times hired* is the number of times the employer hired from the country after the first hire. *Total pay* is the cumulative wages paid by an employer to workers in the given country after the first hire. *Applied* is a dummy that is 1 if workers from the country from which workers applied in the first applicant pool applied *after* the first hire. *No. of applicants* is the number of applicant workers from the country after the first hire. *Top rating* is the number of top rated applicant workers (10/10) from the country after the first hire. *High exp.* is the number of applicants with more than a median amount (17) of completed jobs on the platform. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Robustness: Alternative constructions of the wage IV

	(1)	(2)	(3)	(4)	(5)	(6)
	Dummy = 1 if employer ever hires from country after the first job					
Mean of the dep. var.	0.100	0.100	0.100	0.100	0.100	0.100
First hire country	0.031*** (0.01)	-0.005 (0.01)	0.038*** (0.01)	-0.006 (0.01)	0.046*** (0.01)	0.010 (0.02)
First hire country \times Top rating		0.054*** (0.01)		0.056*** (0.01)		0.101*** (0.02)
First hire country \times Previous: Top rating		0.006 (0.01)		0.027* (0.02)		-0.003 (0.02)
First hire country \times Previous: Experienced		-0.007 (0.01)		-0.009 (0.01)		-0.061*** (0.02)
Instrument	Cheapest		Below average		# Below average	
Employer FEs	X	X	X	X	X	X
Country FEs	X	X	X	X	X	X
Bilateral controls	X	X	X	X	X	X
Kleibergen-Paap F-statistic	5214.54	419.57	9043.14	1007.02	1725.42	395.67
Observations	146273	146273	146273	146273	146273	146273

The unit of observation is the employer-worker country pair. The dependent variable *hire* is a dummy that is 1 if the employer ever hired from the country from which workers applied in the first applicant pool *after* the first hire. *Times hired* is the number of times the employer hired from the country after the first hire. *Total pay* is the cumulative wages paid by an employer to workers in the given country after the first hire. *First hire country* is a dummy that is 1 if the worker hired for the first job came from the country. A *Top rating* is defined as a rating of 10/10. *Previous: Top rating* is a dummy that is 1 if the applicant worker received a top rating in previous jobs. *Previous: Experience* is a dummy that is 1 if the worker has completed more than a median amount (17) of jobs on the platform. Bilateral controls include a dummy for the shared official country language, the absolute time difference in hours, and (log) centroid distance between the employer and worker country. *Cheapest* is the preferred instrument when a worker from the given country submitted the lowest (residual) bid. *Below average wage* is a dummy if at least one worker from the country submitted a wage bid that was below the applicant pool average. *# Below average wage* is a continuous measure that captures the number of workers bidding below the applicant pool average wage. The test for weak instruments is the Kleibergen-Paap Wald F-statistic. Robust SEs, clustered at the employer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: LATE and probabilities of compliance using different instruments

	(1)	(2)	(3)	(4)	(5)	(6)
	LATE		First stage		Compliance probability	
	$\hat{\beta}_1$	$P[D = 1]$	$P[D_1 > D_0]$	$P[z = 1]$	$P[D_1 > D_0 D = 1]$	$P[D_1 > D_0 D = 0]$
Panel A: Wage bid instrument						
Cheapest	0.031*** (0.01)	0.176	0.333	0.184	0.348	0.149
Below average wages	0.038*** (0.01)	0.176	0.200	0.541	0.614	0.525

The table summarizes estimates using different instruments (Column 1) and characterizes the population of compliers. Column 5 shows the size of the complier population relative to the “treated” population (countries first chosen). Column 6 shows the size of the complier population relative to the “untreated” population of countries that were not chosen from the first applicant pool.