Quality & Quantity
https://doi.org/10.1007/s11135-026-02654-9

Discrimination in the university-to-work transition in Italy: a correspondence study

Maria Cristiana Martini[1] · **Luigi Fabbris**[2] · **Rachele Brocco**[3]

Received: 10 March 2025 / Accepted: 13 February 2026

Abstract
In this article, the authors highlight the factors of a graduate's resume that are most relevant in the initial selection of candidates for a job vacancy. Data collection consists of a correspondence study, with a sample of fictitious resumes purposely created for a number of job openings in the Veneto region of Italy. For every vacant position, the authors create five resumes and send them to the companies; in the distribution of the resumes, ten background factors of candidates are controlled. The number of callbacks is used as an indicator to understand preferences and discrimination behaviours of employers. The analyses point out a high level of discrimination against foreigners in the Italian job market, particularly against African and Middle Eastern graduates, but no gender discrimination. Moreover, the results show that, given the non-technical nature of the selected job openings, graduates in humanities and social sciences are preferred to engineers and hard scientists, while there is a preference for master's graduates over bachelor's graduates.

Keywords Discrimination · Ethnic inequality · Correspondence studies · Graduates · Italian labour market · Multilevel logistic regression

1 Introduction

The integration of young adults in the labour market is a problem in many countries, and the youth unemployment rates are high almost everywhere (OECD 2021). Although finding a job is easier for people with higher educational attainment levels, the employment rates of recent graduates differ largely from country to country, with Italy as the EU Member State with the lowest employment rate in 2024 of recent tertiary graduates (ISCED 5–8):

✉ Maria Cristiana Martini
cmartini@unimore.it

[1] Department of Communication and Economics, University of Modena and Reggio Emilia, Viale Allegri 9, 42122 Reggio Emilia, Italy

[2] Tolomeo Studi e Ricerche, Padua, Italy

[3] Independent Researcher, Padua, Italy

Published online: 07 March 2026

only 77.3% against more than 95% in Bulgaria and Estonia, and a European average of 86.7% (Eurostat 2025; table edat_ifse_24). Thus, at least in some countries, the university-to-work transition can be an extremely challenging process, even more so with the increasing uncertainty of the labour market and the changes in the socio-economic context that have followed the 2008 economic crisis (International Labour Office 2017) and the recent pandemic.

The employment rates of recent graduates also differ between men and women, with an average European gender gap steadily in favour of men, varying from 2.7% points in 2022 to 5.4 in 2016 and equal to 4.4% points in 2023; in Italy this gap is even larger, with a difference of 6.8% points in 2023 (72.5% for women, 79.3% for men, see Eurostat 2025; table edat_ifse_24).

Searching for a job is even more ticklish for graduates of foreign origin, who face the double challenge of a precarious job market, and of a discriminating society. Although about 80% of immigrants in Italy, especially those from non-EU countries, work as manual workers (Ministry of Labour and Social Policies 2019), foreign and second-generation graduates are more and more present in the job market, and have reached 4.7% of graduates in 2023 (Almalaurea 2024). Their presence is stronger at the bachelor's level than at the master's one, and their employment rate one year after graduation is slightly lower than for the generality of graduates (Almalaurea 2021).

In this study we aim at analysing the graduates' characteristics that, when underlined in a curriculum vitae, might trigger discriminating mechanisms and a priori refusals or, conversely, can enhance their chances of grabbing attention and favour from the potential employers.

During the university-to-work transition process, in fact, a new graduate usually presents him/herself with his/her curriculum. At the very initial stage of the selection process, there is no direct contact previous to the receipt of the resume. Only at a later stage of the selection process may applicants land an interview and come to a possible agreement with the company. So, knowing employers' attitudes will help new graduates in the thorny transition to work.

This is the justification for the experiment described in the following, that identifies aspects that either improve the candidate's chances of recruitment or potentially compromise the application. The experiment belongs to the category of 'correspondence studies', statistical examinations based on fictitious resumes sent to employers but, instead of applying a typical "matched-pair" design, it relies on a factorial strategy that allows to control for several candidates' characteristics. Its main aims include:

- Estimating the rate of callbacks in response to the sent resumes. The callback rate measures the probability of an applicant getting a response.
- Assessing the existence of inequalities and discriminating behaviours in the preliminary stage of the selection procedure. The considered discriminatory factor is ethnicity and national/regional origin, possibly combined with gender.
- Understanding if possible differences in callback rates depend either on characteristics of enterprises or on the type of activity written in the job opening.
- Pinpointing the aspects that companies value positively or negatively in a resume. We kept under control five applicant's academic achievements: level of higher education, major field of study, final graduation mark and English and computer competencies,

and three social aspects: possessing a personal car, playing music, and volunteering as youth leaders.

In correspondence studies, multiple applications are sent for each job opening, and the correlation among the responses/reactions of the same employer to multiple applications is usually fixed by clustering standard errors at the job-advertisement level. However, some authors claim the superiority of a multilevel regression approach to deal with non-independent data, and recommend a wider use of this method in empirical economic studies (see, e.g., Cheah 2009; Oshchepkov and Shirokanova 2022). Our paper attempts to fill this gap, by addressing the problem of correlated data by means of a multilevel approach with job applications as first level units, and job vacancies as second level units.

We carried out our experiment with reference to companies located in the Veneto region of Italy, and with fictitious graduates from the largest university in that region. The selected region has a strong labour market that includes various manufacturing and service industry opportunities and a large representation of immigrants in lower-level jobs. The graduates' characteristic of coming from the same university aimed to avoid employers guessing about skills achieved in study programs having the same name but delivered by universities with differing reputations. The analysis of such a narrow segment of the Italian business population allowed us to rule out territorial differences and draw more cogent inferences from callbacks regarding the existence of discriminatory factors.

The rest of the paper is organised as follows: Sect. 2 is a short review of correspondence studies; Sect. 3 describes the selection criteria of job openings, the experimental methodology and the survey features; Sect. 4 presents the relevant results of the experiment; Sect. 5 is a discussion of the results in comparison with the relevant literature and the potential limitations of the research design, while Sect. 6 brings some conclusions.

2 Theoretical background

Earlier experiments on the effect of applicant characteristics on callback response have been conducted in a variety of settings: among others, in the U.S., by Newman (1978) and Bertrand and Mullainathan (2004); in Australia, by Riach and Rich (1991); in the U.K., by Jowell and Prescott-Clarke (1970); in Belgium, by Smeesters and Nayer (1998); in the Netherlands, by Bovenkerk et al. (1995); in Sweden, by Carlsson and Rooth (2007); in France, by Duguet et al. (2010) and Adida et al. (2010); in Ireland, by McGinnity and Lunn (2011); in Canada, by Oreopoulos (2011); and in Italy by Busetta et al. (2018). As a whole, these studies signalled race, ethnicity, religion, gender, age and minority-belonging as possible discriminant factors.

In studies focused on race and ethnicity we find almost unanimous evidence of discrimination against minorities, nearly in all continents (see, e.g., the surveys of field experiments conducted by Lippens et al. 2023; Quillian and Midtbøen 2021; Riach and Rich 2002; Rich 2014; Bertrand and Duflo 2017; Neumark 2018). In the Italian job market, self-perceived workplace discrimination is reported significantly more frequently among immigrants (especially Africans) than among Italians (Salvatore et al. 2013; Palumbo and Manna 2019). Also the correspondence study carried out in Italy by Busetta et al. (2018) shows a significant amount of discrimination against first- and second-generation immigrants, but

the number of correspondence studies in the Italian context is limited (Baert 2018). Overall, only few correspondence studies have focused on degree-level job roles and recent graduates (see, e.g., Nunley et al. 2015).

Regarding the discrimination against women, Firth (1982), Neumark et al. (1996), Weichselbaumer (2000), Booth and Leigh (2010), Carlsson (2011) all found that women were discriminated against when they applied for technical and higher-status/higher-paying jobs, but this did not happen for middle- or lower-profile and routine jobs, nor was it evident for traditionally female-dominated jobs. In contrast, Levinson (1975), Weichselbaumer (2000), Arai et al. (2018) and Fullin (2016) found substantial discrimination against men, particularly in women-dominated occupations. Schaerer et al. (2023) claim that hiring discrimination against women has declined significantly and is no longer observed in the last decade.

Even age can be a source of discrimination in an entry-level job (Bendick et al. 1999; Lahey 2008). People of different ages often possess different human capital and different potentials for productivity. Among recent graduates age differences are related to completely different study and life paths, hence employers' discrimination as a direct effect of applicant's age can be difficult to prove and will not be investigated.

Fry (1986) and Graham et al. (1990) tested discrimination on the basis of disability and found substantial discrimination regardless of the type of disability and the type of occupation. Though, this goes beyond the aim of this paper, and a wider set of experiments should be conducted — possibly differentiating by type and level of disability — in order to really understand how to overcome the employers' resistance and favour employability of the disabled.

Riach and Rich (2002), Rooth (2009), Hamermesh (2011) and Galarza and Yamada (2017) have all found that having an attractive body increases the probability of finding a job, while obesity is negatively correlated to labour market outcomes (Busetta et al. 2020a). So, a captivating picture at the top of a curriculum may help at the initial hiring stage. However, since including a photo in a job application is hardly mandatory, examining the so-called "beauty premium" was beyond the scope of this experiment.

Some authors (e.g., Lundberg and Startz 1983; Lahey 2008) distinguish between "statistical" and "taste-based" discrimination. Both types of discrimination derive from preconceptions. Statistical discrimination presumes that certain characteristics of a candidate are negatively related to the quantity and/or quality of his/her human capital. For instance, people who are less educated, or less skilled in a language, or possess other objectively-documented impairments are less productive than a hypothesised worker. As a consequence, if there are enough candidates, applicants having one or more of these negatively perceived characteristics are eliminated from consideration. The taste-based, or animus-based, discrimination (Becker 1957) is instead related to recruiters' values, for instance, of a political or religious or cultural kind. These are values that should not influence either employment or career chances according to constitutions and the human rights declaration.

3 Data and methods

3.1 The selection of openings

We drew the job openings from a specialised website (*subito.it*). From a preliminary comparison with other websites and newspapers, we verified that all job openings advertised through local newspapers were available through the chosen website. Hence, we decided to use just the Internet to collect the ads.

For each vacancy announcement, we composed a set of five resumes and sent them soon after its publication. In case the company's ads included the possibility to work in more than one establishment, only those located in the experimental region were included in the sample. If the announcement was removed before the resumes were sent, another announcement for a similar vacancy was selected and the resumes were sent to this second company. For the purpose of this study, we contacted a total of 142 companies, including substitutions. As a whole, we contacted companies belonging to the service industry sector (43.2%), the other services sector (21.8%), the manufacturing industry (20.5%) and the commerce sector (14.5%). Their size was: micro = 1–9 employees (36.2%); small = 10–49 employees (23.2%); medium = 50–249 employees (27.3%); and large = 250 or more employees (13.3%). The experiment was conducted between October and November 2019.

The job openings considered for the experiment were as follows:

- Job vacancies had to relate to the following five categories: clerical jobs; human resource offices; marketing activities; commercial offices; and information systems. These were jobs for which all hypothesised graduates might be appropriate. By design, each type of activity received the same number of openings, one-fifth of the created resumes.
- All openings explicitly required either a bachelor's or a master's degree. Only vacancies targeted to entry-level jobs, i.e. jobs appropriate for new graduates, were selected.

3.2 The experiment

We wrote six hundred applications allowing the candidate's characteristics to systematically differ according to two experimental conditions:

- Gender (Male; Female);
- Place of origin (Veneto region; Southern Italy; Eastern Europe; North Africa; Central Africa; Middle East); all candidates were hypothesised to reside in the experimental region; foreign-born graduates' birth was placed in the capital city of their origin country, their names were common of nationality so to avoid suspicion of fakery and to avoid social class connotation (see also Zschirnt and Ruedin 2016); for Southern Italy candidates, the place of birth was placed in the regional County Seat.

Any 5-tuple of applications sent to a company contained two candidates of one gender, three of the other and five different countries of origin. This way, each selector had the opportunity to jointly discriminate among genders and all but one nationality. For each job offer, we sent the applications at different times in a two-day period, so as not to create suspicion for the curriculum selectors.

Moreover, we controlled for other eight characteristics, which varied randomly across the applications:

- Academic degree level (Bachelor's; Master's);
- Field of study (Engineering; Sciences; Social sciences; Humanities and education);
- Final degree mark (100–110 cum laude; 85–99; less than 85);
- English knowledge (Mother-tongue or fluent; Intermediate; Basic);
- Computer expertise (Expert or programmer; Intermediate or knowledge of Office; Basic user);
- Car driver (Driving own car; No driving licence);
- Music lover (Not at all; Instrument player or sings in a chorus);
- Youth group volunteer (Not at all; Children's entertainer).

The levels of these eight attributes were determined by the random selection of resumes from the repository. The probability of having more than one resume similar with regard to both human and social capital was very low, about a one in 35 chance for the entire sample (see Sect. 3.3). Moreover, all candidates were about the same age, with two years difference between bachelor's and master's degree holders, for age not to be a discriminating factor.

Also, a set of standard information (ideally unrelated to the experimental attributes) was randomly spread throughout the resumes to improve their realism and perceived credibility. For instance, we took care so that resumes sent to a given company did not indicate an applicant speaking the same second foreign language, nor practising sports at the same engagement level, nor having the same type or duration of odd jobs, and the like. We manually checked all resumes for consistency before sending, and we carefully controlled so that no sentence was repeated nor the writing style was duplicated. This way we aimed to minimise the risk of detection.

Each of the experimental resumes referred to a real person so that the requesting enterprise, when interested in that candidate, could contact him/her either through the telephone or the email address quoted in the resume. Their place and date of birth was explicit in their resumes. We recruited candidates through Facebook and through word of mouth among university students. To allow the candidates to respond to calls/emails, we gave them working mobile phones with prepaid calling cards that enabled them to answer the possible callbacks and used an email specifically created for the experiment. This way, it was possible to control the obtained responses.

The total number of these responding candidates was 12. We inserted each of them in 50 resumes; sex and nationality of candidates corresponded perfectly to those written in the sent job application. The involved individuals consented to participate in the experiment and to observe the confidentiality of the survey data. Of course, people contacted for a job interview declined it politely stating they had already found a job. We compensated the candidates with a basic sum and another amount proportional to their number of responses.

3.3 Random selection of resumes

We created the resumes derived from the full crossing of the levels of the ten characteristics (two experimental factors and eight control variables). This resulted in a repository of

20,736 artificial resumes, each one with at least one attribute level different from the others. We selected the 600 resumes to be sent in a random, but controlled way:

- We randomly sampled the resumes from the repository and progressively cancelled them from the repository after sending them.
- Each level of the 10 experimental characteristics had to appear in the created resumes the same number of times, in order to correctly estimate the experiment main effects. In this way we created 300 resumes with Bachelor's degree and 300 with Masters'; 150 resumes for each of the fields of study; 200 resumes for each class of final grades, English knowledge and computers skills, and so on.
- Also, the experimental levels obtained by crossing pairs of characteristics appeared in the created resumes the same number of times, in order to correctly estimate the interactions between pairs of characteristics; for example, there were 150 males with a master's degree, 150 males with a bachelor's degree, 150 females with a master's degree and 150 females with a bachelor's degree, and the same for each couple of characteristics. To check the equal frequency of each combination of couples of variables, we prepared tables in which the characteristics we wished to control were crossed and controlled manually before sending the resumes.

To the best of our knowledge, this is the only research study that has attempted to control for such a large number of personal factors in order to make inferences about the main and interactive effects of variables that may push in opposite directions while companies take the decision to invite a candidate to an interview.

3.4 A multivariate model for correlated data

The outcome variable Y_{ij} of the analytical model is the occurrence of a callback within seven days for each job application. So, it has only two categories: 1 if the company j called back the job applicant i, and 0 otherwise. The success probability of job application i sent to the job vacancy j is represented by P_{ij}.

Each job vacancy received up to five job applications, and decisions taken by the same firm about the group of received job applications cannot be considered as independent; therefore, we need to handle data correlation. For this purpose, we adapt a two-level random intercept logistic regression model (see, e.g., Snijders and Bosker 1999), with job applications as first level units, and job vacancies as second level units:

$$logit\,(P_{ij}) = \gamma_{\,0} + \sum_{h=1}^{k} \gamma_{\,h} X_{ijh} + U_{0j}$$

where P_{ij} is the probability that the company j called back the job application i, and the logit function of P_{ij} is a sum of the linear function of the experimental variables X_h (h = 1, ... , k) and a random group-dependent deviation U_{0j}[1]. The deviations U_{0j} are assumed to be normally distributed, with zero mean and variance τ_0^2.

[1] The first level residuals are not included because this formula refers to the probability P_{ij}.

We estimated this model by means of the NLMIXED procedure which is part of the statistical software SAS; this procedure attempts to maximise the model likelihood directly by numerical integration methods, more precisely by means of a Dual Quasi Newton optimization algorithm and the adaptive Gaussian quadrature[2] (SAS Institute Inc. 2000).

The first step consists of fitting the empty model with a random intercept, in order to calculate the intraclass correlation coefficient:

$$\rho = \frac{\tau_0^2}{\tau_0^2 + \pi^2/3}$$

The empty model estimate for τ_0^2 is 2.24, which implies an intra-class correlation coefficient $\rho = 2.24/(2.24+3.29) = 0.4054$, indicating that 40.54% of the total variance is due to group variation. This might be partly due to the presence of a large number of small second-level units (142 job vacancies, with an average number of 4.22 job applications each), but also computing the approximate design effect[3] led to a value of 2.31, which is larger than the threshold of 2 suggested by Muthén and Satorra (1995) as an indicator for the need of multilevel modelling.

We partitioned the predictors as follows: the experimental first level variables, whose possible presence in the model pinpoints the idiosyncrasies of the companies calling or not calling back after the resume transmission, and the second level variables, whose role in the model is to explain part of the second level variance.

We estimated a first model that included the preliminarily selected level-one experimental variables and the random intercept, and a second model that included both the level-one experimental variables and the three second level variables (Company activity; Company size; Business sector of vacancy) in addition to the random intercept. We also tried to include 2-way interactions between all possible pairs of levels of the variables place of origin, gender, discipline and degree level, and the cross-level interactions between place of origin and the second level variables, but none was significant, neither at a 90% level.

We repeated the same analyses with all the callback occurrences as the outcome variable; again, the empty model estimate for τ_0^2 showed an intra-class correlation coefficient $\rho = 0.3563$, indicating that more than 35% of the total variance is due to group variation and a multilevel approach is needed.

4 Results

As a whole, we obtained 55 (out of 600: 9.2%) callbacks (Table 1). By a large majority, companies called back through the telephone (36 cases: 6.0%), as an alternative, they used the email (19 cases: 3.2%). For the vast majority of vacancies (71.8%), none of the experimental resumes was called back, while 21.1% called back only one out of the five

[2] The adaptive version of the Gauss-Hermite method consists in first centering the integral on the empirical Bayes estimates of the second level residuals, and then selecting the number of quadrature points adaptively by evaluating the log-likelihood function at the starting values of the parameters; this generally guarantees a more accurate approximation than that obtained by means of a standard Gauss-Hermite quadrature with the same number of quadrature points.

[3] The formula is: DEFF = 1 + (average cluster size − 1) * ρ.

Table 1 Rate of callbacks (%), by nationality and gender of applicants

	Male ($n=300$)		Female ($n=300$)		Total ($n=600$)	
	≤7 days	Overall	≤7 days	Overall	≤7 days	Overall
Veneto, Italy ($n=100$)	12.0	12.0	20.0	20.0	16.0	16.0
South Italy ($n=100$)	8.0	8.0	14.0	18.0	11.0	13.0
Eastern Europe ($n=100$)	6.0	6.0	6.0*	8.0	6.0*	7.0
North Africa ($n=100$)	7.8	7.8	2.0**	6.1	5.0**	7.0
Central Africa ($n=100$)	8.2	12.2	2.0***	3.9*	5.0**	8.0
Middle east ($n=100$)	0.0	0.0	4.0***	8.0	2.0***	4.0*
Total ($n=600$)	7.0	7.7	8.0	10.7[a]	7.5	9.2[a]

* Place of origin is significantly different from Veneto at the 0.10 level

** Place of origin is significantly different from Veneto at the 0.05 level

*** Place of origin is significantly different from Veneto at 0.01 level

[a] Overall rate of callbacks is significantly different from the rate of callbacks within seven days at the 0.01 level

candidates, 4.2% called back two of them, and 2.1% called back three of them; only one of the 142 vacancies for which we sent fictitious resumes generated 4 callbacks. Out of the 55 callbacks, 45 were obtained within seven days from when the resume was sent. No callback exceeded 38 days. For the rest of the analysis, we will pool the obtained responses regardless of the method of communication.

The probability of getting a callback within 7 days (Table 1) was higher for Italians (16% for job applications declaring Veneto origins and 11% for candidates from Southern Italy), than for candidates from Eastern Europe (6%) North Africa (5%), Central Africa (5%), and the Middle East (2%). These differences are significant at a 1% level ($F_{5,494}$=3,82, p=0,002). It appears that, with respect to people born in the same region as companies were located, residents from Southern Italy were not significantly discriminated against. This is partly unexpected, because since the internal migrations of the 1960s–70s the local population showed some level of discrimination against people from the South of Italy (see, e.g., Ginsborg 2003; Ballarino and Panichella 2015).

As compared to local people, foreign-born people were instead very discriminated against. The pattern of discrimination was similar when looking at all the total callbacks, but in this case the differences are smaller ($F_{5, 594}$ = 2,398, p=0,036): responses were less frequent for people from Africa and Eastern Europe than for Italians, but the only significant pairwise comparison is between people from the Middle East and local graduates. We can observe that the rate of callbacks to graduates from Veneto is the same after seven days and after the entire observation period of more than one month, while callbacks of foreigners grow, especially those of candidates from Africa and Middle East, showing that delayed callbacks are mainly directed toward foreigners.

Female graduates are called back slightly more frequently than their male colleagues (8% vs. 7%), but the difference is not significant (p=0,643). However, if we compare male and female callbacks, we can see that discrimination is mainly directed against female foreigners, and in particular against Central Africans. Again, this discrimination lessens when considering the totality of callbacks.

The multilevel logistic regression analysis confirms the same pattern (Tables 2 and 3, Model 1): foreigners are discriminated against, especially Africans and Middle Easterns, but the interaction with gender is non-significant.

Table 2 Multilevel logistic regression models of callbacks within 7 days with first-level variables (Model 1), and with first- and second-level variables (M2)

	Model 0	Model 1	Model 2
Born in South Italy vs. Veneto		− 0.682	− 0.728
Eastern Europe vs. Veneto		− 1.807**	−1.908***
North Africa vs. Veneto		− 1.776***	−1.892***
Central Africa vs. Veneto		− 2.179***	−2.255***
Middle East vs. Veneto		− 2.905***	−3.063***
Gender: Female vs. Male		0.235	0.229
Bachelor vs. Master		− 1.105**	− 1.161**
Field: Hard sciences vs. Engineering		− 0.617	− 0.442
Social science vs. Engineering		1.767***	1.974***
Humanities vs. Engineering		1.154*	1.437**
English skill: Intermediate vs. Expert		0.442	0.484
Basic vs. Expert		− 0.157	− 0.197
ICT skills: Intermediate vs. Expert		0.727	0.773
Basic vs. Expert		0.342	0.360
Final degree mark: High vs. Low		− 0.023	− 0.133
Medium vs. Low		− 0.268	− 0.321
Car driver vs. No car		− 0.482	− 0.654
Music lover		0.008	− 0.011
Youth group volunteer		0.499	0.461
Activity: Commerce vs. Industry			− 0.351
Industry services vs. Industry			0.192
Other services vs. Industry			0.651
Company size: Small vs. Micro			− 0.202
Medium vs. Micro			1.477*
Large vs. Micro			− 0.682
BSV: HR vs. Administration			0.188
Marketing vs. Administration			0.522
ICT vs. Administration			0.800
Commercial office vs. Admin.			0.903
Constant	− 3.309***	− 3.298***	−4.215***
τ_0^2	2.243**	3.634**	3.110*
Proportional change of the variance	Ref.	+62.02%	+38.65%
Intra-class correlation coefficient	0.405	0.525	0.486
AIC	309.3	297.1	306.7
Deviance	305.3	255.1	244.7

Significance level: ***$p<0.01$; **$p<0.05$; *$p<0.10$

Among the control variables, the academic major was significant. The graduates with a title in the humanities or in the social sciences were significantly preferred to engineers and hard scientists for the types of jobs at hand. This may mean that either the jobs retrieved from the Internet were prevalently low-profile, or at least non-technical, or that selectors tended to invite for a job interview graduates they assumed to have lower expectations. Maybe, both mechanisms operated to give these results, though their single relevance is food for thought.

The achieved academic level was significant: *ceteris paribus*, graduates with a master's degree were called back more frequently than those possessing a bachelor's degree, but not

Table 3 Multilevel logistic regression models of callback occurrences with first-level variables (Model 1), and with first- and second-level variables (M2)

	Model 0	Model 1	Model 2
Born in South Italy vs. Veneto		− 0.541	− 0.544
Eastern Europe vs. Veneto		− 1.448**	− 1.490**
North Africa vs. Veneto		− 1.192**	− 1.256**
Central Africa vs. Veneto		− 1.315**	− 1.373**
Middle East vs. Veneto		− 1.948***	−2.066***
Gender: Female vs. Male		0.484	0.494
Bachelor vs. Master		− 0.554	− 0.546
Field: Hard sciences vs. Engineering		− 0.732	− 0.627
Social science vs. Engineering		1.217**	1.331**
Humanities vs. Engineering		0.875*	1.026*
English skill: Intermediate vs. Expert		0.493	0.524
Basic vs. Expert		− 0.245	− 0.302
ICT skills: Intermediate vs. Expert		0.401	0.401
Basic vs. Expert		− 0.033	− 0.068
Final degree mark: High vs. Low		− 0.168	− 0.208
Medium vs. Low		− 0.043	− 0.011
Car driver vs. No car		− 0.450	− 0.535
Music lover		0.208	0.178
Youth group volunteer		0.387	0.349
Activity: Commerce vs. Industry			0.428
Industry services vs. Industry			− 0.198
Other services vs. Industry			0.184
Company size: Small vs. Micro			− 0.363
Medium vs. Micro			0.797
Large vs. Micro			− 0.530
BSV: HR vs. Administration			0.169
Marketing vs. Administration			0.044
ICT vs. Administration			0.557
Commercial office vs. Admin.			0.583
Constant	− 2.926***	− 2.844***	−3.175***
τ_0^2	1.821**	2.080**	1.833*
Proportional change of the variance	Ref.	+ 14.22%	+ 0.66%
Intra-class correlation coefficient	0.356	0.387	0.358
AIC	357.4	355.5	368.4
Deviance	353.4	313.5	306.4

Significance level: ***$p<0.01$; **$p<0.05$; *$p<0.10$

in the models including also the delayed callbacks. This implies that the delayed callbacks are increasingly directed toward bachelor's degree candidates. Neither the final degree mark, nor the English knowledge or computer skills were significant to explain the probability of callbacks, nor did the personal interests or the car ownership.

In Model 2 (Table 2), that includes the second level variables, only the company size effect is marginally significant, showing that medium size companies have a slightly higher rate of callbacks. However, the AIC shows that Model 2 does not perform better than Model 1.

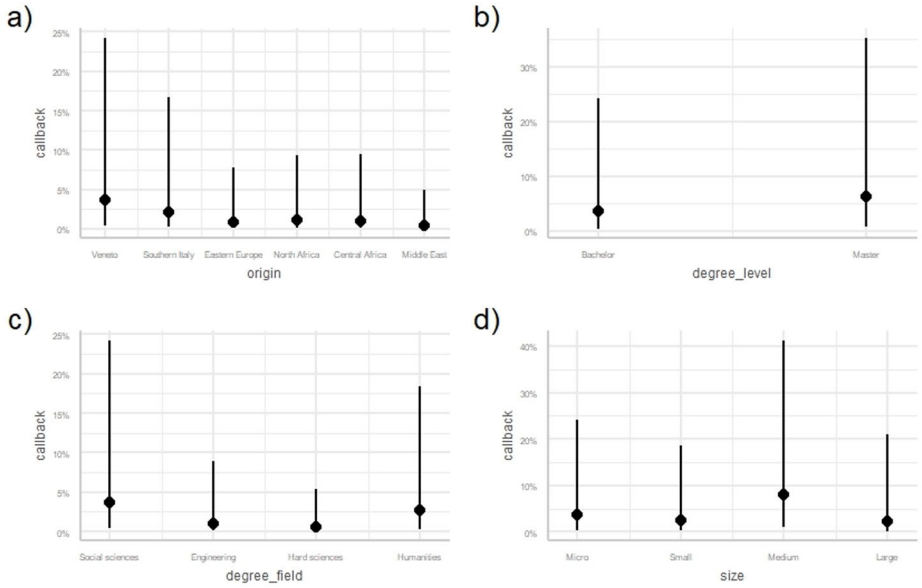

Fig. 1 Predicted probabilities of callback by **a** place of origin, **b** degree level, **c** degree field and **d** company size, with 95% confidence intervals

The comparison of Model 2 and Model 1 deviances also shows a non-significant contribution of the second level variables ($\chi^2_{10} = 255{,}1–244{,}7 = 10{,}4$; *p-value* = 0,406). Moreover, neither the variability of the random intercept, nor the intra-class correlation coefficient, are appreciably reduced by introducing the second level variables[4]. This means that the characteristics of the vacancies and of the companies do not influence the decision to call back a candidate. The same holds true when we estimate models with all the callback occurrences (before and after 7 days) as the outcome variable (Model 3, Table 3).

Figure 1 reports the predicted probabilities of callback with confidence intervals for all the significant variables. We can see once more that the number of callbacks for candidates of foreign origin is noticeably lower than for the Italian candidates, while candidates with a masters' degree or a degree in social sciences or humanities get more callbacks. Although the wide confidence intervals, these results are consistent across different optimization algorithms (Dual Quasi Newton optimization algorithm and Bounded optimization by quadratic approximation), alternative reference categories for categorical variables, different sets of variables (with or without the second level variables, with or without the non-significant predictors) and even with alternative models (standard logistic model, logit multilevel regression or probit multilevel regression).

Figure 2 shows the estimated random intercepts (and 95% CIs) from the full multilevel logistic model. The figure illustrates that, even after adjusting for individual-level and group-level covariates, substantial between-group variation remains. This is consistent with the fact that the second level covariates did not add much to the models. A possible expla-

[4] With a suspicion that the results could be affected by the presence of many non-significant control variables, we tried to fit the models after removing all the non-significant predictors, but the conclusions were totally analogous.

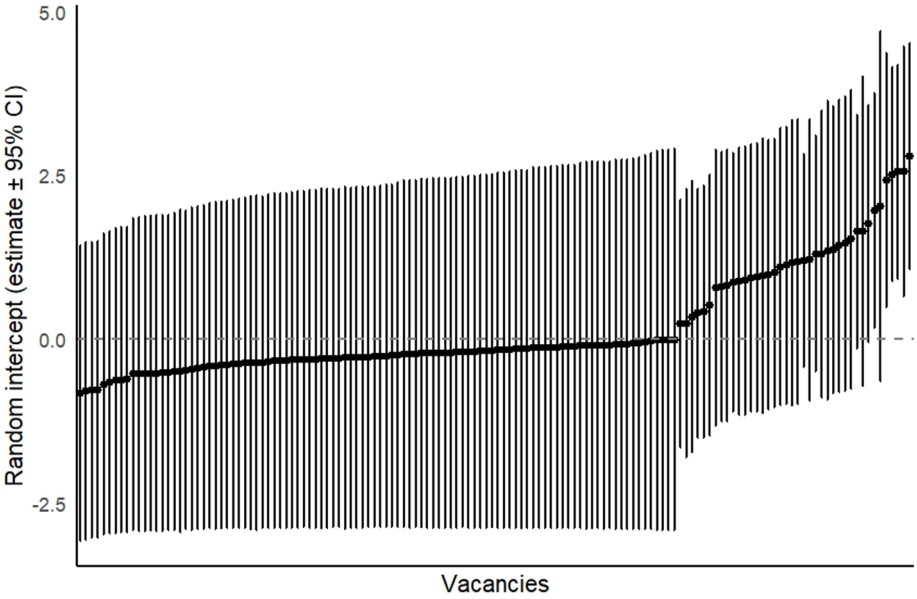

Fig. 2 Caterpillar plot: vacancy-level estimated random effects with 95% confidence intervals

nation resides in the nature of the group-level covariates available in this study, since the decision to call back a candidate is likely to depend on the (unknown) traits of the recruiter more than on the company characteristics. Figure 2 also shows wide confidence intervals, as a consequence of the small size of resumes per vacancy; the CIs are particularly wide for those vacancies that did not produce callbacks for any of the fictional candidates in our experiment, that are the majority of vacancies and are displayed on the left side of the plot. The small group sizes prevented us from adding random slopes to the model.

5 Discussion and limitations

The overall response rate obtained in our experiment was 9.2%. If we consider that this corresponds to less than one out of two job openings, the percentage is rather low. It remains lower than one per job offer even if we parameterize the callbacks to the largest response rate, that of Veneto applicants.

In the literature, response rates differ across populations and methods. Lahey (2008) and Bertrand and Mullainathan (2004) reported experiments similar to ours conducted in the U.S. that obtained an overall 9% and 8.4% callback rate, respectively. In France, Duguet et al. (2010) experimentally tested the potential discrimination of employers against young candidates whose resume made clear their ethnic origin and their residence in the underprivileged suburbs of Paris. The success rate in this experiment was 11.9%. The average response rates were similar to ours.

In Ireland, McGinnity and Lunn (2011) conducted an experiment similar to ours, but sent just two resumes per ad. They obtained 9.6% cases of interviews for both candidates and an

additional 14.6% for just one of the candidates. The proportion of callbacks was higher than ours but not that high even in this study.

Studies conducted in other countries led to similar response rates as ours (for surveys of field experiments, see Riach and Rich 2002; Bøg and Kranendonk 2011; Krause et al. 2012; Quillian et al. 2017). Trying to imagine why this occurred, we can resort to the conjecture in McGinnity and Lunn (2011) that, typically, employers who recruit through job advertisements select only a small number from the candidates who respond, and apply a threshold that is toward the tail of the distribution of suitability across candidates. Thus, the possible responses are few. In a similar vein, Bertrand and Mullainathan (2004) conjecture that the low proportion of callbacks in correspondence studies is due to the fact that employers receive so many resumes to be induced to use quick heuristics in reading the resumes: the authors imagine that, as soon as they see a curriculum adequate to the vacancy, employers stop reading. It is difficult to say which hypothesis is more appropriate.

Preferences are at the core of the selection process, selection being a way to rank people fitting a vacancy. Discrimination may occur if individuals are ranked lower unrelated to their actual performance. The selection rationale may be considered discriminatory if the lower ranks of the list are systematically covered by people belonging to the same social groups. So, in social terms, there is a substantial difference between individual and group discrimination.[5]

The first screening may generate a consistent discrimination toward categories of candidates. An applicant's resume is his/her self-introduction tool. It is expected to picture his/her education and training experience, his/her attitudes and expectations in absence of the possibility to show them to a selector. All experiments evidenced that the real selection, say the limitation of firms' interest to a bunch of applicants, is usually made by employers at the access-to-interview stage. This makes discrimination silent and incontrovertible. Even a small amount of prejudice held by many employers could add up to a high rate of discrimination in the labour market (Cohn 2000). On the other hand, Bøg and Kranendonk (2011) do not exclude that a significant discrimination in the callback rates is consistent with no discrimination at the final offer stage. It depends, they say, on the complexity of the overall recruitment process. The issue of when discrimination may be introduced during the selection process deserves attention by scholars.

Our experiment confirmed the presence of strong discrimination against foreign candidates; if we consider that our study focused on graduates, and Quillian et al. (2019) found that there is less discrimination in jobs that require a college degree than in those that only require a high school degree, we can expect the general labour market to be even more unfair towards foreigners. Consistently with the meta-analysis performed by Lippens et al. (2023), Middle Eastern applicants face the most unfavourable treatment in hiring. Conversely, there is no evidence of gender discrimination, or rather, female candidates are called back slightly more frequently than male candidates, in agreement with Schaerer et al. (2023) who found that, over the last decade, discrimination against female applicants has diminished and sometimes slightly reversed.

[5] This difference was highlighted also in Becker (1957), in which it is stated that the discriminatory effect of a personal characteristic at a certain firm does not imply that discrimination takes place in the market as a whole. Heckman (1998) added that discrimination is when the marginal firm, to which minorities or women actually end up, actualises discrimination. While these are interesting questions in their own right they are beyond the scope of this study.

A methodological question relates to our experiment: is it adequate to highlight the discrimination in the labour market of graduates? McIntyre et al. (1980), Heckman and Siegelman (1993), and Heckman (1998), commenting on discrimination research by employers, argued that other inferences could be deduced from the published conclusions. Since the expected worker's skills cannot be directly observed but only imagined at the hiring stage, we are forced to guess if and to what extent the information contained in a resume is a proxy for relevant factors. Among others, Phelps (1972) states that even skin colour or gender can be taken as proxies for relevant data not sampled. Whatever tool is used to highlight discrimination, it is possible that an indicator of prejudice precedes something else, for instance, the expected productivity of candidates. Heckman asks if the inequalities observed in market outcomes are a proof of discrimination in the market or, instead, does the observed variability reflect the expected difference in candidates' skills? The juxtaposition between taste-based discrimination, led by irrational animosity towards minorities, and statistical discrimination, motivated by a rational optimising behaviour in response to limited information, has informed the economics studies on labour market discrimination for decades (Guryan and Charles 2013). Each response would lead to a different policy: the former might imply the necessity to reduce employers' discrimination, the latter the necessity of policies for the improvement of skills and the social environment of certain categories of candidates.

One common strategy to exclude statistical discrimination is to provide a great amount of information in the fictional resumes. However, other characteristics of the candidates remain unknown to the recruiter and have to be guessed. In fact, a worker's productivity might depend not only on his/her educational path, but also on his or her social participation in a firm's activities and positive attitudes towards a company's success. In other words, the worker's personality traits are to be considered a matter of assessment at the insertion stage. Should a company enrol an applicant who shows a leaving attitude even before recruitment, or who showed in previous work experiences negative attitudes toward collective life, or else who might enforce the dominance of a particular ethnic or linguistic group within the firm? Baert and De Pauw (2014) find that employers' concern that co-workers and customers prefer collaborating with natives drives discrimination. Neumark (2012) suggests a way to address, under some assumptions, the Heckman-Siegelman critique. Alternatively, a research design might include varying levels of information provided in the CVs, in order to manipulate the strength of productivity signals. Anyway, there is no simple answer to this question. Lippens et al. (2022), in a systematic review of the recent literature, claim that ethnic discrimination in the labour market is in general better explained by taste-based discrimination, while Quillian and Midtbøen (2021) question the relevance of this distinction and affirm that "statistical discrimination is no less odious or harmful than other forms of discrimination". In our experiment, we tried to provide a great amount of information in the fictional resumes to exclude statistical discrimination, at least the one based on academic achievements: we declared the final degree mark, the English level and the ICT skills, along with personal characteristics and interests. The high levels of discrimination observed despite this precaution suggest a preponderance of taste-based discrimination, but a definitive solution remains elusive.

Another methodological issue stems from the fact that, for randomization purposes, we created a bank of resumes of essentially equally qualified applicants. As an alternative, Midtbøen and Rogstadt (2012) suggest creating resumes for two subsequent studies,

a correspondence study and an in-depth interview of employers exposed to the fictitious applications. With an analogous purpose, Neumark (2012) upholds that a correspondence study should include some degree of variability for certain characteristics that affect the probability of being hired. So, once the researcher detects that a set of productivity-related characteristics of applicants has affected hiring outcomes, it is possible both to test for bias in the estimates of the two groups (related to the null-discrimination hypothesis and that of the discrimination hypothesis) and also to detect the actual discrimination level.

One interesting result comes from the comparison of the models for the callbacks within seven days and overall: employers tend to call back first Italian candidates and those with a master's degree, while delayed callbacks are concentrated on foreigners and bachelor's degrees. We can hypothesise two alternative reasons for this behaviour: on the one hand, fast reactions are proven to be more governed by stereotyping and discrimination (see, e.g., Casey et al. 2013; Payne 2006), then the employers who call back a candidate a few days after receiving his/her resume might be less likely to discriminate against foreigners; on the other hand, foreigners (as much as bachelors' degrees) might be considered as a second choice, suitable after the job has been declined by other candidates, and these delayed callbacks might be a further symptom of discrimination. A larger sample of vacancies, and consequently a larger number of callbacks, might allow to model the response time and get a deeper understanding of the reasons why, on average, foreigners are called back later than Italians.

The study's design includes some aspects that could represent potential limitations. For example, the focus on graduates from the same university and firms based in only one Italian region has consequences. On the positive side, we can test the existence of regional discrimination, we rule out the possibility that the employers be influenced in their choice by the unlike reputations of equivalent degrees attained in different universities, and we control for differences in recruiting strategies that might be due to the characteristics of the various local economies. On the other hand, although their consistency with the results of analogous studies, we cannot be sure that our findings would be the same at a national level; Busetta et al. (2020b), for example, find that taste-based discrimination in Italy is higher in the Northern and Central regions than in the South.

The study implicitly assumes that all candidates, regardless of national origin, possess full command of Italian, as they are presented as graduates from an Italian university. This reduces the realism of the experimental design. In fact, many Italian universities now offer entire degree programs in English, which means that some foreign graduates may not have strong Italian language skills. In this sense, the study captures discrimination conditional on equal Italian proficiency, rather than reflecting the broader labour market situation where language differences may also play a role.

As previously highlighted, the study focuses on vacancies posted on the website subito.it. Strictly speaking, we had no sampling problem since openings were inserted in the experiment without sampling. Indeed, the type of jobs advertised through the Internet was peculiar, as evidenced in Sect. 3. Although all the selected ads were addressed to graduates, a considerable part of the examined ads concerned medium-to-low profile jobs. In order to limit the observed preference for graduates of social sciences and humanities, we should have sampled firms from a possible list that posted job ads representative of the whole spectrum of insertion wages and positions. In particular, we should have considered a larger quota of jobs from a hypothetical technical and higher-profile list. Would this possibility

have given no evidence of discrimination, or possibly of a different kind of discrimination? This is difficult to say. Heckman himself, after suggesting a similar 'purposive sorting' of firms to solve the problem he saw in the audit-pair method, admitted that there may be "evil lurking in the hearts of firms that is never manifest in consummated market transactions". On the other hand, Auspurg et al. (2020), in a study on racial discrimination in housing, compared results of field experiments drawn from different sources and found that results change very little, suggesting that field experiments be not sensitive to sampling methods. In this sense, the bias induced by this selection of vacancies might be less relevant than feared, especially if we consider that the focus is on recent graduates and entry-level jobs.

The two-level random intercept logistic regression model addressed the problem of the strong dependence of observations within vacancies, though the second-level covariates, referred to the company and the vacancy position, were scarcely effective in explaining the variability of decisions that depend on the recruiters more than on the companies. Finally, the study design itself, with five resumes sent for each vacancy, implies some limitations, as the limited groups size discouraged from introducing random slopes in the model. In this way the potential of multilevel modelling is to some extent weakened. Future studies might consider sending a larger number of CVs for each job vacancy, as in Firth (1981; 1982) or in Busetta et al. (2018), but this would raise the risk of suspicion and might ruin the experiment.

6 Conclusion

Our experiment concerned the detection of the variables that affect the variability of call-back rates obtained in response to the sending of resumes to firms. The analysis showed that firms vary in propensity to contact candidates for a job interview according to nationality, level and major of university degrees. We found evidence of discrimination against foreign candidates and an amount of discrimination in which prejudice operated inversely than expected.

Discrimination against African candidates was higher for females than for males, while females from the Middle East are less discriminated against than males from the same area; this is consistent with the findings of Arai et al. (2018) and Dahl and Krog (2018), who found that Arabic men face stronger discrimination in the labour market than Arabic women. No substantial gender difference was found among candidates from Eastern Europe.

But why are there differences in discrimination? Bail (2008) and Fullin (2016) stress that religion, on top of somatic differences and cultural habits, may explain how applicants are perceived in the labour market. Sniderman et al. (2002) highlight the peculiarity of Italian immigration, which includes that from countries like Albania and Romania, linked to Italy by cultural and political ties that might give these immigrants a more privileged position in the labour market. Koopmans et al. (2019) and Quillian and Midtbøen (2021) claim that the perception of cultural distance increases discrimination. Though, all this explains just part of our results. Applicants from Eastern Europe suffer the same discrimination level as Africans, while Middle Eastern applicants, who are likely perceived as practising a different religion as Africans but are somatically closer to local people, and then should be preferred to black people, are far back in employers' taste.

Following Sniderman et al. (2002), we conjecture that, in Italy as elsewhere (Booth et al. 2012; Quillian et al. 2017; Rinne 2018), there is a hidden hostility to immigrants, whatever their provenance. Insecurity is an additional feeling causing the low callback rate to both Eastern Europeans and people from the Middle East. Prejudices seem more influenced by the pictures of ethnic groups used in the mass media, than by somatic differences.

What is somewhat unexpected is the 'inverse' discrimination against graduate applicants, since we found that social sciences and humanities degrees were preferred to more scientific and technical ones. Although a bachelor's or master's degree was required, most of the selected vacancies concerned non-technical positions, for which social sciences and humanities graduates might be considered better aligned. Nevertheless, if a candidate with a more technical background shows interest for a nonspecific job, what prevents the employer from calling him back and make an interview? We can hypothesise that the higher proportion of callbacks to social sciences and humanities graduates also indicates a preference for lesser-demanding candidates, and the will to avoid losing time in job interviews to candidates that might refuse the job. Although the disadvantaged category be different in this case, this result is analogous to that found by Neumark et al. (1996), where the weaker category (in that case, women) was discriminated against in the higher-level workplaces (in that case, high-price restaurants), while it was favoured in the lower-level ones. The examined job offers considered in this experiment were addressed to less-demanding, less-entitled applicants, and can then be assimilated to the low-price restaurants in Neumark et al. (1996) experiment.

We found that the Internet conveys some entry-level, medium-to-high profile jobs, ideally suitable to graduates, but many more short-term, maternity-leave and the like substitutions, and entry-paid jobs. Based on the assumption that a more expensive ad through a newspaper indicates that employers put more emphasis on that job, we should sample ads separately from that channel and measure how this source differs. If the hypothesis of correlation between ad sources and expected wages holds, other conveyance channels should be sampled (public and private employment agencies, job offers conveyed through university channels, etc.). Indeed, a stratified-by-source sample of ads would merit additional research.

For future experiments, we list some questions that puzzle us. Which realistic characteristics should qualify a resume? We could distinguish part-time positions, season in which sampling was done, distance between home and work, candidate's previous work experience or involvement in political/union/ethnic/environmental/other social activities, and like characteristics that could influence employers' response. Intuitively, these variables are relevant alone but their relevance should be evaluated balancing them with other possibly favourable characteristics. Moreover, since unobservable error variance may mislead inference in correspondence studies, the richer a (fictitious) resume in terms of applicant's characteristics, the more it is reasonable assuming that there are lesser differences in unobservable characteristics across groups of potential employees and across social environments for the estimation of discrimination.

Because a resume is a unique body that includes many aspects of the experience, personality and expectations of applicants, and the selection process of resumes is an employer's choice, one could use conjoint analysis (Louviere 1988) as a method to highlight the aspects that increase and those that decrease the callback probability and the compensation capacity of the former ones to fight discrimination.

Moreover, how should we spread information to circumscribe the risk of suspicion without giving relevant hints that may ruin the experiment? Is that risk correlated to the number of candidates per job opening? In literature, the minimum is one and the maximum varies according to circumstances and researcher's preferences. Among others, Weichselbaumer (2020) sent only one resume per ad; Newman (1978) and McGinnity and Lunn (2011) adopted a number of two resumes; Hubbuck and Carter (1980), Brown and Gay (1985) and Adida et al. (2010) used three; Firth (1981) sent seven resumes; Firth (1982) and Busetta et al. (2018) even more. Which, if any, is an optimum number of resumes for such an experiment? Finally, does a typo or other expressions of inadequacy in a resume have the same effect in a job application of a native than a foreign candidate?

Author contributions All authors contributed to the study conception and design, and share the responsibility for the whole paper. Material preparation and data collection were performed by Rachele Brocco. Data analysis was performed by Maria Cristiana Martini. Luigi Fabbris edited Sects. 2 and 6; Rachele Brocco edited Sects. 3.1, 3.2, and 3.3; Maria Cristiana Martini edited all other sections.

Funding Open access funding provided by Università degli Studi di Modena e Reggio Emilia within the CRUI-CARE Agreement. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Adida C.L., Laitin D.D., Valfort M.-A.: Identifying barriers to muslim integration in France. Proc. Natl. Acad. Sci. USA. (2010). https://doi.org/10.1073/pnas.1015550107

Almalaurea: XXIII indagine - Profilo dei laureati 2020. Consorzio Interuniversitario Almalaurea, Bologna (2021)

Almalaurea: XXVI indagine - Profilo dei laureati 2023. Consorzio Interuniversitario Almalaurea, Bologna (2024)

Arai, M., Bursell, M., Nekby, L.: The reverse gender gap in ethnic discrimination: employer stereotypes of men and women with Arabic names. Int. Migr. Rev. (2018). https://doi.org/10.1111/imre.1217

Auspurg, K., Schnek, A., Thiel, F.: Different samples, different results? How sampling techniques affect the results of field experiments on ethnic discrimination. Res. Soc. Strat Mobil. (2020). https://doi.org/10.1016/j.rssm.2019.100444

Baert, S., De Pauw, A.S.: Is ethnic discrimination due to distaste or statistics? Econ. Lett. (2014). https://doi.org/10.1016/j.econlet.2014.09.020

Baert, S.: Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In: Gaddis, S. (eds.) Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Methodos Series, vol 14, pp.63–77. Springer, Cham (2018).

Bail, C.A.: The configuration of symbolic boundaries against immigrants in Europe. Am. Sociol. Rev. (2008). https://doi.org/10.1177/000312240807300103
Ballarino, G., Panichella, N.: The educational integration of second generation southern Italian migrants to the north. Demogr Res. (2015). https://doi.org/10.4054/DemRes.2015.33.39
Becker, G.S.: The Economics of Discrimination. University of Chicago Press, Chicago (1957)
Bendick, M. Jr., Brown, L.E., Wall, K.: No foot in the door: an experimental study of employment discrimination against older workers. J. Aging Soc. Policy. (1999). https://doi.org/10.1300/J031v10n04_02
Bertrand, M., Duflo, E.: Field experiments on discrimination. In: Banerjee, A.V., Duflo, E. (eds.) Handbook of Field Experiments, pp. 309–393. North Holland, Amsterdam (2017). https://doi.org/10.1016/bs.hefe.2016.08.004
Bertrand, M., Mullainathan, S.: Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am. Econ. Rev. (2004). https://doi.org/10.1257/0002828042002561
Bøg, M., Kranendonk, E.: Labor Market Discrimination of Minorities? Yes, but not in Job Offers. MPRA, Munich (2011)
Booth, A., Leigh, A.: Do employers discriminate by gender? A field experiment in female-dominated occupations. Econ. Lett. (2010). https://doi.org/10.1016/j.econlet.2010.01.034
Booth, A., Leigh, A., Varganova, E.: Does ethnic discrimination vary across minority groups? Evidence from a field experiment. Oxf. Bull. Econ. Stat. (2012). https://doi.org/10.1111/j.1468-0084.2011.00664.x
Bovenkerk, F., Gras, M.J.I., Ramsoedh, D.: Discrimination Against Migrant Workers and Ethnic Minorities in Access to Employment in the Netherlands. International Migration Papers, vol. 4. International Labor Office, Geneva (1995)
Brown, C., Gay, P.: Racial Discrimination 17 Years after the Act. Policy Studies Institute, London (1985)
Busetta, G., Campolo, M.G., Panarello, D.: Immigrants and Italian labor market: statistical or taste- based discrimination? Genus. (2018). https://doi.org/10.1186/s41118-018-0030-1
Busetta, G., Campolo, M.G., Panarello, D.: The discrimination decomposition index: a new instrument to separate statistical and taste-based discrimination using first-and second-generation immigrants. Int. J. Soc. Econ. (2020b). https://doi.org/10.1108/IJSE-02-2020-0055
Busetta, G., Campolo, M.G., Panarello, D.: Weight-based discrimination in the Italian labor market: an analysis of the interaction with gender and ethnicity. J. Econ. Inequal. (2020a). https://doi.org/10.1007/s10888-020-09456-5
Carlsson, M.: Does hiring discrimination cause gender segregation in the Swedish labor market? Fem Econ. (2011). https://doi.org/10.1080/13545701.2011.580700
Carlsson, M., Rooth, D.-O.: Evidence of ethnic discrimination in the Swedish labor market using experimental data. Labour Econ. (2007). https://doi.org/10.1016/j.labeco.2007.05.001
Casey, P.M., Warren, R.K., Cheesman, F.L., Elek, J.K.: Addressing implicit bias in the courts. Court Rev. **49**(1), 64–70 (2013)
Cheah, B.C.: Clustering standard errors or modeling multilevel data? Research note. Columbia University, New York. (2009). https://statmodeling.stat.columbia.edu/wp-content/uploads/2009/08/RevisitMoulton_21.pdf Accessed November 2024
Cohn, S.: Race and Gender Discrimination at Work. Westview, Colorado (2000)
Dahl, M., Krog, N.: Experimental evidence of discrimination in the labour market: intersections between ethnicity, gender, and socio-economic status. Eur. Sociol. Rev. (2018). https://doi.org/10.1093/esr/jcy020
Duguet, E., Leandri, N., L'Horty, Y., Petit, P.: Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area. Annals Econ. Stat. (2010). https://doi.org/10.2307/41219165
Eurostat: Employment rate of recent graduates. Education and training in the EU - Facts and figures. Statistics explained. Eurostat, Luxembourg. (2025). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Employment_rates_of_recent_graduates Accessed November 2025
Firth, M.: Racial discrimination in the British labor market. Ind. Labor Relat. Rev. (1981). https://doi.org/10.2307/2522541
Firth, M.: Sex discrimination in job opportunities for women. Sex. Roles. (1982). https://doi.org/10.1007/BF00287858
Fry, E.: An Equal Chance for Disabled People? The Spastics Society, Campaigns and Parliamentary Departments, London (1986)
Fullin, G.: Labour market outcomes of immigrants in a South European country: do race and religion matter? Work Employ. Soc. (2016). https://doi.org/10.1177/0950017015575867
Galarza, F.B., Yamada, G.: Triple penalty in employment access: the role of beauty, race, and sex. J. Appl. Econ. (2017). https://doi.org/10.1016/S1514-0326(17)30002-8
Ginsborg, P.: A History of Contemporary Italy: Society and Politics 1943–1988. Palgrave Mcmillan, New York (2003)

Graham, P., Jordan, A., Lamb, B.: An Equal Chance? Or no Chance? A Study of Discrimination Against Disabled People in the Labour Market. The Spastics Society, London (1990)
Guryan, J., Charles, K.K.: Taste-based or statistical discrimination: The economics of discrimination returns to its roots. Econ. J. **123**(572), F417 (2013). https://www.jstor.org/stable/42919257
Hamermesh, D.: Beauty Pays: Why Attractive People are More Successful. Princeton University Press, Princeton, NJ (2011)
Heckman, J.J.: Detecting discrimination. J. Econ. Perspect. (1998). https://doi.org/10.1257/jep.12.2.101
Heckman, J.J., Siegelman, P.: The Urban Institute audit studies: Their methods and findings. In: Fix, M., Struyk, R.J. (eds.) Clear and Convincing Evidence: Measurement of discrimination in America, pp. 187–258. Urban Institute, Lanham, MD (1993)
Hubbuck, J., Carter, S.: Half a Chance? A Report on Job Discrimination Against Young Blacks in Nottingham. Commission for Racial Equality, London (1980)
International Labour Office: Global Employment Trends for Youth 2017: Paths to a Better Working Future. ILO, Geneva (2017)
Jowell, R., Prescott-Clarke, P.: Racial discrimination and white collar workers in Britain. Race Cl. (1970). https://doi.org/10.1177/030639687001100401
Koopmans, R., Veit, S., Yemane, R.: Taste or statistics? A correspondence study of ethnic, racial and religious labour market discriminations in Germany. Ethn. Racial Stud. (2019). https://doi.org/10.1080/01419870.2019.1654114
Krause, A., Rinne, U., Zimmermann, K.F.: Anonymous job applications in Europe. IZA J. Eur. Labor Stud. (2012). https://doi.org/10.1186/2193-9012-1-5
Lahey, J.: Age, women, and hiring: an experimental study. J. Hum. Resour. (2008). https://doi.org/10.3368/jhr.43.1.30
Levinson, R.M.: Sex discrimination and employment practice: an experiment with unconventional job inquiries. Soc. Probl. (1975). https://doi.org/10.2307/799750
Lippens, L., Baert, S., Ghekiere, A., Verhaeghe, P.-P., Derous, E.: Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence. J. Ethn. Migr. Stud. (2022). https://doi.org/10.1080/1369183X.2022.2050191
Lippens, L., Vermeiren, S., Baert, S.: The state of hiring discrimination: a meta-analysis of (almost) all recent correspondence experiments. Eur. Econ. Rev. (2023). https://doi.org/10.1016/j.euroecorev.2022.104315
Louviere, J.J.: Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity. J. Transp. Econ. Policy **22**, 93–119 (1988). https://www.jstor.org/stable/20052837
Lundberg, S.J., Startz, R.: Private discrimination and social intervention in competitive labor markets. Am. Econ. Rev. **73**(3), 340–347 (1983). https://www.jstor.org/stable/1808117
McGinnity, F., Lunn, P.D.: Measuring discrimination facing ethnic minority job applicants: an Irish experiment. Work Employ. Soc. (2011). https://doi.org/10.1177/0950017011419722
McIntyre, S.J., Moberg, D.J., Posner, B.Z.: Discrimination in recruitment: an empirical analysis. Comment. Ind. Labor Relat. Rev. (1980). https://doi.org/10.2307/2522700
Midtbøen, A.H., Rogstadt, J.: Discrimination: methodological controversies and sociological perspectives on future research. Nordic J. Migration Res. (2012). https://doi.org/10.2478/v10202-011-0046-5
Ministry of Labour and Social Policies: 9th Annual Report. Foreigners in the Italian labour market. Summary of the main findings and International perspective. Ministry of Labour and Social Policies, Rome: (2019)
Muthén, B.O., Satorra, A.: Complex sample data in structural equation modeling. In: Mardsen, P.V. (ed.) Sociological Methodology, pp. 267–316. American Sociological Association, Washington DC (1995)
Neumark, D., Bank, R.J., Van Nort, K.D.: Sex discrimination in restaurant hiring: an audit study. Q. J. Econ. (1996). https://doi.org/10.2307/2946676
Neumark, D.: Detecting discrimination in audit and correspondence studies. J. Hum. Resour. (2012). https://doi.org/10.1353/jhr.2012.0032
Neumark, D.: Experimental research on labor market discrimination. J. Econ. Lit. (2018). https://doi.org/10.1257/jel.20161309
Newman, J.M.: Discrimination in recruitment: an empirical analysis. Ind. Labor Relat. Rev. (1978). https://doi.org/10.2307/2522415
Nunley, J.M., Pugh, A., Romero, N., Seals, A.: Racial discrimination in the labor market for recent college graduates: evidence from a field experiment. BE J. Econ. Anal. Policy. (2015). https://doi.org/10.1515/bejeap-2014-0082
OECD: OECD Employment Outlook 2021: Navigating the COVID-19 Crisis and Recovery. OECD Publishing, Paris (2021)
Oreopoulos, P.: Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. Am. Econ. J. Econ. Policy. (2011). https://doi.org/10.1257/pol.3.4.148

Oshchepkov, A., Shirokanova, A.: Bridging the gap between multilevel modeling and economic methods. Soc. Sci. Res. (2022). https://doi.org/10.1016/j.ssresearch.2021.102689
Palumbo, R., Manna, R.: A portrait of workplace discrimination in Italy: empirical evidence from a nationwide survey. Evidence-based HRM. (2019). https://doi.org/10.1108/EBHRM-03-2019-0029
Payne, K.B.: Weapon bias: split-second decisions and unintended stereotyping. Curr. Dir. Psychol. Sci. (2006). https://doi.org/10.1111/j.1467-8721.2006.00454.x
Phelps, E.S.: The statistical theory of racism and sexism. Am. Econ. Rev. **62**(4), 659–661 (1972). https://www.jstor.org/stable/1806107
Quillian, L., Heath, A., Pager, D., Midtbøen, A.H., Fleischmann, F., Hexel, O.: Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring. Sociol. Sci. (2019). https://doi.org/10.15195/v6.a18
Quillian, L., Midtbøen, A.H.: Comparative perspectives on racial discrimination in hiring: the rise of field experiments. Annu. Rev. Sociol. (2021). https://doi.org/10.1146/annurev-soc-090420-035144
Quillian, L., Pager, D., Hexel, O., Midtbøen, A.H.: Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proc. Natl. Acad. Sci. USA (2017). https://doi.org/10.1073/pnas.1706255114
Riach, P.A., Rich, J.: Field experiments of discrimination in the market place. Econ. J. (2002). https://doi.org/10.1111/1468-0297.00080
Riach, P.A., Rich, J.: Testing for racial discrimination in the labour market. Camb. J. Econ. (1991). https://doi.org/10.1093/oxfordjournals.cje.a035168
Rich, J.: What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. IZA Discussion Paper No. 8584. Institute of Labor Economics (2014). https://docs.iza.org/dp8584.pdf
Rinne, U.: Anonymous job applications and hiring discrimination. IZA World Labor. (2018). https://doi.org/10.15185/izalwol.48
Rooth, D.-O.: Obesity, attractiveness, and differential treatment in hiring a field experiment. J. Hum. Resour. (2009). https://doi.org/10.1353/jhr.2009.0027
Salvatore, M.A., Baglio, G., Cacciani, L., Spagnolo, A., Rosano, A.: Work-related injuries among immigrant workers in Italy. J. Immigr. Minor. Health. (2013). https://doi.org/10.1007/s10903-012-9673-8
SAS Institute Inc: SAS/STAT® Software: Changes and Enhancements, Release 8.1. SAS Institute Inc., Cary, NC (2000)
Schaerer, M., du Plessis, C., Nguyen, M.H.B., van Aert, R.C.M., Tiokhin, L., Lakens, D., Clemente, G., Pfeiffer, E., Dreber, T., Johannesson, A., Clark, M., Uhlmann, C.J.L.: Gender Audits forecasting collaboration: on the trajectory of discrimination: a meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions. Org. Behav. Hum. Dec. (2023). https://doi.org/10.1016/j.obhdp.2023.104280
Smeesters, B., Nayer, A.: La discrimination à l'accès à l'emploi en raison de l'origine étrangère: Le cas de la Belgique. International Migration Papers, No. 23. International Labor Office, Geneva (1998)
Sniderman, P.M., Peri, P., De Figuereido, R.J.P., Piazza, T.: The Outsider: Prejudice and Politics in Italy. Princeton University Press, Princeton, NJ (2002)
Snijders, T.A.B., Bosker, R.J.: Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. Sage, London (1999)
Weichselbaumer, D.: Is it sex or personality? The impact of sex-stereotypes on discrimination in applicant selection. East. Econ. J. (2000). https://doi.org/10.2139/ssrn.251249
Weichselbaumer, D.: Multiple discrimination against female immigrants wearing headscarves. Ind. Labor Relat. Rev. (2020). https://doi.org/10.1177/0019793919875707
Zschirnt, E., Ruedin, D.: Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. J. Ethn. Migr. Stud. (2016). https://doi.org/10.1080/1369183X.2015.1133279