WORKING PAPER Nº IDB-WP-1611

Exploring Gender Discrimination: A Multi-Trial Field Experiment in Urban Ecuador

Wladimir Zanoni
Suzanne Duryea
Jorge Paredes

Inter-American Development Bank
Country Department Andean Group

August 2024

Exploring Gender Discrimination: A Multi-Trial Field Experiment in Urban Ecuador

Wladimir Zanoni
Suzanne Duryea
Jorge Paredes

Inter-American Development Bank
Country Department Andean Group

August 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Zanoni López, Wladimir, 1975-.
Exploring gender discrimination: a multi-trial field experiment in urban Ecuador / Wladimir Zanoni, Suzanne Duryea, Jorge Paredes.
p. cm. — (IDB Working Paper Series ; 1611)
Includes bibliographical references.
1. Minorities-Economic aspects-Ecuador. 2. Gender mainstreaming-Ecuador. 3. Equality-Ecuador. 4. Discrimination in employment-Ecuador. 5. Labor market-Ecuador. I. Duryea, Suzanne. II. Paredes, Jorge. III. Inter-American Development Bank. Country Department Andean Group. IV. Titles. V. Series.
IDB-WP-1611

http://www.iadb.org

Gender and Discrimination in Urban Labor Markets: Evidence from a Field Experiment in Ecuador*

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

August 28, 2024

Abstract

This study examines the extent and underlying mechanisms of gender-based discrimination in Ecuador's urban labor market through an artifactual field experiment involving 392 human resource recruiters. The experiment assessed whether recruiters showed a differential preference for male versus female candidates with identical observable skills. Our findings reveal a significant 10% preference for female candidates despite recruiters' assessments of job fit being nearly similar for both genders. This preference suggests that factors beyond productivity influence hiring decisions, potentially driven by social norms that, consistent with survey data, favor increased participation of women in the labor force that has been closing the employment gap. Further evidence suggests that the preference for hiring women primarily occurs in roles traditionally held by women, reinforcing rather than challenging occupational gender segregation. Data from the census indicates that this pattern is prevalent across the Ecuadorian labor market, which may help explain the persistent wage gaps and disparities in job quality between men and women. These results contribute to understanding how gender biases manifest in labor market outcomes and underscore the importance of considering societal norms when addressing gender disparities.

JEL Codes: J16, J71, C93
Keywords Gender discrimination, Occupational segregation, Labor market, Stereotyping

*The authors would like to express their thanks to Paloma Acevedo, Osmel Manzano, and Juan Carlos de la Hoz for their invaluable support for this research. Thanks to Hugo Hernández, Omar Zambrano, Emily Díaz, Nicolas Chuquimarca, Albani Granado, Ana Patricia Muñoz, María Caridad Ortiz, as well as to the dedicated teams at Anova Policy Research and Grupo FARO for their work. This project received IRB clearance from Universidad del Rosario.

[†]Inter-American Development Bank. Corresponding author's email: wladimirz@iadb.org.

1 Introduction

Gender inequality in Latin America's labor markets is a well-documented issue, with persistent wage gaps, employment disparities, and occupational segregation being shared across the region. Extensive empirical evidence highlights the significant labor market disadvantages faced by women, mainly due to motherhood, which imposes a more substantial employment penalty compared to other regions. For instance, studies reveal that the employment penalty for mothers ranges from 12% in Brazil to 21% in Chile, with an additional 38% increase in female labor informality after the birth of a first child Kleven et al. (2024); Villanueva and Lin (2020); Berniell et al. (2021). Despite this evidence, the size and mechanisms explaining gender discrimination in Latin America remain under-documented, particularly in terms of how gender biases manifest in recruitment practices in the labor market.

While the parenthood penalty is a key factor driving gender disparities in labor markets of Latin America, these disparities can also result from the interplay between explicit gender prejudices and deeply ingrained social norms that reinforce traditional gender roles. These norms confine women to lower-paying, female-dominated sectors, exacerbating gender inequality of Labor & Development (2023); Nopo and Chong (2010). In Ecuador, these issues are particularly acute, with women facing significant barriers to accessing formal employment, achieving wage equality, and securing job stability. Women in Ecuador are more likely to be employed in informal sectors that offer lower wages and less security Posso (2013); Nopo and Chong (2010). For instance, social norms, prejudice, and stereotypes can play a role in shaping discriminatory behaviors among decision-makers in labor markets; we do not have enough evidence on the magnitude of their relevance and relative importance in explaining gender disparities in labor markets in Latin America.

This paper explores whether human resource agents (recruiters) in Ecuador show a difference in their referral practices when recommending male versus female candidates who have identical observable skills, particularly when these candidates are competing for the same jobs. Specifically, we investigate the underlying behavioral mechanisms driving any differential treatment, focusing on the potential roles of taste-based discrimination (prejudices) and statistical discrimination (stereotyping).[1] Additionally, we examine how evolving social norms related to the growing acceptance of women in the labor force may manifest in gendered job roles and influence hiring decisions, potentially reinforcing occupational gender segregation.

Our study not only explores whether recruiters treat male and female candidates differently but also examines how the behavior of male and female recruiters themselves contributes to this discrimination. Specifically, we found that models with and without fixed effects reported no differences in outcomes for either men or women, indicating that stereotyping is the predominant driver of discrimination, particularly among male recruiters.

[1] Here, prejudice represents the evaluator's innate biases, similar to taste discrimination (Becker, 1957). Stereotypes align with statistical discrimination (Phelps, 1972; Arrow, 1971), where limited information leads recruiters to make statistical inferences about candidates, and with the "social condition approach" by Bordalo et al. (2016), where stereotypes are seen as cognitive shortcuts, rooted in fundamental group differences but subject to contextual distortion.

This suggests that men are more responsive to social norms that encourage female labor force participation, but only in traditionally gendered occupations.

Answering these research questions is crucial for both scientific and policy reasons. Scientifically, the study contributes critical experimental evidence on how gender biases influence hiring decisions in Latin America, a region where such evidence is notably scarce. While research in high-income countries has documented gender bias in labor market decisions, these findings do not always extend across countries, sectors, and career stages Goldin and Rouse (2000). For policymakers in Latin America, understanding the country-specific mechanisms driving gender discrimination in hiring is essential for designing targeted interventions that can effectively reduce gender disparities in those labor markets.

To address these research questions, we rely on data from an artifactual field experiment (AFE) conducted in Quito, Ecuador, that examined how recruiters evaluated job applicants who were women, gays, lesbians, and migrants. This paper focuses on the data from which recruiters assessed applications from women and men competing for the same jobs. The AFE involved 392 human resource professionals recruited via LinkedIn and snowball referral sampling, with support from a local NGO. The recruiters were unaware that they were participating in a field experiment aimed at studying potential gender-based discriminatory behaviors.

We developed a custom web-based platform, designed in consultation with human resource experts, to closely resemble a hiring evaluation software of the type commonly used by firms today. Each recruiter evaluated up to three pairs of job applications, comparing men and women for three different occupations randomly assigned across ten job postings. The candidates in each pair were observationally equivalent in terms of productivity signals, ensuring that any differences in recruiter evaluations could be attributed to random gender assignments.

The data generated by the experiment is stacked at the candidate level and nested within recruiters, with each recruiter having up to six associated observations (i.e., up to three pairs). The primary dependent variables are an indicator for candidate selection and an ordinal variable for job fitness assessment. Ordinary least squares (OLS) coefficient estimates of a gender indicator for females on these outcomes measure discrimination (we call that estimate a "discrimination coefficient"). The longitudinal component of the data (each recruiter sequentially analyzes various pairs of candidates) allows us to measure discrimination using coefficient estimates from OLS regressions with and without recruiters' fixed-effects. Under the assumption that prejudice remains constant during the experiment, comparing discrimination coefficients with and without fixed effects allows us to understand the separate roles that stereotyping and prejudice play in the recruiters' referral decisions.

Our findings reveal that when presented with observationally equivalent candidates, recruiters favored women over men in job referrals. Specifically, women were chosen approximately 10% more often than their male counterparts, with a nearly five percentage point difference in selection rates. This preference for female candidates was more pronounced among male recruiters, who showed a statistically significant 10.35 percentage point difference in favor of women, compared to a smaller and statistically insignificant 2.66 percentage point difference among female recruiters.

Interestingly, our results show that there are no significant differences between models with and without recruiter fixed effects for both male and female candidates. This finding underscores that the primary behavioral mechanism leading to discrimination is stereotyping, particularly by male recruiters. This aspect of gender discrimination, where men are more influenced by social norms that promote female labor participation yet confine women to gendered roles, is not often highlighted in the literature.

The analysis also shows that discrimination coefficients, with and without recruiters' fixed effects, are nearly identical, indicating that stereotyping is the primary driver of discrimination. Significant positive discrimination was observed toward female candidates in stereotypically female-dominated roles, such as cleaning and call center operations. In contrast, negative discrimination persisted in traditionally male-dominated jobs like maintenance. This pattern underscores gender-driven occupational segregation, where women are increasingly hired into traditional roles, reinforcing existing occupational gender structures rather than breaking them down.

Moreover, our findings suggest that the discrimination observed is heavily influenced by male recruiters adhering to societal expectations that women should participate more in the labor force but in roles that align with traditional gender norms. This aspect of gender discrimination—where not only the gender of the candidate matters but also the gender of the recruiter—adds a critical dimension to our scientific understanding of how social norms shape differential hiring practices by gender.

Our study makes a significant contribution to the literature on gender discrimination by documenting that increasing participation of women in the labor force appears to be influenced by occupationally driven stereotypes held by recruiters in a Latin American context, a region and research topic that has received less attention compared to high-income countries. While previous research has documented bias against women in specific sectors and populations Bravo et al. (2008); Arceo-Gomez and Campos-Vazquez (2014) in OECD countries, our study uniquely highlights the gender-driven nature of recruiter choices and the role of social norms in shaping these decisions. By experimentally distinguishing between prejudices and stereotyping, this study offers a deeper understanding of the mechanisms driving gender biases in hiring in Latin America.

The findings have important policy implications. Understanding that stereotyping is a crucial driver of gender discrimination in hiring allows for the development of more targeted interventions. Behavioral interventions aimed at correcting misinformation or challenging entrenched wrong beliefs may be effective in reducing gender discrimination, particularly in higher-wage occupations, thereby contributing to closing the wage gap. Neumark (2018) emphasizes that understanding discrimination in hiring is crucial because it often manifests in gender-segregated occupations where wages are lower for roles typically held by women. This insight is key to better understanding the mechanisms behind persistent wage gaps. These could include nudging recruiters to recognize and counteract implicit biases regarding occupational sectors. Additionally, addressing social norms that reinforce gender roles is crucial for reducing occupational segregation and promoting gender equality in the labor market.

In evaluating the external validity of our findings, we apply the SANS framework (Selection, Attrition, Naturalness, and Scalability) as proposed by List (2020). For *Selection*, our data encompasses 392 experienced human resource recruiters from a convenience sample in

Quito, Ecuador. Compared to national occupational statistics (ENEMDU), our sample is younger, more educated, and predominantly female. Regarding *Attrition*, our study exhibited high participant commitment, with 315 out of 392 recruiters completing all three trials, resulting in 2,176 observations. This high completion rate minimizes the risk of attrition bias.

For *Naturalness*, our experiment was conducted via an online platform, mirroring modern recruitment practices, where remote candidate evaluation is increasingly common.[2] This approach aligns with current trends, where many companies outsource their recruitment services to third-party providers, making it increasingly common for recruiters to interact with such platforms in their daily work.[3] The use of online tools for candidate assessment is not unusual, and the flexibility of these platforms allows for various evaluation methods, including reviewing candidates in pairs. This mirrors actual recruitment practices, where recruiters may assess candidates sequentially or in parallel.[4] Thus, the environment we created is both familiar and credible, enhancing the validity of our findings.

Lastly, concerning *Scalability*, as will be referred to in the document, not only does our experiment's design reflect the trends in Ecuador's hiring practices for formal employment, but the results align with the occupational distribution of jobs. It can be expanded to monitor evolving discrimination dynamics beyond gender, making it a flexible and reliable tool in the context of increasing 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 Labor Market Disparities by Gender: Survey Evidence from Ecuador

Ecuador has experienced notable progress in increasing the labor supply 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

[2] McKinsey Company, "The future of remote work: An analysis of 2,000 tasks, 800 jobs, and 9 countries", https://www.mckinsey.com/future-of-remote-work

[3] Financial Times, "The rise of the platform economy", https://www.ft.com/content/platform-economy-2024

[4] Financial Times, "The rise of the platform economy", https://www.ft.com/content/platform-economy-2024

-1.4. Moreover, the differential in the rate of "adequate employment"—defined as working a minimum of 40 hours across five days in a week—shrank from 33.8 to 24.4. Despite this progress, discrepancies persist in employment outcomes between men and women, as illustrated in figure 1. Understanding the mechanisms that drive those disparities is relevant. Some of those mechanisms might be actionable with public policy.

To examine the disparities in employment rates between men and women within Ecuador's labor force, we use nationally representative household survey data to assess how observable factors—those we can measure and control for—contribute to the employment gap, compared to unobserved factors. This analysis frames the relevance of our field experiment and outlines the 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[5] 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 1 is driven by discrimination exercised by recruiters is an open research question that we seek to address in this paper.

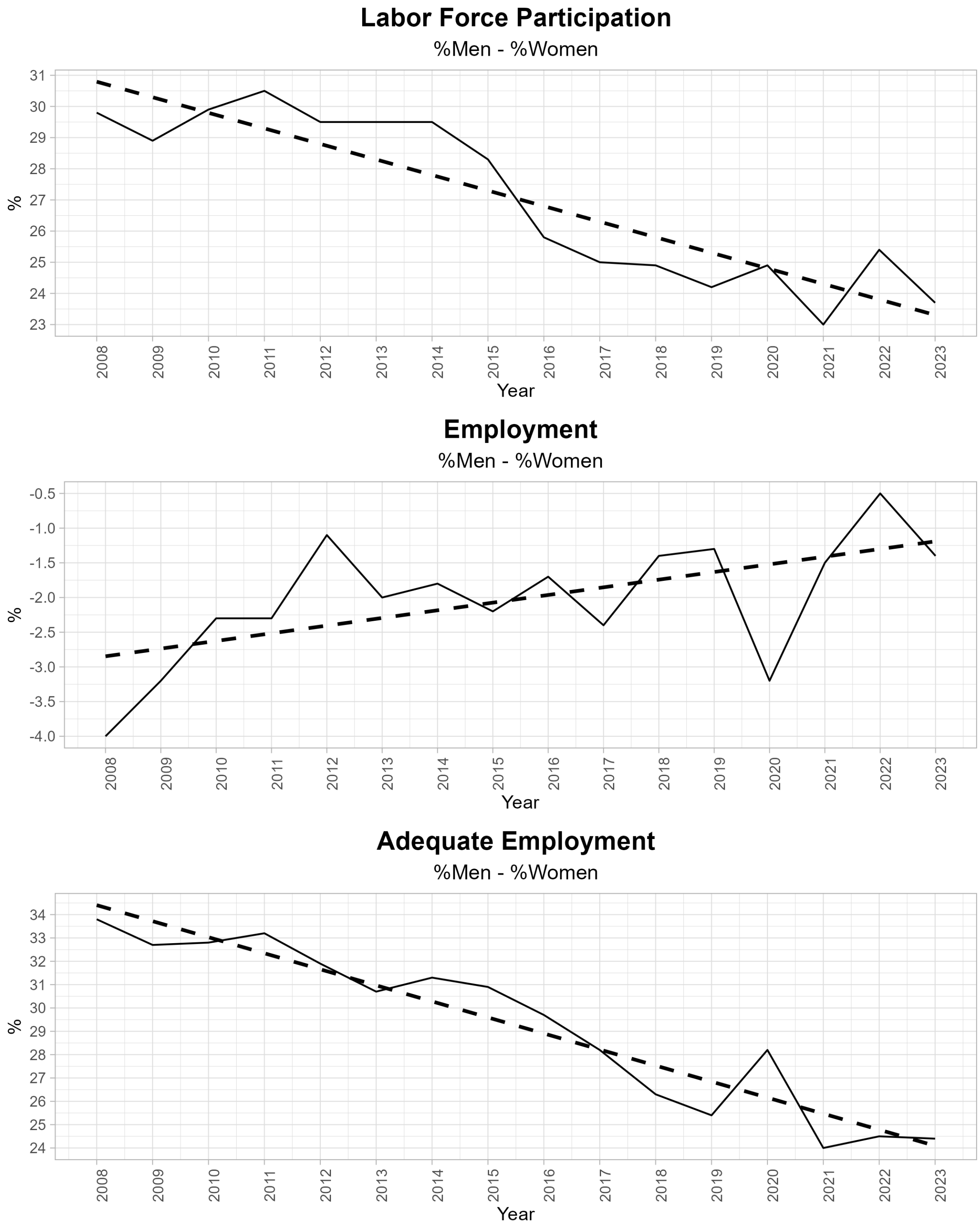
3 Experimental Design

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.[6]

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

[6] 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, Zanoni et al. (2024) in Ecuador, and Zanoni and Díaz (2024) in Colombia. Moreover, Lahey and Beasley (2018) employs 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

Figure 1: Gender Gaps in Ecuador



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

Table 1: 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.

Two companion papers (Zanoni et al. (2024) 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, within 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 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 on a Likert scale from one to ten (where one is low and ten is high job fitness). Here, we analyze the data from the experiment in which recruiters evaluated pairs of applicants, with one candidate being a woman and the other a man.

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,[7] in that it seeks to gauge the revealed

of each candidate in pairs, which is the approach we take.

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

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 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). We worked with them 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 paid work with significant implications for recruitment through the construction of an environment closely resembling platforms used today by HR professionals for job referrals during remote work scenarios. They also helped ensure that the recruiters had proper professional experience and qualifications.

To generate a sample of recruiters, we used two methods: LinkedIn advertisements and referrals via respondent-driven sampling (RDS).[8],[9] The experiment aimed to replicate real-world job recruitment scenarios authentically. To achieve this, we conducted interviews and focus groups with local stakeholders, including NGOs, academia, and hiring firms, to understand the challenges faced by women job seekers in Ecuador's labor market. These consultations served two purposes: validating the observational equivalence of the pairs of profiles in each experimental trial and gathering insights to tailor the online platform to resemble actual job evaluations closely. Grupo FARO, a local organization with experience in making hiring recommendations, facilitated the recruitment of recruiters and assured them they were participating in a genuine hiring exercise.

Using qualitative insights from these consultations and data from the yearly national household survey (ENEMDU, 2022), we crafted job vacancies where men and women were shown to compete for jobs. The synthetic pairs of resumes were designed to ensure comparable characteristics among candidates except for the gender attribute that was randomized by design. Balance tests confirmed equivalence in age, qualifications, and demographics among the fictitious job applicants by gender (see Table 2). The recruitment task was presented as a genuine hiring exercise, emphasizing the need for experienced HR analysts and recruiters in the Quito labor market. Recruiters were tasked with rating productivity and making hiring recommendations 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

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

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

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, and we asked them to provide full descriptions of their demographics, work experience, and knowledge of Ecuador's labor market. They also had to complete a cognitive and socio-emotional development test as well as self-esteem evaluations. After completion of the survey, they proceeded to the online assessment of the candidate pairs. The diagram (Figure 2) visualizes this process, starting with contacting the recruiters, evaluating their eligibility, and moving on to the trial preparation where pairs of candidates were selected and assigned to randomized job positions. Each recruiter conducted ten trials, with the system randomizing the pairings and job assignments in each trial, concluding with the recruiter making a hiring decision in each instance.

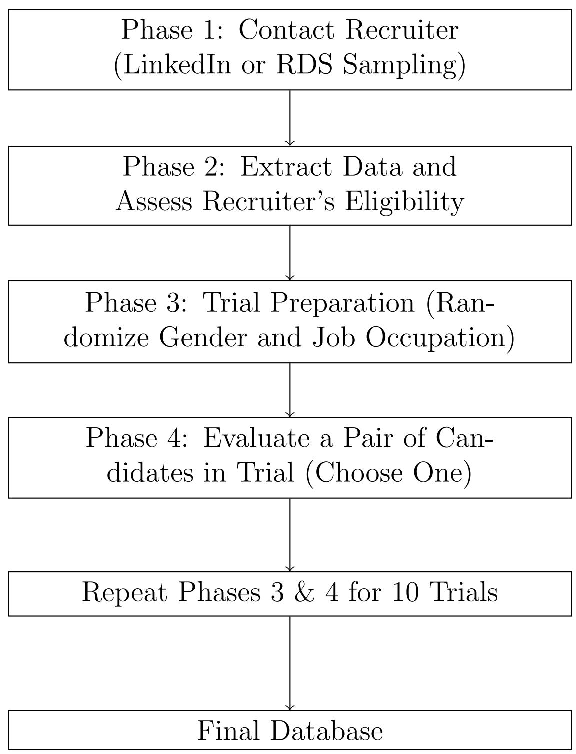


Figure 2: Research Design Diagram: The diagram illustrates the process of contacting recruiters, assessing their eligibility, preparing trials by randomizing pairs and job positions, and evaluating candidates over ten trials.

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 collection, and its posterior analysis, for most CS, takes several months (sometimes more than a year) to complete. 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 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 (i.e., up to ten reviews of pairs of applicants) 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.

Clear instructions, validation of task understanding in pilot tests, and real-stakes remuneration addressed concerns regarding recruiters' choices and potential biases. Ethical considerations were discussed through discussions with an Institutional Review Board (IRB) committee. Post-experiment 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 7 in the appendix, we compare some critical 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 that primarily joins 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 being LinkedIn users in the country.

3.1 Empirical strategy

To estimate the effect of gender on the outcomes of interest, we use OLS regression to estimate the parameters of models of the form:

$$Y_{itr} = \beta_0 + \delta D_{it} + \beta X_{it} + \gamma T_t + \epsilon_{itr}, \tag{1}$$

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) *Choice of Candidate*, an indicator that has a value of one if the candidate is selected by the recruiter (zero if not), and (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 position. 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 1 (the discrimination coefficient henceforth), which we estimate for the entire sample and for selected subsamples by gender of the recruiter, occupation, and trials. The results were validated by studying the coefficient estimates of job fixed effects interacted with the gender indicator.

We proposed two empirical models to analyze the underlying mechanisms of discrimination: one that captures the combined effect of prejudice and stereotyping and another that isolates the impact of stereotyping alone. This distinction is critical for understanding the nuanced dynamics between inherent biases and stereotypical judgments in hiring decisions. To isolate the impact of stereotyping, we add recruiters' fixed effects to the model specification in equation 1. Because we exposed each recruiter to multiple evaluation pairs, the data has a longitudinal dimension, and we can employ a recruiter fixed-effects to differentiate out time invariant prejudice as a source of bias. The model specification now takes the form:

$$Y_{itr} = \beta_0 + \delta_{fe} D_{it} + \beta X_{it} + \gamma T_t + \lambda R_r +_{itr} . \qquad (2)$$

In this formulation, R_r represents a vector of indicators for recruiter fixed effects, and δ_{fe} measures explicitly the impact of stereotyping 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.

3.2 Data

Table 2 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 Table 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 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.

Table 2: 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: * $p < .10$, ** $p < .05$, *** $p < .01$.

3.3 Characteristics of the recruiters

Tables 4, 5, and 6 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) time-performance in the experiment.

As we can see in column 1 of table 4 in appendix A, the recruiters we hired mainly were 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 us using the snowball sampling method were 11% less likely than their counterparts to have a bachelor's degree.

Table 4 in Appendix A shows some socioeconomic variables that characterize 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[10]. 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.

4 Results

In table 3, we present our main results. The table presents OLS discrimination coefficient estimates for δ and δ_{FE} from equations 1 and 2, respectively. They show the average differences in outcomes between women and men. In Panel A, the dependent variable (Choice of Candidate) is an indicator of whether the recruiter chose a candidate. In Panel B, the dependent variable is the candidate's perceived "fit for the job," rated on a 1 to 10 scale. The model controls for trial fixed effects, accounting for the order in which the recruiter reviewed candidate pairs, and occupation fixed effects, which include indicators for the specific occupation associated with the job posting (there were ten possible occupations). Additional controls include the sampling method (whether the recruiter was sourced via LinkedIn or snowball sampling) and indicators for whether the recruiter reviewed specific tabs in the web platform —Personal Information, Experience, Additional Income, and Education—during the evaluation. Standard errors are clustered at the recruiter level and are robust to heteroskedasticity, with significance levels indicated as follows: * $p < .10$, ** $p < .05$, *** $p < .01$.

The results presented in Table 3 reveal that, on average, recruiters showed a preference for women over men in their hiring decisions. Specifically, as indicated by the discrimination coefficient in column 1 of Panel A, women were selected for jobs at a rate 10.4% higher than male applicants when considered as a proportion of male candidates chosen. This bias favoring women persisted even though the recruiters assessed male and female candidates as having nearly equivalent productivity, as evidenced by the small and statistically insignificant differences in the job fit scores presented in Panel B.

Moreover, a comparison between the coefficients δ and δ_{fe} suggests that this preference for hiring women may be driven, at least in part, by stereotyping. The fact that the inclusion of fixed effects does not substantially alter the discrimination coefficients implies that the observed bias is not solely due to measurable factors related to the can-

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

Table 3: Discrimination Rate Differences: Women vs. Men

	(1) Coeff.δ	(2) Coeff.δ_{FE}	(3) Male Recruiter δ	(4) Female Recruiter δ	(5) Male Recruiter δ_{FE}	(6) Female Recruiter δ_{FE}
A. Choice of Candidates						
Discrimination 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)
Mean Male Candidate (on that subsample)	0.4752	0.4752	0.4479	0.4864	0.4479	0.4864
Observations	2176	2176	634	1542	634	1542
B. Fit for the job:						
Discrimination 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 Male Candidate (on that subsample)	8.5754	8.5754	8.6341	8.5512	8.6341	8.5512
Observations	2176	2176	634	1542	634	1542
Model specification:						
Trial and occupation fixed effects (1)	YES	YES	YES	YES	YES	YES
Recruiters' sampling method indicator (2)	YES	YES	YES	YES	YES	YES
Information reviewed (indicators) (3)	YES	YES	YES	YES	YES	YES
Recruiters' FE	NO	YES	NO	NO	YES	YES

Note: The table presents OLS discrimination coefficient estimates of the average differences in outcomes between women and men. In Panel A, the dependent variable is an indicator of whether the recruiter chose a candidate. In Panel B, the dependent variable is the candidate's perceived "fit for the job," rated on a 1 to 10 scale. (1) Trial fixed effects account for the order in which the recruiter reviewed candidate pairs; occupation fixed effects are indicators for the specific occupation associated with the job posting (there were ten possible occupations); (2) Sampling method is an indicator for whether the recruiter was sourced via LinkedIn or snowball sampling. (3) Indicators for whether the recruiter reviewed specific tabs in the web platform —Personal Information, Experience, Additional Income, and Education—during the evaluation. Standard errors are clustered at the recruiter level and are robust to heteroskedasticity, with significance levels indicated as follows: * $p < .10$, ** $p < .05$, *** $p < .01$.

didates' job-relevant characteristics. Instead, it suggests that recruiters may be applying gender stereotypes when making their hiring decisions, leading to a systematic preference for women even when male and female candidates are similarly qualified.

Finally, it is plausible that social norms encouraging the increased participation of women in the labor market also play a role in shaping these outcomes. The positive discrimination toward female candidates could reflect a broader societal effort to correct historical gender imbalances, whereby recruiters derive utility not only from selecting the most productive candidate but also from adhering to these prevailing norms. This interpretation aligns with the notion that recruiters are influenced by more than just classical measures of productivity, incorporating considerations that favor the advancement of women in the workforce.

The analysis also shows that discrimination coefficients, with and without recruiters' fixed effects, are nearly identical, indicating that stereotyping is the primary driver of discrimination. By comparing the discrimination coefficients presented in Table 3, we observe the differences between models with and without occupation and trial fixed effects. The comparison reveals that the majority of the observed discrimination can be attributed to stereotyping rather than direct prejudice. This finding aligns with broader social norms, as evidenced by data from the Gender Social Norm Index for Ecuador, which shows that nearly 61% of the population disagrees with statements such as "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). These societal attitudes suggest that our results may be influenced by a social preference for promoting gender equity, reflecting a broader national commitment to reducing gender disparities in the labor market.

In Table 3, significant positive discrimination was observed toward female candidates in stereotypically female-dominated roles, such as cleaning and call center operations, while negative discrimination persisted in traditionally male-dominated jobs like maintenance. This pattern, as illustrated in Figure 3, underscores gender-driven occupational segregation, where women are significantly favored for roles like General Services Assistant - Cleaning (discrimination coefficient of 0.337***) and Call Center Operator (0.1644*), reflecting a bias aligned with traditional gender roles. Conversely, women face negative discrimination in roles such as Commercial Advisor and Maintenance Technician, with corresponding negative coefficients (-0.1204 and -0.2945**, respectively). In other occupations, like Warehouse Keeper, Certified Public Accountant (CPA), and Software Developer, the discrimination coefficients are close to zero, indicating little to no bias in hiring based on gender.

Notice also that, consistent with the results for referrals, for the outcome of job fit in Figure 3, the coefficients are large and positive for cleaning jobs and call center positions. This pattern suggests that the overall zero average differences between men and women in terms of job fit result from averaging negative and positive fits (along with zeroes) across different occupations. These findings reinforce our findings regarding the persistence of traditional gender roles within the labor market, particularly in roles historically dominated by women.

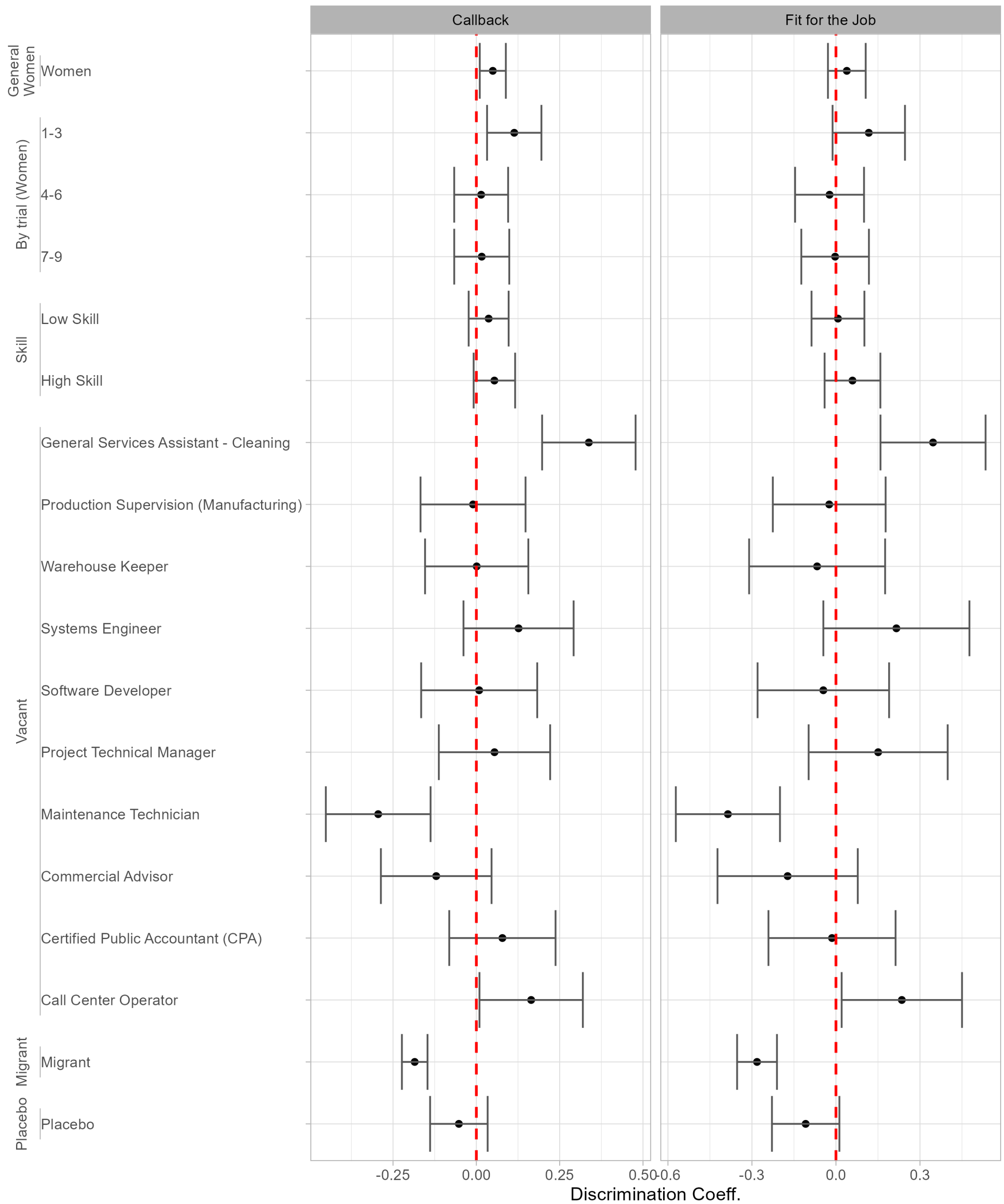
The analysis underscores that while women are increasingly favored in hiring decisions, particularly in traditionally female-dominated roles, this bias does not extend uniformly across all occupations. The similarity between the discrimination coefficients with and without recruiters' fixed effects suggests stereotyping plays a central role, particularly in roles aligned with traditional gender norms. Although positive discrimination toward women may reflect societal efforts to rectify historical gender imbalances, it also reinforces existing occupational segregation by channeling women into stereotypical roles, thereby maintaining rather than dismantling entrenched gender structures in the labor market.

4.1 Robustness tests

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 3. 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,

Figure 3: 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.

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 (Zanoni et al., 2024). In figure 3, we illustrate that migrants also face discrimination in the Ecuadorian labor market (Fabregas and Zanoni, 2024).

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, it varies in complex ways that are likely influenced by stereotypes regarding the attributes judged. This variability is consistent with established psychological theory and suggests that our method captures these nuances effectively.[11]

To assess whether experimenter demand effects could have influenced our findings, 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 3 in this placebo group further supports the credibility of our experimental findings.

Furthermore, 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, the alignment of candidates' skills with the job requirements serves as a proxy variable for the expected value of the marginal product of their labor. Consistency in how recruiters evaluated candidates across that metric, as shown in table 3, suggests the recruiters took the task seriously.

Finally, we highlight that the patterns of recruiters' choices by occupation identified in our field experiment in Figure 3 closely align with the national employment patterns by occupation as analyzed from the 2022 Census microdata in Figure 4. Notably, Figure 4 includes two sets of data: one for the entire country of Ecuador and another for a random sample of individuals from Quito, selected to match the age distribution of job candidates in our experiment. Both the occupational distributions in these datasets and those in our field experiment are similar, underscoring the robustness of our findings. Specifically, the field experiment reveals that discrimination against women varies significantly across occupations, with stronger biases observed in roles such as maintenance technician and systems engineer, which are traditionally male-dominated. This trend is mirrored in the broader employment patterns, where the Census data shows lower employment rates for women in these same occupations. Conversely, in roles such as call center operator and general services assistant (cleaning), where our experiment indicated a preference for female candidates, the Census data similarly reflects higher employment rates for women. The

[11] 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)).

strong alignment between our experimental data and the Census data, both nationally and within Quito stratified by age according to the candidates' data in our experiment, provides compelling support for the external validity of our findings, indicating that the biases observed in our controlled experiment are indeed reflective of broader systemic patterns of occupational segregation in Ecuador's labor market.

(a) Gender Gap in Employment Rate for Ecuador (b) Gender Gap in Employment Rate for Quito Subsample

Figure 4: In this figure, we present the in panel A the gender gap in employment rate for the whole census data and in panel B the gender gap in employment rate for a random sample filtered for Quito matching the distribution of the ages for the synthetic candidates of the field experiment.

5 Discussion and conclusion

6 Conclusions

Our AFE involving recruiters in Ecuador found a preference for female candidates, particularly driven by male participants, reflecting significant gender-driven biases within the hiring process. This trend, aligning with 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. Importantly, our findings highlight that the lack of differences between models with and without fixed effects suggests that stereotyping, rather than explicit prejudice, is the predominant driver of discrimination in this context. This stereotyping behavior is especially pronounced among male recruiters, who appear to be more responsive to societal norms that endorse increased female labor force participation but within traditionally gendered occupations.

Our analysis also underscores the varied nature of gender discrimination across different occupations, indicating that such biases are highly context-dependent and primarily driven by stereotypes rather than objective assessments of productivity. The finding that male recruiters, in particular, are more likely to adhere to traditional gender norms in their hiring decisions adds complexity to the ongoing challenges of achieving genuine gender equality in the labor market. While there has been progress in narrowing gender disparities in labor force participation (though less in wages), our findings suggest that occupational segregation by gender could be a contributing factor to the existing gap in wages.

Our study employs a novel multi-trial 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 than what CS can achieve. Unlike previous studies that primarily focus on high-income countries, our research brings evidence 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 allows for understanding how gender discrimination manifests in different economic environments and across time, particularly in developing countries. By incorporating the dimension of how male recruiters' susceptibility to stereotyping reinforces traditional gender norms, we provide a richer view of the mechanisms behind gender bias in hiring.

By presenting empirical evidence from a developing economy, our study contributes to filling a specific gap in the existing literature, which has predominantly focused on high-income countries (**??**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. 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 can suggest directions to think about the mechanisms of labor market discrimination.

Moreover, the identification of male recruiters as key agents in perpetuating gender stereotypes in hiring decisions underscores the need for interventions that specifically target this group. Addressing the ways in which societal norms influence male recruiters' decisions could be pivotal in breaking down occupational segregation and advancing gender equality in the labor market.

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.

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

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.

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. Urzua (2008). Gender bias in the labor market: Evidence from a correspondence study in chile. *Journal of Economic Inequality 6*(2), 213–230.

Fabregas, R. and W. Zanoni (2024). Discrimination against venezuelan migrants in the ecuadorian labor market. Working Paper.

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.

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

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.

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

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.

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

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.

Nopo, H. and A. Chong (2010). Discrimination in latin america: An economic perspective.

of Labor & Development, I. J. (2023). Gender and ethnic inequalities in lac countries. *IZA Journal of Labor & Development*.

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

Posso, A. (2013). The latin american mystique: Discrimination against women in ecuador's labor market.

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

Schaerer, M., C. du Plessis, M. H. B. Nguyen, R. C. M. van Aert, L. Tiokhin, D. Lakens, E. G. Clemente, T. Pfeiffer, A. Dreber, M. Johannesson, C. J. Clark, and E. L. Uhlmann (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.

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. 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., H. Hernández, O. Zambrano, and G. Quiroz (2024). Discrimination against lesbian and gay job seekers: An artefactual field experiment in urban ecuador. *Labour Economics*. Just Accepted.

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 4: 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.

Table 5: 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 6: 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 7: 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.