NET WORK LISER

Discussion Paper Series

IZA DP No. 19024

October 2026

To Discriminate, or Not to Discriminate, in Online Job Ads

Tho Pham
University of York

Daniel Schaefer
Johannes Kepler University Linz

Carl Singleton
University of Stirling
and IZA@LISER

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

IZA DP No. 19024 October 2026

To Discriminate, or Not to Discriminate, in Online Job Ads*

Abstract

How do employers use and adjust recruitment signals and wage offers when they target applicants of a particular gender? Using online vacancy data from Vietnam and natural language processing methods, we construct an index measuring the direction and intensity of implicit gendered language in job requirements. We document widespread implicit gendered framing and substantial variation across repeated postings for the same firm and job, even where explicit gender requests remain unchanged. When explicit requests do change, we find that the two signals move together: adding a gender request strengthens wording oriented toward that gender, while removing it weakens such wording. Both signals are also associated with posted wages. Within positions, adding a male request is accompanied by 3% higher wages, and adding a female request by 4% lower wages. More male-coded wording is likewise associated with higher wages. A stylised directed-search model can rationalise these patterns when applicants' wage responsiveness differs by gender and inconsistent recruitment signals discourage applicants.

JEL classification

J16, J63, J71

Keywords

gender discrimination, online vacancy data, wage posting, text analysis, directed search

Corresponding author

Tho Pham
tho.pham@york.ac.uk

* We thank seminar participants for the valuable comments and feedback received at Heriot-Watt University, Maynooth University, University of Edinburgh, University of Stirling, University of York, and the Bank of Finland Institute for Emerging Economies (BOFIT), as well as attendees of the Workshop on Labour and Family Economics 2025 (York), the 2026 Asia Meeting of the Econometric Society (Hong Kong), and the European Association for Labour Economists (EALE) Conference 2026 (Barcelona). Pham thanks the Department of Economics and Related Studies (University of York) for financial support.

1. Introduction

Gender inequality in labour markets remains persistent and pervasive in both high-income and developing economies. While much research has focused on wage gaps (Blau and Kahn, 2017; Goldin, 2014), promotion dynamics (Azmat and Petrongolo, 2014; Bertrand, Goldin and Katz, 2010), and sorting across occupations or firms (Blau and Kahn, 2017; Card, Cardoso and Kline, 2016), a growing literature highlights that gender inequality may arise even earlier in the hiring process, through the ways vacancies are advertised and presented to potential applicants (e.g., Card, Colella and Lalive, 2025; Kuhn and Shen, 2013, 2023). In many developed countries, explicit gender requests and profiling in job ads are legally restricted. Yet employers may still convey gender preferences *implicitly* through the wording of required skills, attributes, and personal characteristics. Whether such language carries meaningful preferences or costless signalling remains an open question, with direct implications for our understanding of the incentives that shape how employers advertise their jobs and the design of anti-discrimination policy.

This paper studies how gender requests, implicit wording, and posted wages vary across repeated ads by the same firm for the same job. We use more than 300,000 online postings from one of Vietnam's largest job boards over the 2019–2020 period. This setting enables us to observe both communication channels: over a quarter of ads state an explicit gender request. We construct a *Gender Index* by applying sentence-embedding models to the distinct requirements text of each ad. Unlike approaches that rely on predefined lists of gendered keywords (e.g., Gaucher, Friesen and Kay, 2011), our measure captures the full semantic pattern of an ad's text. The index reveals substantial implicit gendered language even in ads without explicit gender requests.

Our empirical design separates persistent differences across positions from changes within detailed firm-by-job-title cells, which we refer to as 'positions'. We first document considerable variation in implicit gendered framing within positions, including those where explicit gender requests remain unchanged. Much of the association between explicit requests and implicit wording reflects differences across positions: occupations and employers differ in the characteristics they seek and how they describe them. Therefore, we focus on positions whose explicit gender targeting changes during the sample period, using first-difference specifications across consecutive postings of the same position to eliminate persistent employer and occupational differences.

The first key finding is that explicit requests and implicit wording move together within positions. Relative to ads without explicit gender requests, those carrying an explicit male or female request have about 0.15 to 0.19 standard deviations higher implicit framing toward the same gender. Between consecutive postings, removing a request for men is accompanied by less male-coded wording, and removing a request for women by less female-

coded wording. Adding requests is accompanied by movements of similar magnitude in the opposite direction. Changes in individual skill and ascriptive requirements of the ads help to shed light on how the *Gender Index* changes: requirements associated with male- or female-targeted language become more prevalent when the corresponding request is added and generally less prevalent when it is removed. The net pattern is a coordinated adjustment across the two channels, rather than stronger implicit targeting following the withdrawal of an explicit request.

Our second key finding is that both forms of gender signalling described above are reflected in the posted wages, in opposite directions for men and women. Within the same position, dropping a request for women is followed by a 4.0% increase in the posted wage and adding one by a 3.7% decrease, while adding a request for men is followed by a 2.6% increase. A one-standard-deviation increase in male-coded wording between consecutive postings is associated with a 1.3% higher posted wage. Moreover, the female penalty persists even when we condition on detailed job content or between consecutive postings where no measured requirement changes. Gendered signalling, whether explicit or implicit, is therefore not costless: it moves with the wage the firm offers.

To rationalise our findings, we present a directed-search model, in which a firm jointly chooses a posted wage, an explicit gender request (or none), and implicit gendered wording to attract applicants to its vacancy. Workers then direct their search according to the wage and how they interpret the ad's overall gender signal. The model delivers two main predictions. First, explicit requests and implicit wording are complementary channels of gender signalling: when the two are aligned, the ad conveys a coherent recruitment message; when they conflict, the firm's intentions are harder to interpret, discouraging applications from both genders.[1] Firms then have an incentive to align the two channels rather than treat them as substitutes. Second, gender signalling and posted wages are jointly determined by differences in the wage responsiveness of male and female applicants. When male applications are more wage-elastic than female applications, a more female-oriented ad draws a relatively larger share of less wage-responsive applicants, and firms optimally post a lower wage.[2] The model also rationalises the sorting of positions into persistent gender-targeting regimes, consistent with our descriptive evidence.

Our study contributes to several strands of literature. We advance the literature on gendered recruitment by documenting the prevalence and variation of explicit requests and implicit gendered wording in job ads, how employers adjust these signals across repeated postings, and how those adjustments vary along with posted wages. Existing work

[1] This ambiguity channel is consistent with evidence from Kuhn, Shen and Zhang (2020) that job seekers interpret explicit gender requests in conjunction with other information about the employer's likely preference.

[2] The assumption that women's applications respond less to wages is in line with the literature showing that women's firm-level labour supply is less elastic (Hirsch, Schank and Schnabel, 2010; Ransom and Oaxaca, 2010).

establishes the prevalence and content of explicit requests (Kuhn and Shen, 2013), their effects on applications and callbacks (Kuhn, Shen and Zhang, 2020; Kuhn and Shen, 2023), and the relationship between vacancy gender preferences and workplace diversity (Card, Colella and Lalive, 2025). Research on implicit language also connects gendered wording to applicant behaviour and wages (Chaturvedi, Mahajan and Siddique, 2025; Castilla and Rho, 2023; Gaucher, Friesen and Kay, 2011). Our repeated-posting comparisons complement this evidence by distinguishing persistent differences across positions from changes in how the same employer advertises a recurring role. They reveal that the explicit request and the requirements text are coordinated components of a recruitment message, while the wage analysis shows that their associations with advertised pay survive detailed controls for measured requirements.

The paper also connects employer recruitment choices to the literature on gender differences in job search and wage setting. Directed-search studies analyse how advertised wages and vacancy characteristics shape application behaviour (Belot, Kircher and Muller, 2022; Marinescu and Wolthoff, 2020). Evidence on gender differences in applications and labour supply to individual firms provides a channel through which wage and recruitment choices may depend on the group an employer seeks to attract (Fluchtmann et al., 2024; Hirsch, Schank and Schnabel, 2010; Lochner and Merkl, 2026; Ransom and Oaxaca, 2010). Our model brings recruitment messages into this relationship: the firm chooses signals that influence the composition of the applicant pool alongside a wage intended to attract that pool. Our empirical analysis complements this framework by documenting how changes in gender signalling and posted wages are systematically related across repeated ads for the same position.

Additionally, we contribute to the rapidly growing use of online job ads to study labour markets. Prior work shows the value of job ads in revealing employers' skill demand (Deming and Kahn, 2018; Hershbein and Kahn, 2018; Marinescu and Wolthoff, 2020; Modestino, Shoag and Ballance, 2016) and studying employer and job-seeker search behaviour (Brenčič and Norris, 2010, 2012; Faberman and Kudlyak, 2019; Marinescu and Wolthoff, 2018, 2020). Recent work has used online vacancy data to study labour markets in developing countries, including Ukraine and several South Asian economies, where conventional labour force surveys offer limited occupational or real-time detail (e.g., Faryna et al., 2022; Nomura et al., 2017). Our setting adds to this literature by examining a large emerging economy where explicit gender requests are common, and where explicit requests and free-text requirements are observed for the same vacancy, allowing us to compare the two channels within firm-job title cells. Methodologically, whereas much of the existing literature relies on structured vacancy fields or predefined keyword dictionaries, we use Natural Language Processing to recover latent gendered framing from unstructured requirements text. This places our measurement within the broader use of text as data in economics (Gentzkow, Kelly and Taddy, 2019), and alongside

work that extracts latent attitudes rather than topics or sentiment from text (Ash, Chen and Ornaghi, 2024).

Beyond documenting these patterns, our findings speak directly to the design and limits of anti-discrimination policy in labour markets. A natural concern is that, where explicit profiling is restricted, firms simply reallocate targeting from overt labels to implicit wording. Within positions, we find no evidence of such reallocation: when firms withdraw an explicit request, the wording moves away from the previously requested gender. Our theoretical framework predicts the same qualitative response to such a binding restriction: holding a position's gender preference and workers' interpretation of ads fixed, a ban forces firms to post an ad without an explicit request. Because explicit requests and implicit wording are complementary components of a coherent message, firms respond by moving their wording toward the untargeted optimum rather than compensating for the ban by intensifying implicit targeting. This counterfactual, however, remains model-based, since we do not observe a clean policy change in our sample period and a ban itself could alter how workers interpret the absence of an explicit request. Nonetheless, both the evidence and the model suggest that restrictions on explicit profiling need not displace targeting into more gendered implicit language. Our method also offers a scalable tool for monitoring gendered coding in ads that interventions confined to explicit labels or keywords could miss.

The remainder of the paper proceeds as follows. Section 2 describes the data, text processing pipeline, and construction of the Gender Index. Section 3 documents descriptive patterns in explicit and implicit gendered preferences. Section 4 examines the comovement of the explicit and implicit targeting channels and how both are priced in posted wages. Section 5 provides a theoretical framework that can rationalise our empirical findings, and Section 6 concludes.

2. Data and data processing

2.1. Data description

We analyse data from one of the main job posting sites in Vietnam, covering the periods from mid-February 2019 to January 2020 and from March 2020 to mid-December 2020. The online job ads have well-defined fields that include job title, job category, location, firm size, job level, offered wages (if any), and some other specific requirements: age, gender, education, and experience. In addition, the ads include free-text sections for further details of the job descriptions and requirements.

We construct two samples from the raw job ads. First, the *text analysis* sample contains

all Vietnamese-language postings before any sample restrictions (see Section 2.2). Second, we derive our *estimation* sample as follows: We exclude postings made by the platform on behalf of client firms, as these lack reliable firm identifiers. We then keep only full-time, permanent positions to ensure posted wages are comparable across ads, and exclude internships and postings for jobs outside Vietnam. We remove exact duplicate listings (identical on all observable fields). Finally, we exclude postings with requirements written in foreign languages.[3] After cleaning, the estimation sample consists of 306,408 jobs advertised by 10,798 firms. Micro & small (less than 100 employees), medium (100-999 employees), and large (at least 1,000 employees) firms make up 23.4%, 50.9%, and 25.7% of the firms, respectively. Most job locations are in Hanoi or Ho Chi Minh City, which are the two most important cities in Vietnam, both politically and economically. Other provinces with large numbers of vacancies are port cities or areas with large industrial complexes, such as Da Nang, Hai Phong, Binh Duong, or Bac Ninh. A majority of vacancies fall into the following categories: Accounting/Audit (10%), Advertising/Promotion/PR (6%), Clerical/Administrative (7%), Consulting (16%), Customer service (7%), Engineering/Mechanical (6%), IT (5%), and Sales (6%).

As shown in Table 1, most ads (77.9%) are for experienced but non-managerial positions. Some vacancies are seeking candidates for managerial roles, yet the number is small (around 9%). This is not surprising, given that in Vietnam, managerial and senior positions are most likely to be filled through internal promotions rather than external hires. Consistent with typical online vacancies in other developing countries (e.g., see Helleseter, Kuhn and Shen, 2020), those in our dataset tend to target more educated job seekers, with nearly 90% of ads requiring a 3-year diploma or a bachelor's degree, i.e., college or university education. At the same time, most vacancies target early-career candidates, with 96% of ads requiring up to 5 years of experience.

Although the sample is not representative of the entire labour market, the job board's concentration at early career stages makes it particularly well-suited for studying the origins of gender inequality. It is at precisely this point in working life that gender gaps in wages, career progression, and occupational sorting tend to emerge and then compound over time: initial job matches shape subsequent trajectories through both the accumulation of human capital and the signalling of worker type to future employers (Bertrand, Goldin and Katz, 2010; Blau and Kahn, 2017; Goldin, 2014; Kleven, Landais and Søgaard, 2019).[4] The data, therefore, allow us to study how gendered preferences are expressed at the moment when many workers first sort into jobs and firms, and where policy interventions to promote equality may have the greatest long-run impact.

[3]The number of ads dropped after each step is 4,674; 10,083; 35; and 18,030, respectively. Note that our results are robust to the inclusion of some "excluded ads" such as those written in foreign languages.

[4]See also Manning and Robinson (2004) and Pham, Schaefer and Singleton (2026) on the potential importance of relatively small gender differences in hiring wages compounding to account for much larger aggregate wage gaps.

Table 1: Descriptive statistics (sample means & percentages)

	(1)	(2)	(3)	(4)
	All	Men-targeted	Women-targeted	Non-targeting
Wage specified	50.3	46.6	56.1	50.5
Mean wage (USD), when advertised	466.3	441.9	380.6	484.6
Job level				
New Entry	2.4	2.7	3.5	2.1
Experienced (Non-Manager)	77.9	71.8	89.6	77.8
Technical/Engineer	3.8	9.4	0.2	2.8
Team Leader/Supervisor	7.2	7.6	2.6	7.7
Manager or higher	8.8	8.4	4.1	9.5
Firm size				
< 100	23.4	20.3	24.8	24.0
100-499	38.5	38.0	39.0	38.6
500-999	12.4	14.3	14.7	11.6
1,000-4,999	16.5	16.3	11.2	17.3
5,000-9,999	7.1	9.1	9.8	6.3
≥ 10,000	2.0	2.0	0.4	2.2
Gender				
Request women	9.5	0.0	100.0	0.0
Request men	18.3	100.0	0.0	0.0
Age				
Maximum age specified	46.3	58.6	67.6	40.3
Mean max age, when specified	34.2	35.6	31.8	34.2
Minimum age specified	48.4	62.0	68.8	42.3
Mean min age, when specified	22.8	23.3	22.4	22.7
Education				
High school or less	10.1	14.9	7.8	9.1
Associate Degree/Diploma (3 years)	64.2	60.1	71.7	64.2
University	25.3	24.5	20.1	26.1
Others	0.5	0.5	0.4	0.5
Experience				
≤ 1 yr	35.9	34.8	40.8	35.5
1-2 yrs	37.2	35.0	39.4	37.5
2-5 yrs	23.1	24.8	18.3	23.2
≥ 5 yrs	3.8	5.4	1.5	3.7
N of ads	306,408	55,926	29,248	221,234

Notes: This table shows the proportion of ads (in %) in each category, except for the mean wage and mean age. Columns (1)-(4) display the statistics for all ads, explicit men-targeting ads, women-targeting ads, and ads without explicit gender labels, respectively.

Moreover, around 72% of ads do not explicitly require a certain gender. Among the explicitly gender discriminatory ads, the number of male-targeted jobs is about twice as high as the number of female-targeted ones (18.3% versus 9.5%). The overall 28% explicit targeting rate is at the high end compared to other similar datasets: approximately 10% on Chinese vacancy platforms serving high-skill workers (Kuhn and Shen, 2013); 8% in Indian data (Chaturvedi, Mahajan and Siddique, 2025); and 20–34% in Austria (Card, Colella and Lalive, 2025). Half of the job ads (50.3%) specify a wage. This coverage is also comparable,

if not favourable, to the observed wage-posting rates in other online job board datasets used in the literature. For example, Marinescu and Wolthoff (2020) and Azar, Marinescu and Steinbaum (2022) use job postings from CareerBuilder.com where only around 20% include wage information. On Zhaopin.com, the Chinese platform studied by Kuhn and Shen (2013), only 16% of ads post a wage, while the Xiamen job board (XMRC) used in Kuhn and Shen (2023) has a 58% wage-posting rate.

2.2. Text analysis

2.2.1 Title classification and skills extraction

We employ various natural language processing (NLP) techniques to extract relevant information from the job ad titles and text fields, using the text analysis sample that contains all Vietnamese-language postings.

Title classification. We use a 3-step procedure to map the raw job titles to a standardised occupation taxonomy. First, we use GPT-4 to classify the job titles into 10,152 "groups" of English job titles, while preserving both the profession and seniority level. Second, JobBERT, a Sentence Transformers model specifically trained for job title matching and similarity, and the Agglomerative Clustering algorithm are used to group similar English job titles into 5,144 broader title groups. Finally, using GPT-4, we map these broader groups to the standardised, narrowly defined occupations within the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy, resulting in 1,698 occupations. This granularity reflects occupational distinctions preserved by the classification pipeline; for instance, "cost analyst" is distinguished from "accountant", although the former is a sub-category of the latter, which matters for our later within-firm-title empirical identification strategy.[5] The distribution of ads across occupations is highly skewed: the median occupation accounts for just 9 ads, reflecting genuine occupational diversity across the sample, while a small number of common roles account for a disproportionate share of postings. The most frequently advertised titles, such as salesperson (15.6%), accountant (6.7%), financial advisor (4.3%), customer service clerk (2.9%), marketing specialist (2.7%), insurance consultant (2%), and software developer (1.6%), together account for over 36% of all ads. Even these common titles represent substantively distinct occupations with well-defined skill profiles, confirming that the taxonomy does not collapse heterogeneous roles into generic categories.

[5]Throughout the paper, we refer to these narrowly defined ESCO occupations as "titles". Our results are robust to an alternative mapping to the O*NET taxonomy.

Skills extraction. We use KeyBERT to extract the top 10 keywords from each ad's requirements text field.[6] The top keywords that appear in at least 200 ads are then manually screened to classify some additional job requirements. We identify keywords representing 42 distinct requirements related to soft skills and ascriptive characteristics (see Appendix Table D1). Among the various social skills, communication and negotiation are in the highest demand, with 48.4% and 14.6% of ads requesting them, respectively. Similarly, over 30% of job postings require basic computer skills, such as proficiency in Microsoft Office, while 24% require candidates with foreign language proficiency. Employers also state preferences for some ascriptive characteristics, including appearance (e.g., good-looking), height, and voice, which are unrelated to skills but might be job-specific (e.g., jobs that require customer interaction). More interestingly, preferences are also mentioned for certain personality/socially desirable traits, many of which are considered gendered words in the literature (e.g., Gaucher, Friesen and Kay, 2011).

2.2.2 Gender Index: Construction

To generate a text-based gender index, we focus on the text in a job ad's requirements section, as it directly reflects employers' preferences and expectations regarding candidate skills and attributes. In contrast, the job description primarily outlines the tasks, duties, and responsibilities of the role. While informative, this content can often be shaped by occupational norms and historical gender sorting. For instance, descriptions of caregiving tasks in nursing or physical demands in construction may reflect the occupation's gender composition rather than the employer's preferences. This distinction between the two job ad sections matters for identifying employer preferences. Experimental evidence shows that trait-based language in the requirements section is a key channel through which gendered signals in job ads shape applicant behaviour (Born and Taris, 2010). Furthermore, Gaucher, Friesen and Kay (2011) explicitly noted a highly relevant limitation in their study: gendered wording was diffused across all sections of an ad, calling for future work to distinguish between subsections.

To construct the gender index, we preprocess the requirements text to exclude sentences that refer to formal qualifications (e.g., field of study), experience thresholds, and explicit demographic criteria (gender and age), retaining only sentences that primarily describe skills, soft attributes, and personality traits.[7] Qualification and experience requirements often reflect occupational sorting mechanisms, such as credentialing norms and gendered educational pathways, rather than discretionary employer choices over how to frame a vacancy (Goldin, Katz and Kuziemko, 2006; Blau and Kahn, 2017; Castilla and Rho, 2023;

[6] See more about KeyBERT here: Grootendorst (2020).

[7] The filter targets keyword indicators of qualifications and experience. Because field-of-study references typically co-occur with degree-level language in the same sentence, most are excluded in practice, though standalone mentions without accompanying qualification language are retained.

Deming and Kahn, 2018). Removing them allows the resulting index to isolate deliberate gendered framing from occupational sorting.

We then employ Simple Contrastive Learning of Sentence Embeddings with Vietnamese (Gao, Yao and Chen, 2021), a state-of-the-art language model optimised for semantic similarity in Vietnamese. The model converts each requirements text section into a 768-dimensional vector of sentence embeddings that captures its semantic content. Texts with similar meanings are mapped to nearby points in the embedding space, enabling us to quantify subtle differences in how employers describe desired skills and personal characteristics.[8] Unlike dictionary-based approaches, sentence embeddings capture the full contextual meaning of the language of requirements as probably perceived by job searchers reading ads, enabling the index to detect gendered framing in phrases that do not appear in standard gender-coded lexicons, as we demonstrate in Section 2.2.3.

Adapting the approach in Ash, Chen and Ornaghi (2024), who recover latent semantic dimensions by taking vector differences in embedding space, we construct a gender direction using the contrast between the requirements texts of job ads that explicitly target men and those that explicitly target women:

$$\overrightarrow{gender} = \frac{\sum_n \overrightarrow{men\text{-}targeted_n}}{N_{men-targeted}} - \frac{\sum_n \overrightarrow{women\text{-}targeted_n}}{N_{women-targeted}}$$

While Ash, Chen and Ornaghi (2024) define gender slant by comparing two conceptual directions (male-female vs. career-family) within judge-specific embedding spaces from their own legal opinion texts, our approach uses observed vacancy writing behaviour to identify the semantic contrast that employers reveal when targeting different genders.[9] This construction removes wordings that are "equivalently" distributed across male- and female-targeting ads, and isolates the linguistic features that systematically differ by targeted gender. Thus, $\overrightarrow{gender}$ captures the stable linguistic patterns employers use to signal gendered expectations in the requirements text.

We then use $\overrightarrow{gender}$ to compute a gender index for each ad by projecting its embedding vector onto this gender dimension:

$$Gender\ Index = sim(\vec{x}, \overrightarrow{gender}) = \frac{\vec{x} \cdot \overrightarrow{gender}}{\|\vec{x}\| \cdot \|\overrightarrow{gender}\|}, \qquad \in (-1, 1)\,. \tag{1}$$

Economically, this index provides a continuous vacancy-level measure of latent or *implicit* gender targeting. Positive values indicate alignment with the semantic patterns character-

[8] As a robustness check, we also employ multilingual embedding models such as OpenAI's text-embedding-3-large and the M3-Embedding model. Our findings are robust across these alternatives.

[9] See Appendix A for a more detailed comparison of our approach with Ash, Chen and Ornaghi (2024).

istic of men-targeting ads, while negative values indicate alignment with women-targeting patterns. The approach relates to, but advances upon, two strategies used in the existing literature. The first relies on structured gender preference fields recorded directly by employers on job boards (e.g., Card, Colella and Lalive, 2025; Kuhn and Shen, 2013, 2023). While unambiguous, such fields capture only *explicit* preferences and leave the majority of ads that carry no explicit label uncharacterised, such as the 72% in our data. The second strategy uses text analysis to recover implicit gender preferences from ad texts. Some implementations use predefined word lists to count the frequency of masculine or feminine terms (Gaucher, Friesen and Kay, 2011); others use machine-learning classifiers, trained on ads with explicit gender requests (Chaturvedi, Mahajan and Siddique, 2025) or Google News US (Castilla and Rho, 2023), to derive gendered word lists and assign predicted gender-based probabilities to each ad.

Our approach improves upon these two existing strategies in three important ways. First, by focusing on the requirements text rather than the full ad, we analyse the job ad section that most directly reflects employer preferences rather than historical gender sorting, as we argued earlier in this section. Second, by using sentence embeddings trained on Vietnamese text, we capture the full contextual meaning of requirements language holistically, rather than assigning importance scores to individual tokens, which is likely to matter most for job applicants. Third, the resulting gender index is a quantitative measure not only of the direction but also of the intensity of gendered framing at the vacancy level, enabling the within-firm-title identification strategy that is central to our empirical analysis. Taken together, these three improvements provide arguably the most complete large-scale picture of ad-level implicit gender targeting in the literature.

2.2.3 Gender Index: Validity

We evaluate the *Gender Index* along three dimensions: (i) construct validity (does the index load on attributes that are known to be gendered?); (ii) cross-validation (does it recover the same occupational ordering as the explicit gender signal?); (iii) and external validity (does it predict workforce gender composition in an independent data source?).

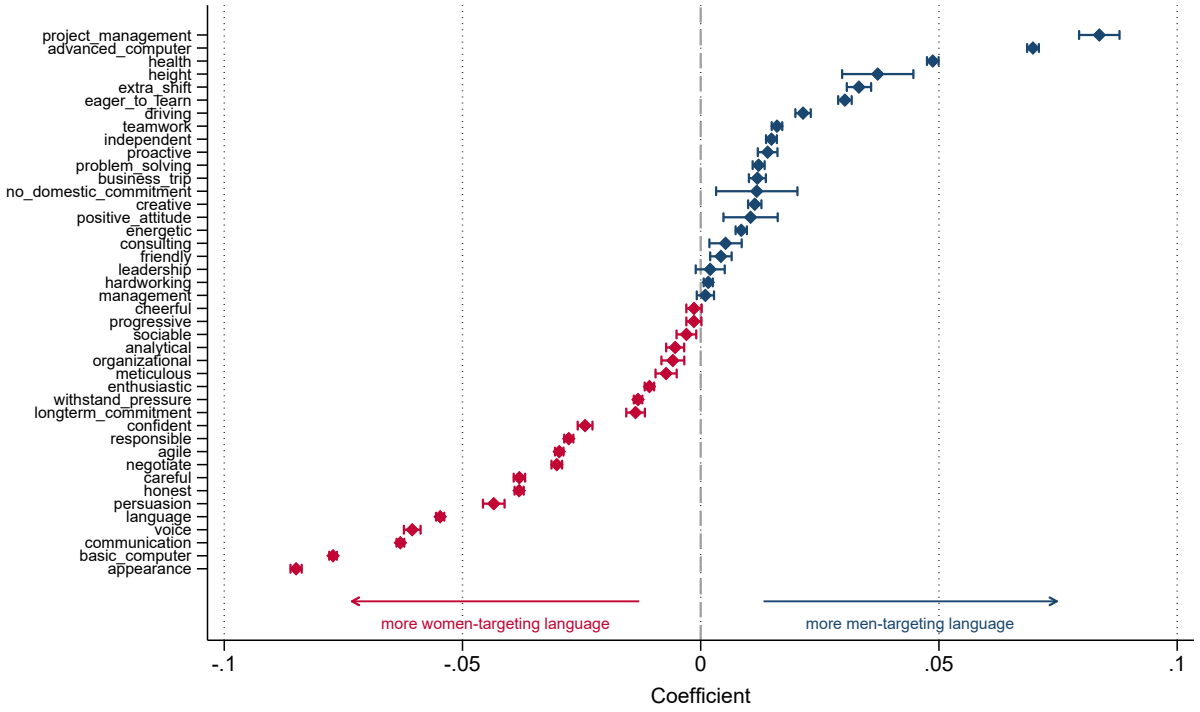
Construct validity. We first verify that the embeddings separate male- and female-targeted language in the semantic space. Reducing the 768-dimensional embedding vectors to two dimensions and plotting them in Appendix Figure D1 reveals that men-targeting and women-targeting ads form distinct clusters, confirming that employers use substantively different wording when targeting different genders.

To illustrate the practical content of this variation in the ads sample, Appendix Table D2 reports representative requirement texts at the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles of

the *Gender Index* distribution. At the most male-coded end, the example ad requirements emphasise specific technical tools and domain knowledge; at the most female-coded end, they foreground interpersonal dispositions and generic competencies. Moreover, the change in language is "smooth" across percentiles, indicating that the index indeed captures continuous variation in gendered framing rather than a binary distinction.

We then regress the *Gender Index* in each ad on 42 dummy variables indicating the presence of each of the skills and ascriptive characteristics that were extracted as described in Section 2.2.1. Figure 1 reports the estimated coefficients. The index loads positively on attributes such as *proactive* and *independent*, and negatively on attributes such as *responsible* and *honest*, which is broadly consistent with the gendered language dimension documented by Gaucher, Friesen and Kay (2011). However, some loadings depart from that English-language classification. For instance, *analytical* and *meticulous* load negatively in the Vietnamese context, suggesting these traits are more strongly associated with female-targeted positions in this labour market. Overall, this regression confirms that the index recovers some known patterns of gendered trait attribution, while also capturing context-specific variation that English-language lexicons would likely misclassify.

Figure 1: The correlations of required skills and ascriptive characteristics with an ad's *Gender Index*



Notes: This figure shows the estimated coefficients and 95% confidence intervals obtained from regressing the *Gender Index* (Equation 1) on 42 dummy variables indicating the presence of required skills and ascriptive characteristics (see Table D3) in an advert ($N = 306,408$).

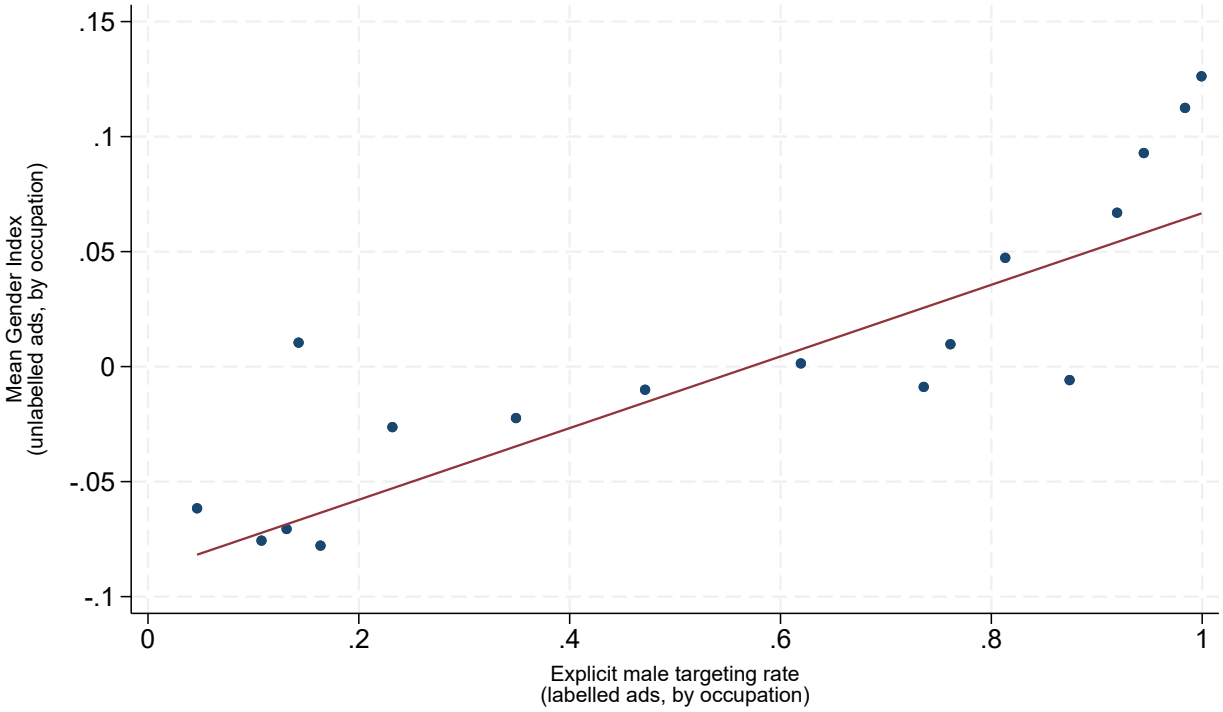
There are two main reasons why we do not simply rely on existing lists of gendered words.

First, the index captures broader semantic patterns beyond canonical gendered terms. For example, it correlates positively with phrases such as "willing to go on business trips" or "no domestic responsibilities" but negatively with references to the ability to "withstand pressure" or "long-term commitment", reflecting deeper stereotypes about gender roles. Studies that rely on predefined word lists (e.g., Gaucher, Friesen and Kay, 2011) or token-level importance scores (e.g., Chaturvedi, Mahajan and Siddique, 2025) would not detect these signals, since such phrases do not appear in standard gender-coded lexicons yet clearly reflect gendered expectations about who performs certain roles. Second, ad requirements text typically represents a bundle of skills and structural job features, whose gender implications are not reducible to simple word counts (Castilla and Rho, 2023). Embedding models help address this by capturing the latent semantic structure of the requirements text. Rather than relying on predefined dictionaries, embeddings learn patterns of word co-occurrence and the contexts in which traits are described, allowing them to detect gendered framing even when employers avoid explicit gendered terms. Phrases such as "willing to travel", "strong analytical mindset", or "meticulous and patient" might carry gendered connotations in practice, despite not appearing in standard gender-coded lexicons. Embeddings place such phrases in a high-dimensional semantic space where their proximity to male- or female-targeted language emerges naturally from usage patterns. This enables our *Gender Index* to recover subtle, context-dependent signals that keyword-based approaches could systematically miss.

Cross-validation. Having established that the index loads as expected on some known gendered attributes, we next assess whether it generalises beyond the individual ads from which it was constructed. First, we test whether the index recovers the same occupational ordering as the explicit gender signal. For each occupation, we compute the mean *Gender Index* from ads *without* explicit gender labels, and correlate this with the occupation-level explicit male targeting rate computed from ads *with* explicit labels. If the index assigns high male-coded scores to occupations that also exhibit high rates of explicit male targeting, then it recovers the same occupational ordering as the overt signal from text alone. Figure 2 presents the binned scatter plot across 433 occupations with at least ten observations in each subset. The Pearson correlation is 0.65, and the Spearman rank correlation is 0.71, and both strengthen as the minimum cell size threshold (minimum number of sample ads) of an occupation increases, indicating that the result is not driven by noisy estimates from sparsely observed occupations.

External validity. Finally, we examine whether the *Gender Index* reflects patterns of occupational gender segregation observed in the broader Vietnamese labour market. We match the standardised occupation titles in our sample to ISCO-08 codes using the CASCOT automated coding tool (Blackwell, Jarman and Siltanen, 2009), retaining only

Figure 2: Implicit *Gender Index* and explicit male targeting rate by occupation



Notes: Each point represents one bin of ESCO occupations (20 equal-frequency bins). The x-axis shows the occupation-level explicit male targeting rate, computed from ads with an explicit gender label. The y-axis shows the occupation-level mean *Gender Index*, computed from ads with no explicit gender label. The two subsets are non-overlapping. Sample: 423 occupations with at least 10 labelled and at least 10 unlabelled ads. Pearson $r = 0.64$, Spearman $\rho = 0.71$ ($p < 0.001$).

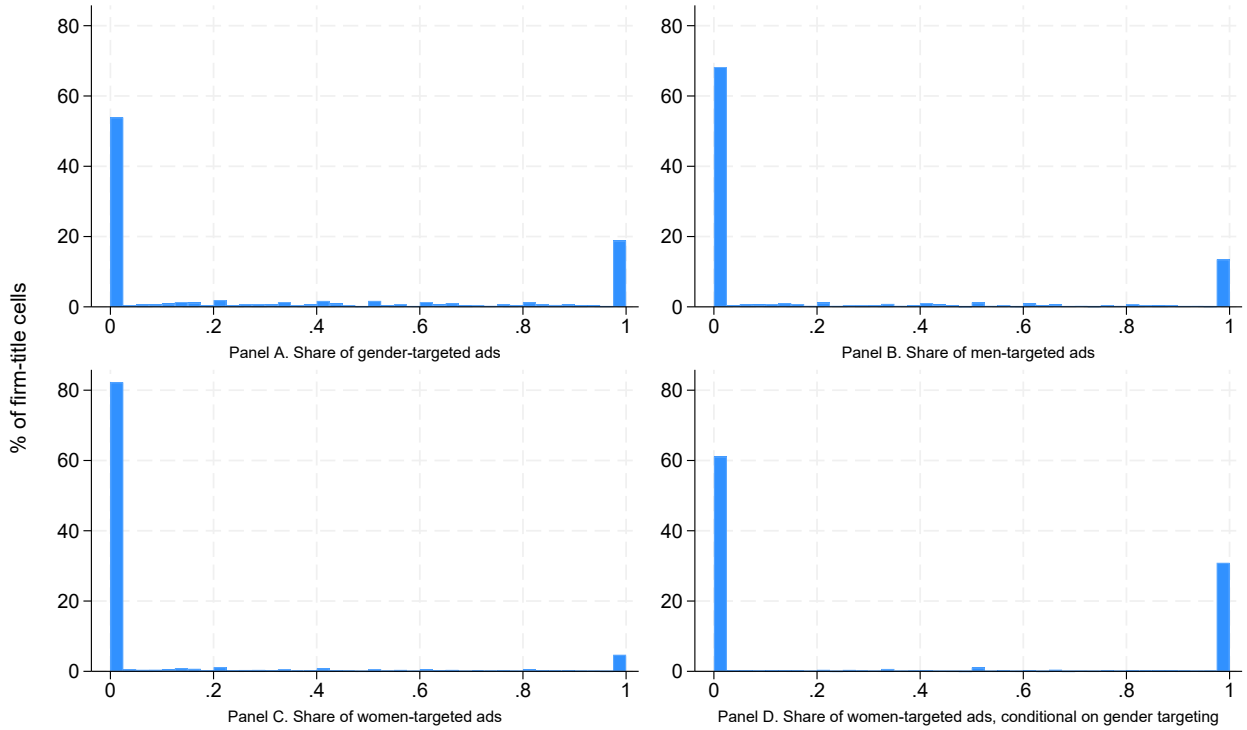
matches with a confidence score of at least 60. This threshold covers 52.5% of ads, spanning 106 distinct 3-digit ISCO-08 occupations.[10] For each 3-digit ISCO-08 occupation with at least 5 ads, we compute the mean *Gender Index* and correlate it with the share of male workers in the same occupation drawn from the Vietnam Population and Housing Census 2019 (Ruggles et al., 2025). We find Pearson and Spearman correlation coefficients both of 0.57. Occupations with high male employment shares in the census (e.g., construction trades, transport operators, and engineering technicians) receive strongly positive *Gender Index* scores in the job ads, while occupations with high female employment shares (e.g., childcare workers, clerks, primary school teachers) receive strongly negative scores. This cross-source correspondence provides external validation that the *Gender Index* captures systematic variation in the gendered content of occupational requirements, rather than idiosyncratic noise in how individual ads are worded. We note that this exercise establishes convergent validity between the text-based measure and an independent labour market outcome. However, it does not, by itself, distinguish whether the underlying variation reflects employer preferences, gendered task content, or occupational stereotypes.

[10]Lower-scoring matches are predominantly hybrid or non-standard titles for which CASCOT identifies no dominant ISCO-08 category. Restricting to high-confidence matches reduces noise in the occupational mapping at the cost of sample coverage.

3. Gender preferences in Vietnamese job ads: Stylised facts

We begin by documenting patterns in gender targeting. Figure 3 plots the distribution of the share of explicitly gender-targeted ads across unweighted firm-title cells (*positions*),[11] restricting to those observed in our sample period with at least five ads (12,811 distinct positions; 209,873 ads). Panel A shows the overall share of explicitly gender-targeted ads; Panels B and C decompose this into men- and women-targeted shares, respectively; Panel D reports the share of women-targeted ads among all explicitly targeted ads in the cell, restricted to positions containing at least one explicitly targeted ad. The distribution in Panel A is sharply skewed: 53.7% of the positions never contain explicit gender targets, while a pronounced right tail indicates that a non-trivial number of positions post a large fraction or all of their ads with explicit gender requirements. At the same time, 18.8% of positions target gender in every ad they post, suggesting that for some firm-title combinations, gender requests are a persistent rather than occasional feature of recruitment.

Figure 3: Distribution over firm-title cells (positions) in the share of their ads that show explicit gender targeting



Notes: Each observation is a firm-title cell (position). The sample is restricted to positions with at least five observed ads (12,811 distinct positions; 209,873 ads). Panel D is restricted to positions with at least one explicitly gender-targeted ad (5,884 positions). The y-axis shows the (unweighted) share of positions in each bin; the x-axis shows the share of ads within the position satisfying the criterion stated in the panel label.

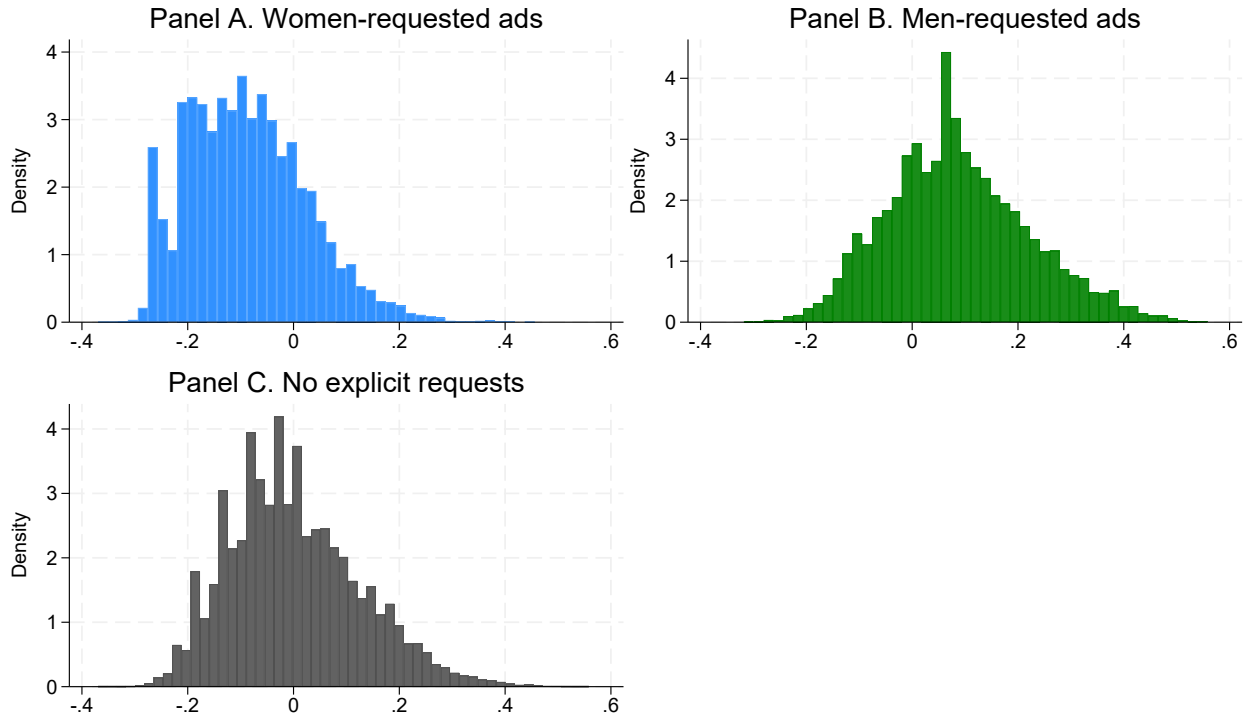
Panels B and C reveal that men- and women-targeting are not mirror images of each other.

[11] Henceforth, we refer to each firm-title cell as a *position*, defined as a recurring role advertised by a given firm under the same job title.

The distribution of men-targeted shares is shifted to the right relative to that of women-targeted shares, reflecting the aggregate pattern that male-targeted ads outnumber female-targeted ads by roughly two to one (18.3% versus 9.6% of all ads). Panel D shows that even among positions with some explicit targeting, the conditional share of women-targeted ads is concentrated at zero and one, suggesting that most positions only target one gender if they target at all, rather than alternating between the two.

Figure 4 plots the distribution over ads of the *Gender Index*, stratified by explicit targeting status. Panels A and B show women- and men-targeting ads, while Panel C shows ads with no explicit targeting. The distributions for gender-requested ads are clearly shifted relative to those without explicit targeting, indicating that explicit requests coincide with distinct linguistic framings in the requirements text. The fact that the "neutral" group clusters tightly around zero is especially informative; the index does not mechanically reproduce labels from the "reference" explicitly-targeted sample, but instead assigns near-zero scores to ads with language that does not systematically lean toward either gender.

Figure 4: Distribution of the *Gender Index* for ads, by type of explicit gender request



Notes: Distributions are histograms of the *Gender Index* by explicit targeting status. Panel A: women-requested ads (29,248 ads); Panel B: men-requested ads (55,926 ads); Panel C: no explicit gender request (221,234 ads). The median values of the *Gender Index* in Panels A–C are −0.0097, +0.072, and −0.014, respectively.

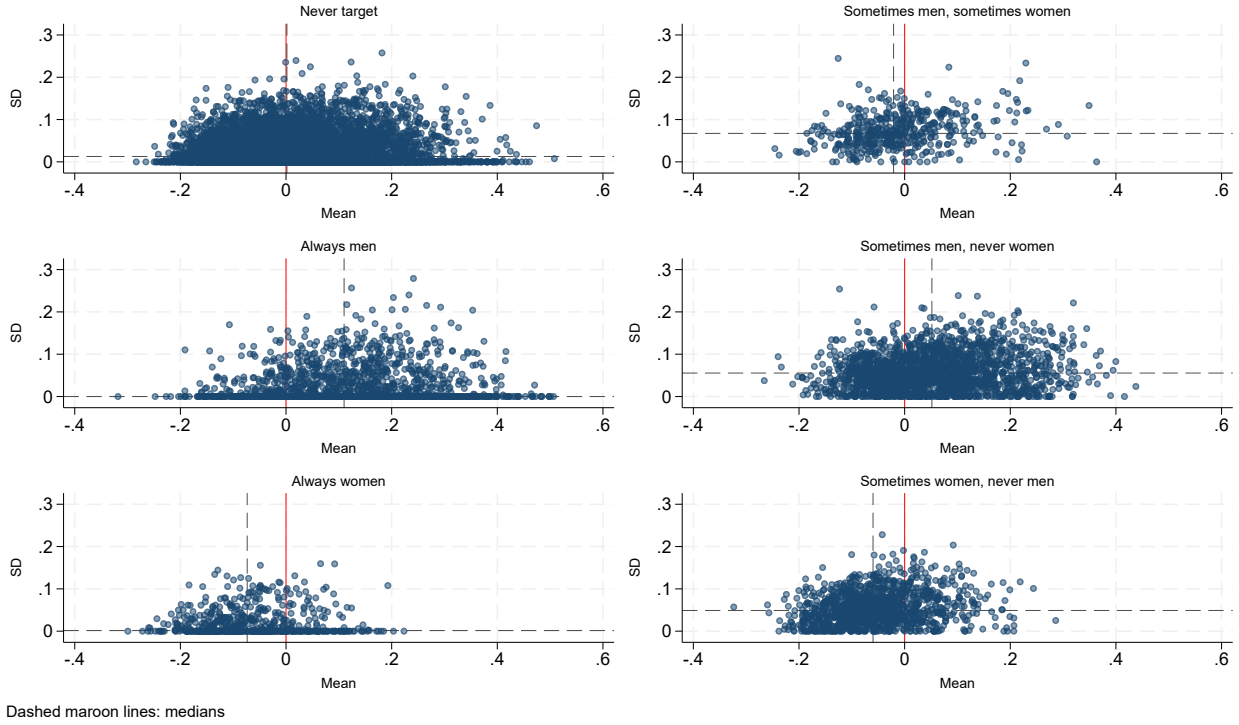
Two features of the targeting-ad distributions are particularly worth noting. The means are nearly symmetric around zero (−0.087 for women-requested and +0.085 for men-requested), confirming that the index is well centred. Men-requesting ads also exhibit a wider spread of index values (SD = 0.134) than women-requesting ads (SD = 0.111),

suggesting that employers draw on a more heterogeneous set of male-coded signals when explicitly targeting men. Beyond the targeted groups of ads, Figure 4 helps establish that the *Gender Index* captures implicit gendered framing even when explicit requests are absent, as some "neutral" ads without explicit gender requests still exhibit meaningful deviations in language in either direction. This implies that employers sometimes embed gendered expectations in the framing of skills and attributes rather than in direct requests.

To further probe whether gendered framing reflects systematic preferences rather than idiosyncratic variation, we again collapse the data to the level of positions with at least five ads. We then further restrict the sample to titles that are mapped to ESCO occupations, as described in the previous section. For each position, we compute the mean and standard deviation of the *Gender Index*. Figure 5 plots these statistics separately for six targeting regimes, defined by each position's observed pattern of explicit gender label use in the sample period: never targeting, always targeting men, always targeting women, sometimes targeting men but never women, sometimes targeting women but never men, and sometimes targeting both. Three key patterns emerge from the plot. First, the regimes are clearly separated in the mean-SD space. Positions that always target men cluster to the right of zero, positions that always target women cluster to the left, and positions that never target are centred near zero. This confirms that explicit targeting is systematically associated with the direction of implicit gendered framing; even after averaging across all ads within a position, the *Gender Index* recovers the expected ordering across regimes.

Second, the sometimes-targeting regimes occupy intermediate positions in the mean dimension and exhibit substantially greater dispersion than the always-targeting regimes. Third, if explicit gender requests were random or uninformative, aggregation to the position level should wash out any signal, producing overlapping clouds across regimes in Figure 5. Instead, the position-level averages show systematic separation, reinforcing the interpretation that the gendered language in job ads reflects structured employer preferences rather than idiosyncratic word choice.

Figure 5: The averages and standard deviations of the *Gender Index* within firm-title cells (positions), by targeting regime



Notes: The x-axis shows the mean *Gender Index* within the position; the y-axis shows its standard deviation. The solid vertical line is drawn at zero. The dashed vertical (horizontal) line marks the median of the position-level means (standard deviations). The six panels correspond to mutually exclusive targeting regimes: never target (6,836 positions; 82,714 ads), sometimes men and sometimes women (436; 28,972), always men (1,707; 15,291), sometimes men and never women (1,815; 45,674), always women (619; 5,310), and sometimes women and never men (1,192; 29,315).

4. Comovement and wage effects of gender signals

The stylised facts in Section 3 raise two economically important questions. First, when firms drop an explicit gender preference, does it resurface in the language of the ad, or does it diminish or disappear there as well? Second, do these gender signals, explicit and implicit, carry economic consequences? This section answers both questions.

Estimation sample and identifying variation. All our estimates in this section use firm-title cells (positions) that vary their explicit gender targeting within the sample period, so that the explicit labels have within-cell variation. In other words, our estimation sample includes three targeting regimes: (1) Sometimes men, never women; (2) Sometimes women, never men; and (3) Sometimes men, sometimes women. All specifications include firm-by-title and time fixed effects, so identification comes from the same employer posting the same position at different times with different explicit targeting choices.

4.1. Are explicit and implicit channels substitutes?

As shown in Figure 4, ads carrying an explicit gender request are written in more gendered language than ads that carry none. However, that association is uninformative about firm behaviour because it confounds the choices firms make with the composition of the vacancies they post. For instance, security and construction vacancies are described in male-coded terms and are also the vacancies for which men are more likely to be requested. Thus, we ask a narrower question: holding the position fixed, does a firm write its job ad differently when it attaches an explicit gender request than when it does not?

4.1.1 Ad-level evidence

Pooling across all three sometimes-targeting regimes, we first estimate:

$$Gender\ Index_{ijkt} = \alpha + \gamma_1 \cdot men_{ijkt} + \gamma_2 \cdot women_{ijkt} + \mathbf{x}'_{ijkt}\boldsymbol{\beta} + \kappa n_{ijkt} + \delta_t + \mu_{jk} + \varepsilon_{ijkt}\,, \quad (2)$$

where $men_{ijkt} = 1$ ($women_{ijkt} = 1$) indicates the presence of an explicit "men" ("women") request in advert i, posted by firm j, for title k, in month t, and the indicator is zero otherwise. The vector $\mathbf{x}_{ijkt}$ contains the human capital requirements (education, experience and job level, as in Table C1), and n_{ijkt} is the number of skill and ascriptive requirements, which separates the direction of the wording from its volume. We do not control for the individual requirements, since they are part of the text from which the index is constructed. With firm–title (μ_{jk}) and month (δ_t) fixed effects, γ_1 and γ_2 measure whether an ad carrying an explicit request is framed differently from other ads for the same position, net of common monthly shifts.

As shown in Column (1) of Table 2, ads carrying an explicit men request score 0.019 higher on *Gender Index* than non-targeting ads for the same position, and ads carrying a "women" request score 0.025 lower. These correspond to shifts of 0.15 and 0.19 *Gender Index* standard deviations (SD = 0.13). The two coefficients carry the signs predicted if firms perceive explicit requests and gendered language as complements: within a position, an explicit request is accompanied by *more* implicit framing in the same direction. As a robustness check, Column (2) drops the positions that request men in some postings and women in others; the estimates are essentially unchanged, which indicates that the pattern does not arise from the minority of positions whose targeting direction itself varies across ads.

The within-position association is nonetheless much weaker than the unconditional one. Estimated on the same sample without fixed effects, men-requested ads score 0.082 above non-targeted ads and women-requested ads 0.066 below. Most of that gap reflects the composition of vacancies rather than firm behaviour: explicit requests and human capital

Table 2: Comovement between explicit and implicit gender targeting

	(1) Ad-level Three regimes	(2) Ad-level Excl. mixed	(3) Event-level Three regimes	(4) Event-level Excl. mixed
Dependent var:	*Gender Index*	*Gender Index*	Δ*Gender Index*	Δ*Gender Index*
men: γ_1	0.019***	0.017***		
	(0.003)	(0.004)		
women: γ_2	−0.025***	−0.028**		
	(0.008)	(0.012)		
$drop^M$: θ_1			−0.021***	−0.018***
			(0.003)	(0.004)
add^M: θ_2			0.019***	0.022***
			(0.003)	(0.003)
$drop^F$: θ_3			0.018***	0.019***
			(0.004)	(0.007)
add^F: θ_4			−0.015***	−0.022***
			(0.004)	(0.005)
N of observations	103,961	74,989	50,907	38,498
N of positions	3,443	3,007	3,423	2,988
R^2	0.761	0.755	0.123	0.144
Human capital	Yes	Yes	Yes	Yes
Number of requirements	Yes	Yes	Yes	Yes
Firm × title FEs	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	No	No
Posting-date FEs	No	No	Yes	Yes

Notes: This table presents estimates from Equation (2) (Columns 1–2) and Equation (3) (Columns 3–4). *Gender Index* is the implicit index. Columns (1) and (3) report the results for the full sometimes-targeting sample; Columns (2) and (4) exclude the mixed regime (cells that sometimes target men and sometimes target women) for comparison. In Columns (1)–(2), *men* and *women* indicate the presence of explicit gender labels. In Columns (3)–(4), *drop* and *add* indicate that an explicit label present in the previous posting is removed or added in the current one. Human capital controls are indicators for categories of education, experience, and job level in Columns (1)–(2), and the changes in the share of ads in each category in Columns (3)–(4); the number of skill and ascriptive requirements enters in levels and in changes, respectively. Standard errors are clustered at the firm level in Columns (1)–(2) and at the firm–title cell level in Columns (3)–(4). Columns (3)–(4) are weighted by the number of ads in each cell-date observation. ***p<0.01, **p<0.05, *p<0.1.

controls alone explain 30 percent of the variance of *Gender Index*, while adding firm-title and month fixed effects raises this to 76 percent. Gendered language is thus primarily a property of the job, and the coefficients in Column (1) isolate the component that firms alter posting-by-posting.

4.1.2 Event-level evidence

Equation (2) compares an ad carrying an explicit gender request against other ads for the same position, however far apart they were posted and whatever else the firm might have changed in between. In this section, we narrow the comparison to consecutive postings and ask how the wording of a position changes between one posting and the next when a request is added or removed. More specifically, we collapse the data to the firm-title-date

level (50,907 ads; 3,423 distinct positions) and construct first differences of the implicit index within each cell c (where c indexes the firm-title pair).[12] Letting z denote the chronological order of postings within cell c, and $\tau = T(c,z)$ the calendar date of the z-th posting, we estimate:

$$\begin{aligned}\Delta Gender\ Index_{c,z} =& \alpha + \theta_1 \cdot drop^M_{c,z} + \theta_2 \cdot add^M_{c,z} + \theta_3 \cdot drop^F_{c,z} + \theta_4 \cdot add^F_{c,z} \\ &+ \Delta \mathbf{x}'_{c,z}\boldsymbol{\beta} + \kappa \Delta n_{c,z} + \delta_\tau + \mu_c + \varepsilon_{c,z}\,,\end{aligned} \tag{3}$$

where $\Delta Gender\ Index_{c,z} = Gender\ Index_{c,z} - Gender\ Index_{c,z-1}$ is the change in the index between consecutive postings within the cell. $drop^M_{c,z} = 1$ if an explicit "men" request present in posting $z-1$ is absent in posting z, and $add^M_{c,z} = 1$ if a "men" request absent in posting $z-1$ is present in posting z. $drop^F_{c,z}$ and $add^F_{c,z}$ are defined analogously for "women" requests. $\Delta \mathbf{x}_{c,z}$ denotes the changes between consecutive postings in the share of ads in each category of required education, experience and seniority, and $\Delta n_{c,z}$ the change in the number of skill and ascriptive requirements. δ_τ and μ_c are posting-date and firm–title fixed effects. Because the specification is in first differences, μ_c absorbs position-specific trends in the index rather than its level. Posting-date fixed effects are well identified: the median date contains 327 firm–title cells, and the median cell contributes 8 consecutive-posting differences. We weight by the number of ads per cell-date, since cell-date means based on more ads are measured more precisely.

Column (3) of Table 2 reports the estimates. When a firm removes an explicit "men" request, the requirements text of that position becomes 0.021 *less* male-coded; when it adds the request, the text becomes 0.019 *more* male-coded. The female margin mirrors this, with coefficients of 0.018 and −0.015. Excluding the mixed regime in Column (4) leaves the pattern intact and, if anything, slightly strengthens the female margin.

Two features of these estimates deserve emphasis. First, adding and dropping a label produce movements of comparable magnitude in opposite directions. Specifically, we cannot reject that the sum of two coefficients is equal to zero on either margin. This reversibility is difficult to reconcile with secular drift in how positions are described, or with a firm rewriting an ad once and leaving the new text in place. Instead, requirements wording appears to be a choice variable that firms adjust from one posting to the next and then undo.

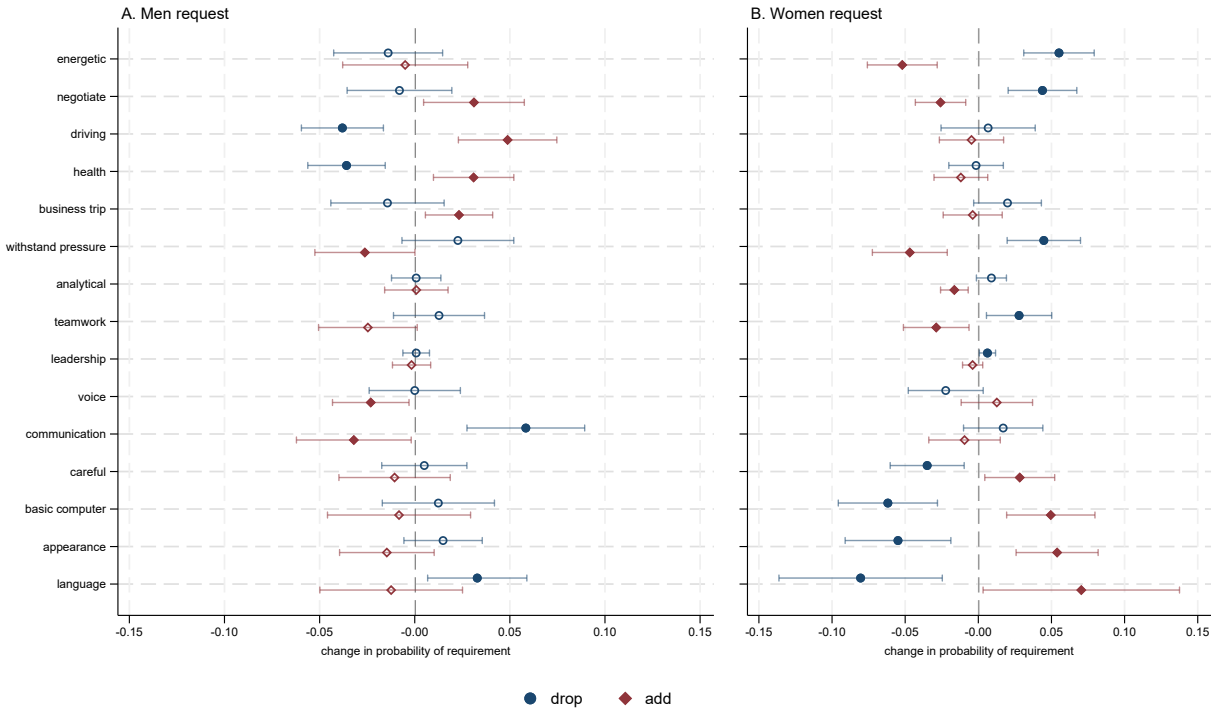
Second, the pattern is not driven by anticipation. Adding leads of the four *add/drop* event dummies to Equation (3), so that label changes occurring at posting $z+1$ enter alongside those at z, leaves the contemporaneous coefficients unchanged and the leads

[12] In the rare instances where multiple ads are posted for the same position on the same day, the index and the human capital shares are cell-date means, and an explicit request is coded as present if any advert carries it. This rule rarely binds: fewer than two per cent of cell-dates mix targeting statuses within a day. Restricting each first difference to postings whose endpoints are both internally homogeneous leaves the estimates unchanged.

jointly insignificant. In other words, label changes that have not yet occurred do not predict wording changes now, which rules out positions drifting toward a change in wording in advance of a change in the explicit request.

To shed light on *how* the *Gender Index* changes when an explicit request is removed or added, we repeat this exercise using the change in each of 42 required skills and ascriptive characteristics as the outcome, and report those with at least one significant coefficient in Figure 6. The requirements that respond are largely those the index treats as most gendered (Figure 1), and they move in the direction of the request. Particularly, adding a "men" request makes male-loading requirements, such as 'driving' and 'health' more likely, and adding a "women" request makes female-loading requirements, such as 'appearance', 'basic computer skills', and 'language', more likely. We also observe reversibility: removing a request largely reverses these language changes.

Figure 6: Specific requirements added and removed along with changes in explicit gender requests



Notes: This figure shows the estimated coefficients obtained from 42 equivalent regressions of Equation (3), with the outcome variables being the change across position ads in each of 42 dummy variables indicating the presence of specific required skills and ascriptive characteristics (see Table D3). Circles denote the coefficients on the *drop* of the gender request and diamonds the *add* of a request; horizontal bars show 95% confidence intervals, and filled markers are significant at the 5% level. Only the 15 requirements with at least one significant *add*/*drop* coefficient are shown.

4.2. Are gendered signals priced?

We now turn to the pricing of the two channels. In a standard directed search framework, restricting the applicant pool through explicit requests thins the expected queue and therefore raises the posted wage, so both labels should carry premia (see Wright et al., 2021, for a recent survey). We assess this prediction in two steps: wage levels and wage changes around label-switching events.

The wage response results in what follows are estimated from the subsample (62,483 ads; 2,240 distinct positions; roughly half of our estimation sample) where we observe a posted wage. The central concern then is not that posting is selective in general, but that it is selective with respect to gender preferences. However, the analysis presented in Appendix C.2 shows no evidence of this: conditional on firm-title and time fixed effects, the implicit index does not predict whether a wage is posted, and the explicit requests do so only marginally and in opposite directions. Thus, we treat the posted-wage sample as informative about pricing on the targeting margins.

4.2.1 Wage levels

We first use within-firm-title variation to estimate variants of:

$$\begin{aligned}\ln(w_{ijkdt}) =& \alpha + \beta_M \cdot men_{ijkdt} + \beta_F \cdot women_{ijkdt} + \gamma \cdot Gender\ Index_{ijkdt} \\ &+ \mathbf{x}'_{ijkdt}\boldsymbol{\theta} + \mathbf{r}'_{ijkdt}\boldsymbol{\phi} + \lambda \ln(\mathrm{HHI}_{-j,kdt}) + \delta_t + \mu_{jk} + \varepsilon_{ijkdt}\ ,\end{aligned} \tag{4}$$

where i indexes the advert, j the firm, k the title, d the district, and t the month; men_{ijkdt} and $women_{ijkdt}$ indicate the explicit gender requests; and $Gender\ Index_{ijkdt}$ is the implicit index. As in Equation (2), $\mathbf{x}_{ijkdt}$ contains the human capital requirements (education, experience, and job level). The vector $\mathbf{r}_{ijkdt}$ contains the skill and ascriptive requirements, which we add progressively across specifications: indicators for work schedule disamenities and for physical and aesthetic requirements, then counts of hard and soft skills and the total number of requirements. We also control for labour market concentration, $\ln(\mathrm{HHI}_{-j,kdt})$, so that the label and index coefficients are not confounded by the component of wages attributable to employer market power. This measure is the standardised log leave-one-out Herfindahl-Hirschman Index of vacancy postings at the title–district–month level, constructed in the spirit of Azar, Marinescu and Steinbaum (2022).[13] All specifications include firm–title fixed effects μ_{jk} and month fixed effects δ_t.

The estimates reported in Table 3 suggest that both channels are priced into posted

[13] The coefficient on $\ln(\mathrm{HHI}_{-j,kdt})$ is negative and significant without firm–title fixed effects, consistent with monopsonistic markdowns in more concentrated labour markets (Azar, Marinescu and Steinbaum, 2022; Manning, 2003). It loses significance once firm–title fixed effects are included, since concentration varies little within a position over time.

Table 3: Posted wages and gendered signals: levels

	(1)	(2)	(3)
Dependent var:		$\ln(w)$	
Gender Index	0.297***	0.217***	0.274***
	(0.058)	(0.050)	(0.059)
men	0.032**	0.029**	0.030**
	(0.014)	(0.014)	(0.014)
women	-0.050***	-0.050***	-0.051***
	(0.010)	(0.011)	(0.010)
N of observations	62,483	62,483	62,483
N of positions	2,240	2,240	2,240
R^2	0.794	0.796	0.796
Human capital	Yes	Yes	Yes
Disamenities, physical, aesthetic	No	Yes	Yes
Skill counts and number of requirements	No	No	Yes
Concentration	Yes	Yes	Yes
Firm × title FEs	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes

Notes: This table presents estimates obtained from Equation (4), pooled across all sometimes-targeting regimes. The dependent variable is the log posted wage, and the sample is restricted to ads that disclose a wage. *Gender Index* is the implicit index. *men* is a dummy variable indicating the presence of an explicit men request. *women* is a dummy variable indicating the presence of an explicit women request. Human capital controls are indicators for categories of education, experience and job level. Disamenities, physical, and aesthetic controls are indicators for work schedule disamenities and for physical and aesthetic requirements; skill counts are the numbers of hard and soft skills, and the number of requirements is the total count of skill and ascriptive requirements. Concentration is the standardised log leave-one-out Herfindahl–Hirschman Index of vacancy postings at the title-district-month level. Standard errors are clustered by firm. ***p<0.01, **p<0.05, *p<0.1.

wages. Within a position, an explicit "men" request carries a premium of about 3%, a "women" request carries a penalty of about 5%, and the two are not statistically different in magnitude. The implicit channel follows the same pattern: a one-standard-deviation shift toward male-coded framing (SD = 0.12 in the wage sample) is associated with approximately a 3% higher posted wage. The estimates are stable as the skill and ascriptive requirements are added across columns, so they are not driven by differences in the observed job requirements. Thus, gendered wording is not cheap talk: wording oriented toward men offers higher wages, whereas wording oriented toward women offers lower ones.

4.2.2 Wage changes around label-switching events

The level estimates compare targeted and untargeted postings of the same position at different dates. To bring the comparison closer to the event of interest (i.e., dropping vs. adding labels), we estimate the wage analogue of Equation (3): the change in the log posted wage between consecutive postings of the same position, no more than sixty days apart, in

which both postings disclose a wage. Requiring both postings to disclose a wage means that the first-difference estimates are internally free of the posting margin, and Appendix C.2 shows that this disclosure is essentially unrelated to the gender signals.

Table 4: Event-level first-difference wage regressions

	(1) Linear	(2) Pos/Neg split	(3) Quadratic
Dep. var:		$\Delta \ln(w)$	
Implicit index			
$\Delta Gender\ Index$	0.166***		0.168***
	(0.061)		(0.061)
$\Delta Gender\ Index^{sqr}$			-0.366*
			(0.204)
$\Delta Gender\ Index^{+}$		0.109	
		(0.076)	
$\Delta Gender\ Index^{-}$		0.225***	
		(0.078)	
Explicit label switches			
$drop^{M}$	−0.021	−0.020	−0.020
	(0.016)	(0.016)	(0.016)
add^{M}	0.026**	0.027***	0.027***
	(0.012)	(0.012)	(0.012)
$drop^{F}$	0.040***	0.040***	0.040***
	(0.011)	(0.011)	(0.011)
add^{F}	−0.037***	−0.036***	−0.036***
	(0.009)	(0.009)	(0.009)
N of observations	21,203	21,203	21,203
N of positions	1,683	1,683	1,683
R^2	0.291	0.291	0.292
Human capital	Yes	Yes	Yes
Job requirements	Yes	Yes	Yes
Concentration	Yes	Yes	Yes
Firm × title FEs	Yes	Yes	Yes
Posting-date FEs	Yes	Yes	Yes

Notes: This table presents first-difference estimates analogous to Equation (3), pooled across all sometimes-targeting regimes. The dependent variable is the change in log posted wage between consecutive postings of the same position, no more than 60 days apart, both disclosing a wage. $\Delta Gender\ Index$ is the change in the implicit index. Column (1) enters $\Delta Gender\ Index$ linearly. Column (2) splits it into positive ($\Delta Gender\ Index^{+}$) and negative ($\Delta Gender\ Index^{-}$) components. Column (3) adds a quadratic term. *drop* and *add* indicate that an explicit label is removed or added relative to the previous posting. Human capital controls are the changes in the share of ads in each category of education, experience and job level. Job requirement controls are the changes in the requirement aggregates, including the total number of requirements. Concentration enters as the change in $ln(HHI)$. Standard errors are clustered at the firm–title cell level. Regressions are weighted by the number of adverts in each cell–date observation. ***p<0.01, **p<0.05, *p<0.1.

Column (1) of Table 4 reports the results. We find that dropping a "women" request is followed by a 4.0% increase in the posted wage and adding one by a 3.7% decrease, while adding a "men" request is followed by a 2.6% increase. Dropping a "men" request is

followed by a 2.1% decrease, although it is not statistically significant. Within each gender, the *add* and *drop* effects are equal and opposite: we cannot reject equal-sized coefficients, implying that the wage change following a gender request removal is undone when the request returns. The implicit channel follows the same sign pattern: a one-standard-deviation increase in the index between consecutive postings (SD=0.08) is associated with a 1.3% higher posted wage. These estimates are also robust to conditioning on job content more finely, both by using the changes in all 42 skill and ascriptive requirements and by restricting the sample to postings in which none of them changes. We will return to these exercises in Section 4.3.[14]

Unlike the explicit request, the implicit index is continuous, so its wage effect could differ by the direction of the change. Column (2) of Table 4 splits the change in the index into its positive part, $\Delta Gender\ Index^{+} = \max\{\Delta Gender\ Index, 0\}$, and its negative part, $\Delta Gender\ Index^{-} = \min\{\Delta Gender\ Index, 0\}$. The point estimates suggest a steeper response to shifts toward female-coded wording (0.23 versus 0.11), but we cannot reject that the two are equal. Column (3) adds a quadratic term, which is negative and marginally significant, although within the observed range the curvature is small relative to the linear effect. In other words, there is weak evidence that the wording premium depends on the direction of the change.

4.3. Discussion

The results reported above describe how firms write and price their postings. In this section, we discuss what the comovement estimates recover and which explanations are consistent with the wage pattern.

What the comovement estimates recover. The estimates in Section 4.1 recover the net of two forces. The first is substitution: for a given gender preference, stating it explicitly could reduce the need to convey it through wording. Thus, withdrawing an explicit request would shift the wording toward the requested gender. The second is a common response: the explicit request and the wording move together, either because the firm's valuation of male relative to female applicants could vary across postings of the same position, or because, as shown in Section 5, workers find an ad easier to interpret when the request and the wording point the same way. The positive comovement reported above indicates that the second dominates, although it does not rule out the first. Distinguishing the two requires variation in the cost of using the explicit channel that is independent of the

[14] To check that wage changes do not precede explicit request changes, we add leads of four request events and of the change in the index (Appendix C.3). The contemporaneous estimates are robust to this check, and the lead of the wording change is close to zero.

underlying preferences, which our setting does not provide.[15]

Potential explanations of the wage pattern. The findings in Section 4.2 suggest that both channels are priced in the same direction: signalling a preference for men, whether through an explicit request or through male-coded wording, is associated with a higher posted wage, and signalling a preference for women with a lower one. Two standard theories, compensating differentials (Rosen, 1986) and taste-based discrimination (Becker, 1957), offer natural candidates for this pattern. Yet, as we discuss below, neither fully accounts for it.

In the compensating differentials framework, gender preferences could proxy for job content. If the jobs for which firms prefer men are also more demanding, the premium and penalty could compensate for job content rather than reflect the gender preference itself. This account predicts the same sign pattern, but it also predicts that the wage effects should disappear once job content is held fixed. The two robustness exercises (Appendix C.3) speak to this. Conditioning on the changes in all 42 individual requirements, the wording coefficient falls from 0.17 to 0.13 and remains significant, and the effects of adding and removing a "women" request are only slightly smaller. For the explicit channel, we can go further by restricting the sample to consecutive postings in which *none of the 42 requirements changes*, so that a request is added or removed while every measured requirement stays the same. Again, the women penalty remains: removing a "women" request is followed by a 4.1% increase in the posted wage and adding one by a 3.6% decrease.[16] For compensating differentials to explain this, the job would have to become lesser in some unmeasured way whenever the "women" request is added and greater again whenever it is removed.

Does taste-based discrimination offer an alternative explanation? If enough employers prefer men, women's market wage falls below men's, so a firm that requests men must offer a higher wage than it would without a request. A posting without a request, however, can already attract women at their lower market wage, so adding a "women" request need not lower the posted wage. It happens nonetheless, including among positions that never switch directly between "men" and "women" requests (Column (4) - Appendix C.3). Thus, taste-based discrimination can only account for the men premium but not for the women penalty.

[15]The revised Labour Code (Law No. 45/2019/QH14) was passed in November 2019 but took effect in January 2021, after our sample ends. It is also a common event affecting all firms simultaneously, and a law regulating discriminatory recruitment could plausibly affect the wording of ads directly, which is our outcome of interest.

[16]The effects of adding and removing a "men" request retain their signs in this subsample but are smaller and imprecisely estimated, so part of the men premium may compensate for the more "demanding" requirements that "men" requests make more likely.

5. A model of explicit and implicit gender requests

To rationalise the main empirical results of the previous section, we propose a static directed-search model. This considers firms with a vacancy and choices over three recruitment instruments: a posted wage, an explicit gender request, and implicit gendered language. Thus, the model combines the applicant-selection problem of Kuhn and Shen (2013) with directed worker applications, as in Marinescu and Wolthoff (2020). This section presents the central mechanisms and predictions. The complete model, including derivations and proofs, is presented in Appendix B.

Recruitment instruments and application supply. Consider a labour-market segment containing many firms, each with one vacancy, and two types of workers, $G \in \{M, F\}$. Each firm posts a recruitment package (w, e, h), where w is a binding posted wage, $h \in \mathbb{R}$ is the *implicit* gender orientation of the ad, and

$$e \in \{-1, 0, 1\}$$

denotes an explicit request for women, no explicit request, or an explicit request for men, respectively. Larger values of h indicate more male-coded language. Explicit and implicit signals jointly determine the total underlying gender orientation of the ad:

$$\tau = h + e. \tag{5}$$

Workers interpret the two signals jointly. Their direction determines whom the ad encourages to apply, while their consistency determines how clearly the ad communicates the firm's recruitment intention. We represent this ambiguity by

$$D(e, h) = \frac{\delta}{2}(h - e)^2, \qquad \delta > 0. \tag{6}$$

For parsimony, we set the coherence benchmark to $h = e$: an explicit request accompanied by implicit wording of the same direction and magnitude supplies a coherent message. Departures from that wording make the firm's intention harder for workers to infer and reduce applications from both men and women.

The closest empirical support for this ambiguity channel comes from Kuhn, Shen and Zhang (2020). They find that explicit gender requests have their largest effects on applicant composition when the employer's likely preference is difficult to infer from the job title. In jobs whose titles are not clearly female-coded, women constitute a relatively large share of applicants primarily when women are explicitly requested; the corresponding response among men is weaker. This suggests that workers combine explicit requests with other

information in an ad when deciding whether an application is worthwhile.

Let n_M and n_F denote the effective equilibrium queue lengths of male and female applicants:

$$n_M(w,e,h) = \exp\left[\eta_M w + (h+e) - D(e,h)\right], \tag{7}$$

$$n_F(w,e,h) = \exp\left[\eta_F w - (h+e) - D(e,h)\right], \tag{8}$$

where $\eta_M, \eta_F > 0$. Ambiguity reduces the common scale and therefore cancels out from the applicant ratio:

$$\log\left(\frac{n_M}{n_F}\right) = (\eta_M - \eta_F)w + 2(h+e). \tag{9}$$

Ambiguity affects the number of applicants rather than the gender composition of the applicant pool. A discrepancy, $h-e$, scales both queues by $\exp[-\delta(h-e)^2/2]$. Thus, the parameter δ measures how strongly workers are discouraged by inconsistency between the explicit request and the implicit wording.[17]

The wage semi-elasticities, η_M and η_F, govern how strongly each group's applications respond to the posted wage. We assume that male applications respond more strongly:

$$\eta_M > \eta_F. \tag{10}$$

The economic intuition is that women's application decisions may depend more strongly on non-wage constraints, such as commuting distance, working hours, and compatibility of a job with family responsibilities. A higher wage at a particular vacancy cannot compensate for these constraints as readily as it can attract applicants whose job search is less geographically or temporally restricted. Equivalently, if men consider a broader set of jobs and rank these opportunities more strongly by pay, a wage increase redirects more male than female applications toward the vacancy. Thus, a more male-oriented applicant pool makes the wage a more effective recruitment instrument.

While this ordering might seem at odds with the literature on aggregate labour-supply elasticities, it is consistent with evidence that women's labour supply to an *individual* firm is less elastic than men's (Hirsch, Schank and Schnabel, 2010; Ransom and Oaxaca, 2010). Search behaviour exhibits the same pattern: women apply more often than men to lower-paying firms in Germany (Lochner and Merkl, 2026) and Denmark (Fluchtmann et al., 2024). In Vietnam, Hoang, Mitra and Pham (2026) likewise find that men's labour supply is more wage-elastic than women's, plausibly reflecting tighter constraints on where women can work, including limited geographic mobility stemming from family responsibilities.

[17] We leave δ unrestricted because it governs the economic strength of the ambiguity channel relative to the possible substitution between the two recruitment instruments; setting $\delta = 1$ would impose a particular strength for this mechanism and could determine whether explicit and implicit targeting move together.

The firm's problem. Each applicant has a group-specific expected value μ_G and an idiosyncratic match component. Under a type-I extreme-value specification for the idiosyncratic component, the applicant-selection problem has an analytical solution (Kuhn and Shen, 2013). The firm's recruitment return can then be written compactly as

$$\mathcal{R}(w, e, h; z) = \mu_F + \beta \log\left[n_M \exp(z) + n_F\right] - w - c(n_M + n_F), \tag{11}$$

where $c > 0$ is the cost of processing an application, $\beta > 0$ captures the dispersion of idiosyncratic applicant quality, and

$$z \equiv \frac{\mu_M - \mu_F}{\beta}$$

is the firm's standardised relative valuation of male applicants.

Using implicit gendered language entails a strictly convex cost $\psi(h)$, which may represent the loss of task-relevant information or credibility when the wording departs from a neutral description. The firm solves

$$\max_{w,h,\,e\in\{-1,0,1\}} \{\mathcal{R}(w, e, h; z) - \psi(h)\}. \tag{12}$$

The two signals create opposing forces. Because both affect total orientation, the firm can use one in place of the other. However, substituting between them makes the ad less coherent and reduces the number of workers willing to apply. When this loss of clarity dominates the underlying substitution force, the firm optimally aligns its explicit request and implicit wording. For regime-specific choices, subscripts $F, 0, M$ denote $e = -1, 0, 1$, respectively.

Proposition 1 (Explicit and implicit targeting are aligned)**.**
If the recruitment gain from a coherent message dominates the substitution between the two targeting instruments, the conditional optima satisfy

$$h_F^* < h_0^* < h_M^*. \tag{13}$$

An explicit request for men is, hence, accompanied by more male-coded implicit wording, while an explicit request for women is accompanied by more female-coded wording. This result arises because workers use the whole ad to infer whether applying is worthwhile. Appendix B states the formal dominance condition and proves the result.

Wage effects across explicit-targeting regimes. The alignment of explicit requests and implicit language implies that the female-request regime is the most female-oriented, the untargeted regime lies in between, and the male-request regime is the most male-oriented.

Because ambiguity also affects how many workers apply, the wage comparison is clearest when the three ad types are all similarly easy to interpret. Therefore, we consider the case in which their conditional optima contain the same degree of ambiguity. This includes fully unambiguous recruitment packages, in which the explicit request and implicit wording convey the same message.

Wage responses. The ordering $\eta_F < \eta_M$ becomes important for the wage response. Define the value-weighted male share of the applicant pool as

$$p_M \equiv \frac{n_M \exp(z)}{n_M \exp(z) + n_F}.$$

A more male-oriented applicant pool raises the return to a higher wage because male applications respond more strongly to it. This gain in applicant value works against the additional processing cost of attracting more applications. When the value gain dominates, wages and male-oriented framing are complements: within an explicit-targeting regime, more male-coded wording raises the optimal wage at a coherent message.

Proposition 2 (Wage effects across explicit-targeting regimes)**.**
Suppose that the applicant-value channel dominates the processing-cost channel, and that the three conditional optima contain the same degree of ambiguity; then

$$w_F^* < w_0^* < w_M^*. \tag{14}$$

Thus, the female-request regime posts a lower wage than the regime with no explicit request, whereas the male-request regime posts a higher wage.

The same ordering also holds when ambiguity differs across regimes, provided that its queue-scale effect is weaker than the wage effect of the change in total orientation. The economic mechanism is the same as within an explicit regime: moving toward a more male-oriented applicant pool makes the wage a more effective recruitment instrument.

This static model does not specify why the same position may appear in different explicit-targeting regimes across postings. Such switching could be rationalised by transitory changes in the legal, reputational, or administrative costs of stating an explicit request, without any change in z. When a temporary change in cost makes an explicit request optimal, complementarity moves implicit wording in the same direction; when the cost returns to its previous level, both choices revert.

Preferences and cross-position sorting. The (unobserved) position-specific preference z provides a separate source of variation across jobs. Within any explicit-targeting regime, a higher z raises the value of attracting men and, therefore, increases the return to more

male-coded wording. Across regimes, the same force makes a male request more attractive as z rises.

Proposition 3 (Preferences and cross-position sorting)**.**
Suppose that the wage and implicit wording satisfy the complementarity conditions described above. Within every explicit-targeting regime,

$$\frac{\partial h_r^*(z)}{\partial z} > 0, \qquad r \in \{F, 0, M\}. \tag{15}$$

If each explicit-targeting regime is optimal for some value of z, there are thresholds $z_F^* < z_M^*$ such that firms with low z request women, firms with intermediate z make no explicit request, and firms with high z request men.

Positions with a larger stable relative valuation of male applicants thus use more male-coded implicit language within a given explicit regime, and are more likely to select a more male-oriented explicit regime. Positions far from the targeting thresholds tend to use the same regime persistently, while positions close to a threshold can switch regimes in response to small transitory changes in the cost of explicit targeting. The targeting regimes reported in Section 3 follow the ordering implied by the model: positions that always request men use more male-coded wording than those that never state a preference, and these more than positions that always request women.

Together, Proposition 1, Proposition 2, and Proposition 3 can rationalise the observed positive comovement of explicit and implicit gender signals when firms re-post ads (Section 4.1.2), their ordering with respect to posted wages, the higher wage associated with male-coded wording (Section 4.2), and the sorting of positions into targeting regimes (Figure 5).

The model also permits a qualitative counterfactual for a ban on explicit gender requests. Holding the position-specific gender preference fixed, such a ban restricts all firms to the regime without an explicit request. Firms that previously requested men consequently adopt less male-coded implicit wording and lower wages, whereas firms that previously requested women adopt less female-coded wording and higher wages. The ban therefore compresses both recruitment signals toward the untargeted optimum, although stable heterogeneity in firms' preferences means that implicit gender differentiation persists. These are partial-equilibrium predictions, assuming that workers' responses to an ad without an explicit request remain stable after the policy change.

6. Conclusion

Restrictions on explicit gender requests in job ads raise an important question: do they also reduce implicit gendered recruitment, or simply displace targeting into the wording of job requirements? Combining online vacancy data from a major Vietnamese job board with sentence-embedding models, we construct a *Gender Index* that measures implicit gendered framing in the distinct requirements text of ads. We document substantial variation in this framing within firm–job-title positions, even among the majority of positions where explicit gender requests remain unchanged.

We then focus on the minority of positions for which explicit gender requests do change between ads. Here, we find that explicit and implicit signalling move together: when an explicit gender request is withdrawn, the wording of an ad for the same title in the same firm shifts away from the withdrawn gender, and reverts when the request returns. We further find that both channels are priced into posted wages: a female request carries a wage penalty, a male request carries a wage premium, and the wage responses to changes in the gendered language follow suit. A directed-search model in which applicants respond negatively to any ambiguity between explicit and implicit gender signals in an ad, and women's labour supply to the firm is less wage-elastic than men's, rationalises both the complementary use by firms of the two signals and their associated wage responses. Thus, the model implies that restrictions on explicit requests need not displace gender preferences into implicit language and may also reduce their expression through it.

Our evidence has limits. First, we observe posted rather than realised wages, and neither applications nor hires, so we speak to the expression and pricing of gendered preferences in ads, not to their consequences for realised sorting or pay. Second, identification rests on within-firm-title comparisons rather than on exogenous variation, and we accordingly interpret the estimates as associations rather than causal effects. Online job board experiments or randomised controlled trials could overcome some of these limitations (e.g., Burbano, 2021; Horton, Rand and Zeckhauser, 2011). For example, a platform experiment that withdrew the explicit gender field option from employers for a random subset of vacancies would establish whether restricting this channel causes a shift away from implicit wording, and allow its consequences to be traced through to applications and hires. Further, complementary correspondence study-type designs, which randomise the gendered framing of otherwise identical ads, would isolate the effect of implicit wording on worker responses (Bertrand and Mullainathan, 2004; Kessler, Low and Sullivan, 2019; Kircher, 2020).

Nevertheless, the findings offer several practical implications. For policymakers, regulations targeting overt profiling might also reach the gendered language within ad text. More broadly, the prevalence of gendered language that we have documented, even when

employers never use explicit gender requests, suggests that ensuring that job ads are genuinely neutral may require more attention to the linguistic framing of requirements, not just the presence or absence of explicit labels. For job platforms, automated tools that flag strongly gender-coded language or encourage more neutral phrasing could support this aim, and the approach presented in this paper could be useful for building such automated tools. For firms, the results underscore that the wording of job ads is not cosmetic: implicit framing is priced into the posted wage, just as explicit requests are, and, in our model, wording that conflicts with the explicit request could discourage applicants of both genders.

References

Ash, Elliott, Daniel L. Chen, and Arianna Ornaghi. 2024. "Gender Attitudes in the Judiciary: Evidence from US Circuit Courts." *American Economic Journal: Applied Economics*, 16(1): 314–350. (Cited on 4, 9, 1, 2)

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources*, 57(S): S167–S199. (Cited on 7, 22)

Azmat, Ghazala, and Barbara Petrongolo. 2014. "Gender and the labor market: What have we learned from field and lab experiments?" *Labour Economics*, 30: 32–40. (Cited on 1)

Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press. (Cited on 26)

Belot, Michèle, Philipp Kircher, and Paul Muller. 2022. "How Wage Announcements Affect Job Search – A Field Experiment." *American Economic Journal: Macroeconomics*, 14(4): 1–67. (Cited on 3)

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, 94(4): 991–1013. (Cited on 32)

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics*, 2(3): 228–255. (Cited on 1, 5)

Blackwell, Louisa, Jennifer Jarman, and Janet Siltanen. 2009. "CASCOT: Computer Assisted Structured COding Tool." Version 5. University of Warwick. (Cited on 12)

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, 55(3): 789–865. (Cited on 1, 5, 8)

Born, Marise P., and Toon W. Taris. 2010. "The impact of the wording of employment advertisements on students' inclination to apply for a job." *Journal of Social Psychology*, 150(5): 485–502. (Cited on 8)

Brenčič, Vera, and John B. Norris. 2010. "Do Employers Change Job Offers in Their Online Job Ads to Facilitate Search?" *Economics Letters*, 108(1): 46–48. (Cited on 3)

Brenčič, Vera, and John B Norris. 2012. "Employers' On-line Recruitment and Screening Practices." *Economic Inquiry*, 50(1): 94–111. (Cited on 3)

Burbano, Vanessa C. 2021. "The Demotivating Effects of Communicating a Social-Political Stance: Field Experimental Evidence from an Online Labor Market Platform." *Management Science*, 67(2): 1004–1025. (Cited on 32)

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics*, 131(2): 633–686. (Cited on 1)

Card, David, Fabrizio Colella, and Rafael Lalive. 2025. "Gender Preferences in Job Vacancies and Workplace Gender Diversity." *Review of Economic Studies*, 92(4): 2437–2471. (Cited on 1, 3, 6, 10)

Castilla, Emilio J., and Hye Jin Rho. 2023. "The Gendering of Job Postings in the Online Recruitment Process." *Management Science*, 69(11): 6912–6939. (Cited on 3, 8, 10, 12)

Chaturvedi, Sugat, Kanika Mahajan, and Zahra Siddique. 2025. "Gendered Language in Job Ads and Applicant Behavior: Evidence from India." *Labour Economics*, 96: 102726. (Cited on 3, 6, 10, 12)

Deming, David, and Lisa B. Kahn. 2018. "Skill Requirements Across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics*, 36(S1): S337–S369. (Cited on 3, 9)

Faberman, Jason R., and Marianna Kudlyak. 2019. "The Intensity of Job Search and Search Duration." *American Economic Journal: Macroeconomics*, 11(3): 327–357. (Cited on 3)

Faryna, Oleksandr, Tho Pham, Oleksandr Talavera, and Andriy Tsapin. 2022. "Wage and unemployment: Evidence from online job vacancy data." *Journal of Comparative Economics*, 50(1): 52–70. (Cited on 3)

Fluchtmann, Jonas, Anita M. Glenny, Nikolaj A. Harmon, and Jonas Maibom. 2024. "The Gender Application Gap: Do Men and Women Apply for the Same Jobs?" *American Economic Journal: Economic Policy*, 16(2): 182–219. (Cited on 3, 28)

Galenianos, Manolis, and Philipp Kircher. 2009. "Directed Search with Multiple Job Applications." *Journal of Economic Theory*, 114(2): 445–471. (Cited on 3)

Gao, Tianyu, Xingcheng Yao, and Danqi Chen. 2021. "SimCSE: Simple Contrastive Learning of Sentence Embeddings." *arXiv preprint arXiv:2104.08821*. (Cited on 9)

Gaucher, Danielle, Justin Friesen, and Aaron C. Kay. 2011. "Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality." *Journal of Personality and Social Psychology*, 101. (Cited on 1, 3, 8, 10, 11, 12)

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–74. (Cited on 3)

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119. (Cited on 1, 5)

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, 20(4): 133–156. (Cited on 8)

Grootendorst, Maarten. 2020. "KeyBERT: Minimal keyword extraction with BERT." (Cited on 8)

Helleseter, Miguel Delgado, Peter Kuhn, and Kailing Shen. 2020. "The Age Twist in Employers' Gender Requests: Evidence from Four Job Boards." *Journal of Human Resources*, 55(2): 428–469. (Cited on 5)

Hershbein, Brad, and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review*, 108(7): 1737–1772. (Cited on 3)

Hirsch, Boris, Thorsten Schank, and Claus Schnabel. 2010. "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany." *Journal of Labor Economics*, 28(2): 291–330. (Cited on 2, 3, 28)

Hoang, Trang, Devashish Mitra, and Hoang Pham. 2026. "The effect of export market access on labor market power: Firm-level evidence from Vietnam." *Journal of Development Economics*, 182: 103826. (Cited on 28)

Horton, John, David Rand, and Richard Zeckhauser. 2011. "The online laboratory: conducting experiments in a real labor market." *Experimental Economics*, 14(3): 399–425. (Cited on 32)

Kessler, Judd B., Corinne Low, and Colin D. Sullivan. 2019. "Incentivized Resume Rating: Eliciting Employer Preferences without Deception." *American Economic Review*, 109(11): 3713–3744. (Cited on 32)

Kircher, Philipp. 2020. "Search design and online job search – new avenues for applied and experimental research." *Labour Economics*, 64(C): 101820. (Cited on 32)

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11(4): 181–209. (Cited on 5)

Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics*, 128(1): 287–336. (Cited on 1, 3, 6, 7, 10, 27, 29, 4, 5, 12, 13)

Kuhn, Peter, and Kailing Shen. 2023. "What Happens When Employers Can No Longer Discriminate in Job Ads?" *American Economic Review*, 113(4): 1013–1048. (Cited on 1, 3, 7, 10)

Kuhn, Peter, Kailing Shen, and Shuo Zhang. 2020. "Gender-targeted job ads in the recruitment process: Facts from a Chinese job board." *Journal of Development Economics*, 147: 102531. (Cited on 2, 3, 27)

Lochner, Benjamin, and Christian Merkl. 2026. "Gender-Specific Application Behaviour, Matching, and the Residual Gender Earnings Gap." *Economic Journal*, 136(673): 97–124. (Cited on 3, 28)

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets.* Princeton University Press. (Cited on 22)

Manning, Alan, and Helen Robinson. 2004. "Something in the way she moves: a fresh look at an old gap." *Oxford Economic Papers*, 56(2): 169–188. (Cited on 5)

Marinescu, Ioana, and Ronald P. Wolthoff. 2018. "Mismatch Unemployment and the Geography of Job Search." *American Economic Journal: Macroeconomics*, 10(3): 42–70. (Cited on 3)

Marinescu, Ioana, and Ronald P. Wolthoff. 2020. "Opening the Black Box of the Matching Function: The Power of Words." *Journal of Labor Economics*, 38(2): 535–568. (Cited on 3, 7, 27, 14)

Michelacci, Claudio, and Javier Suarez. 2006. "Incomplete Wage Posting." *Journal of Political Economy*, 114(6): 1098–1123. (Cited on 14)

Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. 2016. "Downskilling: Changes in Employer Skill Requirements over the Business Cycle." *Labour Economics*, 41: 333–347. (Cited on 3)

Nomura, Shinsaku, Saori Imaizumi, Ana Carolina Areias, and Futoshi Yamauchi. 2017. *Toward Labor Market Policy 2.0: The Potential for Using Online Job-Portal Big Data to Inform Labor Market Policies in India.* World Bank Policy Research Working Paper No. 7966. (Cited on 3)

Pham, Tho, Daniel Schaefer, and Carl Singleton. 2026. "Unequal Hiring Wages and their Impact on the Gender Pay Gap." *Economic Inquiry*. (Cited on 5)

Ransom, Michael R., and Ronald L. Oaxaca. 2010. "New Market Power Models and Sex Differences in Pay." *Journal of Labor Economics*, 28(2): 267–289. (Cited on 2, 3, 28)

Rosen, Sherwin. 1986. "The Theory of Equalizing Differences." In *Handbook of Labor Economics*. Vol. 1, ed. Orley Ashenfelter and Richard Layard. Elsevier: North Holland, Amsterdam and Boston. (Cited on 26)

Ruggles, Steven, Lara Cleveland, Rodrigo Lovaton, Sula Sarkar, Matthew Sobek, Derek Burk, Dan Ehrlich, Jane Lee, and Nate Merrill. 2025. "Integrated Public Use Microdata Series, International: Version 7.6." (Cited on 13)

Shimer, Robert. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions." *Journal of Political Economy*, 113(5): 996–1025. (Cited on 3, 4)

Wolthoff, Ronald P. 2018. "Applications and Interviews: Firms' Recruiting Decisions in a Frictional Labor Market." *Review of Economic Studies*, 85(2): 1314–1351. (Cited on 4)

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri. 2021. "Directed Search and Competitive Search Equilibrium: A Guided Tour." *Journal of Economic Literature*, 59(1): 90–148. (Cited on 22)

Online Appendix

Appendix A. Technical notes

This brief appendix explains how our construction of the Gender Index relates to and builds on the approach used by Ash, Chen and Ornaghi (2024) to measure gender slant in judicial opinions.

Ash, Chen and Ornaghi (2024) measure gender attitudes using *judge-specific* word-embedding spaces. Their procedure involves:

- training a separate GloVe model on the corpus of majority opinions authored by each judge;
- constructing, *within each judge-specific embedding space*, a **gender direction** (male-female) and a **career-family direction** using predefined word lists;
- defining a judge's gender slant as the cosine similarity between these two concept vectors.

The resulting measure captures how strongly a judge's language associates men with career and women with family. It is a *concept-concept association* computed at the *judge level*.

Our objective is different: we aim to measure *gendered framing in individual job ads*. Accordingly, we:

- embed each requirements text using a sentence-level model (SimCSE-Vi);
- construct a **data-driven gender direction** using the observed contrast between ads that explicitly target men and those that explicitly target women;
- compute a **vacancy-level Gender Index** by projecting each ad's embedding onto this direction.

This index captures the direction and intensity of gendered framing in the requirements text, even when no explicit gender label is present.

Although both approaches use vector differences to recover latent semantic structure, they differ in purpose, implementation, and suitability for the underlying text:

- **Unit of analysis.** Ash, Chen and Ornaghi (2024) construct a single slant measure per judge, based on the association between two conceptual dimensions in that judge's

writing. Our setting requires variation at the vacancy level to study substitution behaviour, monopsony responses, and wage posting. A document-level index is therefore essential.

- **Source of the gender direction.** In Ash, Chen and Ornaghi (2024), the gender and career-family directions are defined using *predefined* word lists and constructed separately within each judge-specific embedding space. In contrast, we identify the gender direction from *employer-revealed* contrasts in the text of actual job ads. This is crucial in a context where gendered language is culturally specific, often implicit, and not well captured by English-based lexicons.

- **Type of association measured.** Ash, Chen and Ornaghi (2024) measure the similarity between two abstract semantic concepts (gender and career/family). Our goal is different: we measure how *each ad* aligns with a gendered semantic direction. This allows us to detect meaningful variation *within* firm-title cells and across vacancies.

- **Embedding model.** Ash, Chen and Ornaghi (2024) train separate word-embedding models for each judge. We use sentence embeddings pre-trained on a large Vietnamese corpus, which can capture the contextual meaning of short requirements texts and allow for a unified embedding space across all ads.

- **Applicability to the setting.** Judicial opinions are long, homogeneous, and authored by identifiable individuals, making judge-level embeddings feasible and concept–concept comparisons meaningful. In contrast, requirements texts are short, heterogeneous, and vary at the vacancy level. A document-level projection approach is therefore more appropriate for recovering fine-grained, context-dependent gendered framing.

Our approach adapts the underlying idea of semantic directions in Ash, Chen and Ornaghi (2024) to a context where the relevant variation occurs across individual vacancies, and where gendered language is revealed through employers' own wording choices rather than pre-defined lexicons.

Appendix B. Conceptual model – details and proofs

This Appendix provides the complete model, derivations, and proofs for the model in Section 5 in the main text.

B.1. Environment

Market participants and timing. We consider a static labour market divided into segments containing vacancies and workers who are sufficiently close substitutes to apply for the same jobs. We study one such segment in isolation.

There is a large number of firms, each with one vacancy, and two types of workers, indexed by $G \in \{M, F\}$. Workers and firms are risk neutral. Each worker submits one application. Each firm posts a recruitment package (w, e, h), where w is a binding posted wage, $e \in \{-1, 0, 1\}$ is the explicit-targeting regime, and $h \in \mathbb{R}$ is the implicit gender orientation of the advertisement. The values $e = -1, 0, 1$ denote an explicit request for women, no explicit request, and an explicit request for men, respectively. A larger value of h denotes more male-coded language. The posted wage cannot depend on the type of worker eventually hired. Workers observe all three recruitment instruments before choosing where to apply.

Gender orientation and application supply. Explicit and implicit signals jointly determine the underlying gender orientation of the vacancy:

$$\tau = h + e. \tag{B1}$$

Their consistency determines the ambiguity of the recruitment message:

$$D(e, h) = \frac{\delta}{2}(h - e)^2, \qquad \delta > 0. \tag{B2}$$

The derivatives used below are

$$D_h = \delta(h - e), \qquad D_e = -\delta(h - e), \qquad D_{hh} = \delta, \qquad D_{eh} = -\delta. \tag{B3}$$

We call a recruitment package coherent when $h = e$, so that $D = D_h = D_e = 0$.

We normalise the baseline application pools of men and women to unity. As is standard in the literature (e.g., Marinescu and Wolthoff, 2020; Shimer, 2005), identical workers use identical mixed strategies in the large market.[18]

[18] The assumption that each worker submits only one application per period captures the opportunity cost of applying and induces workers to choose applications selectively (see Belot, Kircher and Muller, 2022). Models allowing homogeneous workers to submit multiple applications include Galenianos and Kircher

Let n_M and n_F denote the equilibrium effective queue lengths:

$$n_M(w,e,h) = \exp\left[\eta_M w + (h+e) - D(e,h)\right], \tag{B4}$$

$$n_F(w,e,h) = \exp\left[\eta_F w - (h+e) - D(e,h)\right], \tag{B5}$$

where $\eta_M, \eta_F > 0$. The exponential specification guarantees positive queues and imposes constant wage semi-elasticities.[19]

For later use, define the queues in the absence of ambiguity,

$$\overline{n}_M(w,h,e) \equiv \exp[\eta_M w + (h+e)], \qquad \overline{n}_F(w,h,e) \equiv \exp[\eta_F w - (h+e)]. \tag{B6}$$

Then

$$n_G(w,e,h) = \exp[-D(e,h)]\overline{n}_G(w,h,e), \qquad G \in \{M,F\}. \tag{B7}$$

Because

$$\frac{\partial^2 \log n_G}{\partial (h-e)^2} = -\delta,$$

δ is the curvature of log application supply with respect to signal inconsistency. Equivalently, a one-unit discrepancy reduces each queue by the factor $\exp(-\delta/2)$. We retain δ because its magnitude determines whether the recruitment gain from greater clarity dominates the underlying substitution between explicit and implicit targeting.

Ambiguity reduces the common scale of the two queues. It does not affect their ratio:

$$\log\left(\frac{n_M}{n_F}\right) = (\eta_M - \eta_F)w + 2(h+e). \tag{B8}$$

Assumption 1 (Wage semi-elasticities)**.**
The parameters satisfy

$$\eta_F < \eta_M. \tag{B9}$$

Assumption 1 means that a higher wage attracts relatively more men. Together with equation (B8), this makes a higher wage and a more male-oriented signal complementary recruitment instruments under the condition stated below.

Applicant values. Following Kuhn and Shen (2013), the gross value to the firm of applicant j from group G is

$$U_{G,j} = v_G + \varepsilon_j, \qquad G \in \{M,F\}, \tag{B10}$$

(2009) and Wolthoff (2018).

[19] In standard directed-search models, symmetric mixed application strategies in a large market generate Poisson-distributed applicant counts, with the equilibrium queue length as the endogenous mean (e.g., Shimer, 2005). Our analysis only uses the reduced-form expected queue lengths.

where the ε_j are independent type-I extreme-value draws with scale parameter $\beta > 0$. Hence $\mathbb{E}[\varepsilon_j] = \beta\gamma$, where γ is Euler's constant. Define

$$\mu_G \equiv v_G + \beta\gamma \tag{B11}$$

and

$$z \equiv \frac{\mu_M - \mu_F}{\beta}. \tag{B12}$$

As in Kuhn and Shen (2013), z is the firm's standardised relative valuation of men for the vacancy. It may reflect expected productivity, turnover, customer or co-worker preferences, or employer tastes.

Lemma 1 (Expected value of the best applicant)**.**
Let

$$\omega \equiv \frac{n_M}{n_M + n_F}.$$

The expected value of the best applicant in the pool is

$$U^*(n_M, n_F) = \beta \log\left[n_M \exp\left(\frac{\mu_M}{\beta}\right) + n_F \exp\left(\frac{\mu_F}{\beta}\right)\right]. \tag{B13}$$

Equivalently,

$$U^*(n_M, n_F) = \mu_F + \beta \log\left[\omega \exp(z) + (1-\omega)\right] + \beta \log(n_M + n_F). \tag{B14}$$

Proof. This follows from the standard extreme-value result used in Kuhn and Shen (2013, Lemma 1 and Online Appendix 1). Substitute $n_M = \omega(n_M + n_F)$, $n_F = (1-\omega)(n_M + n_F)$, and equation (B12) into equation (B13). □

B.2. The Firm's Objective

The firm hires the most valuable applicant, incurs a processing cost $c > 0$ per application, and pays the posted wage. It also incurs a cost $\psi(h)$ from using implicit gendered language. We assume

$$\psi(0) = \psi'(0) = 0, \qquad \psi''(h) > 0, \qquad \lim_{|h|\to\infty} \frac{\psi(h)}{|h|} = \infty. \tag{B15}$$

The recruitment return before the implicit-language cost is

$$\mathcal{R}(w, e, h; z) = \mu_F + \beta \log\left[n_M(w, e, h) \exp(z) + n_F(w, e, h)\right] - w - c\left[n_M(w, e, h) + n_F(w, e, h)\right]. \tag{B16}$$

To isolate the ambiguity channel, define the clear-message return

$$\begin{aligned}\mathcal{R}_0(w,h,e;z) \equiv \mu_F + \beta \log\left[\overline{n}_M(w,h,e)\exp(z) + \overline{n}_F(w,h,e)\right] - w \\ - c[\overline{n}_M(w,h,e) + \overline{n}_F(w,h,e)].\end{aligned} \tag{B17}$$

Equations (B7)–(B17) imply the exact decomposition

$$\mathcal{R}(w,e,h;z) = \mathcal{R}_0(w,h,e;z) - \beta D(e,h) + c\overline{N}(w,h,e)\left[1 - \exp[-D(e,h)]\right], \tag{B18}$$

where $\overline{N} \equiv \overline{n}_M + \overline{n}_F$. The first ambiguity term is the loss in expected applicant quality from a smaller pool. The second is the processing-cost saving from receiving fewer applications. Thus

$$\mathcal{R}_D = -\beta + cN, \qquad N \equiv n_M + n_F. \tag{B19}$$

Ambiguity lowers recruitment returns whenever $\beta > cN$.

The firm solves

$$\max_{w,h,\,e\in\{-1,0,1\}} \{\mathcal{R}(w,e,h;z) - \psi(h)\}. \tag{B20}$$

B.3. Optimal Recruitment Choices

The optimal posted wage. Define the value-weighted male share

$$p_M(w,e,h;z) \equiv \frac{n_M(w,e,h)\exp(z)}{n_M(w,e,h)\exp(z) + n_F(w,e,h)}. \tag{B21}$$

Because the common ambiguity term cancels from this share, the wage first-order condition is

$$\mathcal{R}_w = \beta\left[\eta_M p_M + \eta_F(1-p_M)\right] - 1 - c\left[\eta_M n_M + \eta_F n_F\right] = 0. \tag{B22}$$

Assumption 2 (Interior wage choice)**.**
The parameters satisfy $\beta\eta_F > 1$, *and*

$$\mathcal{R}_{ww}(w,e,h;z) < 0 \tag{B23}$$

over the relevant range.

Assumption 2 gives a unique interior wage

$$w^*(e,h;z) \equiv \arg\max_w \mathcal{R}(w,e,h;z). \tag{B24}$$

Define the return after optimising the wage by

$$\widehat{\mathcal{R}}(e,h;z) \equiv \mathcal{R}\Big(w^*(e,h;z),e,h;z\Big). \tag{B25}$$

Conditional on explicit regime e, the firm chooses

$$h^*(e;z) \in \arg\max_h \left\{ \widehat{\mathcal{R}}(e,h;z) - \psi(h) \right\}. \tag{B26}$$

We focus on strict interior conditional optima satisfying

$$\psi''(h^*) - \widehat{\mathcal{R}}_{hh}(e,h^*;z) > 0. \tag{B27}$$

For regime-specific choices, subscripts $F, 0, M$ denote $e = -1, 0, 1$, respectively. For the comparative static, extend e locally to a continuous variable. The wording first-order condition is

$$\widehat{\mathcal{R}}_h(e,h;z) - \psi'(h) = 0. \tag{B28}$$

Implicit differentiation gives

$$\frac{\partial h^*(e;z)}{\partial e} = \frac{\widehat{\mathcal{R}}_{eh}(e,h^*;z)}{\psi''(h^*) - \widehat{\mathcal{R}}_{hh}(e,h^*;z)}. \tag{B29}$$

At a coherent message, the cross-partial has the transparent decomposition

$$\widehat{\mathcal{R}}_{eh} = \widehat{\mathcal{R}}_{0,\tau\tau} + \delta\left(\beta - c\overline{N}\right). \tag{B30}$$

Here $\widehat{\mathcal{R}}_{0,\tau\tau}$ is the curvature of the wage-optimised clear-message return. The first term captures substitution: because e and h both increase τ, diminishing returns to orientation tend to make them substitutes. The second term captures clarity: raising e shifts the ambiguity-minimising value of h upward.

Assumption 3 (Clarity dominates substitution)**.**
Along the conditional optima connecting $e = -1, 0, 1$,

$$\widehat{\mathcal{R}}_{eh}(e, h^*(e;z);z) > 0. \tag{B31}$$

At coherent optima, a sufficient primitive condition is

$$\delta\left(\beta - c\overline{N}\right) > -\widehat{\mathcal{R}}_{0,\tau\tau}. \tag{B32}$$

Proposition B1 (Explicit and implicit targeting are aligned; Section 5, Proposition 1)**.**
Under Assumption 2, equation (B27)*, and Assumption 3,*

$$h_F^* < h_0^* < h_M^*. \tag{B33}$$

Moreover,

$$\tau_F^* < \tau_0^* < \tau_M^*. \tag{B34}$$

Proof. The denominator of equation (B29) is positive by equation (B27), while the numerator is positive by Assumption 3. Hence $\partial h^*/\partial e > 0$. Evaluating this ordering at $e = -1, 0, 1$ gives equation (B33). Since $\tau^*(e;z) = e + h^*(e;z)$,

$$\frac{\partial \tau^*}{\partial e} = 1 + \frac{\partial h^*}{\partial e} > 0,$$

which gives equation (B34). □

Wage-orientation complementarity. Differentiating the value-weighted male share while holding w and e fixed gives

$$\left.\frac{\partial p_M}{\partial h}\right|_{w,e} = 2p_M(1-p_M). \tag{B35}$$

Define

$$\begin{aligned}\mathcal{B}(w,e,h;z) \equiv\ & 2\beta(\eta_M - \eta_F)p_M(1-p_M) \\ & - c\left[\eta_M n_M - \eta_F n_F\right]\end{aligned} \tag{B36}$$

and

$$\mathcal{C}(w,e,h) \equiv \eta_M n_M + \eta_F n_F. \tag{B37}$$

The wage-wording cross-partial is

$$\mathcal{R}_{wh} = \mathcal{B}(w,e,h;z) + cD_h(e,h)\mathcal{C}(w,e,h). \tag{B38}$$

The second term is the wage consequence of the queue-scale response to ambiguity. At a coherent message, $D_h = 0$, leaving only the composition channel:

$$\mathcal{R}_{wh}|_{h=e} = \mathcal{B}(w,e,h;z). \tag{B39}$$

Assumption 4 (Underlying wage-orientation complementarity)**.**
Over the relevant range,

$$\mathcal{B}\Big(w^*(e,h;z),e,h;z\Big) > 0. \tag{B40}$$

This condition requires the applicant-value gain from combining a higher wage with a more male-oriented applicant pool to dominate the associated processing-cost increase. It is more likely to hold when $\eta_M - \eta_F$ or β is large and less likely when c is high. For the local within-regime result, the condition need only hold at the relevant coherent optimum. For the comparisons across regimes, it must hold along the fixed-ambiguity intervals connecting the conditional optima.

Implicit differentiation of equation (B22) yields

$$\left.\frac{\partial w^*}{\partial h}\right|_e = -\frac{\mathcal{R}_{wh}}{\mathcal{R}_{ww}}. \tag{B41}$$

At a coherent message, define

$$\mathcal{A}(e,h;z) \equiv \frac{\mathcal{B}(w^*,e,h;z)}{-\mathcal{R}_{ww}(w^*,e,h;z)} > 0. \tag{B42}$$

(B39) and (B41) then imply

$$\left.\frac{\partial w^*}{\partial h}\right|_e = \mathcal{A}(e,h;z) \tag{B43}$$

Wage effects across explicit-targeting regimes. Because e and h enter the queues through τ and D, write the optimal wage as $w^*(\tau, D; z)$. Holding D fixed,

$$\mathcal{R}_{w\tau} = \mathcal{B}(w,\tau,D;z). \tag{B44}$$

Under Assumption 4 and wage concavity,

$$\left.\frac{\partial w^*(\tau,D;z)}{\partial \tau}\right|_D = -\frac{\mathcal{R}_{w\tau}}{\mathcal{R}_{ww}} > 0. \tag{B45}$$

Let $D_r^* \equiv D(e_r, h_r^*)$ for $r \in \{F, 0, M\}$. The following transparent condition isolates the orientation comparison.

Assumption 5 (Equal ambiguity across regimes)**.**
The three conditional optima satisfy

$$D_F^* = D_0^* = D_M^*. \tag{B46}$$

This condition holds in particular if all three conditional optima are coherent. More generally, the result below continues to hold when ambiguity differs, provided its queue-scale effect does not overturn the wage response to total orientation.

Proposition B2 (Wage effects across explicit-targeting regimes; Section 5, Proposition 2)**.**
Under Assumption 1, Assumption 2, Assumption 4, and Assumption 5,

$$w_F^* < w_0^* < w_M^*. \tag{B47}$$

If the relevant packages lie on the male-coded margin, the response to the difference in total orientation is attenuated by κ.

Proof. Proposition B1 gives

$$\tau_F^* < \tau_0^* < \tau_M^*.$$

Assumption 5 holds D fixed across these comparisons, and equation (B45) implies that the optimal wage is strictly increasing in τ. This yields equation (B47). □

The choice of whether to target explicitly. We first establish how the fixed position preference z shifts optimal implicit wording. By the envelope theorem,

$$\widehat{\mathcal{R}}_z(e,h;z) = \beta p_M. \tag{B48}$$

Differentiating with respect to h, accounting for the optimal wage response, gives

$$\widehat{\mathcal{R}}_{hz} = \beta p_M(1-p_M)\left[2 + (\eta_M - \eta_F)\left.\frac{\partial w^*}{\partial h}\right|_e\right]. \tag{B49}$$

Assumption 6 (Preference-wording complementarity)**.**
At the relevant conditional optima,

$$2 + (\eta_M - \eta_F)\left.\frac{\partial w^*}{\partial h}\right|_e > 0. \tag{B50}$$

The condition holds whenever the optimal wage is locally increasing in male-coded wording, including at coherent messages under Assumption 4. It allows modest ambiguity-induced wage effects away from coherence as long as they do not overturn the direct composition effect. Under this condition, equation (B49) is positive. Implicit differentiation of equation (B28) therefore yields

$$\frac{\partial h^*(e;z)}{\partial z} = \frac{\widehat{\mathcal{R}}_{hz}}{\psi''(h^*) - \widehat{\mathcal{R}}_{hh}} > 0. \tag{B51}$$

Let the conditional value in regime $r \in \{F, 0, M\}$ be

$$\overline{\Pi}_r(z) \equiv \max_h \left\{\widehat{\mathcal{R}}(e_r, h; z) - \psi(h)\right\}. \tag{B52}$$

The envelope theorem gives

$$\overline{\Pi}_r'(z) = \beta p_{M,r}^*. \tag{B53}$$

By Proposition B1 and Proposition B2, total orientation and the wage both increase from F to 0 to M. Equation (B8) and $\eta_M > \eta_F$ therefore imply

$$p_{M,F}^* < p_{M,0}^* < p_{M,M}^*. \tag{B54}$$

Consequently, the pairwise conditional-value differences satisfy single crossing in z.

Proposition B3 (Stable preferences and cross-position sorting; Section 5, Proposition 3)**.** *Under Assumption 6 and the conditional second-order condition* (B27)*,*

$$\frac{\partial h_r^*(z)}{\partial z} > 0, \qquad r \in \{F, 0, M\}. \tag{B55}$$

If each explicit-targeting regime is optimal for some value of z, and the assumptions of Proposition B2 also hold, there exist thresholds $z_F^* < z_M^*$ such that

$$e^*(z) = \begin{cases} -1, & z < z_F^*, \\ 0, & z_F^* \le z \le z_M^*, \\ 1, & z > z_M^*. \end{cases} \tag{B56}$$

Proof. Equation (B51) proves the within-regime result. For the regime choice, equations (B53) and (B54) imply

$$\frac{d}{dz}(\overline{\Pi}_M - \overline{\Pi}_0) > 0, \qquad \frac{d}{dz}(\overline{\Pi}_F - \overline{\Pi}_0) < 0, \qquad \frac{d}{dz}(\overline{\Pi}_M - \overline{\Pi}_F) > 0.$$

The pairwise value differences therefore satisfy single crossing. If every regime is optimal for some z, the optimum progresses from a female request to no request and then to a male request as z increases. □

The model treats z as a stable position characteristic. Transitory legal, reputational, or administrative costs of stating an explicit request can nevertheless move a position across regimes without changing z. Such costs alter the level of a regime's conditional value and can rationalise observed switching while leaving all within-regime and cross-position comparative statics unchanged.

Appendix C. Additional analyses

This Appendix presents some additional analyses that complement the main findings.

C.1. Explicit gender requests and skills targeting

We follow Kuhn and Shen (2013) and examine whether general human capital and skill-related requirements are systematically associated with gender targeting in job ads, by estimating linear regression models of the following form:

$$y_{ijt} = \alpha + \mathbf{x}'_{ijt}\boldsymbol{\beta} + \delta_t + \mu_j + \varepsilon_{ijt}\,, \tag{C1}$$

where y_{ijt} is one of four outcome variables in ad i posted by firm j in month t: (1) *GenderReq* equals one if an ad includes an explicit gender request and zero otherwise; (2) *GenderTarget* takes values −1 (female-targeting), 0 (no explicit targeting), or 1 (male-targeting); (3) *Gender Index* measures the implicit gender target in job ads; (4) |*Gender Index*| is the absolute value of the *Gender Index*, capturing the overall intensity of gender signalling in the requirements text irrespective of the targeted gender. The vector $\mathbf{x}_{ijt}$ includes indicators for explicit characteristics of the advert: minimum education and experience requirements, and job-level seniority (see also Table 1). Month fixed effects δ_t are included in all specifications, and firm fixed effects μ_j are added in certain specifications to isolate within-firm variation.[20]

Table C1 reports the estimation results. Job seniority plays a significant role in predicting gender targeting. Compared to entry-level positions, ads for experienced non-managerial roles are 4 to 6.5 percentage points *less* likely to include explicit gender requests (Columns 1 & 5) and show a 0.01 unit (≈ 0.1 SD) reduction in implicit gender requests, regardless of the direction (Columns 3 & 7). Managerial and supervisory roles show even stronger effects: ads for managers and team leaders are 8 to 13 percentage points less likely to include explicit gender requests and exhibit a 0.02 to 0.03 unit (≈ 0.35 SD) reduction in |*Gender Index*|. Experience requirements tell a somewhat different story from job level. Ads requiring five or more years of experience are more likely to include explicit gender requests and lean significantly toward male targeting (Columns 2 & 6), whereas shorter experience categories show weaker, less consistent associations. This suggests that experience and job level capture distinct dimensions of skill demand, with differing associations depending on the direction of gender targeting.

[20] Unlike the descriptive analysis in Figures 3 and 5, which operated at the firm–title level, the regression specifications here include firm fixed effects rather than firm–title fixed effects. Firm–title fixed effects would eliminate all cross-occupation variation, which is the primary source of identifying variation for the question of whether explicit skill level and human capital requirements predict gender targeting.

In contrast, technical and engineering roles are consistently associated with higher levels of gender requests and targeting. Ads for these roles are 7.5 to 12.7 percentage points more likely to include an explicit gender request (Columns 1 & 5), shift targeting toward men and away from women in the *GenderTarget* specification (Columns 2 & 6), and show a 0.11 unit ($\approx$ 0.9 SD) increase in the signed *Gender Index*, even after controlling for firm fixed effects (Column 8). The direction of targeting also shifts systematically with job level. While lower-level roles show little association with the direction of explicit targeting, supervisory and managerial positions are strongly associated with male targeting: the team leader and manager coefficients for *GenderTarget* are 0.08 to 0.13 and 0.04 to 0.12, respectively, and statistically significant once firm fixed effects are included (Column 6). This suggests that within-firm variation in targeting direction is concentrated at senior levels.

Higher education requirements present a more nuanced pattern. Associate and university degree requirements are strongly associated with female targeting in the *GenderTarget* specification (coefficients of -0.11 to -0.18; Columns 2 & 6), yet show near-zero effects on $|Gender\ Index|$. This suggests that education-intensive roles explicitly target women, but the requirement text itself is relatively gender-neutral in its framing. This might reflect the concentration of degree requirements in feminised professional occupations, such as accounting and administration, where female targeting is conventional, while the language is gender-neutral.

These results support the hypothesis proposed by Kuhn and Shen (2013) that more demanding jobs are less likely to be gender-targeted, even conditional on the requested level of experience. The persistence of these patterns, after controlling for firm fixed effects, indicates that they are not solely driven by firm-level heterogeneity but reflect meaningful within-firm variation in how job requirements are framed. As such, this regression exercise complements the descriptive evidence in Figures 3 to 5 and reinforces the stylised fact that gendered framing in job ads is structured, systematic, and tied to occupational characteristics.

C.2. Selection into wage posting

The wage results in Section 4.2 are estimated on the subsample of advertisements that post a wage. Obviously, posting a wage is a firm's choice, so the concern is that the labelled and unlabelled postings being compared are drawn from different parts of the wage distribution. The relevant threat is not that posting a wage is selective in general, which is well documented, but that selection is *differential* with respect to the gendered signals whose wage effects we estimate. To assess this, we estimate linear probability models of wage posting on the explicit labels, the residualised index, and the full set of

controls, within the same fixed-effects structure as the wage regressions. Table C2 reports the results for the pooled estimation sample in Column (1) and, for completeness, for the always-targeting and never-targeting cells in Columns (2 & 3).

We find that wage posting is unrelated to gender targeting on both margins. Conditional on the fixed effects, labelled and unlabelled postings of the same position disclose wages at statistically and economically indistinguishable rates. Similarly, the implicit margin is statistically insignificant. As such, the selection into wage posting is uncorrelated with the (gender) signals whose effects we estimate. By contrast, wage posting varies systematically with skill requirements. For example, postings requiring a higher level of education are 9-24 percentage points less likely to disclose a wage, and those requiring more experience are 3-12 percentage points less likely. This is in line with evidence from CareerBuilder, where jobs requiring a graduate degree are significantly less likely to post a wage than jobs requiring a 2-year college degree (Marinescu and Wolthoff, 2020), and with the theoretical prediction that employers may forgo wage posting where bargaining better accommodates worker-specific productivity differences (Michelacci and Suarez, 2006). This finding also underlines that the null effects of the gendered signals are not an artefact of an uninformative specification: the same regressions detect selection where theory predicts it, and detect none on the targeting margins.

C.3. Wage results - Robustness

Table C1: The human capital and skill requirements of jobs and gender requests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	GenderReq	*G'Target*= {−1, 0, 1}	|*Gender Index*|	*Gender Index*	*GenderReq*	*G'Target*= {−1, 0, 1}	|*Gender Index*|	*Gender Index*
Min. education (excl., ≤ High school):								
Associate degree	−0.068***	−0.157***	−0.000	−0.010	0.008	−0.108***	−0.001	−0.022**
	(0.025)	(0.041)	(0.007)	(0.017)	(0.035)	(0.036)	(0.005)	(0.009)
University degree	−0.114***	−0.184***	−0.000	−0.006	−0.044	−0.161***	−0.001	−0.034***
	(0.025)	(0.040)	(0.008)	(0.019)	(0.037)	(0.038)	(0.005)	(0.009)
Others	−0.105***	−0.135***	−0.005	0.023	−0.052	−0.110**	−0.005	−0.018
	(0.040)	(0.046)	(0.009)	(0.022)	(0.039)	(0.048)	(0.008)	(0.013)
Min. experience, years (excl., < 1):								
1-2	0.011	0.016	0.011*	0.003	−0.004	−0.026	0.014*	−0.015*
	(0.016)	(0.039)	(0.006)	(0.005)	(0.013)	(0.039)	(0.007)	(0.009)
2-5	0.036*	0.054*	0.010**	0.018***	−0.019*	0.006	0.012*	−0.012
	(0.020)	(0.030)	(0.004)	(0.004)	(0.011)	(0.029)	(0.006)	(0.008)
5+	0.092***	0.137***	0.017***	0.038***	0.011	0.067**	0.021**	0.004
	(0.022)	(0.037)	(0.007)	(0.006)	(0.016)	(0.033)	(0.008)	(0.010)
Job level (excl., Entry):								
Experienced, Non-manager	−0.065**	−0.012	−0.009*	−0.021**	−0.039*	0.041	−0.008*	0.008
	(0.030)	(0.041)	(0.005)	(0.010)	(0.022)	(0.032)	(0.005)	(0.006)
Technical/Engineer	0.130***	0.388***	0.044***	0.118***	0.077***	0.345***	0.033***	0.111***
	(0.030)	(0.041)	(0.006)	(0.010)	(0.026)	(0.035)	(0.006)	(0.008)
Team leader/Supervisor	−0.122***	0.078*	−0.021***	−0.001	−0.081***	0.133***	−0.021***	0.026***
	(0.029)	(0.042)	(0.006)	(0.008)	(0.030)	(0.032)	(0.006)	(0.007)
Manager or higher	−0.131***	0.040	−0.027***	−0.020**	−0.084**	0.118***	−0.028***	0.017**
	(0.032)	(0.047)	(0.006)	(0.009)	(0.040)	(0.033)	(0.007)	(0.008)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	No	No	No	No	Yes	Yes	Yes	Yes
N of ads	306,408	306,408	306,408	306,408	304,839	304,839	304,839	304,839
R^2	0.016	0.033	0.025	0.051	0.301	0.240	0.213	0.328

Notes: Linear regression model estimates of Equation (C1). Columns (1) & (5) use an indicator equal to 1 if the advertisement contains an explicit gender request and 0 otherwise. Columns (2) & (6) uses a variable equal to −1 for ads requesting women, 0 for ads with no explicit label, and 1 for ads requesting men. Columns (3) & (7) use the absolute value of the estimated gender index of the ad (see Equation 1), and Columns (4) & (8) use the signed value of this index. Standard errors (in parentheses) are clustered at the firm level. ***p<0.01, **p<0.05, *p<0.1.

Table C2: Selection into wage posting

	(1) Sometimes target	(2) Always target	(3) Never target
Gender Index	0.160	0.119	-0.034
	(0.183)	(0.088)	(0.060)
men	-0.016		
	(0.013)		
women	0.027*		
	(0.016)		
N of observations	98,769	19,543	78,788
N of positions	3,442	2,325	6,833
R^2	0.577	0.861	0.770
Human capital	Yes	Yes	Yes
Job requirements	Yes	Yes	Yes
Concentration	Yes	Yes	Yes
Firm × title FEs	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes

Notes: This table reports linear probability estimates of whether an ad discloses a wage, using all ads regardless of disclosure. Column (1) uses positions that sometimes state a gender preference, the estimation sample for the wage analysis in Section 4.3; Columns (2) and (3) use positions that always and never state one, respectively. Human capital controls are indicators for categories of education, experience and job level. Job requirement controls are indicators for work schedule disamenities and for physical and aesthetic requirements, the counts of hard and soft skills, and the total number of skill and ascriptive requirements. Concentration is the standardised log leave-one-out Herfindahl–Hirschman Index of vacancy postings at the title-district-month level. Standard errors are clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C3: Event-level first-difference wage regressions: robustness

	(1) With leads	(2) All 42 requirements	(3) No requirement change	(4) Excluding mixed regime
Dep. var:		$\Delta \ln(w)$		
$\Delta Gender\ Index$	0.176**	0.128**	−0.040	0.182***
	(0.070)	(0.063)	(0.085)	(0.065)
$drop^M$	−0.019	−0.019	−0.014	−0.045*
	(0.018)	(0.015)	(0.015)	(0.025)
add^M	0.038**	0.023*	0.016	0.014
	(0.016)	(0.013)	(0.015)	(0.017)
$drop^F$	0.037***	0.035***	0.041***	0.046***
	(0.012)	(0.011)	(0.010)	(0.014)
add^F	−0.036***	−0.031***	−0.036***	−0.034***
	(0.013)	(0.009)	(0.010)	(0.011)
N of observations	17,191	21,203	9,254	15,153
N of positions	1,258	1,683	1,252	1,432
R^2	0.283	0.306	0.222	0.324
Human capital	Yes	Yes	Yes	Yes
Requirement aggregates	Yes	No	—	Yes
Individual requirements (42)	No	Yes	—	No
Leads of add/drop	Yes	No	No	No
Concentration	Yes	Yes	Yes	Yes
Firm × title FEs	Yes	Yes	Yes	Yes
Posting-date FEs	Yes	Yes	Yes	Yes

Notes: This table reports robustness checks for Column (1) of Table 4. The dependent variable is the change in log posted wage between consecutive postings of the same position, no more than 60 days apart, both disclosing a wage. Column (1) adds leads of the change in the index and of the four request events, measured at the next posting of the same position. Column (2) replaces the requirement aggregates with the changes in all 42 individual requirement indicators. Column (3) restricts the sample to consecutive postings in which none of the 42 requirements changes. Column (4) excludes positions that sometimes request men and sometimes women, so that every event adds or removes a single request rather than switching between them. $\Delta Gender\ Index$ is the change in the implicit index. *drop* and *add* indicate that an explicit request is removed or added relative to the previous posting. Human capital controls are the changes in the share of ads in each category of education, experience and job level. Requirement aggregates are the changes in indicators for work schedule disamenities and for physical and aesthetic requirements, in the counts of hard and soft skills, and in the total number of requirements. Concentration is the change in the standardised log leave-one-out Herfindahl–Hirschman Index. Standard errors are clustered at the firm–title level. Regressions are weighted by the number of ads in each cell–date observation. ***p<0.01, **p<0.05, *p<0.1.

Appendix D. Additional tables and figures

Table D1: The frequency of additional requirements extracted from job ads ($N = 306,408$)

Requirement:	% of ads
Problem solving	9.4
Teamwork	13.9
Project management	0.7
Consulting	1.2
Communication	48.4
Persuasion	2.6
Organisational	2.2
Negotiate	14.6
Management	4.1
Basic computer	30.4
Advanced computer	9.6
Leadership	1.4
Language	24.0
Analytical	3.8
Positive attitude	0.4
Long-term commitment	3.4
Eager to learn	7.5
Driving	5.5
Business trip	4.1
No domestic commitment	0.2
Extra shift	2.0
Voice	5.7
Appearance	12.7
Height	0.2
Health	9.9
Withstand pressure	22.7
Cheerful	6.7
Responsible	18.6
Agile	22.9
Meticulous	2.9
Sociable	3.2
Progressive	5.6
Proactive	3.1
Energetic	11.8
Careful	13.1
Independent	13.1
Honest	22.3
Hardworking	18.9
Friendly	3.4
Enthusiastic	17.2
Creative	7.7
Confident	6.1

Notes: This table reports the proportion of ads (in %) that include the listed requirement or characteristic.

Table D2: Examples of job ad requirements text

Gender Index	Example requirement text (translated)
5^{th} percentile	Possesses strong communication, persuasion, and negotiation skills. Able to work independently and under pressure. Preference given to candidates with experience in sales or technical positions. Able to read and understand English documents. Positive, energetic, honest, and collaborative. Proficient in Microsoft Office.
25^{th} percentile	Knowledge of CAD and Office. Preference given to candidates fluent in Chinese.
50^{th} percentile	Preference given to candidates with N2 Japanese language skills. Proficient in using computers and Office software. Energetic, able to handle high pressure, and possessing good teamwork and communication skills.
75^{th} percentile	Majors in quality management, business administration, etc. Knowledge of ISO standards. Proficient in using specialised software. Problem-solving skills. Planning and teamwork skills. Foreign language not required.
95^{th} percentile	Knowledge of testing processes and techniques. Knowledge of C+/C++ programming. Proficient in using code management tools such as Git, SVN, etc. Ability to work in a team and handle high-pressure situations.

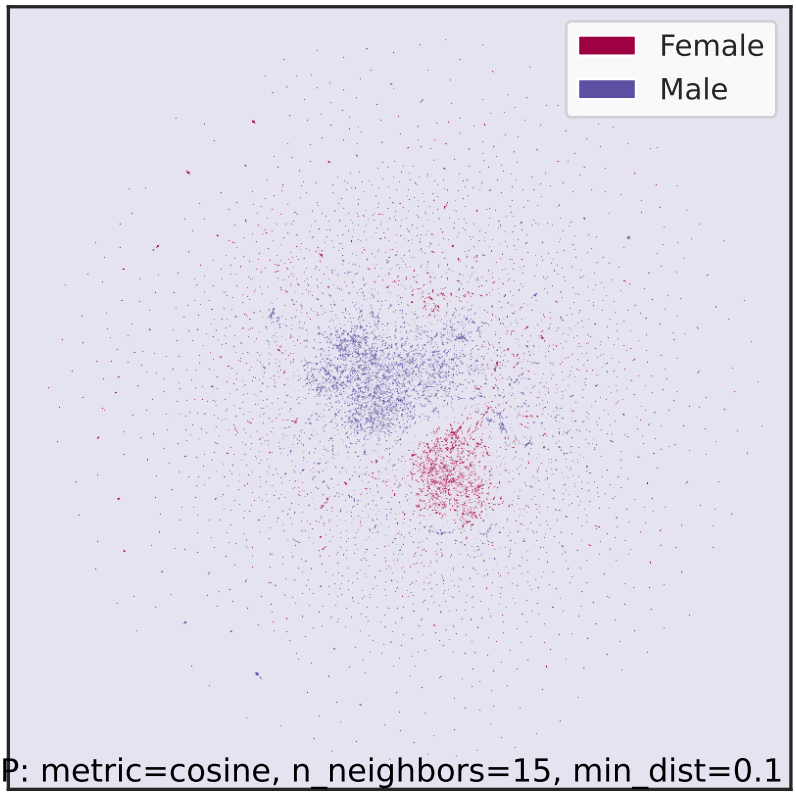
Notes: This table reports examples of the requirements text observed at selected percentiles of the *Gender Index* for the estimation sample of ads ($N = 306,408$), higher values imply more male-targeting language. See Section 2.2.3.

Table D3: Gender Index regressed on required skills and ascriptive characteristics

	$\widetilde{GI}$		$\widetilde{GI}$
agile	-0.030^{***}	language	-0.055^{***}
analytical	-0.005^{***}	leadership	-0.001
appearance	-0.089^{***}	longterm commit.	-0.014^{***}
basic computer	-0.072^{***}	management	-0.001
business trip	0.013^{***}	meticulous	-0.007^{***}
careful	-0.037^{***}	negotiate	-0.034^{***}
cheerful	-0.003^{***}	no domestic commit.	0.012^{***}
communication	-0.067^{***}	organizational	-0.009^{***}
confident	-0.027^{***}	persuasion	-0.046^{***}
consulting	0.005^{***}	positive attitude	0.011^{***}
creative	0.019^{***}	proactive	0.015^{***}
driving	0.014^{***}	problem solving	0.011^{***}
eager to learn	0.032^{***}	progressive	-0.002^{***}
energetic	0.007^{***}	project management	0.095^{***}
enthusiastic	-0.011^{***}	responsible	-0.027^{***}
extra shift	0.033^{***}	sociable	-0.002
friendly	0.004^{***}	teamwork	0.019^{***}
hardworking	-0.001	voice	-0.063^{***}
health	0.047^{***}	withstand pressure	-0.013^{***}
height	0.036^{***}		
honest	-0.039^{***}		
independent	0.017^{***}		
N of ads: 306,408	R^2: 0.419		

Notes: Each entry is the coefficient from regressing the *Gender Index* on 41 dummy variables indicating the presence of required skills and ascriptive characteristics in an advert. Standard errors are omitted for brevity. This regression underlies Figure 1. ***p<0.01, **p<0.05, *p<0.1.

Figure D1: 2-D embedding vectors



Notes: Each point represents a job ad's requirement text embedded in 2-D space using dimensionality reduction. Purple dots indicate explicitly male-targeting ads; red dots indicate explicitly female-targeting ads.