What Do Online Job Vacancies (OJVs) Measure? A Critical Assessment Using Matched OJVs – Administrative Data*

Massimiliano Bratti[1], Alessandro Corvasce[2], Agata Maida[3], and Filippo Passerini[4]

[1]University of Milan; GLO; IZA; LdA
[2]University of Milan
[3]University of Milan; LaboR; IZA
[4]University of Milan; LaboR

April 13, 2026

Abstract

Online job vacancies (OJVs) are widely used to proxy firms' labor demand, yet their relationship with realized hirings remains unclear. We construct a novel plant-level panel by matching the universe of Italian online postings (Lightcast) with mandatory administrative records for firms in Piedmont and Veneto (2014–2019). We document strong selection into online posting: larger, younger, and more productive firms are significantly more likely to advertise. While aggregate postings and hires are positively correlated, this relationship weakens sharply across education and occupation groups. The distribution of the demand gap (postings minus hires) is highly dispersed, and posting durations are often exceptionally long, consistent with signaling and talent-pooling practices. Overall, OJVs capture a mix of labor demand and recruitment strategies rather than direct vacancy creation.

Keywords: Online job vacancies; labor demand measurement; recruitment strategies
JEL Codes: J21 J23

*Corresponding author: Alessandro Corvasce, University of Milan, 20122, alessandro.corvasce@unimi.it.
We thank the DEMM - University of Milan for the funding used to purchase the data and Mauro Pelucchi of Lightcast for the helpful assistance. All errors are our own.

1 Research Question and Motivation

Over the last decade, Online Job Vacancies (OJVs) have become a prominent data source in applied labor economics. Large-scale collections of online postings are now widely used to study labor demand, occupational change, skill requirements, job-filling frictions, and employer market power, among other outcomes (Deming and Kahn, 2018; Marinescu and Wolthoff, 2020; Azar et al., 2022; Acemoglu et al., 2022; Grasso and Tatsiramos, 2023). Their appeal lies in the high level of granularity they provide and in their ability to track firms' recruiting behavior at high frequency.

At the same time, a growing methodological literature has emphasized that OJVs reflect multiple layers of firm behavior, recruitment strategies, and platform-specific dynamics. Fabo and Kureková (2022) document substantial non-representativeness and unstable coverage of online vacancies, while OECD (2024) shows that OJV counts diverge systematically from official vacancy statistics across regions, sectors, and occupations. Importantly, even influential applications using OJVs for structural analysis acknowledge that postings capture a mixture of hiring needs and organizational responses to shocks rather than a direct measure of vacancies (Acemoglu et al., 2022). Contributions focusing on data construction further highlight persistent challenges related to deduplication, misclassification, and missing information. Taken together, these insights suggest that postings should not be interpreted mechanically as direct measures of vacancy creation.

Despite these caveats, much of the applied literature continues to rely on implicit assumptions that postings correspond to genuine vacancies, that posting intensity scales proportionally with hiring needs, and that posting duration captures job-filling difficulty. To the best of our knowledge, these assumptions have not been systematically evaluated using linked administrative hiring data at the micro level. This paper seeks to address this gap by exploiting a unique empirical setting in Italy that allows the universe of Lightcast postings to be matched to the universe of mandatory employer–employee records (COB) for all firms headquartered in Piemonte and Veneto. The resulting plant-level panel data enable a direct comparison between advertised and filled vacancies and provide a rigorous basis for reassessing what OJVs measure in practice. To the best of our knowledge, this is among the first studies to systematically match large-scale online job posting data to administrative employer–employee records at the plant level.

2 Literature

Labor Demand, Wages, and Employer Adjustment A first strand of the literature uses OJVs to study labor demand and wage-setting behavior. Marinescu and Wolthoff (2020) show that job advertisement content, including wage information, shapes applicant flows, while others use vacancy postings to measure labor market concentration and employer market power (Azar et al., 2022). Recent contributions further explore these dynamics; Choi and Marinescu (2024) utilize OJVs to analyze involuntary part-time employment as a measure of labor market slack and employer power. OJVs are also instrumental in tracking demand dynamics during shocks, such as the COVID-19 crisis

(Forsythe et al., 2020), capturing firm-level strategic adjustments to changing economic conditions (Babina et al., 2024) and the general equilibrium effects of unemployment insurance (Marinescu, 2016).

Skills, Task Content, and Technological Change A vast literature exploits the textual richness of job ads to study the evolution of skill requirements. Deming and Kahn (2018) document substantial skill heterogeneity across firms, while Hershbein and Kahn (2018) show how recessions accelerate routine-biased technological change by raising requirements. Recent studies provides more granular insights: Ziegler (2024) finds a robust link between complex skill requirements, higher wages, and longer vacancy durations, while Galassi et al. (2024) document that the pandemic's impact on long-term digitalization was more modest than previously hypothesized. Further research highlights the necessity of country-specific taxonomies and advanced Natural Language Processing (NLP) methods to capture skill-gender associations and diverging occupational contents in emerging economies (Escudero et al., 2024; Chaturvedi et al., 2024; Antonie et al., 2024). Other studies use OJVs to analyze task content across markets (Atalay et al., 2024), skill obsolescence in STEM (Deming and Noray, 2020), and institutional reforms (Grasso and Tatsiramos, 2023; Modestino et al., 2016; Fabo et al., 2017).

Matching, Frictions, and Recruitment Behavior OJVs are widely used to study search behavior and matching frictions. While geographic mismatch explains only a limited share of unemployment (Marinescu and Rathelot, 2018), digital platforms can reduce information barriers. Morales et al. (2024) show that mandatory vacancy reporting to public services can shift the Beveridge Curve inward, enhancing efficiency. Beyond matching, online ads reveal insights into recruitment practices, including explicit gender preferences (Kuhn and Shen, 2013; Kuhn et al., 2018) and the impact of online candidate information on hiring outcomes (Acquisti and Fong, 2019; Faberman and Kudlyak, 2019). Bratti et al. (2026) finally exploits OJVs data to compute different measures of firms' demand and investigate whether a GMI scheme introduced in Italy in 2019 may induce higher frictions.

Measurement, Representativeness, and Methodological Limits Alongside the rapid expansion of empirical applications, a growing methodological literature has questioned what online job vacancies (OJVs) actually measure. A central concern relates to the representativeness of online postings and the potential discrepancies between these data and underlying hiring activity. Several studies highlight structural challenges in the construction of OJV datasets: Fabo and Kureková (2022), for instance, document substantial issues related to deduplication, misclassification, and unstable coverage across different sources, which can distort both the levels and the dynamics of measured labor demand. These technical limitations are compounded by measurement errors inherent in automated classification and scraping procedures (Joint Research Centre, European Commission, 2024; Tzimas et al., 2024).

Beyond technical issues, the statistical representativeness of OJVs remains a subject of debate.

Comparisons between online data and official labor market statistics reveal systematic divergences across sectors, occupations, and regions (OECD, 2024). Recent evidence from Shen and Zhu (2024) provides a nuanced view: while Shen and Zhu (2024) find that job postings exhibit predictive power comparable to traditional unemployment rates for labor transitions, they also acknowledge that representativeness varies systematically by sector and firm type. As Lovaglio and Mezzanzanica (2026) demonstrate using generalized sample selection models, online job ads suffer from substantial over-representation and endogenous selection bias that traditional post-stratification techniques fail to correct fully. This evidences suggests that online postings may reflect a selective segment of recruitment activity rather than the full universe of vacancies.

Importantly, even influential empirical applications recognize that OJVs do not necessarily correspond one-to-one to realized vacancies or hiring decisions. As Acemoglu et al. (2022) emphasize, job postings may capture a combination of hiring needs, organizational adjustments, and strategic responses to technological shocks, reflecting broader firm strategies rather than pure vacancy creation.

This study contributes to this debate by providing, to the best of our knowledge, the first direct micro-level linkage between the universe of online job postings with administrative employer–employee records. This unique matching approach allows for a systematic assessment of what online job vacancies actually measure in practice, addressing the representativeness concerns raised by the most recent literature.

3 Data, Sample Selection, and Descriptive Analysis

The analysis combines three main data sources. First, we exploit the universe of Lightcast Online Job Vacancies (OJVs) for Italy, covering 2013–2024. Restricting to postings with complete information on occupation, education, contract type, location, and duration yields approximately nine million advertisements. Second, we merge the OJV data with CERVED, at the firm level, to incorporate balance-sheet information for our sample of firms. Third, we link Lightcast to the administrative employer–employee compulsory communications archive (COB) from the Ministry of labor, which records all hiring events at plants headquartered in Piemonte and Veneto between 2010 and 2019 — roughly 1.6 million hires across 161,000 plants.

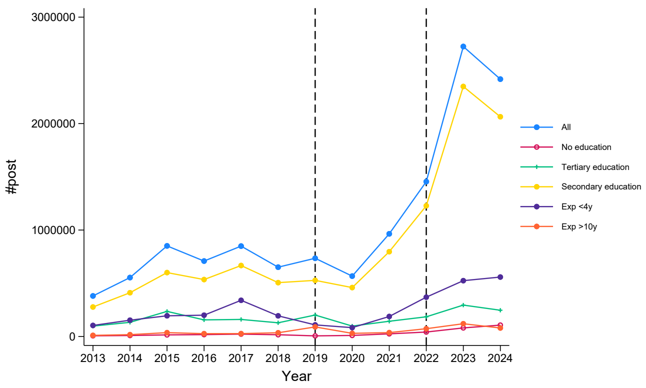
Job vacancy data. The primary data source is the universe of Online Job Vacancies (OJVs) for Italy, 2013–2024, provided by Lightcast. Restricting to postings with complete information on occupation, education, contract type, location, and duration yields approximately nine million postings associated with approximately 281,000 distinct firms. These data allow us to characterize, at a very granular local level, the skill and contractual requirements that firms signal to the labor market. Crucially, the panel dimension of the dataset — spanning over a decade and covering the near-universe of online postings in Italy — allows us to track how labor demand evolves over time and across space, up to the municipality of posting, documenting shifts in occupational composition, educational requirements, and contract flexibility. At the firm level, repeated observations over time allow us to distinguish secular trends in skill demand from cyclical fluctuations, and to identify whether changes in posted requirements reflect economy-wide shifts or firm-specific adjustments in recruitment strategy.

Merge with CERVED. We merge OJVs with CERVED, which provides harmonized accounting data — revenues, value added, capital stock etc.. — for the universe of Italian incorporated firms, conceptually equivalent to Orbis. The merge is at the firm level, as CERVED contains no plant-level information, and yields 2,011 unique firms and 12,066 firm–year observations over 2014–2019. This linkage allows us to examine whether vacancy posting behavior correlates with firm-level productivity, profitability, size, age, and growth, and whether it varies systematically across industries. Among these dimensions, firm size deserves particular emphasis. The COB archive records only hiring flows, with no information on the stock of incumbent workers; firm size can therefore not be recovered from the administrative hiring records alone. CERVED instead reports employment levels drawn from INPS (Italian Institute of Social Prevision) administrative records, providing the most precise measure of firm size available.

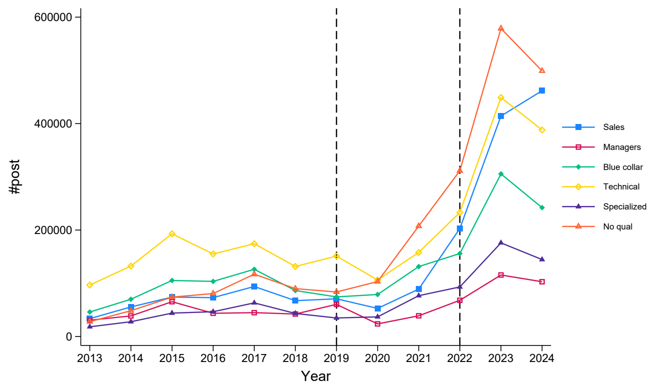
Merge with COB. The most novel feature of our research is the linkage between Lightcast OJVs and the administrative COB archive, covering all hirings at plants headquartered in Piemonte and Veneto between 2010 and 2019 (roughly 1.6 million hires across 161,000 plants). The match was performed by Lightcast using firm names; firms with generic or duplicate names were excluded, rendering the match virtually exact. The resulting sample comprises 13,530 plant–year observations with simultaneous vacancy and hiring information, corresponding to 2,041 distinct fiscal codes and 2,256 plants.[1] To our knowledge, this is the first dataset enabling a systematic plant-level comparison between advertised and realized hiring for two large, urbanised, and economically dense regions. The reduction relative to the full archives reflects four factors: only a minority of firms both post and hire within the same year; COB covers only Piemonte and Veneto; duplicate-name firms are excluded; and we restrict to direct recruiters, dropping staffing agencies. All analyses aggregate postings and hirings to the annual level. Annual aggregation minimizes spurious timing mismatches arising from the asynchronous nature of the two data sources.

Merge the three archives. For a final analysis, we merge the three archives by fiscal code, yielding an unbalanced panel of 12,066 firm–year observations across 2,011 firms. This triple-matched sample allows us to characterize baseline differences between firms that have balance-sheet records and hire workers, and firms that, in addition, post vacancies online at least once during the period. The richness of the merged dataset further allows us to condition on a broad set of firm- and local-level controls, capturing systematic differences in posting dynamics across and within industries and provinces. The analysis is presented in Section 4.2.

[1] Defined as standard in the literature as a tuple fiscal code by region.



(a) By Education and Experience.

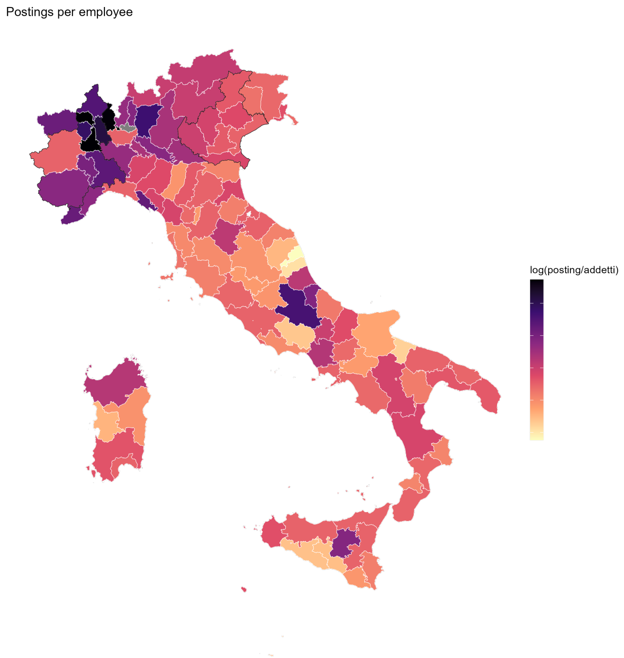


(b) By Occupation.

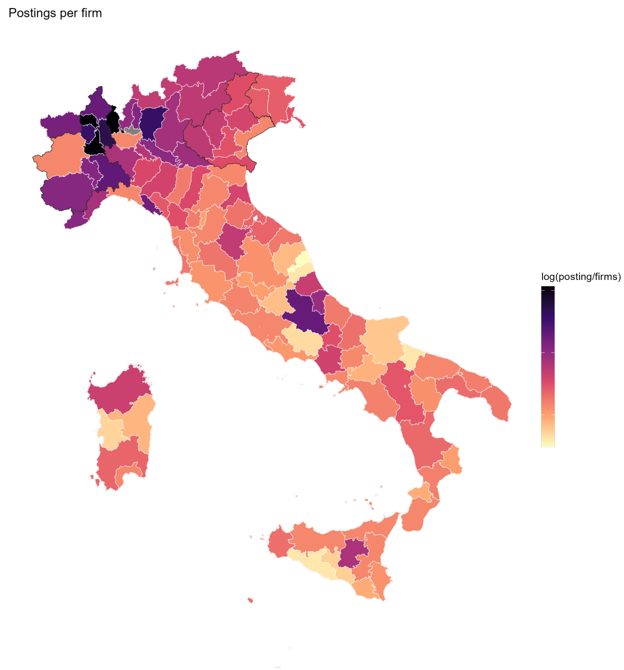
Figure 1: **OJVs trend by type.** *Notes:* The figure plots the annual number of online job postings in the Lightcast archive for Italy between 2013 and 2024. Panel (a) disaggregates postings by education requirement and experience level; panel (b) by broad occupational category. The vertical dashed lines indicate the COVID19 period. Elaborations were made on the dataset containing approximately 9 million job postings associated with 281,000 firms.

Figure 1 displays the evolution of online job postings between 2013 and 2024, by contract type, education level, and occupation. Posting activity grows moderately until 2019 and rebounds strongly after the COVID-19 shock, with particularly pronounced increases in full-time and medium-educated positions. Figure 2 displays the geographic distribution of online job postings across Italian municipalities and provinces over the same period. We normalize the number of OJVs by the number of employees (panel a) and firms (panel b) recorded in 2013. Posting activity is highly concentrated in Northern and urban areas, with a marked North–South divide, suggesting that online postings may reflect the geography of digitally intensive and economically stronger firms rather than the full territorial distribution of hiring activity. Importantly, the maps reveal that Piedmont and Veneto together account for a substantial share of total posting activity, even after normalizing by local employment and firm counts. This concentration suggests that, while our analysis is geographically restricted to these two regions, it nonetheless captures a large and representative segment of Italian online recruitment activity, particularly for the Northern industrial economy where OJV adoption is most widespread.

Table A.1 reports plant–year descriptive statistics for the matched Lightcast–COB sample. On average, plants publish approximately 8 online job postings per year, compared with roughly 20 hires recorded in the administrative data, suggesting that online postings capture only a fraction of realized hiring activity. Table A.2 reports descriptive statistics for the CERVED subsample. Firms display substantial heterogeneity across all financial dimensions, with means systematically exceeding medians, consistent with the right-skewed distribution of firm size and financial outcomes.



(a) #employees.



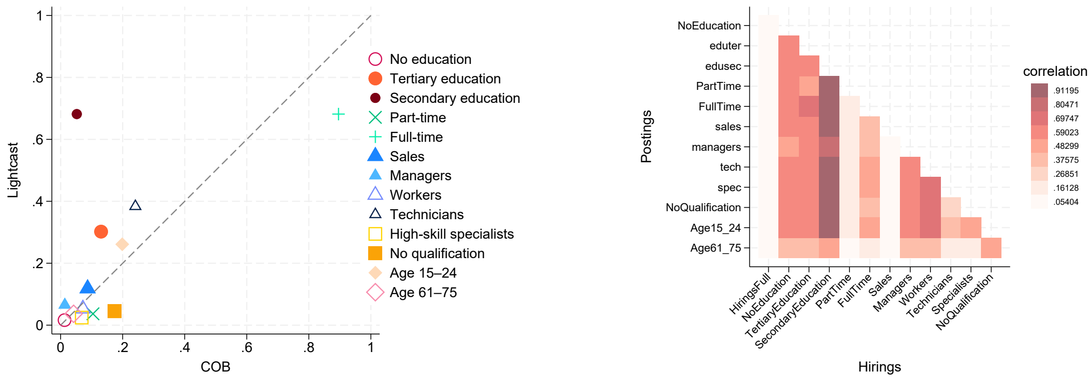
(b) #firms.

Figure 2: **Distribution of OJVs weighted by provinces' employees and firms.** *Notes:* Both employees and firms are measured at the province level as the total number across industries of employees and firms measured in 2013 and are drawn from the publicly available ASIA ISTAT archive. Maps plot the log transformation of each outcome. Elaborations were made on the dataset containing approximately 9 million job postings associated with 281,000 firms.

4 Research Design and Results

4.1 Comparing OJVs and COB

Figure 3a compares the percentage composition of hirings (COB, x-axis) and OJVs (Lightcast, y-axis) across education, occupation, contract type, and age categories. For each data source, we compute the share of each job category over total postings and total hirings respectively, and plot these compositional shares against each other. Points above the 45-degree line indicate categories over-represented in postings relative to realized hirings, while points below indicate the reverse. Full-time positions dominate both datasets but are roughly proportional, while part-time contracts are substantially under-represented in OJVs relative to their share in actual hirings. Among educational groups, secondary-educated positions are over-represented in postings, whereas tertiary-educated hirings account for a larger share of COB records than of OJVs. Occupationally, technical profiles and managers are relatively more visible in posting data, while sales workers, who account for the largest single occupational category in the COB data, are markedly under-represented in OJVs, suggesting that sales recruitment relies disproportionately on non-digital channels. Younger workers (age 15–24) are also more prominent in postings than in realized hirings, while the opposite holds for older age groups. These compositional differences suggest that online postings do not constitute a random draw from the universe of hiring activity, but instead reflect a selective representation of firm or recruitment strategies.



(a) Composition shares by job types.

(b) Raw correlations.

Figure 3: **Overview of (a) composition shares by job types, and (b) raw correlations between postings (Lightcast) and hirings (COB).** *Notes:* On the y-axis, there are the posting outcomes, while on the x-axis, there are the hiring outcomes. Results obtained from the dataset of 13,530 plant–year observations associated with 2,041 distinct fiscal codes and 2,256 plants.

Figure 3b examines the correspondence between postings and realized hires more formally, reporting raw correlations at the plant-year level for both aggregate and disaggregated job categories. At the aggregate level, total postings and total hires are strongly correlated, which may superficially suggest that OJVs are informative proxies for hiring activity. However, this relationship weakens sharply once disaggregated by education or occupation. Correlations remain positive and economically meaningful for high-skilled positions, but drop substantially for secondary-educated and low-skill jobs, with some categories displaying weak or even negative correlations. This pattern reveals that the aggregate correlation is largely driven by a subset of job types, and that proportionality between postings and hires does not hold uniformly across the labor market. Aggregation therefore masks substantial heterogeneity in the informational content of OJVs, with important implications for studies that use posting counts as a proxy for labor demand across occupational or educational groups.

4.2 Selection

We begin by examining selection into posting online job vacancies. We estimate the following linear probability model:

$$D_{jt} = \mathbf{X}_j'\boldsymbol{\beta} + \gamma_s + \gamma_p + \varepsilon_{jt} \tag{1}$$

where $D_{jt} \in \{0, 1\}$ indicates whether plant j posts at least one online vacancy in year t, $\mathbf{X}_j$ is a vector of standardized firm-level financial characteristics, profitability measures, and size, γ_s and γ_p

denote 2-digit industry and province fixed effects respectively, and standard errors are clustered at the firm level. Figure 4 reports the estimated $\boldsymbol{\beta}$. Larger and more productive firms are significantly more likely to advertise online, as are younger firms: the right panel shows that each additional year of age reduces the probability of posting by a small but precisely estimated margin. Profitability measures (ROS, ROI, ROE, ROA) display weaker and less precisely estimated associations. These patterns are consistent with the diffusion of digital recruitment technologies among more dynamic and organizationally sophisticated firms, and indicate that OJVs represent a positively selected subset of hiring firms rather than the full population of employers.

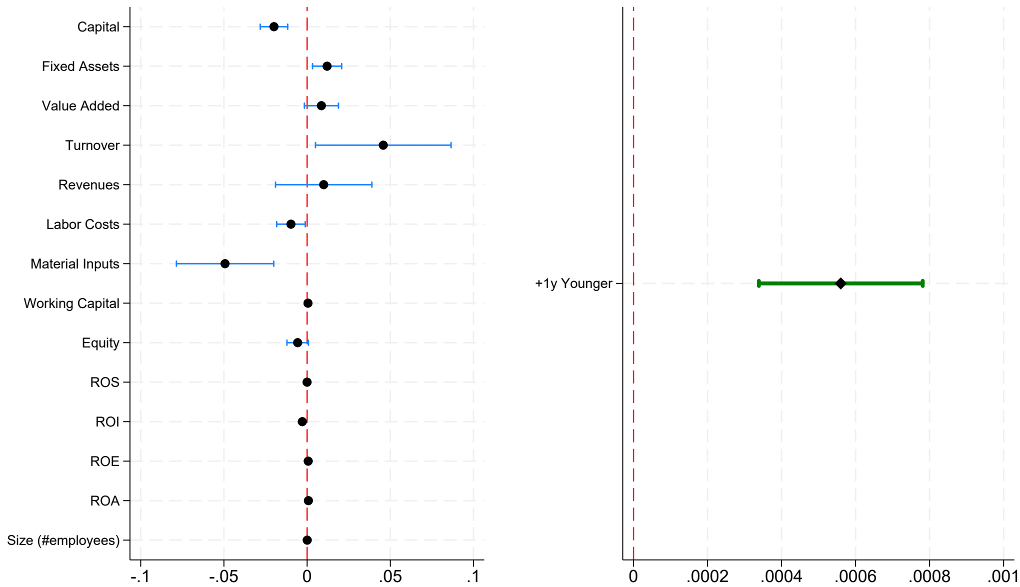
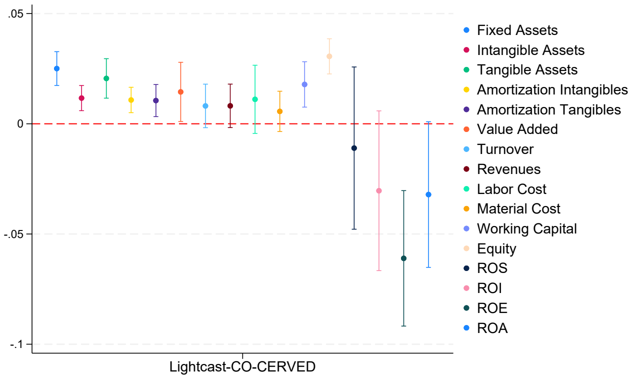


Figure 4: Likelihood to post OJVs by firms' records. *Notes:* The figure reports OLS estimates from linear probability models in which the dependent variable indicates whether a plant posts at least one online vacancy in a given year. The left panel reports coefficients on standardized financial outcomes; the right panel reports the coefficient on firm age. Results were obtained from the dataset of 2,011 unique firms and 12,066 firm–year observations. over 2014–2019. 95% confidence intervals displayed. Firm-level clustered standard errors.

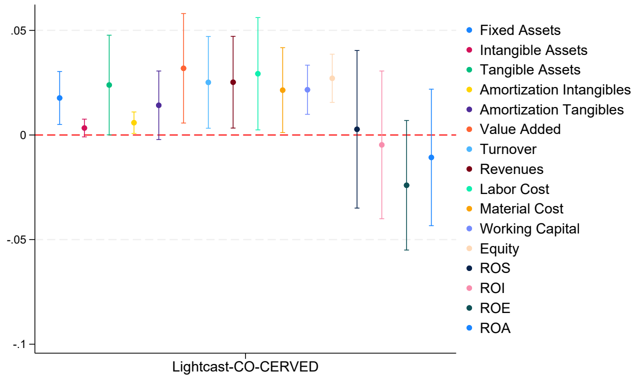
We further assess whether firms appearing only in the administrative data differ systematically from firms observed across all three archives. For each standardized financial outcome Y_j, we estimate:

$$Y_j = \delta S_j + \varepsilon_j \tag{2}$$

where $S_j \in \{0, 1\}$ indicates whether firm j appears in the matched Lightcast–COB–CERVED sample, with $S_j = 0$ denoting firms present only in COB–CERVED, and standard errors are clustered at the plant level. Figure 5 reports $\hat{\delta}$ for each outcome, without controls in panel (a) and controlling for province-by-year trends and 2-digit industry fixed effects in panel (b). The differences are statistically significant for nearly all financial outcomes, confirming that firms in the matched sample are systematically larger, more capitalized, and more productive than firms present only in the administrative records. Importantly, these differences persist after the inclusion of fixed effects, indicating that selection into online posting reflects firm-level heterogeneity beyond sectoral or geographic sorting. Taken together, these findings imply that the universe of OJVs captures a positively selected segment of the firm population, with direct implications for the external validity of empirical applications that rely on posting data as a proxy for aggregate labor demand.



(a) Baseline.



(b) Industry and province by year fixed effects.

Figure 5: Balance test of posting vs. non-posting firms' outcomes. *Notes:* All outcomes have been standardized to have a mean of 0 and an SD of 1. Panel (a) displays the baseline differences between firms appearing in all three archives (Lightcast–COB–CERVED) and those appearing only in COB–CERVED. Panel (b) displays the same differences controlling for province-by-year trends and 2-digit industry (ATECO code) fixed effects. Estimates based on 879,540 firm-year tuples and 146,590 distinct firms in the CO-CERVED archives from Piedmont and Veneto, of which 12,066 firm-year tuples and 2,011 fiscal codes exist in the three archives simultaneously. Firm-level clustered standard errors. 95% confidence intervals displayed.

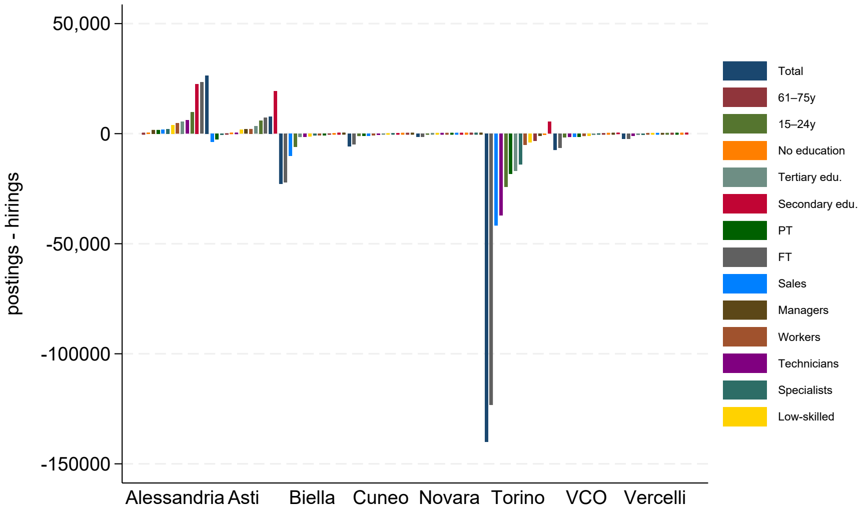
4.3 Demand Gap

Figure 6 plots the distribution of the Demand Gap at the plant-year level, disaggregated by province and job type. The distribution is highly dispersed and features a negative mean, indicating that many firms hire workers without advertising the corresponding positions online. At the same time, the right tail of the distribution reveals that a non-trivial share of plants posts far more vacancies than they ultimately fill. These are not small deviations around zero but systematic and persistent differences that vary across firms, provinces, and occupational categories. Taken together, this evidence points to the coexistence of multiple recruitment channels and heterogeneous posting strategies, and cautions against interpreting raw posting counts as direct measures of labor demand.

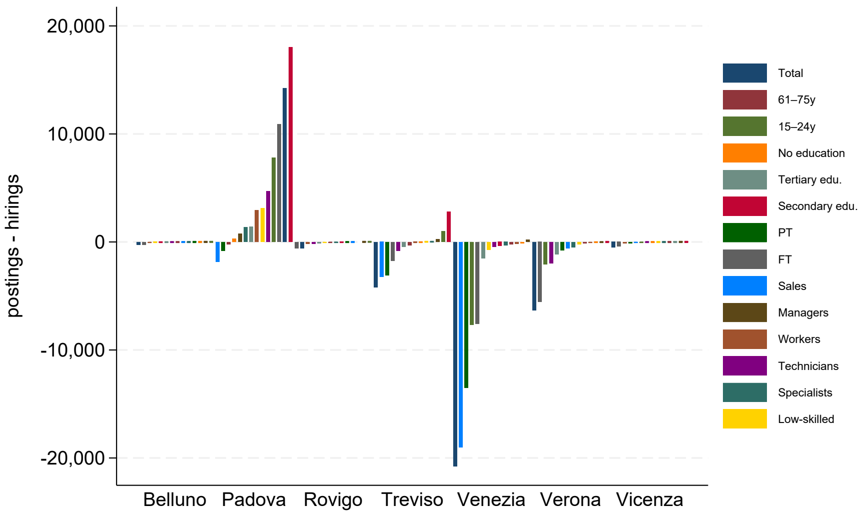
We match postings and hirings at the job-type level and define Demand Gap_{jt}=$Postings_{jt}$-$Hirings_{jt}$, comparing OJVs requiring a given qualification with hirings of workers with the corresponding qualification. Figure 7 reports the probability of posting at least one online ad and of having a positive demand gap (conditional on posting) by industry and province, using Torino as the baseline. A positive demand gap indicates that a firm posts vacancies but does not hire within the same year. Formally, we estimate:

$$Y_{it} = \alpha_s + \delta_p + X'_{m(i)t}\beta + P'_{m(i)t}\gamma + \varepsilon_{it} \tag{3}$$

, where $Y \in \{0, 1\}$ indicating if the firm does not post and either the firm posts but does not hire. i indicates the plant, t the year, and $m(i)$ the posting municipality of plant i. α_s and δ_p are industry 1d and province fixed effects and $X'_{m(i)t}$ and $P'_{m(i)t}$ are municipality-by-year-level covariates describing population by sex and cohorts and economic variables we employ as controls.[2]



(a) Piedmont.



(b) Veneto.

Figure 6: **Mismatch across provinces, by region.** *Notes:* Bars indicate the total postings gap, in number, defined as postings minus hirings, across provinces over time. Starting from the local unit-by-year-level panel dataset, we collapse it to the province level by summing all outcomes across firms and years. VCO indicates the province of Verbania-Cusio-Ossola. Results obtained from the dataset of 13,530 plant–year observations associated with 2,041 distinct fiscal codes and 2,256 plants.

[2] They are number of taxpayers, income, income per capita, and number of citizenship income demands to capture local labor market characteristics.

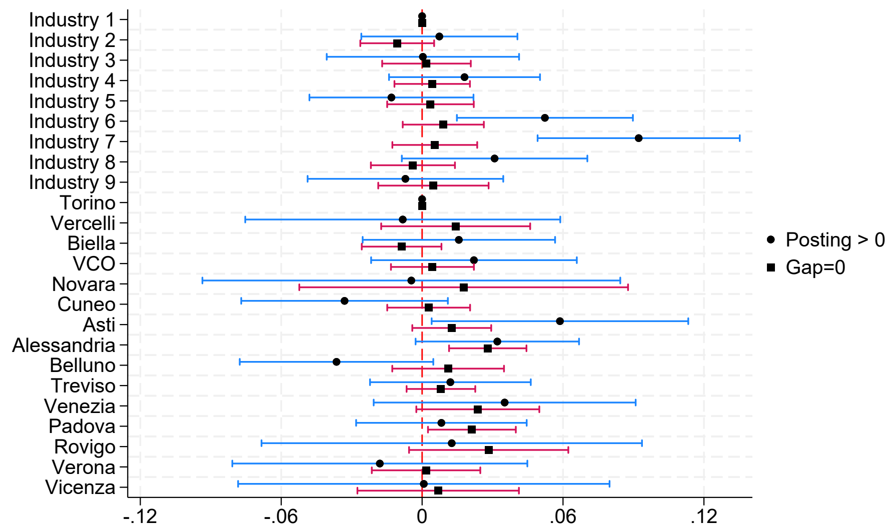


Figure 7: **Posting and Gap Likelihood.** *Notes:* Baselines are respectively Industry 1 and Torino. We control for municipality by year-level controls for population and economic dynamics. 95% confidence interval displayed. Firm-level clustered standard errors are estimated. VCO indicates the province of Verbania-Cusio-Ossola. Results obtained from the dataset of 13,530 plant–year observations associated with 2,041 distinct fiscal codes and 2,256 plants.

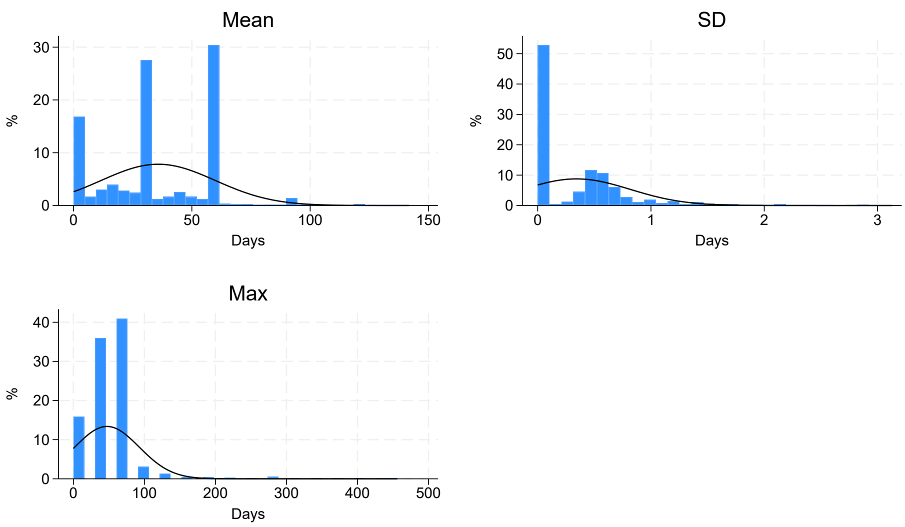


Figure 8: OJVs Duration. *Notes:* The figure plots the distribution of posting durations (in days) for all ads in the Lightcast archive active between 2013 and 2019. The three panels report respectively the distribution of the mean, standard deviation, and maximum posting duration computed at the firm-year level. Elaborations were made on the dataset containing approximately 9 million job postings associated with approximately 2 million firm-year tuples and 281,000 firms between 2013 and 2019.

4.3.1 Ghost Jobs

A particularly striking manifestation of the demand gap is the phenomenon of so-called ghost jobs[3], postings that remain active over extended periods without generating corresponding hiring flows.

[3] *Beware of 'ghost job' listings*, 2022, Financial Times.

Rather than reflecting genuine vacancy creation, such postings may serve strategic purposes: maintaining employer visibility, building applicant pools, or signaling organizational dynamism to the market. Figure 8 reports the distribution of posting durations. While the average posting remains online for approximately forty days, the distribution is heavily right-skewed, with some postings persisting for more than one thousand days. Such extreme durations are difficult to reconcile with standard interpretations of posting length as a measure of job-filling difficulty, and suggest instead that duration reflects heterogeneous recruitment practices and communication strategies that go beyond job characteristics alone. Moreover, Figure A.1 illustrates the demand gap in detail, breaking it down by educational level, work experience, job type, and occupation.

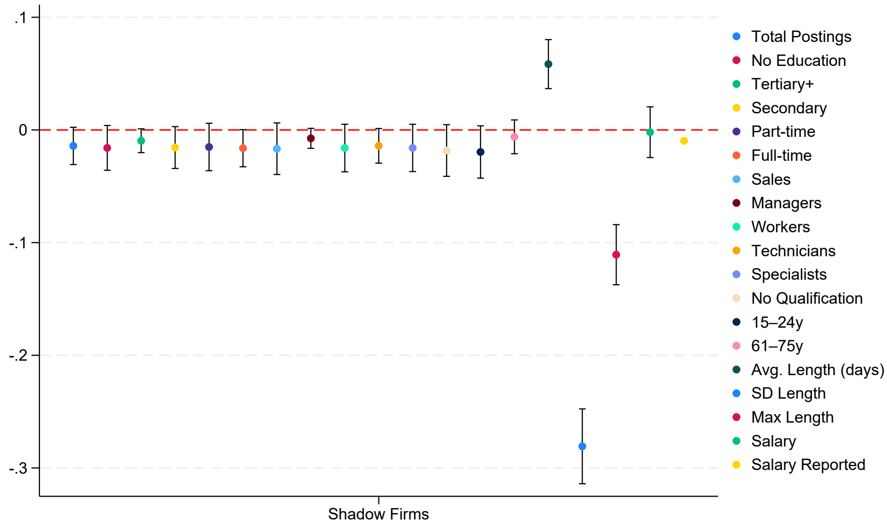


Figure 9: Baseline differences in postings between shadow and non-shadow firms. *Notes:* Shadow firms are those firms appearing in the Lightcast archive between 2013 and 2019 but never in the COB records — i.e., firms that post online but never register a hiring event. All outcomes are standardised to have mean 0 and SD 1. Results were obtained from a dataset of 392,189 firm-year tuples associated with 57,596 firms, of which 54,336 plant-year, 8,839 firms, and 9,056 plants also post online. 95% confidence intervals displayed. Firm-level clustered standard errors.

To formally characterize shadow firms, we estimate:

$$Y_j = \delta \, \text{Shadow}_j + \varepsilon_j \tag{4}$$

where y_j denotes a standardized posting outcome (as total posting volume, disaggregated counts by job type, average posting duration, and advertised wages), $\text{Shadow}_j \in \{0, 1\}$ indicates firms appearing in the Lightcast archive between 2013 and 2019 but never in the COB records, i.e. firms that post online but never register a hiring event, and standard errors are clustered at the firm level. Figure 9 reports the estimated $\hat{\delta}$ for each outcome. Shadow firms post significantly fewer ads across all job categories and advertise substantially lower wages. The pattern on posting duration is more nuanced: while shadow firms display longer average posting durations, the standard deviation and maximum duration are significantly lower, suggesting that their posting activity is more repetitive and standardized rather than reflecting persistent unfilled vacancies. Taken together, this evidence indicates that shadow firms engage with online platforms in a qualitatively different way, using postings as an occasional signaling device rather than as an active and sustained recruitment tool.

5 Conclusion

This paper provides new evidence on what online job vacancies (OJVs) measure, exploiting a unique empirical setting: the first, to our knowledge, plant-level linkage between the universe of Lightcast postings and the universe of mandatory employer-employee records (COB) for firms headquartered in Piemonte and Veneto over 2014-2019. Our findings suggest that OJVs are an informative but imperfect proxy for labor demand, whose reliability varies systematically across contexts and job types.

First, selection into online posting is substantial and non-random: larger, younger, and more productive firms are significantly more likely to advertise online, and this selection persists after controlling for industry and geographic sorting. OJVs therefore represent a positively selected subset of hiring firms and do not capture the full population of employers.

Second, the compositional structure of postings diverges systematically from that of realized hirings. While aggregate posting counts correlate strongly with total hires, this relationship breaks down once disaggregated by education, occupation, or contract type, with some categories displaying weak or negative correlations. OJVs are considerably more informative for high-skill positions than for low-skill or routine jobs, and aggregation masks this heterogeneity.

Third, the demand gap (defined as the difference between posted vacancies and realized hires at the plant-year level) is large, persistent, and heterogeneous across firms, provinces, and job types. The presence of shadow firms, which post online but never register hiring events, and the prevalence of unusually long posting durations further suggest that a non-trivial share of posting activity reflects signaling, talent-pooling, or visibility strategies rather than contemporaneous labor needs.

Taken together, these results imply that OJVs capture a combination of labor demand and firm recruitment strategies, and should not be interpreted as direct measures of vacancy creation. For applied work, the implications are clear: OJVs can be informative when studying trends, high-skill demand, or within-firm variation, but may be misleading when used to measure levels of labor demand or to compare across heterogeneous groups. In such cases, combining posting data with administrative hiring records remains the most reliable approach to assessing what online vacancies actually measure in practice.

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acquisti, A. and Fong, C. (2019). An experiment in hiring discrimination via online social networks. *Management Science*, 66(3):1005–1024.

Antonie, L., Gatto, L., Oloumi, S., and Plesca, M. (2024). Personality characteristics demanded by employers: Analysis of job descriptions from university job boards. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 213–255. Emerald Publishing Limited.

Atalay, E., Sotelo, S., and Tannenbaum, D. (2024). The geography of job tasks. *Journal of Labor Economics*, 42(4):979–1040.

Azar, J., Marinescu, I., and Steinbaum, M. (2022). Labor market concentration. *Journal of Human Resources*, 57(S):S167–S199.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745.

Bratti, M., Passerini, F., and Maida, A. (2026). Firm recruitment under guaranteed minimum income: Evidence from italy's online job vacancy data. *in progress*.

Chaturvedi, S., Mahajan, K., and Siddique, Z. (2024). Using domain-specific word embeddings to examine the demand for skills. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 345–381. Emerald Publishing Limited.

Choi, H. and Marinescu, I. (2024). The labor demand side of involuntary part-time employment. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 111–147. Emerald Publishing Limited.

Deming, D. J. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Deming, D. J. and Noray, K. L. (2020). Earnings dynamics, changing job skills, and stem careers. *Quarterly Journal of Economics*, 135(4):1965–2005.

Escudero, V., Liepmann, H., and Podjanin, A. (2024). Using online vacancy and job applicants' data to study skills dynamics. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 257–299. Emerald Publishing Limited.

Faberman, R. J. and Kudlyak, M. (2019). The intensity of job search and search duration. *American Economic Journal: Macroeconomics*, 11(3):327–357.

Fabo, B., Beblavý, M., and Lenaerts, K. (2017). The importance of foreign language skills in the labour markets of central and eastern europe: Assessment based on data from online job portals. *Empirica*, 44(3):487–508.

Fabo, B. and Kureková, L. M. (2022). Methodological issues related to the use of online labour market data. Technical Report ILO Working Paper 68, International Labour Organization, Geneva.

Forsythe, E. C., Kahn, L. B., Lange, F., and Wiczer, D. (2020). Labor demand in the time of covid-19: Evidence from vacancy postings and ui claims. *Journal of Public Economics*, 189:104238.

Galassi, G., Bellatin, A., and Chu, V. (2024). Letting job postings talk: Recent trends in digitalization. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 301–344. Emerald Publishing Limited.

Grasso, G. and Tatsiramos, K. (2023). The impact of restricting fixed-term contracts on labor and skill demand. Technical Report IZA Discussion Paper No. 16496, Institute of Labor Economics (IZA), Bonn.

Hershbein, B. and Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–1772.

Joint Research Centre, European Commission (2024). Online job ads: Promising yet biased data for employment and education policy. Technical report, Joint Research Centre, European Commission.

Kuhn, P. and Shen, K. (2013). Gender discrimination in job ads: Evidence from china. *Quarterly Journal of Economics*, 128(1):287–336.

Kuhn, P., Shen, K., and Zhang, S. (2018). Gender-targeted job ads in the recruitment process: Evidence from china. Technical Report Working Paper No. 25365, National Bureau of Economic Research, Cambridge, MA.

Lovaglio, P. G. and Mezzanzanica, M. (2026). Analyzing non-random selectivity in online job advertisements using eurostat benchmark data and generalized sample selection models: An application to eu regional labor markets. *Labour*, 0:1–31.

Marinescu, I. (2016). The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. Technical Report Working Paper No. 22447, National Bureau of Economic Research, Cambridge, MA.

Marinescu, I. and Rathelot, R. (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, 10(3):42–70.

Marinescu, I. and Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, 38(2):535–568.

Modestino, A. S., Shoag, D., and Ballance, J. (2016). Downskilling: Changes in employer skill requirements over the business cycle. *Labour Economics*, 41:333–347.

Morales, L. F., Ospino, C., and Amaral, N. (2024). Online vacancies and their role in labor market performance. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 45–79. Emerald Publishing Limited.

OECD (2024). How well do online job postings match national sources? benchmarking lightcast data against official vacancy statistics across regions, sectors and occupations. Technical Report 2024/01, Organisation for Economic Co-operation and Development, Paris.

Shen, K. and Zhu, Y. (2024). Labor force transition dynamics: Unemployment rate or job posting counts? In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 3–44. Emerald Publishing Limited.

Tzimas, G., Zotos, N., Mourelatos, E., Giotopoulos, K. C., and Zervas, P. (2024). From data to insight: Transforming online job postings into labor-market intelligence. *Information*, 15(8):496.

Ziegler, L. (2024). Skill demand and wages: Evidence from linked vacancy data. In *Big Data in Labor Economics*, volume 52 of *Research in Labor Economics*, pages 149–183. Emerald Publishing Limited.

A Additonal Tables and Figures

Table A.1: Descriptive Statistics in the Lightcast-CO matched sample.

	N	Mean	SD	Median	Min	Max
Online Job Postings (OJV) *Source*: Lighcast						
all	13530	8.104508	110.5637	0	0	4277
no education	13530	.1348115	2.637989	0	0	147
tertiary	13530	1.724538	23.38801	0	0	872
secondary	13530	6.245159	86.82819	0	0	3310
PT	13530	.6634146	9.496338	0	0	358
FT	13530	6.568884	96.89411	0	0	3926
sales	13530	.7275684	9.928623	0	0	419
managers	13530	.5294161	6.502902	0	0	230
workers	13530	1.082483	19.05141	0	0	761
technicians	13530	1.8915	25.88729	0	0	915
specialized	13530	.5133777	9.357551	0	0	403
no qualification	13530	.9998522	17.22103	0	0	751
15-24y	13530	2.692239	40.03897	0	0	1798
61-75y	13530	.1151515	1.301153	0	0	60
Hirings *Source*: Comunicazioni Obbligatorie (CO)						
all	13530	20.21715	185.0478	0	0	6225
no education	13530	.1519586	2.002873	0	0	169
tertiary	13530	2.693792	23.76993	0	0	924
secondary	13530	1.240872	11.57166	0	0	464
PT	13530	3.676866	48.35727	0	0	2188
FT	13530	16.54028	172.5056	0	0	6188
sales	13530	6.735994	105.5926	0	0	4546
managers	13530	.2538064	5.695818	0	0	342
workers	13530	1.003917	17.78992	0	0	1063
technicians	13530	4.422321	64.06274	0	0	2592
specialized	13530	1.547081	12.72849	0	0	457
no qualification	13530	.9335551	11.03796	0	0	531
15-24y	13530	4.087583	48.74329	0	0	2328
61-75y	13530	.5181079	6.901652	0	0	524

Table A.2: Descriptive Statistics on the CERVED-CO-Lighcast matched sample.

Variable	**N**	**Mean**	**SD**	**Median**	**Min**	**Max**
Capital	10686	9443365	1.42e+08	278769.5	0	6.43e+09
Intangible Assets	10723	1.68e+07	2.74e+08	284888	0	1.21e+10
Total Intangibles	10723	1737059	2.55e+07	18819	0	9.08e+08
Tangible Assets	10723	6533739	1.07e+08	140111	0	5.10e+09
Intangible Amortization	10686	316371.4	4941546	5782	0	1.77e+08
Tangible Depreciation	10686	829460.7	1.48e+07	22833.5	0	7.27e+08
Value Added	10722	5966014	4.94e+07	611144	-4962000	2.27e+09
Sales	10723	3.61e+07	6.09e+08	1753662	-1260439	2.86e+10
Operating Revenue	10723	3.61e+07	6.09e+08	1752378	0	2.86e+10
Labor Costs	10723	3826638	3.92e+07	394375	0	1.62e+09
Material Costs	10723	2.42e+07	5.15e+08	366290	0	2.46e+10
Net Working Capital	10723	1.47e+07	1.73e+08	1092085	321	8.03e+09
Equity	10723	1.21e+07	1.63e+08	329890	-1.22e+08	9.71e+09
Return on Sales (ROS)	10602	.1679117	66.77445	4.05	-1000	1000
Return on Investment (ROI)	10711	5.349234	21.43475	4.94	-757.94	140.34
Return on Equity (ROE)	10439	5.62208	78.02194	10.28	-1000	1000
Return on Assets (ROA)	10718	5.996069	20.44523	4.805	-1000	105.1

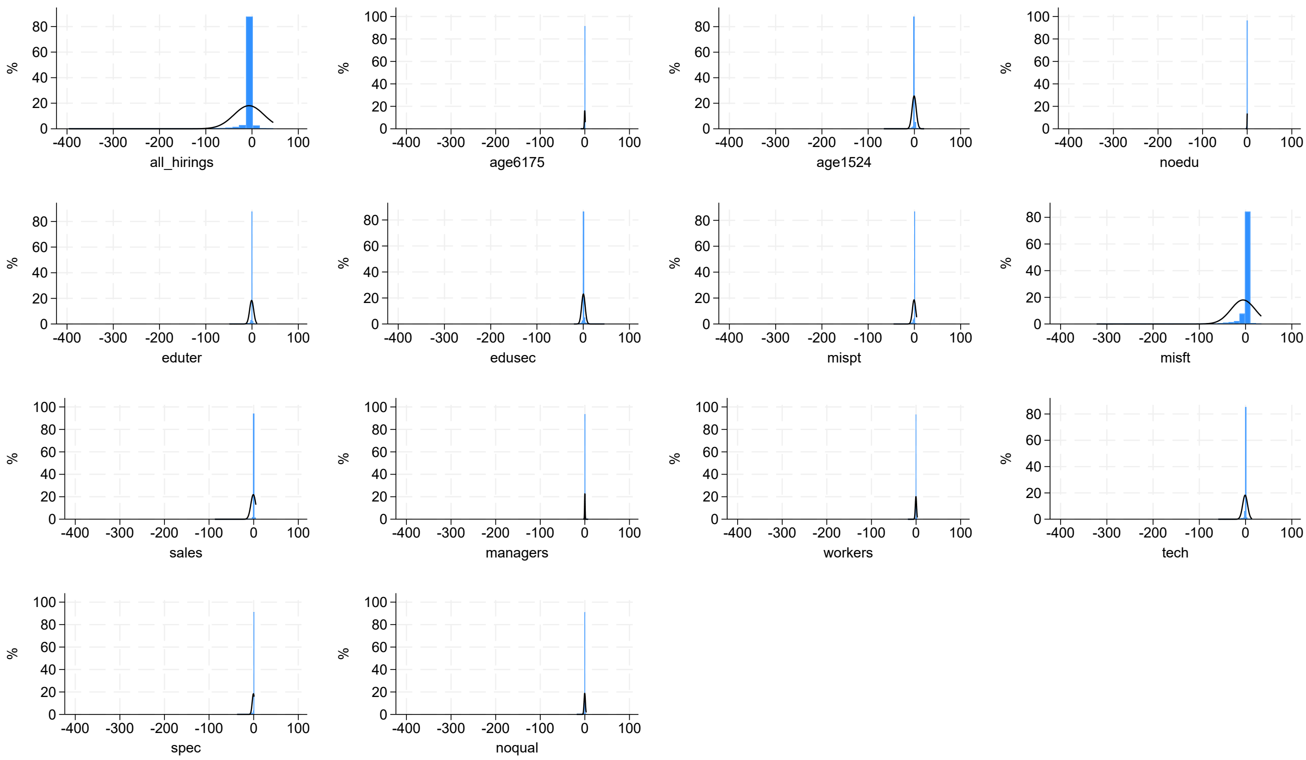


Figure A.1: Histograms of demand gaps by educational level, experience, job type, and occupation.
Notes: Gap is computed as define Demand Gap$_{jt}$=$Postings_{jt}$-$Hirings_{jt}$ at the plant j and year t level.