Exploring Gender Discrimination:
A Multitrial Field Experiment in Urban Ecuador[*]

Wladimir Zanoni[†], Suzanne Duryea[†], Jorge Paredes[†]

May 16, 2025

Abstract

In this study, we investigate the extent and mechanisms of gender-based discrimination in urban Ecuador's hiring practices. Using an artifactual field experiment with 392 recruiters evaluating observationally equivalent male and female job candidates, we find a significant preference of 10% for female candidates, despite comparable productivity assessments between genders. This suggests that hiring decisions are influenced by factors beyond assessed productivity differentials. We hypothesize that social norms advocating for gender equality are significant drivers of the preference we identify, a claim that is further supported by the observed trend of narrowing the employment gender gap in survey data.

1 Introduction

In this paper, we aim to quantify and analyze the extent to which gender-based discrimination contributes to disparities in hiring rates in urban areas of Ecuador. Specifically, we investigate whether equally qualified applicants face unequal treatment in hiring because of their gender, thus opening the black box on the factors driving the employment outcome gaps between men and women. Besides generating evidence on the sign and magnitude of the phenomenon, we propose a theoretical model that provides an explanation of potential mechanisms behind the results.

In recent years, Ecuador has experienced a notable transformation in its labor market dynamics, particularly regarding the engagement and employment of women (Mahé et al., 2022). Data from yearly household surveys produced by the Ecuadorian Census and Statistics Office (Instituto Nacional de Estadística y Censos; INEC) from 2008 to 2023 show that the disparity in labor force participation rates between genders decreased from 29.8 to 23.7 percentage points, while the gap in employment rates narrowed from -3.2 to -1.4. Moreover, the differential in the rate of "adequate employment"—defined as working a minimum of 40 hours per week—shrank from 33.8 to 24.4. Despite this progress, discrepancies persist in employment outcomes between men and women, as illustrated in figure 2. In Latin America as a whole, women face significant disadvantages in the labor market due to parenthood. According to Kleven et al. (2024), the impact of motherhood on employment prospects is more pronounced in Latin America than in any other global region, while fatherhood seems to have little to no effect. Villanueva and Lin (2020) demonstrate that, even after adjusting for factors such as employment and education, the penalties mothers face in five Latin American countries range from 12% in Brazil to 21% in Chile. Furthermore, Berniell et al. (2021) reveals that in Chile, female labor informality surged by 38% following the birth of a first child. The existing disparities and trends motivate further investigation into the barriers to equal employment opportunities in Ecuador.

Answering this question is important for several reasons. From a scientific perspective, doing so contributes to the existing body of knowledge on gender disparities in labor economics. It offers an opportunity to apply, test, and possibly refine theoretical frameworks regarding labor market discrimination, gender norms, and economic participation. The scope of these frameworks is not limited to developing countries because although many developed countries have experienced similar patterns of economic inclusion of women, the kinds of gender inclusion challenges Ecuador faces persist in those countries OECD (2023). From a policy-making standpoint, this empirical insight provides a foundation for formulating public policies that equitably promote human capital, enhance productivity, and foster innovation, while addressing the persistent challenges of gender equity and inclusion.

To study discrimination against women, we conducted an artifactual field experiment (AFE) that examined whether current hiring practices involved unequal treatment of job applicants who were women, gays, lesbians, and migrants in Quito, Ecuador. Here, we focus on the trials of that experiment where recruiters analyzed pairs of applications that only compared men to women. Both ? and Zanoni and Díaz (2024) use data from the same experiment to study other manifestations of discrimination, in the first case against gays and lesbians and the second against Venezuelan migrants.

The experimental data were collected in close collaboration with Grupo FARO, a local nongovernmental organization (NGO) that provides employment referral services in Quito. Leveraging LinkedIn for recruitment and employing respondent-driven sampling (RDS), we hired 392 experienced human resource recruiters to participate in the study under the pretense of their providing human resources (HR) professional advisory services as remote consultants for a multinational corporation looking to hire in Ecuador. To ensure their committed involvement, recruiters were offered competitive compensation above market rates. The recruiters were unaware that their participation was part of a field experiment designed to scrutinize employment referral processes in the context of gender discrimination.

In the task assigned to the recruiters, they were to evaluate and make employment referrals based on the review of ten pairs of job applications (or trials) that were designed to reflect realistic scenarios across different occupations extracted from actual job postings. Each pair was designed to showcase candidates who, while being observationally equivalent in terms of productivity signals, differed in either gender, sexual orientation, or nationality. The full experiment comprised ten trials: three pairs contrasting a male and a female candidate, three pairs differentiating between a gay or lesbian candidate and their heterosexual counterpart, and three pairs comparing a Venezuelan migrant to an Ecuadorian national. A placebo trial was included in which candidates without any distinguishable prejudice-based differences served as a control for testing the fidelity and professionalism of the recruiters with respect to the task assigned. Besides allowing for the answering of questions about discrimination toward various groups, the multiple trials that comprised the experiment helped to obscure the study's focal point on any single group. By including several disadvantaged populations in the experiment, the objective of measuring gender discrimination was concealed, preventing recruiters from adjusting their evaluation practices based on anticipated scrutiny. This methodological aspect was key in ensuring that the recruiters' responses were genuine and reflective of their discrimination behaviors in the regular course of work. Recruiters carried out their assessments using an online platform that we had developed to closely mimic real-world hiring evaluation software used in remote work, emphasizing the realism of the employment task in the post-pandemic era.

Counter to our initial hypothesis, the results revealed that when presented with observationally equivalent candidates, recruiters favored women over men in their job referrals. The analysis of the experimental data revealed a difference of nearly 5 percentage points in the selection rate between women and men, translating to women being chosen around 10% more often than their male counterparts. This preference for female job applicants is not consistent in terms of magnitude across all recruiters, however, it is predominantly influenced by male recruiters. Male recruiters demonstrated a statistically significant preference for female applicants, with a 10.35 percentage point difference in their likelihood of selecting women over men. Female recruiters also showed a preference for female applicants, but to a lesser extent, with a 2.66 percentage point difference (which was not statistically significant). This preference for women was observed in relation to the majority of positions, with female candidates being favored in seven out of ten job roles examined. However, there also was a conspicuous manifestation of traditional gender roles. For instance, women were strongly preferred for cleaning positions, whereas men were more likely to be selected for maintenance roles. Interestingly, recruiters chose women candidates despite their not re-

porting discernible differences in their assessments of job fitness among the candidate pairs evaluated.

To provide a formal rationale for our findings, we developed a model of choice in which recruiters gain more utility when they choose the more productive of two job candidates. Those choices are shaped by discrimination, a behavior that we model as a weighted average between the influence of prejudices and stereotypes. If we assume that a recruiter's prejudice remains constant across the trials of the experiment, we can estimate discrimination coefficients with and without recruiter fixed-effects, and disentangle the impact of discrimination that runs through prejudice from that which runs through stereotyping. In our formulation, prejudice is akin to taste-based discrimination (Becker, 1957). Stereotyping differs from statistical discrimination, as elucidated by Arrow (1971) and Phelps (1972), in that it involves the influence of prejudice on recruiters' beliefs, which in turn guide their biased choices. In our data, we can "control" for prejudice and isolate a discrimination parameter that quantifies the independent impact of statistical discrimination—that is, drawing inferences from the genuine population distribution of unobserved attributes—and its interplay with prejudice by drawing inferences from a skewed population distribution of unobserved attributes that reflect the recruiter's biases and choice context. This interplay characterizes what we define as stereotyping.

Interestingly, although there is a noticeable preference for hiring women, the evaluations of job fitness between male and female candidates are nearly identical. The discrimination coefficient, which measures assessments of job fit, is very small—close to zero—and statistically indistinguishable from zero. This indicates that despite the similar perceptions of productivity between genders, the preference for hiring women is likely influenced by factors other than on-the-job productivity assessments. According to our model, when differences in productivity assessments do not dictate recruiter choices, social norms may play a significant role. Specifically, we suggest that prevalent social norms advocating for gender equality could be driving the dynamics that, as we showed, are narrowing the employment outcomes gap between men and women in Ecuador. This is consistent with the observation that closing this gap may require providing women with preferentially favorable access to job opportunities, a key finding from this research.

Our paper extends the existing body of research on discrimination against women by documenting patterns of both discrimination and occupational segregation in Latin American labor markets, a region that to date has been less emphasized in this context. This study is unique in highlighting the gender-driven nature of recruiter choices and the complex role social norms play in hiring decisions. We contribute to the empirical evidence on gender discrimination generated by research that has assessed gender differences in contexts such as hiring, employment, and referrals such as Neumark et al. (1996); Goldin and Rouse (2000); Black and Strahan (2001); Bertrand and Mullainathan (2004); Moss-Racusin et al. (2012); Reuben et al. (2014); Baert et al. (2016); Bohnet et al. (2016); Sarsons (2017), and Blau and Kahn (2017), the overarching finding of which is that differences in occupation and industry are significant drivers of the gender wage gap in the United States.

Our field experiment not only allows us to enhance the empirical literature by applying alternative methods to analyze the process of discrimination against women, but it also provides actionable insights that can inform policy and intervention strategies that effec-

tively address gender disparities in the labor market. However, it is important to note that our research does not encompass all possible manifestations of discrimination against women in the labor market, such as issues related to the glass ceiling, promotion gaps, work-life balance challenges, sexual harassment, and complications arising from flexible working hours and limited benefits. [1]

While path-breaking research has documented gender bias in labor market decisions (Goldin and Rouse, 2000), systemic discrimination against women is not found across all sectors and career stages in high-income countries (where most of the empirical evidence comes from). A large-scale correspondence study in the United States found that, on average, CVs with names of women are as likely to be contacted as those with names of men, with some firms exhibiting bias against hiring women and others against hiring men (Kline et al., 2022, 2024)). The gender composition of the firm has been shown to be related to the level of hiring bias ((Hangartner et al., 2021; Galos et al., 2023)). Job candidates in Sweden faced a 7% callback penalty in professions in which persons of their gender comprise a small share of the profession (Hangartner et al., 2021). Rigorous studies of the job promotion process have tended to find bias against women employees, after controlling for productivity, including at financial institutions and large retail chains (Huang et al., 2024; Benson et al., 2023). Huang et al. (2024) find through study of marginally promoted bank managers at over 1000 financial institutions that the marginally promoted male manager subsequently reduced the productivity of their subordinates by 3% compared to an average manager, whereas the marginally promoted woman manager increased productivity by 2% and Benson et al. (2023) find that despite receiving higher job performance ratings, women were rated as having lower potential, which explained half of the gap in promotions. Women subsequently outperform men with the same potential ratings and are not more likely to leave the firm.

Correspondence studies in Latin America have not documented bias against hiring women, except in specific sectors and for specific population groups. A study in Chile (Bravo et al., 2008) found similar callback rates by gender, while women received fewer callbacks among the less-skilled applicants. In their correspondence study in Mexico, Arceo-Gomez and Campos-Vazquez (2014) find that women, on average, are 3 percentage points more likely to receive a callback invitation than men, using CVs that control for education and experience. However, married indigenous women have lower call-back rates than both married indigenous men and their married white peers.

Our paper, while focused on Ecuador, includes empirical findings that contribute to the broader labor economics literature. At the same time, we extend the current scope of field experimental designs to isolate the effects of gender discrimination in hiring practices, a methodology that can be replicated in diverse geographic and economic contexts. Moreover, the findings broaden existing theories about gender biases in labor markets grounded almost exclusively in distinguishing taste from statistical discrimination, suggesting a new approach to understanding and addressing these issues globally, with insights that are applicable in both developed and other developing economies. Thus, the paper's scope

---

[1] Examples of scholarly research on those topics include Cotter et al. (2001) on the glass ceiling effect, Kelly and Kossek (2011) on work-life balance challenges, McLaughlin et al. (2012) on sexual harassment, and Chung (2020) on complications arising from flexible working hours and limited benefits.

extends well beyond its specific study area, offering implications that are relevant for a general audience interested in the role of gender discrimination in labor market dynamics and economic inclusion.

In assessing the external validity of our findings, we rely on the SANS framework (standing for "selection," 'attrition," "naturalness," and "scalability") suggested by List (2020). With reference to the *selection* aspect of our research, we highlight that our data joined 392 experienced human resource recruiters, drawn from a convenience sample of individuals currently working or looking to work in recruitment roles across various industries in Quito, Ecuador. A comparison of the recruiters in our sample with those in national representative occupational statistics (ENEMDU) showed that the recruiters whom we sampled were younger, with higher education levels, and more likely to be women than recruiters from the general population, sampled in the national statistics of Ecuador.

Attrition rates in our study underscore the commitment of participants to the experimental process, with a substantial majority completing the full suite of tasks. Out of the 392 engaged in the study, 11 completed one trial, 66 two trials, and 315 all three, yielding 2,176 observations for analysis. This high completion rate suggests minimal bias from dropout.

Concerning the *naturalness* of our experiment compared to working for a firm, we highlight our experiment was conducted through an online platform, reflecting modern recruitment practices where remote candidate evaluation and referral have become increasingly commonplace. This alignment with current industry practices, coupled with previous evidence from field experiments Zanoni et al. (2022, 2023) that show consistency in discrimination patterns across different methodologies, underscores the real-world applicability of our findings. The observed variance in discrimination coefficients across different groups—women, migrants, and gay and lesbian candidates—further underscores the seriousness with which recruiters approached the task, so that we can be confident the results genuinely reflect biases present in actual hiring scenarios.

Regarding *scalability*, our experiment's design and execution imply that the discriminatory patterns identified may reflect wider trends within Ecuador's formal employment sectors. The methodological approach utilized in our field experiment mimics market practices, given the increasingly widespread practice of working remotely. In fact, our tool can be adapted to monitor discrimination dynamics in the labor market that extend beyond gender as a basis for prejudice. This approach is both flexible and reliable, being able to accommodate the evolving nature of workplace discrimination and the growing prevalence of remote recruitment practices.

This paper is organized as follows: section 2 provides a review of the relevant literature, laying the groundwork for the study's theoretical framework. Section 3 describes the methodology employed in our field experiments, including the design, sampling techniques, and data collection processes. Section 4 presents the results of our experiments, offering detailed statistical analyses and interpretations of the findings. Section 5 discusses the implications of the results for understanding gender discrimination in hiring practices, integrating insights from both the empirical data and established theories. Finally, Section 6 concludes with a summary of our key findings, recommendations for future research, and potential policy implications arising from our study.

2 A Model of Discrimination in the Recruitment Process

In this section, we introduce a basic microeconomic model designed to describe the behavior of recruiters when tasked with making a hiring decision between two candidate profiles. We believe presenting this model to the reader is crucial, because it helps to describe the recruiters' decision-making processes and assesses the field experiment's realism versus real-world hiring scenarios. By modeling the recruiters' behavior, we provide insights into the sources of potential biases regarding their gender preferences. The model also supports the investigation of how altering decision-making parameters in an experimental setting might influence recruiters' responses, offering further insights into the experiment's external validity.

In our formulation, a recruiter is tasked with making a hiring referral after assessing the CVs of two job applicants $i(1, 2)$, where $i = 1$ is a man and $i = 2$ a woman. The recruiter's utility function depends on whether the recruiter accurately refers the better candidate in that pair to the firm. In so doing, the recruiter makes a productivity assessment of each candidate, which can misrepresent the latent productivity of the applicants. Calling $Q_i(D_i)$ the productivity assessment and Q_i^*, the latent productivity, we relate both additively through the measurement error $\epsilon_i = Q_i^* - Q_i$. The recruiter realizes higher utility through the referral of the candidate with the highest latent productivity Q_i^* for the job:

$$U = U(\max(Q_1^*, Q_2^*)) = U(\max(Q_1(D_1) - \epsilon_1, Q_2(D_2) - \epsilon_2)). \tag{1}$$

We introduce a discrimination function D_i as an argument of the assessment function $Q_i(D_i)$ to suggest the idea that the recruiter's assessments of the productivity of the applicants can be influenced by discrimination on the recruiter's part. D_i enters the productivity assessment function as the core component of the behavioral model; this variable quantifies the discrimination level a candidate, i, experiences from the recruiters' productivity assessments. The discrimination function D_i is a weighted average that balances through α the impacts of inherent prejudice, P_i, against stereotype-based judgments, S_i, as in the following equation:

$$D_i = \alpha P_i + (1 - \alpha)S_i(R, K, \omega, B_i(O_i, D_{fi}, D_{si}|P_i)). \tag{2}$$

The prejudice component, P_i, reflects the evaluator's innate biases; it is akin to the concept of taste discrimination (Becker, 1957). Stereotypes, represented by S_i, are conceptually linked to the idea of statistical discrimination Phelps (1972); Arrow (1971), whereby bounded rationality and/or the limited availability of information with which to form rational expectations about the unobserved productivity traits of candidates lead recruiters to make statistical inferences to fill the void. Our concept of stereotype is close to the "social condition approach" explained by Bordalo et al. (2016), where stereotypes are conceptualized as "intuitive generalizations that individuals routinely use in their everyday life, and they entail savings on cognitive resources." These authors explain that while stereotypes are rooted in true underlying differences between groups, they are also subject to variation

related to context and can be distorted or influenced. Here stereotypes emerge or "are produced" from the behavior of the recruiter that combines resources (R—time and money) with the recruiter's human capital $(K)^2$ in a setting where the recruiter utilizes a specific technology, ω, to assess an applicant's productivity.

The key component of the stereotyping function S_i is the belief B_i that the recruiter formulates about the productivity of the applicant i. A productivity-related belief results from the drawing of a distribution of productivity-related unobserved attributes from a broader "belief system" that encompasses characterizations of applicants' types. When resorting to beliefs, the attributes and the distribution in the system from which the former are retrieved to complement a candidate's unobserved productivity will depend on the attributes the recruiter has available in the CV O_i (that is, the set of observable characteristics in the CV). In addition, the unobserved productivity construct will also be shaped by the recruiter's inherent prejudices P_i. Prejudice and stereotyping thus interact. Our framework, which examines how beliefs shape discrimination, mirrors the discussion in Bohren et al. (2023, 2019); Agan et al. (2023), all of whom conceptualize how recruiters' beliefs can often be misaligned with a group's actual characteristics and explore some economic implications of these misconceptions.

Two additional parameters D_{fi} and D_{si}, incorporated into the belief function B_i, broadly represent the recruiter's second-order beliefs regarding the extent of discrimination exhibited by both firms and society toward individuals with attribute i. These parameters can be conceptualized as weighted sums of discriminatory behaviors observed within firms (including formal and informal policies and organizational culture) and within society (encompassing prevailing social norms). Because individuals exert varying degrees of influence within firms and society, the mentioned weights shape the differences in their influence within the recruiter's organization and within society at large, constructing the analytical narratives (as defined by Shiller (2017)) that contribute to stereotypes about women's productivity and the challenges they face in the labor force. D_{fi} speaks to the idea that the recruiter makes a referral considering what the recruiter thinks the firm might be looking for in a candidate (the productivity of the applicant), including the probability the applicant will be able to integrate within the firm. D_{si} is included to characterize the recruiter's second-order beliefs about what the social norm is, assuming that this would also influence the recruiter's referral choices, as has been documented in previous studies of discrimination **?**.

3 Labor Market Disparities by Gender: Survey Evidence from Ecuador

In this section, we examine the disparities in employment rates between men and women within Ecuador's labor force. Specifically, we study to what extent these disparities can be attributed to observable factors by us (researchers), as opposed to those unobserved in household survey data. This analysis sets the stage for our experiment and outlines the

---

^{2}Human capital encompasses both cognitive and noncognitive abilities.

motivation behind the present study of discrimination. We utilize the 2022 Ecuadorian National Household Survey (ENEMDU-INEC, 2022) data for conducting an Oaxaca decomposition to analyze the variance in employment probabilities between men and women. The ENEMDU-INEC dataset contains 358,096 observations. By narrowing our focus to Ecuador-born individuals ages 18 to 65 who are either formally employed[3] or unemployed, our analytic sample joins 64,973 individuals. Properly weighted, that sample portrays a population of 2,688,909 individuals (59% men and 41% women; 13% of that subpopulation are unemployed).

Performing a standard Oaxaca-Blinder decomposition exercise, we produce conditional comparisons in employment rates, holding constant differences in ages (and age squared), educational achievement, fixed effects by province, race/ethnicity, place of birth, the individual's relationship with the head of household, and marital status. We find statistically significant differences in employment probabilities between the two groups. Conditional on those factors, the predicted probability of employment for men was 0.893, compared to 0.835 for women. The 0.058 percentage points gap in employment probabilities indicates a disparity disfavoring women in the labor market. When analyzing the components of this disparity using the Oaxaca-Blinder method, we can attribute about 0.020 of the gap between genders to differences in the observable characteristics for which we controlled. Note, however, that the decomposition also shows that a larger portion of the gap, approximately 0.037, remains unexplained by the model's predictors, which could reflect the influence of discrimination. Exploring whether the employment gap in table 2 is driven by discrimination exercised by recruiters is an open research question that we seek to address in this paper.

4   Experimental Design

In 2022, we implemented a multipurpose AFE in Quito, Ecuador's capital, to measure discriminatory behaviors of recruiters toward women, migrants, gays, and lesbian job seekers.[4] Two companion papers (? and Zanoni and Díaz (2024)) utilize data from the same AFE experiment conducted here to study discrimination against gay and lesbian job seekers and Venezuelan migrants, respectively, in the same market.

We hired 392 human resource recruiters and provided them with 10 pairs of applications for fake job postings (trials), presented in random order. Each recruiter was asked to evaluate up to three pairs of profiles of job applicants, matched for productivity equivalence, with the only differing factor being the applicants' randomly assigned gender. The remainder of the experiment entailed comparisons between migrants and locals, lesbians

---

[3]We coded formality if the employee has a job with access to social security benefits.

[4]Examples of research conducted using similar methods are outlined in Bertrand and Duflo (2017), Gaddis (2018), and Neumark (2018). Recent instances of similar AFE experiments measuring discrimination in Latin America include Zanoni et al. (2023) in Argentina, ? in Ecuador, and Zanoni and Díaz (2024) in Colombia. Moreover, Lahey and Beasley (2018) employ recruiters to investigate discrimination against African American job applicants. However, their methodology involves iteratively rating multiple attributes of candidates within brief time windows, as opposed to the dedicated evaluation of each candidate in pairs, the approach we take.

and gays versus straight candidates by gender, and an additional placebo trial devoid of any potentially prejudiced attribute (resulting in a sum of ten experimental trials per recruiter). The recruiters were tasked with recommending that one candidate in each pair be hired, as well as with rating the fit for the job and making wage recommendations for all of them.

We developed an online platform for presenting candidates to recruiters, collecting their attributes, and gathering their evaluations of the job candidates securely. By randomly varying gender across the job applicants in each pair, we aimed to isolate the effects of the gender of the applicant on the recruiters' discrimination behaviors. Our AFE is similar to a correspondence study (CS) of the type popularized by Bertrand and Mullainathan (2004) and extensively replicated across the world,[5] in that it seeks to gauge the revealed preferences of actual recruiters in the field. It differs from CS studies in that the environment where the recruiter makes choices is also simulated; in this simulation, we utilize a virtual online platform designed to closely resemble authentic remote work environments in human resources. Our methodology is similar to other papers that analyze a set of recruitment decisions made by specialists who are outsourced to firms (Agan et al., 2023). The practice of contracting third-party recruiters rather than using in-house procedures has become more common in recent years (Agan et al., 2023).

To optimize the outcomes of the AFE, we partnered with two research organizations, Grupo FARO (a local NGO with expertise in research and policy advocacy in relation to minorities and recruiting of personnel in Ecuador) and ANOVA (a consulting firm with field expertise measuring discrimination toward minorities in Latin America). They guided us in selecting, hiring, and paying local recruiters and in framing the experiment online to ensure its credibility and relevance. Their role was to guarantee that the recruiters perceived the task they had to carry out as genuine work with significant implications for recruitment through the construction of an environment closely resembling platforms used today for job referrals during remote work scenarios. They also helped ensure that the recruiters had proper professional experience and qualifications.

To generate samples of the recruiters, we employed two methods, LinkedIn advertisements and referrals via respondent-driven sampling (RDS)[6].[7] The experiment aimed to authentically replicate real-world job recruitment scenarios. To achieve this, we conducted interviews and focus groups with local NGOs, academia, and firms. These consultations served two purposes: one, we validated the observational equivalence of the pairs of profiles in each experimental trial, and two, we gathered insights on how to better tailor the online platform to resemble actual job evaluations. As a local organization with prior experience in making hiring recommendations, Grupo FARO facilitated the recruitment of recruiters and subsequent interactions during the experiment to assure the recruiters they were participating in genuine hiring exercises.

Grupo FARO also conducted interviews and focus groups with local stakeholders (NGOs, minorities' rights organizations, academics, and hiring firms) to understand the challenges faced by women job seekers in Ecuador's labor market. In addition to making use of the

---

[5]Evidence of gender discrimination using field experiments is summarized in Schaerer et al. (2023)

[6]On the RDS method, see Heckathorn (1997, 2002)

[7]In 2023 there were 4 million LinkedIn users in Ecuador.

qualitative insights those interviews yielded, we drew on the yearly national household survey (ENEMDU, 2022) to craft job vacancies where the data showed women and men competed for jobs. The synthetic pairs of resumes were designed to closely resemble actual job applications, as portrayed in samples of CVs we gathered, so that we could both produce credible job applications and ensure comparable characteristics among candidates except for the gender attribute. Balance tests of equivalence in attributes of the applicants (see table 6) confirmed balance in age, qualifications, and demographics among the fictitious job applicants we designed. We presented the recruitment task as a genuine hiring exercise, highlighting the need for human resources (HR) analysts and/or HR recruiters experienced (with at least two years) in the Quito labor market. We emphasized their roles would entail rating productivity as well as making hiring recommendations of and salary referrals for preselected candidates for jobs with an international consultancy company.

The experiment unfolded as follows: First, the recruiter was contacted and requested to work for us via LinkedIn or by personal referral within the RDS design. Those who accepted were subject to eligibility scrutiny (they needed to be older than 18 and have at least 2 years of experience in HR). Second, we hired those applicants who were eligible with payment conditional on the completion of the task, which was to make hiring and salary recommendations and assess fit for the job of ten pairs of candidates applying for ten jobs. They were asked to go online to complete a survey about themselves where we asked them provide full descriptions of their demographics, work experience, and knowledge of Ecuador's labor market. They also had to complete a test of cognitive and socio-emotional development as well as self-esteem evaluations. After completion of the survey, they proceeded to the online evaluation of the candidate pairs.

Our research design has some benefits compared to a CS, which is the most widespread method to measure discrimination in the field. It is more time and cost-efficient, bypassing the extensive monitoring and customization of applications required for a CS. Additionally, we are able to mitigate the potential impact of job market seasonality on data interpretation. Our AFE also provides access to diverse information beyond callbacks, for example wage data, assessments of candidates' productivity, and scale ratings of productivity attributes (as in Lahey and Beasley (2009)), enabling a richer analysis. Unlike CS, AFEs avoid the issues stemming from universally low response rates (which can skew results if correlated with experimental conditions), thereby enhancing the reliability of findings Heckman and Siegelman (1993). Finally, AFEs address an ethical concern associated with CSs: while participants may still be deceived in evaluating fake candidates, they are compensated fairly for their time with a competitive wage. Note that when compared to a CS, where recruiters make decisions within established firms with tangible implications for their own careers, the reliability of AFE experiments depends on recruiters' believing that their task mirrors a genuine job with significant hiring implications. The resemblance of the online platform to the familiar choice environment encountered by recruiters is a crucial factor in the success of our experiment.

One key advantage of our AFE is that it incorporates multiple evaluation trials within the same experiment, adding a longitudinal dimension that is typically absent in CSs. This longitudinal aspect enables us to control for what we consider a fundamental recruiter-specific, time-invariant factor: prejudice. We operate under the assumption that prejudice

toward specific groups remains stable throughout the experiment's duration. By implementing a recruiter fixed-effects estimator, we are able to isolate the contributions to discrimination that stem directly from prejudice, which are distinct from those arising from stereotyping (as discussed in section 2).

Concerns regarding recruiters' choices and potential biases were addressed through clear instructions, validation of task understanding in pilot tests, and real-stakes remuneration. Ethical considerations were meticulously addressed through discussions with an Institutional Review Board (IRB) committee. Postexperiment focus groups were conducted to ensure participants' understanding of the experiment's purpose; the feedback attested to the integrity and reliability of the exercise.

In table 8 in the appendix, we compare some key attributes of the recruiters in our sample with those from national occupation statistics. We joined occupational statistics from INEC (yearly data from the National Household Survey ENEMDU in 2021, 2022, and 2023) to characterize the attributes of recruiters from nationally representative data for Ecuador. The data in the table show that our sample joins younger, more women and more university graduate recruiters than does the market. The age and educational attainment differences are not surprising, provided the sampling strategy departs from LinkedIn, which is a social platform joining primarily young professionals. We attribute the gender differences (proportionally more women recruiters than men in our sample than in the household surveys) to more women than men's being LinkedIn users in the country.

4.1 Empirical strategy

To estimate the effect of gender on the outcomes of interest we use ordinary least squares (OLS) regression to estimate the parameters of models of the form:

$$Y_{itr} = \beta_0 + \delta D_{it} + \beta X_{it} + \gamma T_t + \epsilon_{itr}, \tag{3}$$

where Y_{itr} is a vector of responses for recruiter r coding labor market variables associated to candidate i in trial t, with the variable T_t being trial fixed effects. There are three dependent variables: (1) *Callbacks*, an indicator that has a value of one if the candidate is selected by the recruiter (zero if not); (2) *Fit for the job*, which is a ranking on a Likert-type scale of 1 to 10 summarizing assessments of the candidate suitability for the advertised positionl and (3) *Good fit for the job*, a dummy variable that indicates if the candidate received a score of 8 or more on the Likert-type scale in *Fit for the job*. The variable D_{it} is an indicator for whether the candidate being evaluated was a woman, in which case it has the value of one (and zero if this is not the case). X_{it} is a vector of controls, which we will explain in detail in the empirical section and the variable ϵ_{itr} represents unobserved heterogeneity.

Our subject of interest is the δ coefficient from equation 3 (the discrimination coefficient henceforth), which we estimate for the full sample and for selected subsamples by gender of the recruiter and the level of skills required to perform the job. We also conduct heterogeneity analysis of the δ coefficient by splitting the data into relevant subsamples (for example, by position advertised, trials, skills, etc.). The results were validated by studying the coefficient estimates of job fixed effects interacted with the gender indicator.

In our analysis of gender discrimination in the hiring process, we construct two empirical models to dissect the underlying mechanisms of discrimination: one that captures the combined effect of prejudice and stereotyping and another that isolates the effect of stereotyping alone. This distinction is critical for understanding the nuanced dynamics between inherent biases and stereotypical judgments in hiring decisions.

To isolate the effect of stereotyping, in line with the model developed in section 2 we extend our analysis by incorporating recruiter and trial fixed effects in equation 3. Note that because we expose each recruiter to multiple trials of the experiment, we can employ a recruiter fixed-effects model. In this case it takes the form below:

$$Y_{itr} = \beta_0 + \delta_{fe}D_{it} + \gamma T_t + \lambda R_r + \beta X_{it} + \epsilon_{itr}. \tag{4}$$

In this formulation, R_r represents dummy variables for recruiter fixed effects and δ_{fe} specifically measures the impact of stereotyping (as defined in section 2) on discrimination, adjusting for both recruiter preferences and trial-specific factors. By comparing δ from the first model and δ_{fe} from the second model, we can deduce the direct influence of prejudice on discrimination, with δ_{fe} isolating the stereotyping effect. This approach provides a clearer picture of how stereotypical judgments, independent of inherent biases, influence recruiters' choices of candidates and assessments of job fitness.

4.2   Data

Table 6 compares male and female applicants across various attributes such as age, previous jobs, employment experience, and educational qualifications. Consistent with the balancing of observable attributes built into our research design, the data in the stable show that there are negligible differences between the genders in most categories, indicating a balanced representation (none of the differences are statistically significant at conventional levels of precision). Specifically, the average age, number of previous jobs, and years of employment experience are nearly identical between male and female candidates, demonstrating that the pool from which candidates are drawn is uniform with regard to experience and age. Additionally, professional status and levels of education (ranging from secondary to professional degrees) show no significant variance between genders. This balance in observables suggests that the recruitment process is equitable and that both male and female candidates have similar qualifications and backgrounds. The absence of discrimination and randomization of gender should lead to unbiased selection, similar wages, and assessments of productivity on the part of the recruiters. Of the 392 recruiters, 11 completed one trial, 66 recruiters completed two trials, and 315 completed all three trials, yielding an analytic sample of 2176 observations.

4.3   Characteristics of the recruiters

Tables 3, 4, and 5 in the appendix A provide a comprehensive overview of the characteristics of recruiters divided into three types: (1) demographics and education, (2) scores on standardized tests, and (3) performance in the experiment.

As we can see in column 1 of table 3 in appendix A, the recruiters we hired were mostly young (averaging 31 years old) and the majority were women (70% of the sample). Most of them had college degrees (92%) and some had master's degrees (21%). In terms of their work experience, they averaged around 7 years, with nearly 5 of those years spent working as HR recruiters.

When examining differences in those dimensions according to the sampling method—columns 2–4—we first observe that recruiters hired using RDS were, on average, two years younger than their LinkedIn counterparts. Recruiters in the former group also had 1.6 fewer years of overall work experience, a difference that almost matches the differences in years of experience working as HR recruiters between them and those hired using the LinkedIn method. In terms of educational credentials, recruiters hired by the RDS method were 11% less likely than their counterparts to have a bachelor's degree.

Table 3 in Appendix A shows some socioeconomic variables that characterized recruiters by hiring method. As can be seen, regardless of the method, recruiters were similar in terms of gender, age, nationality, and proportion having an HR-focused university degree[8]. The LinkedIn recruiters had slightly higher levels of education, more years of experience, and better knowledge of the Quito labor market. However, a higher proportion of RDS recruiters responded that they were employed.

5   Results

In table 7 we show our main results. Columns refer to specifications. Columns 1–3 show the discrimination coefficient estimates for δ from 3 as the model specification changes according to the set of variables described in the panel at the bottom of the table. Columns 4 and 5 show the estimates by the level of skills required for the advertised job and columns 6 and 7 do the same by the gender of the recruiter. The row panels split results by the outcomes of interest—A. *Callbacks*, B. *Good fit for the job*, and C. *Fit for the job*. Our preferred specification is model 3, where we include an indicator variable for whether the recruiter was hired using LinkedIn or RDS (there were some differences in some attributes across recruiters in each sample and that indicator attempted to deal with those discrepancies). In addition, we include a set of indicator variables for whether the recruiter opened information tabs, the position advertised, the trial of the experiment, and recruiters' fixed effects with standard errors robust to heteroskedasticity.

As illustrated in row A of table 7, women are selected for jobs at a rate 10.4% higher than male applicants, when considered as a proportion of male candidates chosen. This discrepancy is even more pronounced when the behaviors of male recruiters specifically are considered, as shown in column 6; the impact here is more than double that of the average discrimination coefficient in model 3. Furthermore, a comparison of the discrimination coefficients shows that the positive bias exhibited by male recruiters toward female applicants is nearly four times larger than that shown by female recruiters (for male recruiters the value is 0.1035 and for female recruiters it is 0.0266). Interestingly, as shown by the

---

[8]The proportions of the LinkedIn and the RDS samples holding a degree in human resources (or related) were 75% and 74%, respectively

results in columns 4 and 5 of the top panel of the table, the skills required for the job did not significantly alter the weighted average represented by the discrimination coefficient, indicating that the observed biases are consistent across skills.

The observed bias favoring the hiring of women occurs despite recruiters' having assessed the productivity of male and female candidates as equivalent. This pattern is substantiated by the data presented in rows B and C of table 7. Assuming that the assessments of job fit serve as proxies for productivity, the table reveals a critical finding: there are no statistically significant differences in either measure between male and female candidates, with the coefficient estimates being very small and near zero. This preference for hiring women, documented in the experimental data, suggests that factors beyond mere productivity assessments are influencing recruiter decisions. The result aligns with the theoretical framework developed in section 2, where we integrated social norms into the utility function that drives the choice of the recruiter. In our model, recruiters gain utility not only from selecting the most productive candidate, with productivity understood in the classical sense of the value of the marginal product of labor, but also from adhering to prevailing social norms, which may include the promotion of gender equality. These norms potentially influence hiring practices by encouraging positive discrimination toward women, which we hypothesize is a societal response to historical gender imbalances in the job market.

The panel to the left of figure 1 presents the discrimination coefficient for women across the ten different occupations evaluated in the experiment. The most noteworthy observation is that the data reveal women are favored significantly for jobs such as General Services Assistant - Cleaning (with a discrimination coefficient of 0.337***) and Call Center Operator (0.1644*), suggesting a gender bias that aligns with traditional gender roles. Conversely, there is negative discrimination against women in relation to jobs such as Commercial Advisor and Maintenance Technician, as indicated by negative coefficients (-0.1204 and -0.2945**, respectively). However, with regard to several other jobs, such as Warehouse Keeper, Certified Public Accountant (CPA), and Software Developer, the discrimination coefficient is close to zero, suggesting little to no bias in hiring based on gender. These figures underscore the complexity of gender discrimination in the labor market, with significant variation across different job types.

We conduct some robustness tests to bolster confidence in the reliability of our results. First, we evaluate whether the behavior of recruiters was consistent throughout the ten trials of the experiment by testing the differences in magnitudes of the discrimination coefficients when assessed across the experiment (in trials 1–3, 4–6, and 7–9). These results are also shown in figure 1. Setting aside the results for trials 1–3, associated with the greatest effects, we find no statistically significant differences in discrimination coefficients for women throughout the experiment. This speaks to the professionalism with which recruiters approached the task; broader and inconsistent estimates would have suggested a lack of seriousness. Despite the duration of the experiment, recruiters remained focused and efficient in their evaluations.

At the same time, it should be noted that throughout the data collection experiment, discrimination coefficients did exhibit variation in sign and magnitude across the evaluated groups (namely, women, migrants, gays, and lesbians). A companion paper utilizing the same data source finds distinct discrimination coefficients for gays and lesbians (-0.0796

and 0.1681, respectively), both statistically significant at the 95% confidence level (**?**). In figure 1, we illustrate that migrants also face discrimination in the Ecuadorian labor market (**?**).

The finding that the discrimination coefficients vary across different groups when all the groups are subject to prejudice further underscores the validity of our results. By observing different rates of discrimination across various groups, our study provides evidence that discrimination is not a monolithic, uniform behavior, instead varying in complex ways that are likely influenced by the prejudiced attribute judged. This variability is consistent with established psychological theory and suggests that our method captures these nuances effectively.[9]

To assess whether our findings could have been influenced by experimenter demand effects, we included a placebo round in the experiment, which revealed no evidence of discrimination among identical applicants. One of the ten experimental trials served as a placebo group, where pair members were equally qualified for the job and of the same gender. We compute the discrimination coefficient for the trials consisting of placebo groups and find no statistically significant coefficients. The absence of discrimination observed in the discrimination coefficient estimate at the bottom of 1 in this placebo group further supports the credibility of our experimental findings.

Finally, a robust indicator of recruiters' dedication to the task can be derived from their approach to evaluating candidates' *Fit for the job*. As previously mentioned, both the alignment of candidates' skills with the job requirements and the offered wages serve as proxies for the anticipated value of their contributions (that is, the value of their marginal product of labor). Consistency in how recruiters evaluated candidates across these two metrics, as shown in table 7, suggests the recruiters took the task seriously. Moreover, the alignment of the proposed wages with prevailing market rates for candidates' experience and education levels underscores the thoroughness and professionalism with which the recruiters approached their tasks.

Interpreting the results through the conceptual model we introduced provides insights into the nature of gender discrimination in the hiring process. The finding that female candidates were selected for jobs at a significantly higher rate than male candidates, especially by male recruiters, challenges traditional expectations of gender bias's favoring of men. This preference exists despite recruiters' assessing the productivity of male and female candidates as equivalent, as evidenced by similar wages and job fit metrics. This suggests that the recruiters' preference for hiring women over men stems from factors beyond mere productivity assessments. According to our model, such preferences might be influenced by a mix of inherent prejudice (P_i) and stereotype-based judgments (S_i), pointing toward taste-based discrimination where recruiters' personal biases or societal norms regarding gender roles subtly influence their decisions. However, in light of the labor market trends we described, according to which women are gaining more access to the labor market, the experimental result is reasonable: without the favoring of women during the path-to-equality period, those gaps would never close.

---

[9]Levitt and List (2007) cite multiple examples in the early psychological literature on behavioral consistency that suggest behavior, including discrimination, is not uniformly consistent across different situations or contexts (Mischel (1968); Ross and Nisbett (1991); Hartshorne and May (1928)).

Moreover, the discrimination coefficients across different occupations highlight the role of stereotypes S_i in shaping hiring decisions. The significant preference for women for jobs traditionally dominated by females and the negative bias against them for male-associated jobs illustrates how societal stereotypes about gender roles influence recruiter behavior. This variability underscores the complexity of gender discrimination, suggesting it is highly contingent on the context and perceived gender appropriateness of the job. The model's discussion on stereotypes as "intuitive generalizations" offers a theoretical basis for understanding these patterns, suggesting that in the absence of clear productivity differentials recruiters may rely on societal cues about gender roles when making hiring decisions.

5.1 Prejudice and stereotypes

This section ends with an exercise to separate the direct effect of prejudice from the effect of stereotypes. To do so, we compare the δ coefficients from model 3 with the δ_{fe} coefficient estimates from model 4. Table 1 below compares those estimates, computing their differences and testing the statistical significance of those differences. Interpreting the magnitude of the δ with the δ_{fe} coefficients, we can attribute most of the impact of discrimination to stereotyping S_i.

Table 1: Differences between Model 2 and Model 3

	No recruiter fixed effect	Recruiter fixed effect	Difference
A. Callback			
Discrimination Coeff.	0.0494**	0.0494**	0.0001***
	(0.0215)	(0.0238)	
Mean Male (on that subsample)	0.4752	0.4752	
Observations	2176	2176	
B. Good fit for the job			
Discrimination Coeff.	-0.0068	-0.0066	-0.0002**
	(0.0151)	(0.0139)	
Mean Male (on that subsample)	0.8539	0.8539	
Observations	2176	2176	
C. Fit for the job			
Discrimination Coeff.	0.0383	0.0388	-0.0005***
	(0.0503)	(0.041)	
Mean Male (on that subsample)	8.5754	8.5754	
Observations	2176	2176	

Note: This table exhibits the discrimination coefficients across the regression models that include recruiter fixed-effects and those that do not. The last column displays the difference between coefficients; to test whether their difference is statistically significant, we perform a Wald test where we build the W statistic as $\left(\frac{\beta_{\text{No FE}} - \beta_{\text{FE}}}{se(\beta_{\text{No FE}})}\right)^2$, which under the null hypothesis follows an asymptotic χ^2 distribution with one degree of freedom. Stars indicate the statistical significance of differences in means across groups at various significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

As explained in section 2, we model S_i as a function of social norms $D_s i$ that affect the recruiters' decision-making processes. In fact, data from the Gender Social Norm Index for Ecuador shows that nearly 61% of the population do not agree with the statements "Men

should have more right to a job than women" and "Men make better business executives than women do"UNDP (United Nations Development Programme) (2023). This perspective gives margin for our results as being driven by a social preference of a country seeking gender equity.

6 Discussion and conclusion

In conclusion, our AFE involving recruiters in Ecuador indicates a preference for female candidates, particularly driven by male participants, reflecting significant gender-driven biases within the hiring process. This trend, aligning with certain traditional gender roles, reveals a form of positive discrimination favoring women in relation to most of the occupations that were included in the AFE. Notably, these biases seem less related to inferences about unobserved productivity, as evidenced by negligible differences in job fitness assessments between genders. This suggests that current hiring preferences may stem more from societal norms rather than classical views of marginal productivity.

As outlined in section 2, our theoretical model incorporates social norms into the utility function guiding recruiter decisions. Recruiters derive utility not only from selecting candidates based on productivity, but also from conforming to social norms, such as the promotion of gender equality. These norms may drive the observed positive discrimination toward women, which, as we have reiterated in the paper, appears to reflect a societal effort to correct historical gender imbalances in the job market.

Moreover, our analysis underscores the varied nature of gender discrimination across different job sectors, indicating that such biases are highly context dependent. This complexity highlights the ongoing challenges encountered in the effort to achieve genuine gender equality in the labor market and underscores the necessity for targeted measures to address these issues. While there has been progress in narrowing gender disparities in labor force participation (though less in wages), our findings suggest that hiring bias at the entry level is not the main contributing factor to the existing gender gap in wages.

Our study employs a novel multitrial field experiment methodology, where recruiters working remotely and by contract rated observationally equivalent male and female candidates. This approach enables us to analyze gender discrimination across different job roles and with reference to recruiters' characteristics, yielding a deeper understanding of gender biases in hiring practices. Unlike previous studies that primarily focus on high-income countries, our research brings a crucial perspective from urban Ecuador, thus broadening the geographical and cultural scope of labor market discrimination research (Neumark et al. (1996); Bertrand and Mullainathan (2004)). This context is essential for understanding how gender discrimination manifests in different economic environments and across time, particularly in developing countries.

By presenting empirical evidence from a developing economy, our study remedies a specific gap in the existing literature, which has predominantly focused on high-income countries (?Blau and Kahn (2017); Schaerer et al. (2023)). Our findings challenge traditional views of gender bias by demonstrating a significant preference for female candidates in the Ecuadorian labor market. This suggests that awareness of the specificity of context

and time should be incorporated into the analysis of discrimination. More specifically, the evolution of societal norms advocating for gender equality may play a more substantial role in hiring decisions than previously thought. By linking these findings to broader theoretical frameworks, our study provides new insights into the mechanisms of labor market discrimination, supporting, challenging, and expanding current theories that give almost exclusive preponderance to taste-based and statistical discrimination as the mechanisms driving discriminatory behaviors.

Finally, the practical implications of our research are significant for policymakers aiming to design interventions that promote gender equity in the labor market. Understanding the dynamics of hiring preferences as a utility maximization problem influenced by social norms is crucial for creating effective policies that address both overt and subtle forms of discrimination. Our study's insights into the preference for female candidates can inform policy decisions and interventions, contributing to the ongoing efforts to achieve equity in employment outcomes.

References

Agan, A. Y., D. Davenport, J. Ludwig, and S. Mullainathan (2023). Automating automaticity: How the context of human choice affects the extent of algorithmic bias. Working Paper w30981, National Bureau of Economic Research.

Arceo-Gomez, E. O. and R. M. Campos-Vazquez (2014). Gender and ethnic discrimination in hiring: Evidence from mexico. *Journal of Human Resources 49*(3), 667–694.

Arrow, K. J. (1971). *Some Models of Racial Discrimination in the Labor Market*, Volume 6253. Rand San Monica, CA.

Baert, S., A.-S. De Pauw, and N. Deschacht (2016). Do employer preferences contribute to sticky floors? *ILR Review 69*(3), 714–736.

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

Benson, A., F. Li, and K. Shue (2023). Gender bias in performance evaluations and promotions. *Management Science 69*(4), 1057–1080.

Berniell, I. et al. (2021). The impact of motherhood on female labor informality in chile. *World Development 139*, 105328.

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

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*(4), 991–1013.

Black, S. E. and P. E. Strahan (2001). The division of spoils: Rent-sharing and discrimination in a regulated industry. *The American Economic Review 91*(4), 814–831.

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

Bohnet, I., A. van Geen, and M. Bazerman (2016). When performance trumps gender bias: Joint versus separate evaluation. *Management Science 62*(5), 1225–1234.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, 1–45.

Bohren, J. A., A. Imas, and M. Rosenberg (2019, October). The dynamics of discrimination: Theory and evidence. *American Economic Review 109*(10), 3395–3436.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016, 07). Stereotypes*. *The Quarterly Journal of Economics 131*(4), 1753–1794.

Bravo, D., C. Sanhueza, and S. S. Urzua (2008). An experimental study of labor market discrimination: Gender, social class and neighborhood in chile. Technical Report Working Paper No. 226, Inter-American Development Bank. Presented May 2008; IDB Working Paper.

Chung, H. (2020). Gender, flexibility stigma, and the perceived negative consequences of flexible working in the uk. *Social Indicators Research 151*(2), 521–545.

Cotter, D. A., J. M. Hermsen, S. Ovadia, and R. Vanneman (2001). The glass ceiling effect. *Social Forces 80*(2), 655–681.

Gaddis, S. M. (2018). An introduction to audit studies in the social sciences. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, pp. 3–44. Springer.

Galos, D. et al. (2023). Gender diversity and bias in recruitment decisions. *Industrial and Labor Relations Review 76*(3), 456–490.

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

Hangartner, D. et al. (2021). Gender composition and hiring bias: Evidence from sweden. *Journal of Labor Economics 39*(2), 287–320.

Hartshorne, H. and M. A. May (1928). *Studies in Deceit*. New York: Macmillan.

Heckathorn, D. D. (1997, May). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems 44*(2), 174–199.

Heckathorn, D. D. (2002). Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Social Problems 49*(1), 11–34. Publisher: [Oxford University Press, Society for the Study of Social Problems].

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

Huang, W., A. Mayer, and R. Miller (2024). Bias in job promotions: Evidence from financial institutions and retail chains. *Journal of Human Resources 59*(1), 112–145.

Kelly, E. L. and E. E. Kossek (2011). Getting there from here: Research on the effects of work–family initiatives on work–family conflict and business outcomes. *The Academy of Management Annals 5*(1), 305–349.

Kleven, H. et al. (2024). The impact of motherhood on employment prospects in latin america. *Journal of Economic Perspectives 38*(1), 123–145.

Kline, P. et al. (2022). Gender bias in hiring: Evidence from a large-scale correspondence study. *Quarterly Journal of Economics 137*(4), 2053–2094.

Kline, P. et al. (2024). The dynamics of gender bias in the hiring process. *Journal of Labor Economics 42*(1), 145–178.

Lahey, J. and R. Beasley (2018). *Technical Aspects of Correspondence Studies*, pp. 81–101. Springer, Cham.

Lahey, J. N. and R. A. Beasley (2009). Computerizing audit studies. *Journal of Economic Behavior and Organization 70*(3), 508–514.

Levitt, S. D. and J. A. List (2007, June). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives 21*(2), 153–174.

List, J. A. (2020, July). Non est disputandum de generalizability? a glimpse into the external validity trial. Working Paper 27535, National Bureau of Economic Research.

Mahé, C., W. Zanoni, and M. L. Oliveri (2022). Women's informal labor market participation in ecuador. IDB Technical Note 2617, Inter-American Development Bank. Available at Inter-American Development Bank.

McLaughlin, H., C. Uggen, and A. Blackstone (2012). Sexual harassment, workplace authority, and the paradox of power. *American Sociological Review 77*(4), 625–647.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Moss-Racusin, C. A., J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences 109*(41), 16474–16479.

Neumark, D. (2018, September). Experimental research on labor market discrimination. *Journal of Economic Literature 56*(3), 799–866.

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

OECD (2023). *Joining Forces for Gender Equality: What Is Holding Us Back?* Paris: OECD Publishing.

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

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Ross, L. and R. E. Nisbett (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.

Sarsons, H. (2017). Interpreting signals in the labor market: Evidence from medical referrals. *Working Paper*.

Schaerer, M., C. du Plessis, M. H. Nguyen, R. C. van Aert, L. Tiokhin, D. Lakens, E. G. Clemente, T. Pfeiffer, A. Dreber, M. Johannesson, et al. (2023). On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions. *Organizational Behavior and Human Decision Processes 179*, 104280.

Shiller, R. J. (2017). Narrative economics. *The American Economic Review 107*(4), 967–1004.

UNDP (United Nations Development Programme) (2023). 2023 Gender Social Norms Index (GSNI): Breaking Down Gender Biases: Shifting Social Norms towards Gender Equality.

Villanueva, E. and S. Lin (2020). Motherhood wage penalties in latin america: The significance of labor informality. *Social Forces 99*(1), 59–88.

Zanoni, W., P. Acevedo, and H. Hernandez (2022). Job market discrimination against slum dwellers in urban argentina. *SSRN*. http://dx.doi.org/10.18235/0004179.

Zanoni, W. and L. Díaz (2024). Discrimination against migrants and its determinants: Evidence from a multi-purpose field experiment in the housing rental market. *Journal of Development Economics 167*.

Zanoni, W. and R. F. Díaz (2024). The migrant penalty in latin america: Experimental evidence from job recruiters. *IDB Working Papers Series*.

Zanoni, W., A. Paloma, Z. Giulia, and H. Hugo (2023). Exploring the prevalence, causes, and solutions to discrimination against workers from slums. *SSRN*. https://dx.doi.org/10.2139/ssrn.4479907.

A   Figures & tables

Table 2: Oaxaca-Blinder Decomposition Results for Employment Status by Gender

Component	Coefficient	Standard error
Differential		
Prediction for men (1)	0.8929	(0.0016)
Prediction for women (2)	0.8353	(0.0022)
Difference (1)-(2)	0.0576	(0.0028)
Decomposition		
Explained (3)	0.0202	(0.0033)
Unexplained (4)	0.0375	(0.0040)

Note: This table presents the Blinder-Oaxaca decomposition analysis outcomes, explaining differences in the employment probability between women and men. Column 1 shows he predicted employment probability for men, indicating a baseline comparison point; column 2 the predicted employment probability for women, showcasing the comparison group; column 3 the explained component, representing the portion of the employment status gap attributable to differences in observed variables; and column 4 the unexplained component, reflecting the portion of the gap not accounted for by the model and potentially indicating discrimination or other unobservable factors. Coefficients are presented alongside their respective standard errors. Values are weighted using the ENEMDU-2022 survey weights.

Table 3: Recruiters' Demographics and Education by Sampling Method

Variable	(1) All	(2) Linkedin	(3) RDS	(4) Difference (2)-(3)
Demographics and education				
Age (years)	31.1168	31.5867	30.8633	-0.7234
	(7.319)	(6.5246)	(7.7246)	(0.9984)
Gender (Female == 1)	0.693	0.6582	0.7114	0.0532
	(0.4623)	(0.4773)	(0.4546)	(0.0654)
Nationality (Ecuadorian == 1)	0.9693	0.9241	0.9933	0.0692**
	(0.1729)	(0.2666)	(0.0819)	(0.0307)
Work Experience (years)	7.0154	7.1203	6.9597	-0.1605
	(5.8208)	(4.8093)	(6.3062)	(0.7481)
Experience as an HR Recruiter (years)	5.3032	5.9557	4.927	-1.0287
	(5.0382)	(4.9447)	(5.071)	(0.7051)
Does the recruiter have a college degree? (Yes == 1)	0.9123	0.9873	0.8725	-0.1149***
	(0.2835)	(0.1125)	(0.3347)	(0.0302)
Education: Secondary Education	0.0482	0	0.0738	0.0738***
	(0.2148)	(0)	(0.2624)	(0.0215)
Education: Post-secondary Education	0.0351	0.0127	0.047	0.0343
	(0.1844)	(0.1125)	(0.2123)	(0.0215)
Education: University	0.7149	0.8228	0.6577	-0.1651***
	(0.4524)	(0.3843)	(0.4761)	(0.0582)
Education: Masters	0.193	0.1646	0.2081	0.0435
	(0.3955)	(0.3731)	(0.4073)	(0.0536)
Education: Doctorate	0.0044	0	0.0067	0.0067
	(0.0662)	(0)	(0.0819)	(0.0067)
Observations	228	79	149	228

Note: Time reviewing applications is conditional on the recruiter's opening those tabs.* $p < .10$, ** $p < .05$, *** $p < .01$. (a) these are 4 indicator variables, each for whether the recruiter opened the "Personal Information," "Work Experience," "Schooling/Training," and "Additional Information" tabs; (b) this is the expected time, conditional on the recruiter's opening the tab.

Figure 1: Discrimination against Women Coefficient Plot
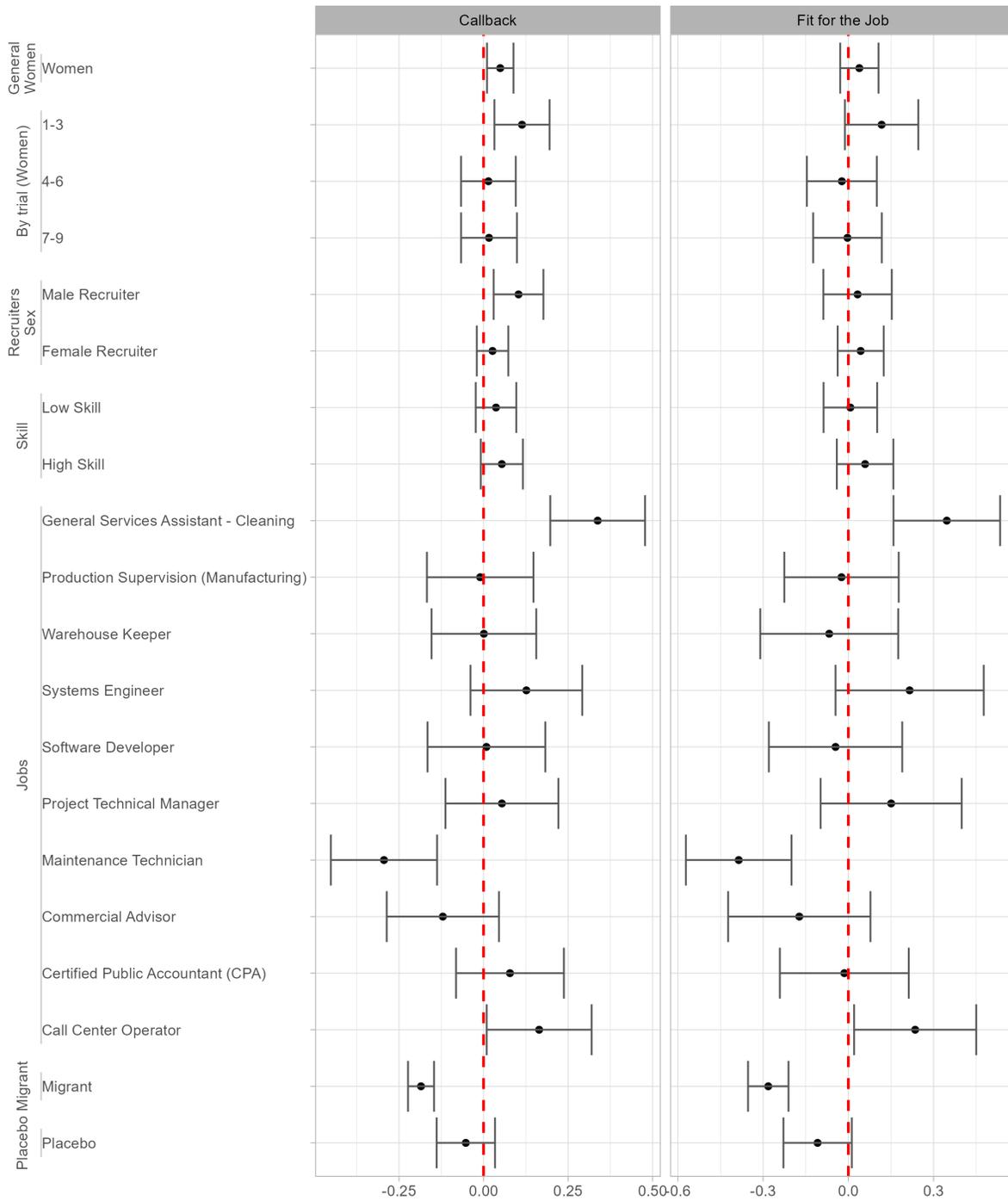
Note: This figure presents estimates of discrimination coefficients for various subgroups, accompanied by 90% confidence intervals. The discrimination coefficients presented come from the third specification, which contains covariates as design covariates (recollection method, if the recruiter opened the information tabs, the trial fixed effects), position fixed effect, and a recruiter fixed effect. Dimensions analyzed include trial groups, Venezuelan migrants, and placebo group.

Table 4: Recruiters' Balance Table: Scores on Standardized Tests

Variable	(1) All	(2) Linkedin	(3) RDS	(4) Difference (2)-(3)
Scores in standardized tests				
Standarized values of neuroticism	0.0646	-0.119	0.1619	0.2808**
	(0.9755)	(0.899)	(1.003)	(0.1303)
Standarized values of extroversion	0.1907	0.1543	0.21	0.0557
	(0.7638)	(0.8724)	(0.7019)	(0.1138)
Standarized values of openness	0.1808	0.1413	0.2018	0.0605
	(0.8093)	(0.83)	(0.8001)	(0.1141)
Standarized values of agreeableness	0.1425	0.1261	0.1512	0.0251
	(0.7784)	(0.8506)	(0.7401)	(0.1133)
Standarized values of conscientiousness	0.1669	0.1261	0.1886	0.0625
	(0.748)	(0.8335)	(0.7005)	(0.1099)
Score in Neoffi test (std.)	0.1826	0.1236	0.2139	0.0902
	(0.7219)	(0.8319)	(0.6571)	(0.108)
Score in Rosenberg test (std.)	0.1803	0.0881	0.2291	0.1411
	(0.7865)	(0.8981)	(0.7189)	(0.117)
Score in Wonderlic test (std.)	0.097	0.027	0.1342	0.1072
	(0.9377)	(0.9795)	(0.916)	(0.1333)
Observations	228	79	149	228

Note: Stars indicate the statistical significance of differences in means across groups at various significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$. Columns (2) and (3) display the attributes of recruiters based on whether they were sampled and hired using the RDS or the LinkedIn method.

Table 5: Recruiters' Characteristics: All and by Sampling Method

Variable	(1) All	(2) Linkedin	(3) RDS	(4) Difference (2)-(3)
Performance in the experiment: (a)				
Opened Personal Information tab	0.7818	0.8523	0.7444	-0.1079***
	(0.2873)	(0.2188)	(0.3119)	(0.0355)
Opened Work Experience tab	0.9485	0.9768	0.9334	-0.0433**
	(0.1578)	(0.0874)	(0.183)	(0.0179)
Opened Schooling/Training tab	0.8827	0.942	0.8512	-0.0908***
	(0.2381)	(0.1233)	(0.2758)	(0.0265)
Opened Additional Information tab	0.7259	0.8006	0.6862	-0.1144***
	(0.3053)	(0.2362)	(0.3301)	(0.0379)
Time Reviewing Applications: (b)				
Total time (min)	74.0302	90.6384	65.0436	-25.5948***
	(55.4207)	(62.6761)	(48.9911)	(8.1342)
Time on Personal Information tab (min)	7.9042	7.9589	7.8753	-0.0836
	(14.257)	(7.6987)	(16.7487)	(1.6226)
Time on Work Experience tab (min)	44.2997	54.8829	38.6885	-16.1944**
	(64.3273)	(49.9276)	(70.2963)	(8.0448)
Time on Schooling/Training tab (min)	20.0764	18.6435	20.8361	2.1927
	(35.6139)	(17.119)	(42.2995)	(3.9646)
Time on Additional Information tab (min)	14.8874	10.5032	17.212	6.7088
	(101.2281)	(12.2775)	(124.987)	(10.3321)
Observations	228	79	149	228

Note: Stars indicate the statistical significance of differences in means across groups at various significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$. Columns 2 and 3 display the attributes of recruiters based on whether they were hired using the RDS or the LinkedIn method.

Table 6: Candidates' Balance Table

Variable	(1) Male	(2) Female	(3) Difference (1) - (2)
Demographics and Education			
Age (years)	29.6912	29.8171	0.1259
	(3.8306)	(3.8326)	(0.1643)
Previous Jobs	2.9715	2.9715	0
	(0.7805)	(0.7805)	(0.0335)
Is candidate considered a professional? (Yes == 1)	0.5827	0.5827	0
	(0.4933)	(0.4933)	(0.0212)
Employment Experience (years)	4.6504	4.6431	-0.0074
	(1.4697)	(1.4154)	(0.0619)
Education: Secondary Education	0.114	0.114	0
	(0.3179)	(0.3179)	(0.0136)
Education: Technical Degree	0.3033	0.3033	0
	(0.4599)	(0.4599)	(0.0197)
Education: Professional	0.5827	0.5827	0
	(0.4933)	(0.4933)	(0.0212)
Applied Job Position:			
Job Position: Comercial Advisor	0.0956	0.0956	0
	(0.2942)	(0.2942)	(0.0126)
Job Position: General Services Assistant - Cleaning	0.114	0.114	0
	(0.3179)	(0.3179)	(0.0136)
Job Position: Warehouse Keeper	0.1002	0.1002	0
	(0.3004)	(0.3004)	(0.0129)
Job Position: Certified Public Accountant (CPA)	0.1039	0.1039	0
	(0.3052)	(0.3052)	(0.0131)
Job Position: Software Developer	0.0928	0.0928	0
	(0.2903)	(0.2903)	(0.0124)
Job Position: Systems Engineer	0.0956	0.0956	0
	(0.2942)	(0.2942)	(0.0126)
Job Position: Project Technical Manager	0.0983	0.0983	0
	(0.2979)	(0.2979)	(0.0128)
Job Position: Call Center Operator	0.1048	0.1048	0
	(0.3064)	(0.3064)	(0.0131)
Job Position: Production Supervision (Manufacturing)	0.0965	0.0965	0
	(0.2954)	(0.2954)	(0.0127)
Job Position: Maintenance Technician	0.0983	0.0983	0
	(0.2979)	(0.2979)	(0.0128)
Observations	1088	1088	2176

Note: This table exhibits the attributes for male and female synthetic candidates. No statistical significance in the attributes means that male and female candidates are equivalent in terms of a specific characteristic. Stars indicate the statistical significance of differences in means across groups at various significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7: Discrimination Rate Differences: Women vs. Men

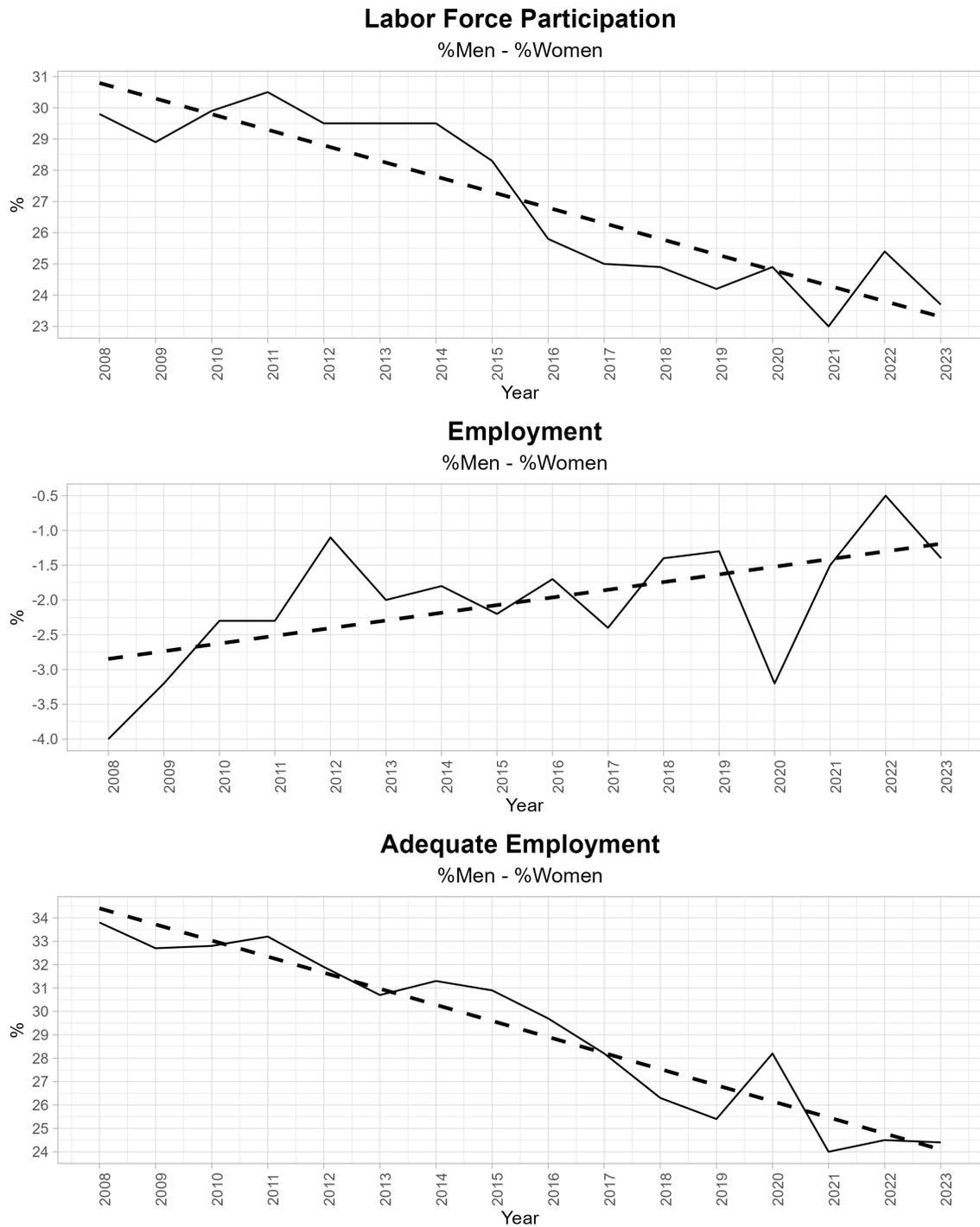|	(1) Model 1	(2) Model 2	(3) Model 3	(4) Low skill	(5) High skill	(6) Male recruiter	(7) Female recruiter
A. Callbacks							
Women Applicant Indicator Coeff.	0.0494**	0.0494**	0.1037**	0.0267	0.1035**	0.0266	
	(0.0215)	(0.0238)	(0.0403)	(0.0257)	(0.0447)	(0.0283)	
% of Men Applicants Chosen	0.4752	0.4752	0.4479	0.4864	0.4479	0.4864	
Observations	2176	2176	634	1542	634	1542	
B.Good Fit for the job:							
Discrimination Coeff.	-0.0068	-0.0066	0.0011	-0.0087	0.0013	-0.0087	
	(0.0151)	(0.0139)	(0.0276)	(0.0179)	(0.0259)	(0.0165)	
Mean Male Candidate (on that subsample)	0.8539	0.8539	0.8580	0.8521	0.8580	0.8521	
Observations	2176	2176	634	1542	634	1542	
C. Fit for the job:							
Women Applicant Indicator Coeff.	0.0383	0.0388	0.0312	0.0466	0.0326	0.0434	
	(0.0503)	(0.041)	(0.0867)	(0.061)	(0.0732)	(0.0493)	
Mean Fit Among Men Applicants	8.5754	8.5754	8.6341	8.5512	8.6341	8.5512	
Observations	2176	2176	634	1542	634	1542	
Model specification:							
Candidates' characteristics	NO	YES	YES	YES	YES	YES	YES
Recruiters' sampling method	YES	YES	YES	YES	YES	YES	YES
Opened tabs	YES	YES	YES	YES	YES	YES	YES
Recruiters' FE	NO	YES	NO	NO	NO	NO	NO
Clusterized SE	YES	NO	YES	YES	YES	YES	YES

Note: The values in the table are OLS coefficients that identify the mean difference in the outcomes associated with candidates who belong to a minority group and those who do not. All regressions include observations for trials where the vignette did not appear. In horizontal panel A, the dependent variables are dummy variables (*Callbacks* indicating whether the candidate was selected ("called back") for the position. In panel B, the dependent variable is *Good fit for the job*, indicating when *Fit for the job* was 8 or more. In panel C, the dependent variable represents the candidate's Fit for the job, measured on a scale of 1 to 10. Standard errors clustered at the recruiter level and robust are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8: Mean Differences in Recruiter Characteristics: INEC's Household Survey vs. Our Experiment

	ENEMDU	Experiment	Difference	pvalue
Age (years)	37.73	31.39	6.34	0.00
Women (%)	58.26	70.15	-11.89	0.00
College (%)	80.29	91.84	-11.54	0.00
Observations	1202	392	.	.

Note: This table presents means of demographic characteristics of recruiters, comparing the sample from Ecuador's Household Survey with that from our experiment. The statistical significance of the difference in the means is assessed by means of a *t*-test, the *p*-value of which we show in the table.

Figure 2: Gender Gaps in Ecuador

Labor Force Participation
%Men - %Women

Employment
%Men - %Women

Adequate Employment
%Men - %Women

Note: The data source is the official employment series published by INEC.