

Do Online Job Postings Capture Job Vacancies?*

Michael R. Dalton
Bureau of Labor
Statistics

Lisa B. Kahn
University of Rochester,
NBER

Andreas I. Mueller
University of Zurich,
CEPR and IZA

August 22, 2026

Abstract

We ask whether online job postings capture U.S. job vacancies, matching near-universal Burning Glass (BG) postings to the representative Job Openings and Labor Turnover Survey (JOLTS). BG converged toward JOLTS over 2007–2022, and its representativeness varies across establishment characteristics. We reweight BG to align postings with JOLTS openings along observables, and assess impacts on prior BG findings. Because large establishments account for most openings and are well represented, the level and cyclicalities of skill demand survive our adjustment. But BG overstates labor market concentration (the Herfindahl-Hirschman Index) nearly threefold, because small and mid-sized establishments post less online.

JEL Classification: E24, E32, J23, J63, J42

Keywords: job vacancies, labor demand, online job postings, vacancy measurement, labor market concentration, skill demand

*We thank Serdar Birinci, Jason Faberman, Claudia Macaluso, Christina Patterson and seminar participants at the AEA Annual Meeting, the New York and Richmond Feds, and the TCMB Conference on Recent Developments in Macro Labor for helpful comments. We are especially indebted to Matt Sigelman and Bledi Taska for providing the Burning Glass data. All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

1 Introduction

Job vacancies play a central role in the theory and measurement of the state of the labor market. The job vacancy rate is an important gauge of firms’ labor demand and its relationship to the unemployment rate is the key determinant of job-finding rates in the canonical search-and-matching model. Increasingly, firms have posted their vacancies online, leading to the development of aggregate indicators based on online job postings as well as a surge in research activity using the underlying micro data. A key advantage of online job vacancy data is the rich information it contains, such as that on skill requirements and posted wages, enabling analysis of firms’ labor demand in unprecedented detail. At the same time, little is known about the representativeness of online job ads data.

We link online job postings to representative data on job vacancies at the establishment level to understand alignment of the two series. We use the near-universal U.S. database of online job ads Burning Glass (BG), now collected by Lightcast, and the Job Openings and Labor Turnover Survey (JOLTS) administered by the Bureau of Labor Statistics. We document differential representation of BG postings and JOLTS openings over time and across establishment characteristics. We find BG data converged to the JOLTS series over our study period of 2007 and 2010 to 2022, with the largest changes occurring early in the sample period. We also document substantial heterogeneity in representativeness across industries, locations, and establishment wage and size. We reweight BG to align postings with the representative JOLTS openings series along establishment characteristics, and examine how conclusions previous researchers have drawn from BG data are impacted by our adjustment.

To conduct these analyses, we first link BG postings to business registry data for the U.S. (the Quarterly Census of Employment and Wages, QCEW), matching 91.6% of postings with an identifiable employer name and location to a QCEW establishment. This linkage provides characteristics for BG posting establishments, such as size and average wage, which were previously unknown. It also allows us to link to the JOLTS, a representative establishment-level survey on vacancies and hires. To compare to BG, we define the flow of JOLTS openings in a month as the number of hires plus the net of new unfilled vacancies over the month.

We develop a conceptual framework that relates BG postings to JOLTS openings through a multiplicative adjustment factor comprising: (1) an extensive margin – the probability of posting a job vacancy online, (2) an intensive margin – the number of vacancies captured by one online job posting, and (3) a nuisance factor – the share of postings that originate from establishments that actually have JOLTS openings. The adjustment factor and its components provide a useful lens for describing variation in the postings-openings relationship over time and across establishment-level characteristics – variation that could not be examined with BG data alone. We also propose a prediction method that reweights BG postings to generate moments that are representative of JOLTS openings. This prediction lets us move beyond the BG-JOLTS merge to the full business registry (QCEW, which also contains all the model covariates) and address questions for which JOLTS is underpowered.

Our analysis of the matched micro-data reveals several facts. First, representation of BG improved steadily from 2007 to 2022, most sharply in the earliest years – the adjustment factor fell three-fold between 2007 and 2011, then declined more gradually from about 10 in 2011 to 4 by 2019. Convergence was driven primarily by the extensive margin, the likelihood that an establishment with openings posts them online. Over this more stable 2011-19 period, which coincides with many applications of the BG data, the average adjustment factor is 6.2: postings from the average establishment must be multiplied by 6.2 to match JOLTS openings, with larger values indicating greater misalignment.

Second, the adjustment factor decreases strongly in an establishment’s average wage and urbanicity:

higher-paying establishments and urban areas are better represented online. It also decreases, more weakly, in employment size – a relationship central to the postings-openings link. Large establishments are well represented online and account for most job openings, while small and mid-sized establishments are less well represented but account for most establishments; BG thus captures the distribution of openings well but not the distribution of hiring establishments.

We then apply our adjustment factor to revisit two empirical patterns previously studied with BG data. First, researchers have used BG to measure the skill requirements of postings, though BG is thought to overrepresent skilled professions. Reweighting by our adjustment factor does lower measured skill requirements, but only modestly, since most openings come from large, well-represented establishments. We also replicate [Hershbein and Kahn \(2018\)](#), who show that skill requirements related to automation rose after the Great Recession. Despite the fact that they benchmark against a single pre-recession year (2007), when BG representation was changing rapidly, we find their results remain similar after reweighting.

Second, researchers have used the richness of BG to study labor market concentration at highly disaggregated geographic and occupational levels ([Azar et al., 2020](#); [Hershbein et al., 2022](#); [Azar et al., 2024](#); [Schubert et al., 2025](#)) – work that has helped shift academic and policy debates toward employer market power. Using a Herfindahl-Hirschman Index (HHI), [Azar et al. \(2020\)](#) find that local labor markets are surprisingly concentrated (an index above 2500). Applying our adjustment factor, we find that postings-based HHIs overstate the concentration of job openings nearly threefold, because small and medium establishments are underrepresented. Rankings across markets, however, are preserved well after adjusting. Focusing on the level of concentration is therefore ill-advised without our adjustment, though ordinal comparisons across markets remain reliable.

Our paper relates to several literatures. Vacancy rates play a central role in canonical search and matching models ([Pissarides, 1985](#); [Mortensen and Pissarides, 1994](#)) and a long empirical literature on the matching function, worker flows over the cycle, and the Beveridge curve ([Petrongolo and Pissarides, 2001](#); [Lange and Papageorgiou, 2020](#); [Abraham, 1987](#); [Blanchard and Diamond, 1989](#); [Shimer, 2005, 2007](#); [Şahin et al., 2014](#); [Elsby et al., 2015](#)). [Davis et al. \(2013\)](#) leveraged JOLTS micro data to study recruiting intensity at the firm level, but JOLTS as a survey is not powered to analyze labor market dynamics at fine levels of disaggregation.¹ Online job listings data, particularly BG, have risen in popularity and helped fill this gap, supporting research on demand for skill ([Deming and Kahn, 2018](#); [Braxton and Taska, 2023](#); [Hemelt et al., 2023](#); [Acemoglu et al., 2022](#); [Hershbein and Kahn, 2018](#)), job attributes like work-from-home ([Hansen et al., 2023](#)), public policy impacts ([Dillon et al., 2025](#); [Clemens et al., 2021](#); [Azar et al., 2024](#); [Callaci et al., 2024](#)), market concentration ([Azar et al., 2020, 2022](#); [Hershbein et al., 2022](#); [Schubert et al., 2025](#)), and rapid cyclical movements during the COVID pandemic ([Forsythe et al., 2020, 2022](#)).²

While online ads track aggregate vacancy *trends* well ([Carnevale et al., 2014](#)) – a comparison that pits an ad flow against the JOLTS opening stock – the same studies note that ads overrepresent skilled, high-wage positions and capture only a subset of openings. To date, comparisons to representative vacancy data have been possible only at coarse levels of aggregation (e.g., one-digit NAICS), and little is known about the composition of establishments represented in BG. Our BG-JOLTS linkage offers a great deal of insight by characterizing the relationship between online postings and representative vacancy data. We conclude that BG is apt for many research questions related to the distribution of openings in the economy. However, a

¹[Barnichon \(2010\)](#) observes that the Help Wanted Index, available for earlier periods, closely tracks aggregate movements in JOLTS. See also [Mongey and Violante \(2026\)](#).

²For online job board data from other sources and countries, see [Barron et al. \(1997\)](#), [Banfi and Villena-Roldan \(2019\)](#), [Marinescu \(2017\)](#), [Marinescu and Wolthoff \(2020\)](#), [Mueller et al. \(2024\)](#), and [Mueller et al. \(2025\)](#).

key finding urges caution for an active literature using online job vacancy data to measure the level of labor market concentration.

2 Data Sources and Matching Procedure

We use three data sources. Online job postings come from BG, now Lightcast. We match these at the employer-location level to two Bureau of Labor Statistics datasets: the QCEW, the business registry that serves as the sampling frame for all BLS establishment surveys, and the JOLTS, a survey of employers' hiring needs. These matches let us identify establishment characteristics absent from BG and compare BG posting behavior to self-reported vacancies and hires.

Online Job Postings. BG collects online job postings on a rolling basis by scraping over 40,000 websites and uses proprietary machine-learning to code fields such as location, industry, and skill requirements and to de-duplicate ads. We use data for 2007 and 2010 to 2022.³ We measure the flow of new postings from BG, though the number of jobs a posting represents is unclear: an employer may seek many hires from one ad, or may keep a semi-permanent posting up (e.g., high-churn fast-food) even after hiring.⁴ We make progress on this question here.

QCEW. The QCEW is derived from wage records, covers more than 95% of all jobs and is the sampling frame for all BLS establishment surveys. For each establishment it gives the address, NAICS industry, monthly employment, the quarterly wage bill, and, through tracking over time, establishment age – characteristics absent from BG. Merging to QCEW also links BG to any other BLS survey.

JOLTS. JOLTS is a monthly survey of about 16,000 establishments reporting end-of-month vacancies, monthly hires and separations, and employment, with a panel component in which most establishments are surveyed for 24 consecutive months. As internal BLS datasets, QCEW and JOLTS link easily; being a survey, JOLTS does not cover all QCEW establishments or all establishments represented in BG.

We compare establishment responses on vacancies to BG postings, which requires aligning the two definitions of openings. JOLTS defines openings as “*all positions that are open (not filled) on the last business day of the month*” that meet three requirements: a specific position with work available exists, the job can start within 30 days, and there is active recruiting (which includes posting ads online). JOLTS also records hires – “*all additions to payroll during a month,*” including recalls – which need not meet the opening requirements, so a hire may occur without active recruiting or with a long lag between posting and start date.

JOLTS thus measures a *stock* of open positions at a point in time, whereas BG measures a *flow* of new postings. To compare the two, we use the law of motion of unfilled positions to infer the flow of new openings for establishment i in month t :

$$N_{i,t} = J_{i,t} - J_{i,t-1} + H_{i,t}, \quad (1)$$

where J is end-of-month unfilled vacancies and H is hires over the month. Because some hires occur without a vacancy, N overstates the true flow of openings, while J alone understates it (counting only positions that remain unfilled). Hereafter we call J the JOLTS *vacancy* and N the JOLTS *opening*. We compare BG

³There is a gap for the years 2008 and 2009.

⁴To limit the latter, BG treats any ad up for 60 days as a *new* posting.

postings to N , which reflects both hires and unfilled openings and is therefore closer to what BG measures than the vacancy stock.

Merging BG to QCEW. Supplemental Appendix A describes the merge of BG ads with an identifiable employer name and location to QCEW establishments at the name-location level.⁵ After cleaning names, we look for exact name matches within a narrow geography, falling back on fuzzy matching validated by common geography and industry codes. We are able to match 91.6% of such ads. We further restrict our sample to the private sector and establishments with at least 2 employees, as the data appeared misaligned for government agencies and single-employee establishments.

3 Methodology

We connect BG postings to JOLTS openings through a multiplicative “adjustment factor.” Estimating it across time and establishment-level covariates – including size and average wage, which are absent from BG but recovered through our BLS linkages – yields new insights. We then derive and validate a prediction method that reweights BG postings to generate moments representative of JOLTS openings, useful where JOLTS is underpowered. See the Supplemental Appendix B for details, derivations, and validation exercises.

3.1 Conceptual Framework

We assume a data-generating process in which the number of openings N and postings B of an establishment depend only on its type j and time t . An establishment has a positive number of openings N^+ with probability p^N , and zero otherwise:

$$N_{ijt} = \begin{cases} N_{jt}^+ & \text{with probability } p_{jt}^N \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

If the establishment has openings, it posts online with probability p^B and, if so, posts B^+ ads. Even with zero openings an establishment may post B^0 ads with probability p^{B0} , for instance because a posting need not meet the JOLTS definition of a vacancy, or because of time aggregation or other measurement issues. The BG posting decision is therefore

$$B_{ijt} = \begin{cases} B_{jt}^+ & \text{with probability } p_{jt}^N p_{jt}^B \\ B_{jt}^0 & \text{with probability } (1 - p_{jt}^N) p_{jt}^{B0} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Taking expectations and combining the two decisions, expected openings are a fixed multiple of expected postings,

$$E_{jt}(N_{ijt}) = \frac{n_{jt}}{p_{jt}^B} s_{jt}^B E_{jt}(B_{ijt}) = a_{jt} E_{jt}(B_{ijt}), \quad (4)$$

⁵The restriction on firm name is common in papers that use Burning Glass data. Firm name is missing 30% of the time, usually because a recruiter keeps the company confidential. Geography is rarely missing during this time period.

where $n = N^+/B^+$ and $s^B = \frac{p^N p^B B^+}{p^N p^B B^+ + (1-p^N)p^{B^0} B^0}$. This defines a multiplicative *adjustment factor*, a , the product of three margins:

1. The *extensive margin* ($1/p^B$), the inverse likelihood of having any BG postings in the same time period as a JOLTS opening.
2. The *intensive margin* (n), openings per posting, conditional on having both, reflecting either multiple openings per posting or an establishment posting only a fraction of its openings.
3. *Nuisance* (s^B), the fraction of postings corresponding to an opening; a nuisance parameter from, e.g., time aggregation, measurement errors, or conceptual differences between JOLTS and BG.

We also derive an analogous adjustment $a^{B,r}$ for higher-order moments $E(N^r)$, which we use below to study labor market concentration.

3.2 Estimation and Prediction Outside JOLTS

Six moments combine into our three margins of adjustment and the overall factor a :

$$\mathcal{M} = \{p^N, p^B, p^{B^0}, N^+, B^+, B^0\}. \quad (5)$$

We calculate these moments and use them to compare along dimensions of heterogeneity, such as time or establishment size categories. We make these comparisons in section 4. Because time-aggregation issues may prevent BG from perfectly capturing monthly JOLTS openings, we average BG postings over a 3-month window centered on the JOLTS measurement month.

To move outside JOLTS, we derive a predictor of openings leveraging the adjustment factor. We estimate regressions for each moment (5) using as covariates a broad set of establishment characteristics, defining a type j , and yielding rich predictions.⁶

The resulting predicted adjustment factor, $\hat{a}$, is a function of the establishment characteristics and time, and we predict openings as follows:

$$\hat{N}_{ijt} = \hat{a}_{jt} B_{ijt}. \quad (6)$$

We can calculate this adjustment for any contemporary sample with postings, B , and the same set of covariates. Because the same covariates appear in both, we compute $\hat{a}$ for JOLTS and for the full QCEW. The latter facilitates analyses for which JOLTS is under-powered.

Our approach assumes that a is constant within a group of establishments of the same type j in a given time period. Given this, our predictor has the property that it will equal N in expectation, when taking expectations over j and/or t within the estimation sample. The appendix details validation exercises that motivate model selection. Using a k-fold cross-validation method, we settle on this model for being predictive

⁶Specifically, we estimate regressions separately for the year groups 2007, 2010, 2011-13, 2014-16, 2017-19, and 2020-22 and include fixed effects for year, 2-digit NAICS sector, average wage quintile (total wage bill divided by average monthly employment), urbanicity category, age groupings, employment growth rate groupings, and size – indicators for 2-5, 6-9, and ≥ 5000 , plus linear spline terms at 10, 50, 250, and 1000 that apply only below 5000. We also include the following time-varying location characteristics: quintiles of the unemployment rate, the share of residents with a bachelor's or greater, the share of employment in professional occupations, and broadband penetration. All establishment-level covariates are measured in the previous calendar year to avoid reclassification bias.

yet parsimonious. We show the model captures well predictable variation in firms’ recruiting behavior.⁷

A natural alternative would be to predict openings directly from a regression of N on a flexible polynomial in postings, the establishment covariates, and their interactions. In practice, such regressions carry little predictive power – especially the postings terms – because the establishment-month relationship between postings and openings is noisy: establishments with openings often do not post, while many postings occur at establishments without openings. Our adjustment factor models these two margins explicitly – through the extensive margin and the nuisance share, respectively – and uses postings only through group-level means, where idiosyncratic variation due to both forces averages out.

4 Do online job postings capture job openings?

We assess the representativeness of the BG data across time and establishment characteristics using the adjustment factors from the previous section. We limit the analysis to the JOLTS sample (the analytic sample, in which we calculate a and its components) and summarize the adjustment factor for establishment-months with postings.⁸

4.1 The aggregate adjustment factor

Figure 1 plots time series for our adjustment and its components. We find substantial convergence over time. The full adjustment factor averages near 30 in 2007, falls to 10 by 2011 and settles around 4 in the later time period. Convergence is especially rapid until 2011, with more gradual declines from there. We see similar convergence when holding constant establishment characteristics at those of a representative establishment (dashed line), so it is not driven by shifts in the composition of hiring establishments captured in the data. Because the BG postings and JOLTS openings series are quite misaligned in the earliest years, we restrict attention to the more stable 2011-19 period for most of the analyses.⁹ We will, however, revisit the earlier years when evaluating previous researchers’ use of 2007 as a pre-Great Recession benchmark (Hershbein and Kahn, 2018).

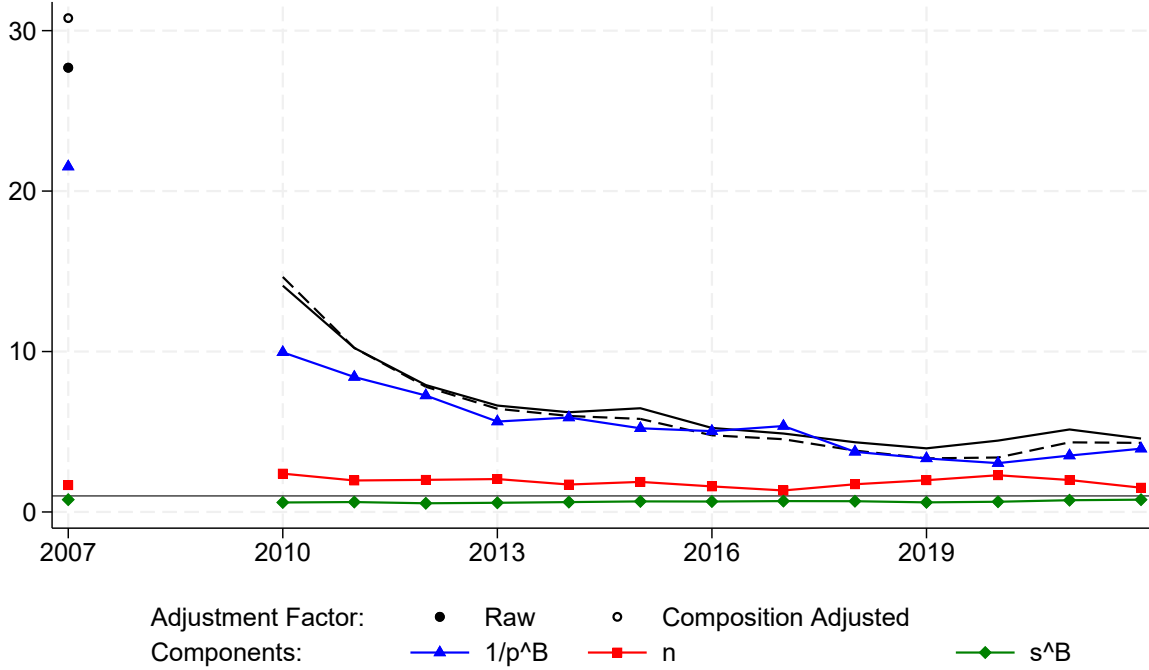
The extensive margin, $\frac{1}{p^B}$ – the inverse likelihood of having BG postings conditional on openings – shows similar convergence to the overall adjustment factor, declining roughly 7-fold over our sample period and ending around 3. The intensive margin n , the number of openings per posting, is stable over this period, averaging roughly 2; a given posting may correspond on average to 2 openings, or companies with openings may post only half of them online. Our nuisance parameter s^B is steady at about 0.6 for the whole time period, so about 40% of postings come from establishment-months with no corresponding JOLTS opening. The implied p^B even in the later sample years – roughly one in three ads is posted online – and an s^B of 60% may both seem low. The apparent mismatch here – many openings without postings and vice versa – implies substantial measurement error or that many postings fall outside the JOLTS definitions of hires and vacancies. However, we show below that these issues are entirely confined to the small firms, while both terms approach 1 for larger firms. Importantly, our conclusions on market concentration are robust to excluding the smallest firms where mismatch is most pronounced.

⁷In supplemental figure S.1, we estimate the model on two-thirds of the sample and use it to predict moments for the held-out third. Adjusted postings track expected openings with a slope near one both in and out of sample, versus only 0.15 for raw postings.

⁸As the adjustment will only be applied to postings data, the sample conditioned on non-zero postings is the relevant one, and we make this restriction for all the descriptives in this section.

⁹We exclude the COVID period because of the large and rapid shifts in labor demand and also because of declining response rates in the JOLTS.

Figure 1: Adjustment Factor and its Components over Time



Notes: The black solid line shows the average adjustment factor by year. The dashed line adjusts for establishment composition. To do so, we compute fitted values from the regressions for a representative establishment and only allow the time component to vary. To place the two series on the same scale, we normalize both to the unadjusted 2011 value by multiplying all values by the 2011 ratio of the two series. We also plot the adjustment factor's components, which multiply to create the full adjustment: $\frac{1}{p^B}$ is the inverse of the extensive margin, n is the intensive margin, and s^B is the share of postings with non-zero N , reflecting measurement error. We calculate the 6 moments, $\mathcal{M}$, in equation 5 for each year and create the adjustments from there (equation 4). Estimates are from the BG-JOLTS matched sample, restricting to private sector establishments with at least 2 employees. Values are averaged annually for establishments with non-zero postings using JOLTS sampling weights.

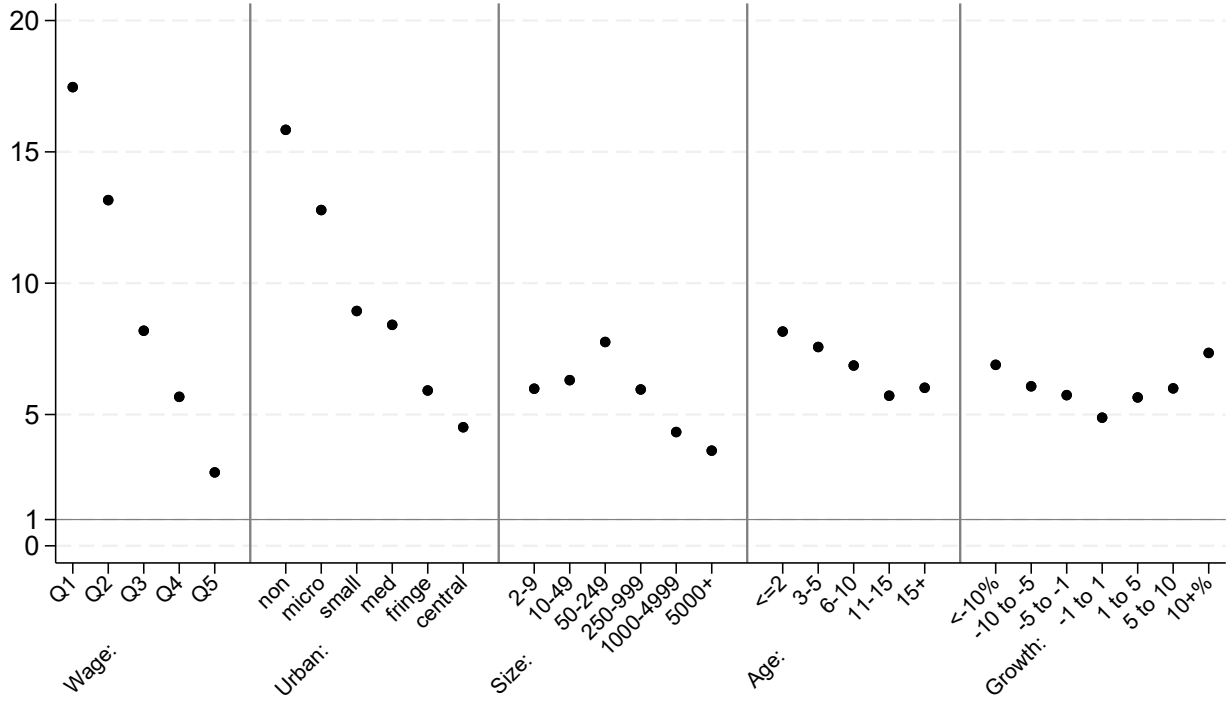
4.2 Variation across establishment characteristics

Figure 2 compares adjustment factors for establishment characteristics – averaging moments along an indicated dimension. We also explored gradients when controlling for all other characteristics and found similar conclusions. From the left, total wage bill per worker quintiles and urbanicity categories exhibit the strongest gradients. The lowest paying establishments' adjustment is 6 times larger than the highest paying establishments'. A posting from the lowest wage quintile must be scaled up 17 times to match the JOLTS distribution of observables, while the highest quintile needs only 2.8 times. Urbanicity shows a 3.5-fold gradient: large central metro areas are best represented online, with postings scaled by 4.5 to match the JOLTS distribution, while noncore areas require a factor of about 16.¹⁰

Next, the adjustment factor shows a gradient with respect to establishment size, with a fairly steep decline in adjustments between the mid-sized 50-249 category and the largest sizes. The largest establishments need to be scaled by about 3.6 to match the JOLTS, while the adjustment factor for mid-sized establishments is

¹⁰The urbanicity gradient is not altered with the inclusion of time-varying location characteristics measuring the skill level of the population and the economic health of the location.

Figure 2: Adjustment Factor by Establishment Characteristics



Notes: The figure plots the adjustment factor a (equation 4) by establishment characteristics. We calculate the 6 moments, $\mathcal{M}$, in equation 5 within each category along a given characteristic and use these to generate adjustment factors. We use the 2011-2019 BG-JOLTS sample, conditioning on establishments with postings and private sector establishments with at least two employees.

double that. The gradient does not continue upward for the smallest size classes, which also show adjustment factors around 6. We find that the various components of the adjustment offset one another for small establishments, resulting in a flatter gradient (appendix figure S.2). Specifically, the smallest establishments have low values for s^B (a low s^B means a higher share of BG postings come from establishments with no JOLTS opening). Their postings must be scaled down because they are less likely to correspond to JOLTS openings; meanwhile large firms have values close to 1. This force offsets a large gradient in the extensive margin – small establishments are much less likely to post online, conditional on having a JOLTS opening, compared to large establishments.

The overall size gradient matters for interpreting BG moments: the largest size classes receive only a small adjustment (they are well represented), while small and mid-sized establishments receive double the adjustment. Most establishments are small, so questions related to the distribution of establishments that are hiring will vary substantially in BG versus a more representative dataset. However, most workers are employed at large establishments, and therefore most vacancies are posted by large establishments. Since BG represents these establishments quite well, it represents the distribution of openings well.

The final panels of figure 2 show only modest patterns with respect to establishment age and growth. Differential representation of growing versus shrinking establishments could bias many inferences; for instance, if expanding establishments posted more vacancies online, researchers could exaggerate the rise in labor

demand during expansions. Given the symmetry of the adjustment factor across growth rates, however, this is likely not a concern.

There are also strong gradients across industries (appendix figure S.3), with construction and leisure and hospitality sectors requiring large adjustments to make BG representative of JOLTS, while professional service sectors require much smaller adjustments. We also examined gradients for time-varying location characteristics, including local demographics and economic variables, but did not find strong variation in adjustments once controlling for firm characteristics (not shown).

5 Adjusting online postings data to capture vacancies

We now assess whether the differences in firm composition highlighted above matter for drawing inferences from BG. We use our adjustment factor as a weight, rescaling BG to match the true distribution of openings on observables. For these exercises, we focus on the full BG-QCEW matched sample, which is helpful when JOLTS is underpowered. We can estimate adjustment weights regardless of whether the establishment responded to JOLTS, since our prediction model only uses establishment characteristics that are available in the QCEW.

5.1 Application I: The Demand for Skills

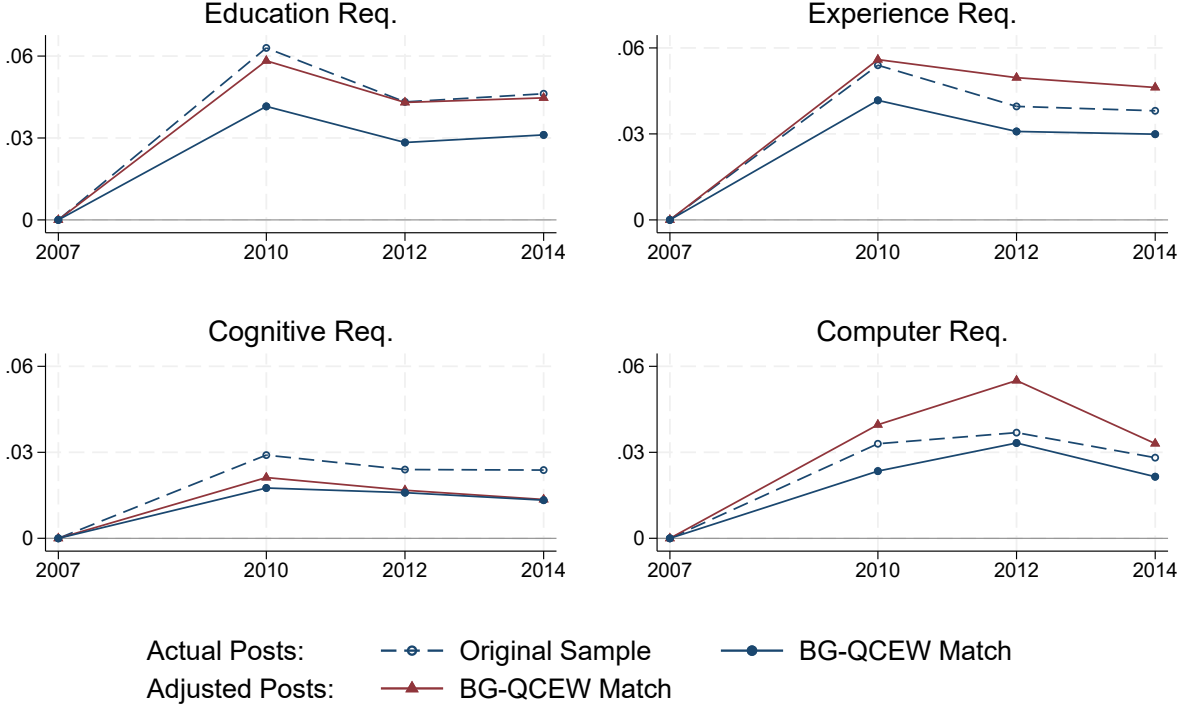
Researchers have made extensive use of the skill requirements measured in BG to track demand for skill. But because BG overrepresents higher-paying establishments in higher-skilled industries and labor markets, its ads may overstate demand for high-skilled positions, and correcting this could change conclusions about skill demand and how it varies across time and space.

Applying our adjustment helps match the distribution of BG establishments to that of hiring establishments in JOLTS, leveraging the ad characteristics of establishments that do post to infer skill demand for those that do not. We do indeed find that BG over-represents skilled positions. As two examples, 45% of ads are in a professional occupation and 56% specify an education requirement; when applying our adjustment, we find a modest downward correction in skill demand of 7 ppts (16%) for professional share and 6 ppts (11%) for education requirements. Supplemental table S.2 reports statistics for these and other measures of skill, reaching similar conclusions. The movements are modest because the distribution of postings is fairly representative of the adjusted distribution; most openings come from large establishments, which are well represented.

Hershbein and Kahn (2018, HK) show that skill requirements related to automation increased after the Great Recession. Using spatial variation in recession severity, they find evidence of episodic restructuring toward labor-saving technologies. They use the 2007 datapoint as their pre-period, but as we showed above, BG substantially underrepresents JOLTS openings in the earlier years and especially in 2007, so a compositional shift in BG establishments could impact their inference.

In equation 7, we regress the change from 2007 in the share of ads in Metropolitan Statistical Area (MSA) m and year t requiring a given skill on shock_m , which measures the size of the Great Recession shock

Figure 3: Hershbein-Kahn Skill Requirements over the Cycle



Notes: We summarize coefficients $\alpha_{1,t}$ from equation 7, regressing a change in skill requirements in BG from 2007 to the indicated year on the Great Recession shock experienced at the MSA level. The blue dashed line replicates [Hershbein and Kahn \(2018\)](#); the blue solid line replicates the results in the BG-QCEW match sample; the maroon line adjusts posts based on our predictor (equation 6) and restricts to the BG-QCEW match.

experienced by the MSA, as well as a vector of MSA-level controls (C_m):¹¹

$$Skill_{mt} - Skill_{m07} = [shock_m * I_t] \alpha_1 + I_t + C_m \beta + \varepsilon_{mt}, \quad (7)$$

where I_t is a vector of year indicators, so α_1 collects the year-specific effects of $shock_m$, which are plotted in figure 3. MSA-year level observations are weighted by the size of the MSA labor force in 2006 in order to upweight larger areas for greater precision.

Figure 3 compares the Hershbein-Kahn results to those that apply our adjustment. We plot the coefficients on the shock-year interactions.¹² HK focus on education and experience requirements, the cognitive skill index of [Deming and Kahn \(2018\)](#), and computer requirements – skills they argue firms demand more of when shifting toward automation technology.

The blue dashed lines replicate HK’s results showing upskilling in the immediate post-recession period. Crucial for the HK argument, this upskilling persists well after local labor markets recover from the Great Recession. So purely cyclical explanations, such as opportunistic hiring in a slack labor market, cannot

¹¹HK use a shift-share that predicts the change in employment growth from 2006-09 by applying national industry employment changes to baseline MSA shares. Controls C_m are standard demographics, education, age, and labor-market characteristics measured in the American Community Survey (2005-06 averages; [Ruggles et al. \(2021\)](#)).

¹²The outcome is the postings-weighted average of an employer’s share of ads with a skill requirement, as a change from 2007. Our adjustment instead weights the share of ads with a skill requirement by adjusted posts.

explain the upskilling. When paired with evidence of capital deepening in harder-hit MSAs, HK argue that the Great Recession accelerated routine-biased technological change. Blue solid lines show this persistence holds when restricting to the BG-QCEW matched sample. This persistence also holds when applying our adjustment (maroon line). Since HK’s story concerns the composition of postings over the cycle, and BG captures large establishments and most postings well, our adjustment makes little difference and, if anything, strengthens the results.

5.2 Application II: Market Concentration

Researchers have leveraged the richness of BG data to study labor market concentration at highly disaggregated levels (Azar et al., 2020; Hershbein et al., 2022; Azar et al., 2024; Schubert et al., 2025). These advances in measurement have been crucial for moving academic and policy conversations toward a broader discussion of firms’ wage-setting power. However, concentration measures can be quite sensitive to whether the long tail of small and mid-sized establishments is captured in the data. Here we explore the consequences of our adjustment for researchers’ conclusions on labor market competition.

The standard measure is the Herfindahl-Hirschman Index (HHI), the sum of squared market shares. Applied to online postings (B) for establishment i of type $j(i)$ across the S BG-QCEW establishments in a market-month, it is (suppressing market and time subscripts):

$$HH^B = \sum_{i=1}^S \left(\frac{B_{i,j(i)}}{\sum_{k=1}^S B_{k,j(k)}} \right)^2. \quad (8)$$

We instead want the HHI of actual openings, HH^N .

We derive a higher-order adjustment in Supplemental Appendix B to compute $E(N^2)$ – the adjustment $a^{B,r}$ from section 3.1 with $r = 2$ – and denote it $a^{B,2}$.¹³ We divide this moment by the square of predicted market size. Our adjusted HHI thus is:

$$\hat{H}H^N = \sum_{i=1}^S a_{j(i)}^{B,2} \left(\frac{B_{i,j(i)}}{\sum_{k=1}^S a_{j(k)} B_{k,j(k)}} \right)^2, \quad (9)$$

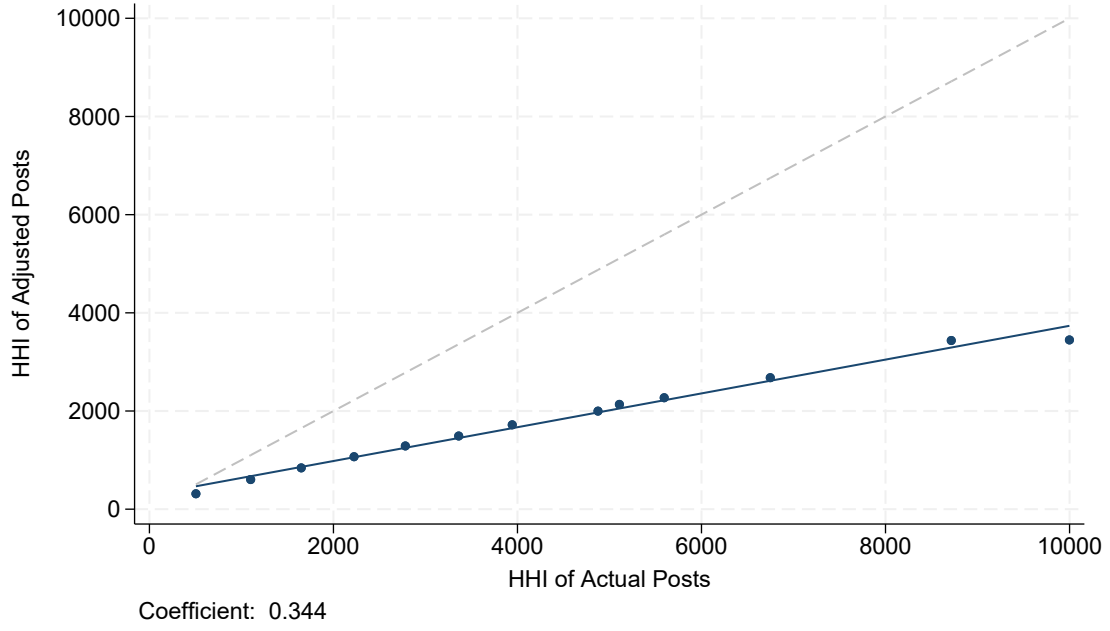
where the numerator applies establishment i ’s second-order factor $a_{j(i)}^{B,2}$ to estimate $E(N^2)$, while the denominator applies each establishment k ’s first-order factor $a_{j(k)}$ to form predicted market size. Under our assumptions, conditional on predicted market size, the expectation of the predicted HHI equals the expectation of the actual HHI of openings; we derive this property in Supplemental Appendix B. In our analysis below, we probe the robustness of our results to conditioning on predicted market size.

We compare HHIs of postings (8) to adjusted B s (9) across local labor markets in the matched BG-QCEW sample. Markets are typically defined as a skill grouping crossed with geography; because QCEW has industry but not occupation, we use the 4-digit NAICS industry by commuting zone (CZ) level. This aggregation is common in the literature, but data are thin for some markets, so we present HHIs across all market-months with at least one posting, as well as versions that restrict on market size.

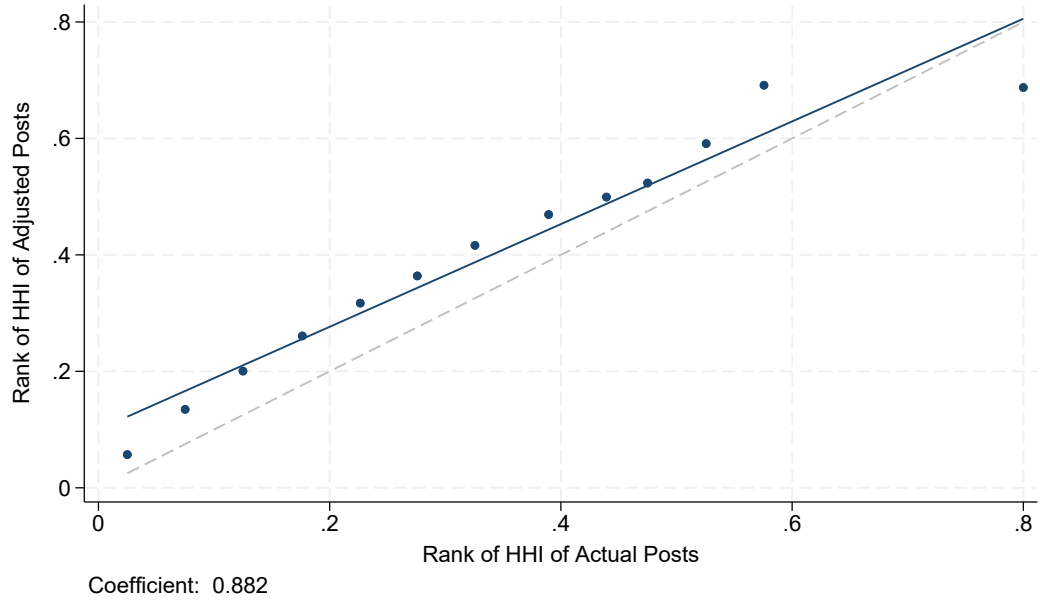
Supplemental figure S.5 validates our HHI adjustment, showing it does a good job when using JOLTS data where we can measure the HHI of actual openings. Due to the much smaller sample sizes in JOLTS, we

¹³In supplemental figure S.4, we show that the adjustment factor of second-order moments (based on equation 13) shows similar patterns by establishment characteristics as our main adjustment factor a .

Figure 4: HHIs of BG Postings: Adjusted vs. Actual



(a) Levels



(b) Ranks

Notes: Binscatter of adjusted and actual HHIs in BG postings, along with best fit line (solid) with slope listed at the bottom and 45-degree line (dashed). Each observation is a market-month defined by commuting zone, four-digit NAICS, and month. Observations are divided into 5-percent bins based on the HHI of BG postings (equation 8). Panel (a) plots the average actual HHI and adjusted HHI (equation 9) within each bin. Panel (b) plots the average percentile within each bin of HHIs versus adjusted HHIs. The figure pools eligible monthly labor-market observations from 2011-19.

Table 1: HHI Summary Statistics

Panel A.	Market-Month level summary statistics	
	Actual Postings	Adjusted Postings
Average HHI	6289.14	2327.44
Median HHI	5926.51	1936.66
Fraction above 2500 HHI	0.80	0.38
Fraction 1500-2500 HHI	0.09	0.23
Panel B.	Worker-weighted summary statistics	
	Actual Postings	Adjusted Postings
Average HHI	2666.41	1133.07
Median HHI	1461.48	627.23
Fraction above 2500 HHI	0.35	0.13
Fraction 1500-2500 HHI	0.14	0.12

Notes: We summarize monthly labor-market concentration measured using actual BG postings (equation 8) and adjusted postings (equation 9). Labor markets are defined at the four-digit-NAICS-by-Commute-Zone level. Panel A gives each market-month equal weight, while Panel B weights by monthly employment. The sample is the BG-QCEW match, restricted to private-sector establishments with at least two employees. Monthly HHIs are pooled over 2011-19.

conduct this validation by simulating QCEW-sized markets drawn from JOLTS establishments. The HHI of postings relates to that of actual openings with a slope of only 0.39, far below the 45-degree line; the adjusted HHI relates to the postings HHI with roughly this same slope, successfully capturing the true concentration of job openings. Because the simulated markets are drawn from actual JOLTS establishments, they embed realistic within-establishment-type dispersion in openings and postings, so this exercise also validates the approximations underlying the higher-order adjustment.

Turning to our main results, panel (a) of figure 4 shows a binscatter of HHIs of actual B s against the adjusted HHIs. Adjusted HHIs are much less concentrated: a one-unit increase in concentration measured by actual posts translates into only a 1/3-unit increase once adjusted for composition. The reason is that small and mid-sized establishments are underrepresented in BG, and the HHI is particularly sensitive to this long tail. Our finding that BG overstates HHI-based concentration by a factor of nearly three is robust across a range of checks: excluding establishments with fewer than 10 employees, weighting markets by employment rather than equally, and conditioning on predicted market size by quartile (supplemental figures S.6, S.7, and S.8). Our conclusions are not driven by the smallest establishments and markets which could generate outliers.

Table 1 summarizes concentration across markets. Panel A gives each market-month equal weight, while Panel B weights markets by employment. Average HHI falls from about 6,300 to about 2,300 in Panel A and from about 2,700 to 1,100 in Panel B, reductions of 63 and 57 percent, respectively. The same pattern emerges under the standard cutoffs from the 2010 DOJ/FTC Horizontal Merger Guidelines, which Azar et al. (2020) adopt: an HHI above 2500 indicates a highly concentrated market and 1500–2500 a moderately concentrated one. From panel A, 80% of markets are highly concentrated when considering actual posts.¹⁴ Adjusting for representativeness yields very different conclusions: only 38% of markets are highly concentrated based on

¹⁴This share is not directly comparable to the 60% in Azar et al. (2020), who impose a restriction on market size before calculating HHIs (and define markets by occupation rather than industry); our universe of all market-months with at least one posting includes single-poster markets at the maximal HHI of 10,000, pushing the unadjusted share up. Again, our robustness analyses show that our conclusions hold with stronger restrictions on market eligibility.

adjusted posts. As has been noted, the worker-weighted statistics are substantially less concentrated (Berger et al., 2022), so most workers live in labor markets that are not highly concentrated.

While HHI levels in BG are about three times too large, BG captures *ordinal* comparisons across markets well. Panel (b) of figure 4 plots a rank-rank binscatter based on actual and adjusted HHIs across the quantile bins: the slope across percentiles is 0.9, so one can fairly accurately compare HHIs across markets in BG (the employment-weighted relationship has an even stronger slope of 1). BG data can therefore still be useful for understanding establishment-level concentration in local labor markets, but researchers should focus on relative comparisons; drawing strong conclusions from the levels of HHIs in BG is ill-advised.

6 Conclusion

Large-scale online job postings data such as Burning Glass-Lightcast offer a great deal of richness beyond surveys or administrative data, but little is known about their representativeness. By linking these data to the representative JOLTS vacancy survey, we show that the job openings of small and low-paying establishments, certain industries, and non-urban areas are not well represented online. We derive an approach to reweight BG postings so that establishment characteristics match those in the representative JOLTS data.

Reassuringly, our reweighting reproduces qualitatively similar results for many research questions. Since BG represents large establishments well – and most vacancies are posted by large establishments – questions about the level and cyclicalities of skill demand are robust to our adjustment. However, most establishments are small, so questions sensitive to the distribution of hiring establishments – most notably measures of establishment-level concentration – can vary substantially in BG versus a more representative dataset. We advise against drawing conclusions from the *level* of HHIs measured in online postings data; the HHI is sensitive to whether the long tail of small establishments is captured, while ordinal comparisons of concentration across labor markets in BG are far more reliable.

More broadly, online postings reliably represent the vacancy distribution for posting-level questions. Researchers can continue to benefit from large-scale online vacancy data sources for many research questions, even without our linkage.

References

- Abraham, K. G. (1987). Help-Wanted Advertising, Job Vacancies, and Unemployment. *Brookings Papers on Economic Activity* 18(1), 207–248.
- Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics* 40(S1), S293–S340.
- Azar, J., E. Huet-Vaughn, I. Marinescu, B. Taska, and T. von Wachter (2024). Minimum wage employment effects and labor market concentration. *The Review of Economic Studies* 91(4), 1843–1883.
- Azar, J., I. Marinescu, and M. Steinbaum (2022). Labor market concentration. *Journal of Human Resources* 57(S), S167–S199.
- Azar, J., I. Marinescu, M. Steinbaum, and B. Taska (2020). Concentration in US labor markets: Evidence from online vacancy data. *Labour Economics* 66, 101886.
- Banfi, S. and B. Villena-Roldan (2019). Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market. *Journal of Labor Economics* 37(3), 715–746.
- Barnichon, R. (2010). Building a composite Help-Wanted Index. *Economics Letters* 109(3), 175–178.
- Barron, J. M., M. C. Berger, and D. A. Black (1997). Employer Search, Training, and Vacancy Duration. *Economic Inquiry* 35(1), 167–192.
- Berger, D., K. Herkenhoff, and S. Mongey (2022). Labor market power. *American Economic Review* 112(4), 1147–93.
- Blanchard, O. J. and P. Diamond (1989). The Beveridge Curve. *Brookings Papers on Economic Activity* 20(1), 1–76.
- Braxton, J. C. and B. Taska (2023). Technological change and the consequences of job loss. *American Economic Review* 113(2), 279–316.
- Callaci, B., M. Gibson, S. Pinto, M. Steinbaum, and M. Walsh (2024). The effect of franchise no-poaching restrictions on worker earnings. *The Review of Economics and Statistics*, 1–35.
- Carnevale, A. P., T. Jayasundera, and D. Repnikov (2014). Understanding online job ads data: A technical report. Technical report, Washington, DC: Georgetown University Center on Education and the Workforce.
- Clemens, J., L. B. Kahn, and J. Meer (2021). Dropouts need not apply? the minimum wage and skill upgrading. *Journal of Labor Economics* 39(S1), S107–S149.
- Davis, S. J., R. J. Faberman, and J. C. Haltiwanger (2013). The Establishment-Level Behavior of Vacancies and Hiring. *The Quarterly Journal of Economics* 128(2), 581–622.
- Deming, D. and L. B. Kahn (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics* 36(S1), S337–S369.
- Dillon, E. W., L. B. Kahn, J. Venator, and M. Dalton (2025). Do workforce development programs bridge the skills gap. Working Paper 34012, National Bureau of Economic Research.

- Elsby, M. W. L., R. Michaels, and D. Ratner (2015). The Beveridge curve: A survey. *Journal of Economic Literature* 53(3), 571–630.
- Forsythe, E., L. B. Kahn, F. Lange, and D. Wiczer (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and ui claims. *Journal of Public Economics* 189, 104238.
- Forsythe, E., L. B. Kahn, F. Lange, and D. Wiczer (2022). Where have all the workers gone? recalls, retirements, and reallocation in the COVID recovery. *Labour Economics* 78, 102251.
- Hansen, S., P. J. Lambert, N. Bloom, S. J. Davis, R. Sadun, and B. Taska (2023). Remote work across jobs, companies, and space. Working Paper 31007, National Bureau of Economic Research.
- Hemelt, S. W., B. Hershbein, S. Martin, and K. M. Stange (2023). College majors and skills: Evidence from the universe of online job ads. *Labour Economics* 85, 102429.
- Hershbein, B. and L. B. Kahn (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review* 108(7), 1737–72.
- Hershbein, B., C. Macaluso, and C. Yeh (2022). Monopsony in the U.S. labor market. *American Economic Review* 112(7), 2099–2138.
- Lange, F. and T. Papageorgiou (2020). Beyond Cobb-Douglas: Flexibly estimating matching functions with unobserved matching efficiency. Technical report, National Bureau of Economic Research.
- Marinescu, I. E. (2017). The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board. *Journal of Public Economics* 150, 14–29.
- Marinescu, I. E. and R. Wolthoff (2020). Opening the Black Box of the Matching Function: The Power of Words. *Journal of Labor Economics* 38(2), 535–568.
- Mongey, S. and G. L. Violante (2026). Macro recruiting intensity from micro data. *American Economic Journal: Macroeconomics* 18(2), 1–33.
- Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies* 61(3), 397–415.
- Mueller, A. I., D. Osterwalder, and J. Zweimüller (2025). The Ins and Outs of Vacancies. Mimeo., University of Zurich.
- Mueller, A. I., D. Osterwalder, J. Zweimüller, and A. Kettemann (2024). Vacancy Durations and Entry Wages: Evidence from Linked Vacancy–Employer–Employee Data. *The Review of Economic Studies* 91(3), 1807–1841.
- Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39(2), 390–431.
- Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *The American Economic Review* 75(4), 676–690.
- Ruggles, S., S. Flood, R. Goeken, J. Grover, and M. Sobek (2021). IPUMS USA: Version 11.0 [dataset].

- Şahin, A., J. Song, G. Topa, and G. L. Violante (2014). Mismatch unemployment. *American Economic Review* 104(11), 3529–64.
- Schubert, G., A. Stansbury, B. Taska, and M. Dey (2025). Employer concentration and wages: Evidence from large firms’ hiring shocks. Conditionally accepted, *American Economic Journal: Economic Policy*.
- Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95(1), 25–49.
- Shimer, R. (2007). Mismatch. *American Economic Review* 97(4), 1074–1101.

Supplemental Appendix

A BG-QCEW Match

We use information on firm name, location, and imputed industry in the Burning Glass (BG) dataset to match to establishments in the Quarterly Census of Employment and Wages (QCEW). Our BG extract was obtained from Emsi-Burning Glass in late 2022. The company has since merged and changed its name to Lightcast. Our postings data include ads from January through December of 2007, and from January 2010 through October 2022.

We restrict attention to 69% of postings that include a firm name and a geographic identifier – at least as detailed as the state where the job is located. Importantly, BG does regularize firm names and this algorithm has gotten better over time. More recent extracts may therefore have different (cleaner) firm names than ours. However, all updates to their algorithms are applied to the entire dataset so they are internally consistent.

We begin by regularizing employer names by removing common components such as “LLC” or “the,” singularizing, removing geographic information, and removing punctuation. A BG cell is constructed for each unique edited employer name and city-county-state. We then find the most similar employer name in QCEW within the same commuting zone (CZ).

We determine exact matches to be those where the ad shares exact cleaned firm name with a QCEW in the same substate geography. When an exact match is not available, we use fuzzy matching techniques to find similar names across datasets, while relying on common substate geography and industry codes to identify higher quality matches. Our fuzzy match algorithm is as follows:

1. Identify 30 best matches based on Term Frequency-Inverse Document Frequency (TF-IDF) text match score (score of 0 to 1)
2. Create Levenshtein text match score across all 30 matches (normalized to be scored from 0 to .15). We add points to this baseline match score as follows:
 - +0.11 if match has same sector
 - +0.2 if match is in same city
 - +0.15 if match is in same county
 - +0.1 if match is in same state and commuting zone
 - +0.16 if match exists in years consistent with BG employer name - city
3. Among matches that score at least 0.9, we keep the highest value

These match score components were chosen based on a maximization routine on a training dataset of employer names and geography matched to QCEW that minimizes false positives.

If a posting does not have an exact firm name match nor a high enough fuzzy match score, we make one final attempt to link them to a QCEW establishment. These “sibling” matches share exact firm name but only match to an establishment in the same state (and not sub-state geography).

It is entirely possible for a single BG cell (firm name-geography) to be correctly matched to multiple establishments. If an employer in a particular city has multiple establishments (e.g., a large coffee chain), then the BG cell will capture the online postings for all of those establishments - with no easy way to

determine which singular establishment the postings should be assigned to. For this reason, when a BG cell has multiple establishment matches, we divide the BG postings proportionally to the relative employment of each establishment matched to the BG cell. This can mean that occasionally some establishments are assigned a fractional posting. In our analyses, we treat a positive number of postings as that establishment being assigned at least 1 posting. We could instead aggregate the QCEW to the employer-location level as well, but that approach creates problems when we move to comparisons with the JOLTS. The JOLTS may only capture one of many establishments within a given name-location cell and we would not want to attribute all BG postings to that one observation.

Table S.1 summarizes the breakdown of match types, by percent of BG postings. Matches are grouped into 3 different categories for the firm name linkage and across levels of geography based on the most detailed level of common geography the match attains (i.e., if a match has the same city then it also has the same county, CZ, and state but will be summarized in the city column). Most of the postings are exactly matched on (cleaned) employer name and most are matched to an establishment in the same city (the smallest geography available in BG). Overall, 91.6% of BG postings match to at least one establishment in the QCEW.

Table S.1: Distribution of BG Postings across QCEW Match Type

	City	County	CZ	State	Total
Exact Match	29.6	8.7	8.4	-	46.7
Fuzzy Match	13.8	5.3	6.1	-	25.3
Sibling Match	-	-	-	19.7	19.7
Total	43.4	14.0	14.5	19.7	91.6

Notes: We summarize the percent of BG postings across the type of match to establishments in the Quarterly Census of Employment and Wages. Exact matches share the same cleaned firm name and sub-state geography. Fuzzy matches are based on the algorithm described in this section. Sibling matches share exact firm names but only a geography match at the state level. 46.7% of BG postings match to an establishment in QCEW based on the exact match type. Overall 91.6% of postings successfully match.

B Methodology Details

This section provides the full derivation of the adjustment factor summarized in Section 3. We also detail the estimation and validation exercises.

B.1 Conceptual Framework: Derivations

We assume the data generating process as described in equations 2 and 3 in the text. Importantly, this model assumes that the number of openings (N) and postings (B) depend only on an establishment’s type, j . We treat the parameters of the model as fixed, i.e. there is no strategic behavior on the part of establishments, and assume that there is no variation in these variables within type j .

First-order moments. Given these assumptions, we can derive equation 4 in the text as follows. The expected number of openings and postings for establishment i of type j at time t are:

$$\begin{aligned} E_{jt}(N_{ijt}) &= p_{jt}^N N_{jt}^+ \\ E_{jt}(B_{ijt}) &= p_{jt}^N p_{jt}^B B_{jt}^+ + (1 - p_{jt}^N) p_{jt}^{B0} B_{jt}^0. \end{aligned}$$

Define

$$\begin{aligned} n_{jt} &= \frac{N_{jt}^+}{B_{jt}^+}, \\ s_{jt}^B &= \frac{p_{jt}^B p_{jt}^N B_{jt}^+}{p_{jt}^N p_{jt}^B B_{jt}^+ + (1 - p_{jt}^N) p_{jt}^{B0} B_{jt}^0}. \end{aligned}$$

Combining these definitions with the two equations above gives the expected number of N s as a function of expected B s.¹⁵

$$E_{jt}(N_{ijt}) = \frac{n_{jt}}{p_{jt}^B} s_{jt}^B E_{jt}(B_{ijt}). \quad (10)$$

This defines an adjustment factor of the form

$$a_{jt} = \frac{n_{jt}}{p_{jt}^B} s_{jt}^B. \quad (11)$$

Higher-order moments. Building on the first-order derivation above, we derive expectations for higher-order transformations of the form $f(N) = N^r$, where r is a scalar, again leveraging our assumption that openings and postings vary only with type j at time t :

$$\begin{aligned} E_{jt}((N_{ijt})^r) &= p_{jt}^N (N_{jt}^+)^r \\ E_{jt}((B_{ijt})^r) &= p_{jt}^N p_{jt}^B (B_{jt}^+)^r + (1 - p_{jt}^N) p_{jt}^{B0} (B_{jt}^0)^r. \end{aligned}$$

Following the same steps as for the first-order moments, we obtain:

$$E_{jt}((N_{ijt})^r) = \frac{(n_{jt})^r}{p_{jt}^B} s_{jt}^{B,r} E_{jt}((B_{ijt})^r), \quad (12)$$

where $s_{jt}^{B,r} = \frac{p_{jt}^N p_{jt}^B (B_{jt}^+)^r}{p_{jt}^N p_{jt}^B (B_{jt}^+)^r + (1 - p_{jt}^N) p_{jt}^{B0} (B_{jt}^0)^r}$ and thus defines an adjustment factor of the form:

$$a_{jt}^{B,r} = \frac{(n_{jt})^r}{p_{jt}^B} s_{jt}^{B,r}. \quad (13)$$

These higher-order expectations underlie the adjusted HHI used in the main text. Note that for $r > 1$ these expressions treat N^+ , B^+ , and B^0 as constant within type jt ; the validation exercise in supplemental figure S.5 assesses the resulting approximation on data with realistic within-type dispersion.

¹⁵For this derivation, we use the fact that $s_{jt}^B = \frac{E_{jt}(B_{ijt}) - (1 - p_{jt}^N) p_{jt}^{B0} B_{jt}^0}{E_{jt}(B_{ijt})} = \frac{p_{jt}^B p_{jt}^N B_{jt}^+}{p_{jt}^N p_{jt}^B B_{jt}^+ + (1 - p_{jt}^N) p_{jt}^{B0} B_{jt}^0}$.

Adjusted HHI. Consider a market with S establishments, suppressing market and time subscripts, and let $j(i)$ denote establishment i 's type, $\bar{N} = \sum_{i=1}^S N_{i,j(i)}$ denote actual market size, and $\hat{\bar{N}} = \sum_{i=1}^S a_{j(i)} B_{i,j(i)}$ denote predicted market size. The actual HHI of openings is $HH^N = \sum_{i=1}^S (N_{i,j(i)}/\bar{N})^2$, and the adjusted HHI in the main text (equation 9) applies the adjustment factor in equation 13 for $r = 2$ (denoted $a_j^{B,2}$ in the main text). Conditional on predicted market size, the expectation of the adjusted HHI equals the expectation of the actual HHI of openings:

$$E(\hat{H}H^N | \hat{\bar{N}}) = E\left(\sum_{i=1}^S \left(\frac{N_{i,j(i)}}{\hat{\bar{N}}}\right)^2 \middle| \hat{\bar{N}}\right) = E(HH^N | \hat{\bar{N}}). \quad (14)$$

The first equality applies equation 12 with $r = 2$ term by term, so that $\frac{s_{j(i)}^{B,2} (n_{j(i)})^2}{p_{j(i)}^B} E((B_{i,j(i)})^2 | \hat{\bar{N}}) = E((N_{i,j(i)})^2 | \hat{\bar{N}})$, while $\hat{\bar{N}}$ in the denominator is fixed by the conditioning. For the second equality, note that $\sum_{i=1}^S (N_{i,j(i)}/\hat{\bar{N}})^2 = (\bar{N}/\hat{\bar{N}})^2 HH^N$. We use the approximation $E(\bar{N}^2 HH^N | \hat{\bar{N}}) \approx E(\bar{N} | \hat{\bar{N}})^2 E(HH^N | \hat{\bar{N}})$, which is innocuous given that we condition on expected market size, together with the fact that $E(\bar{N} | \hat{\bar{N}}) = \hat{\bar{N}}$: applying the first-order relation (equation 11) term by term gives $E(\bar{N} | \hat{\bar{N}}) = \sum_{i=1}^S E(N_{i,j(i)} | \hat{\bar{N}}) = \sum_{i=1}^S a_{j(i)} E(B_{i,j(i)} | \hat{\bar{N}}) = E(\hat{\bar{N}} | \hat{\bar{N}}) = \hat{\bar{N}}$. Supplemental figure S.8, which splits markets by quartile of predicted size, probes these approximations empirically.

B.2 Estimation Details

Estimated moments. To form the adjustment factor, we estimate the six moments from equation 5 in the JOLTS sample as follows, where j is establishment type and t is time:

$$\begin{aligned} p_{jt}^N &= \Pr(N_{ijt} > 0 | j, t) \\ p_{jt}^B &= \Pr(B_{ijt} > 0 | N_{ijt} > 0, j, t) \\ p_{jt}^{B0} &= \Pr(B_{ijt} > 0 | N_{ijt} = 0, j, t) \\ N_{jt}^+ &= E[N_{ijt} | N_{ijt} > 0, j, t] \\ B_{jt}^+ &= E[B_{ijt} | B_{ijt} > 0 \text{ and } N_{ijt} > 0, j, t] \\ B_{jt}^0 &= E[B_{ijt} | B_{ijt} > 0 \text{ and } N_{ijt} = 0, j, t]. \end{aligned} \quad (15)$$

In practice, we compute these moments over a 3-month rolling window centered on the JOLTS measurement month. For example,

$$\begin{aligned} p_{jt}^B &= \Pr\left(\sum_{k=-1}^{+1} B_{ij(t+k)} > 0 \mid N_{ijt} > 0, j, t\right), \\ B_{jt}^+ &= E\left[\frac{1}{3} \sum_{k=-1}^{+1} B_{ij(t+k)} \mid \sum_{k=-1}^{+1} B_{ij(t+k)} > 0, N_{ijt} > 0, j, t\right]. \end{aligned}$$

The other terms are defined analogously. At sample end points (January and December 2007, January 2010, and October 2022) the rolling mean uses only the available current and adjacent months rather than a fixed denominator of three.

Estimation. We define a type j as a set of establishment characteristics X and estimate a series of regressions as in equation 16, where outcomes Y_{ijt} are the six moments described above and i denotes an establishment whose type is j :

$$Y_{ijt} = \alpha_t + X_{jt}\beta_g + \varepsilon_{ijt}. \quad (16)$$

Given our establishment-level match, we regress an outcome on year indicators (α_t) and a vector of establishment characteristics (X_{jt}) that depend on its type j , including size, age, average wage, industry dummies, urbanicity, and time-varying location characteristics. We estimate these regressions at the establishment-month level using JOLTS sampling weights. We allow the coefficients β to vary by year group g (2007, 2010, 2011-13, 2014-16, 2017-19, 2020-22) and control for year fixed effects within each group, as data become too sparse if we estimate effects separately by year. To avoid reclassification bias, we define the time-varying characteristics at a lag with respect to the month of JOLTS measurement.¹⁶ The advantage of this approach is that we can include just one dimension of establishment heterogeneity at a time to summarize the relationships, or include many to understand whether controlling for one establishment characteristic accounts for variation in another.

For the binary variables (p^N , p^B , p^{B0}), we estimate the equation with a Probit model, whereas for the other variables (N^+ , B^+ , B^0), we estimate it with the exponential model, which respects the non-negativity constraint of these variables.¹⁷

B.3 Validation

Outside the prediction sample, differences in the underlying postings-openings relationship or sampling variation may generate discrepancies in the predictions relative to true openings. We employ a k-fold cross-validation to validate our model on an “out” sample within JOLTS. This approach has the advantage that we can still measure true JOLTS openings even in the out sample and it helps avoid problems with over-fitting. To do so, we split the JOLTS sample into thirds. For each iteration, we estimate our model on a two-thirds “in” sample. The model consists of the probit and exponential regressions used to estimate the 6 moments outlined above. We can also predict fitted values for both in and out samples. We repeat this process 3 times, thus generating a complete set of out-of-sample predictions covering the entire JOLTS sample.

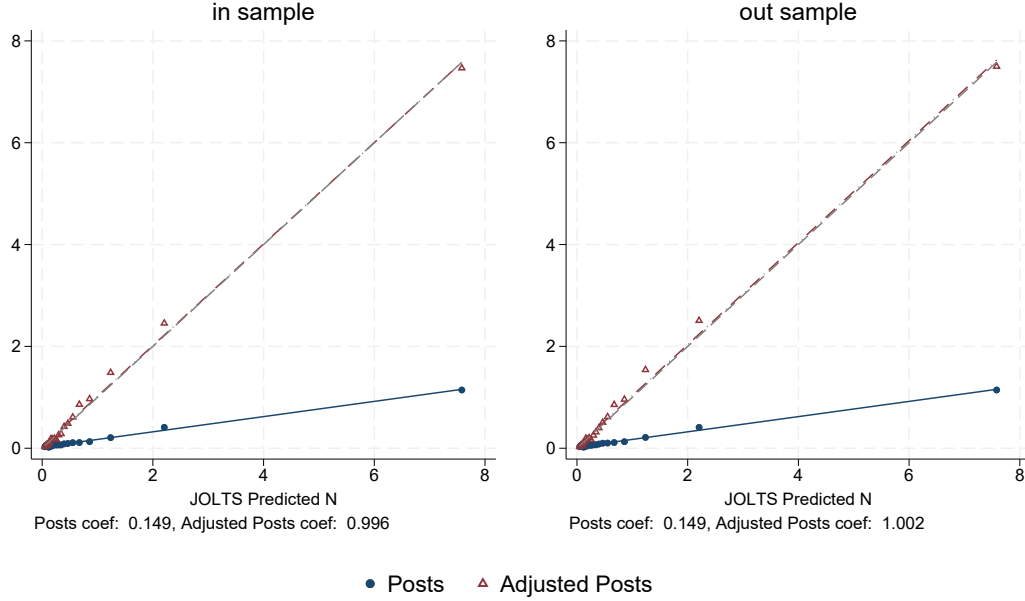
We conduct several validation exercises on the predicted $\hat{N}_{it}$. First, along dimensions X that our model uses in its prediction, we match expected values of $\hat{N}_{it}$ and N well both in- and out-of-sample. That is, the average $\hat{N}_{it}$ equals the average N conditional on a given wage bin, say. We used this approach to choose our preferred model: more flexible models that interacted the key covariates flexibly with year or with each other generated noisier predictions, and as a result the fit was worse.

To test the fit of our model with more granularity, we examine binscatters of expected values of N on the x-axis and adjusted postings on the y-axis in the in and out samples. This exercise tests in a non-parametric way whether our predictors can match expected variation in N . To calculate expected N we use the same

¹⁶Covariates include year fixed effects, 16 2-digit NAICS sector indicators, indicators for lagged average wage quintile, urbanicity category, age bins, and lagged employment growth rate bins, and a flexible specification of lagged employment size (indicators for size 2-5, 6-9, and ≥ 5000 , plus linear spline terms at 10, 50, 250, and 1000 that apply only below 5000). Time-varying location characteristics include quintiles of the unemployment rate, the share of residents with a bachelor’s degree or higher, the share of employment in professional occupations, and broadband penetration.

¹⁷Our estimate is closely related to the Poisson count model, but since we sometimes allocate BG postings across multiple establishments, the monthly number of postings can take non-integer values. Since we use these variables for predictive analysis, it is crucial that the predicted values respect the non-negativity constraints, or else the predictions may become implausible, especially when combined into the overall adjustment factor.

Figure S.1: Validating our Model in JOLTS



Notes: This figure presents binscatter plots, dividing JOLTS establishments into 20 evenly sized bins based on predicted N , and plotting average values of postings and adjusted postings within each bin. We predict N based on the full set of covariates in our model. Both samples are drawn from JOLTS. We split the JOLTS sample into thirds and estimate our model on each 2/3 sample. We predict fitted values based on these estimates for the full JOLTS sample where the remaining 1/3 contributes to the “out sample”. We repeat 3 times and aggregate to make up the full out sample.

covariates X as in our model. Figure S.1 shows for both the in and out samples that while the relationship between actual posts and expected N is quite muted (a slope of 0.15), our predictor achieves slopes close to 1. That is, our adjustments do a good job capturing expected variation in N , even in this non-parametric illustration.

Our adjustment factor applies on a sub-population level. We use establishments with postings as an indicator of what behavior would be like among establishments with similar covariates that do not have postings. Our adjustment then reweights the data to scale up establishments with less well represented characteristics. As such, we do not expect our estimates to predict well idiosyncratic establishment-level variation. And therefore goodness-of-fit measures such as R-squares are not informative. Instead, these validation exercises are reassuring that we are capturing predictable variation in JOLTS recruiting behavior well.

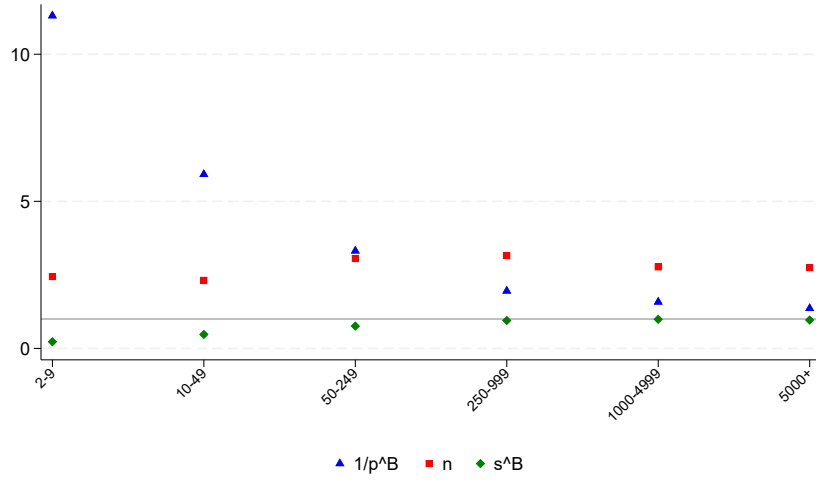
C Additional Tables and Figures

Table S.2: Average Skill Requirements in BG with and without our Adjustment

	Actual B	Adjusted B
<i>Occupation Distribution:</i>		
Professional occupations	0.45	0.38
Sales and administrative occupations	0.24	0.23
Service occupations	0.14	0.19
Blue collar occupations	0.12	0.16
<i>Skill Requirements:</i>		
Education	0.56	0.50
Experience	0.49	0.44
Cognitive	0.31	0.26
Social	0.48	0.43
Computer	0.37	0.30
<i>Salary Listing:</i>		
No Salary Listed	0.89	0.89
Average Salary, Conditional	49,405.29	46,234.43

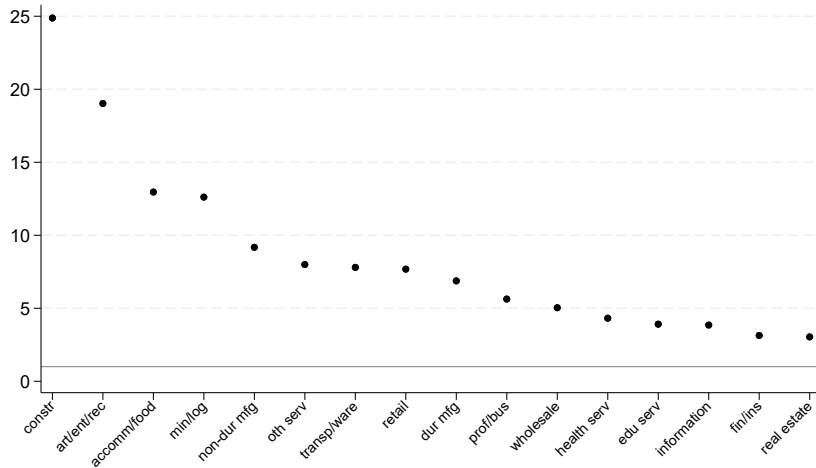
Notes: Postings-weighted summary statistics for 2011-2019 in the BG-QCEW sample, restricting to private sector establishments with at least 2 employees. Column 1 weights each ad equally. Column 2 applies our adjustment in equation 6 and uses these adjusted postings as weights. Occupation ad shares aggregate listed SOC occupation codes into 4 broad categories. Skill requirements include social and cognitive indices from [Deming and Kahn \(2018\)](#) as well as BG-created measures for education, experience, and computer requirements.

Figure S.2: Margins of Adjustment by Size



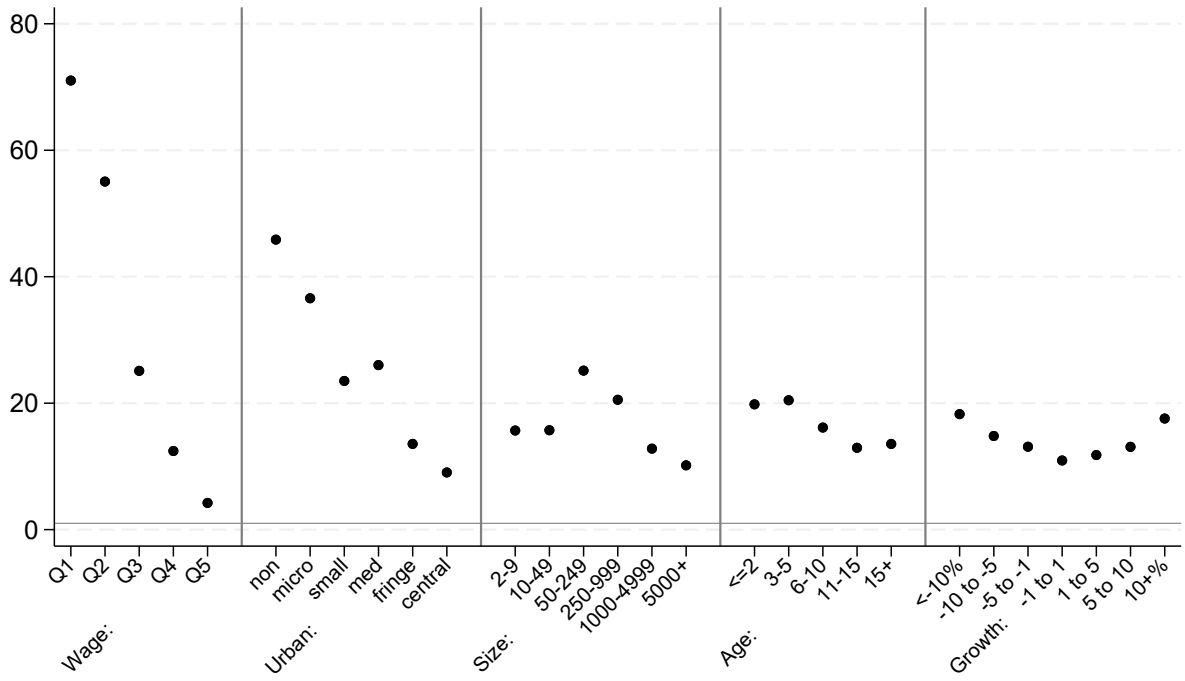
Notes: The figure shows the decomposition of the adjustment factor a in its three main components: blue triangles for (the inverse of) the extensive margin (p^B), red squares for the intensive margin (n) and green diamonds for the share of postings with non-zero N , reflecting measurement error (s^B). We calculate establishment-level averages within each year and then take an unweighted average across years from 2011-2019. We use the BG-JOLTS sample, restricting to private sector establishments with at least two employees.

Figure S.3: Adjustment Factor by Sector



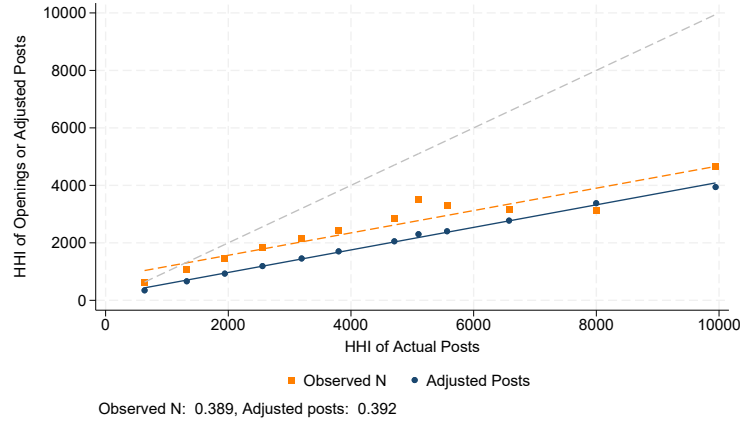
Notes: The figure shows the overall adjustment factor a for different two-digit NAICS industries. Values are averaged using JOLTS sampling weights within year and equal weights across years for 2011-19. Estimates are from the BG-JOLTS matched sample, restricting to private sector establishments with at least 2 employees.

Figure S.4: Adjustment Factor for $E[N^2]$



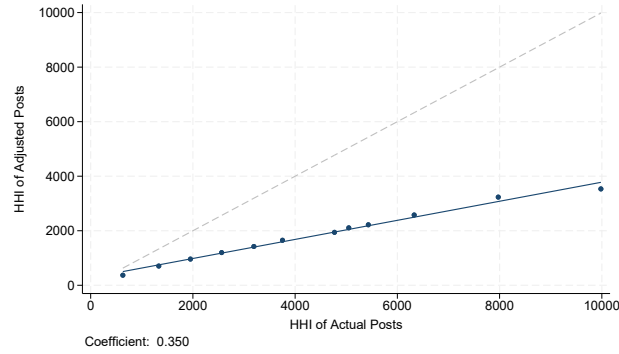
Notes: The figure plots the adjustment factor $a^{B,2}$ by establishment characteristics. See equation 13. Results are shown for the baseline sample with data from 2011 to 2019.

Figure S.5: JOLTS Validation of HHI Adjustment



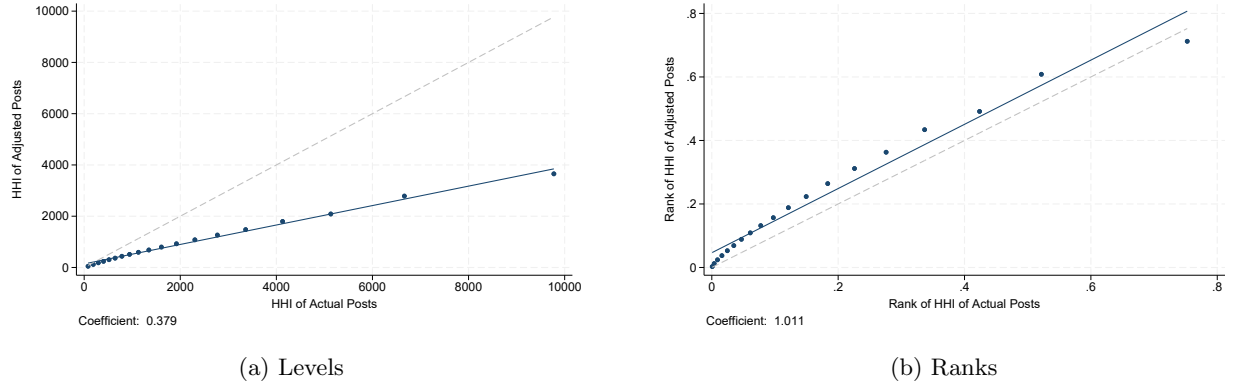
Notes: We provide evidence that our adjustment does a good job of matching the HHIs of actual openings, N . For validation purposes, we use the JOLTS to simulate markets matching the number and size in the QCEW. (Our main exercises are conducted in the larger QCEW, where we cannot observe actual openings.) Specifically, we randomly sample JOLTS establishments (using JOLTS sampling weights) with replacement to generate the number of markets in QCEW and matching the distribution of the number of establishments in each market. We then calculate HHIs of B , N , and adjusted B in these markets. We plot binscatters of these relationships based on the HHI of actual posts in the market. The fact that the relationship between HHIs of observed N and actual posts is approximated by that of adjusted posts validates our use of adjusted HHIs in the QCEW.

Figure S.6: HHI: Excluding Small Establishments



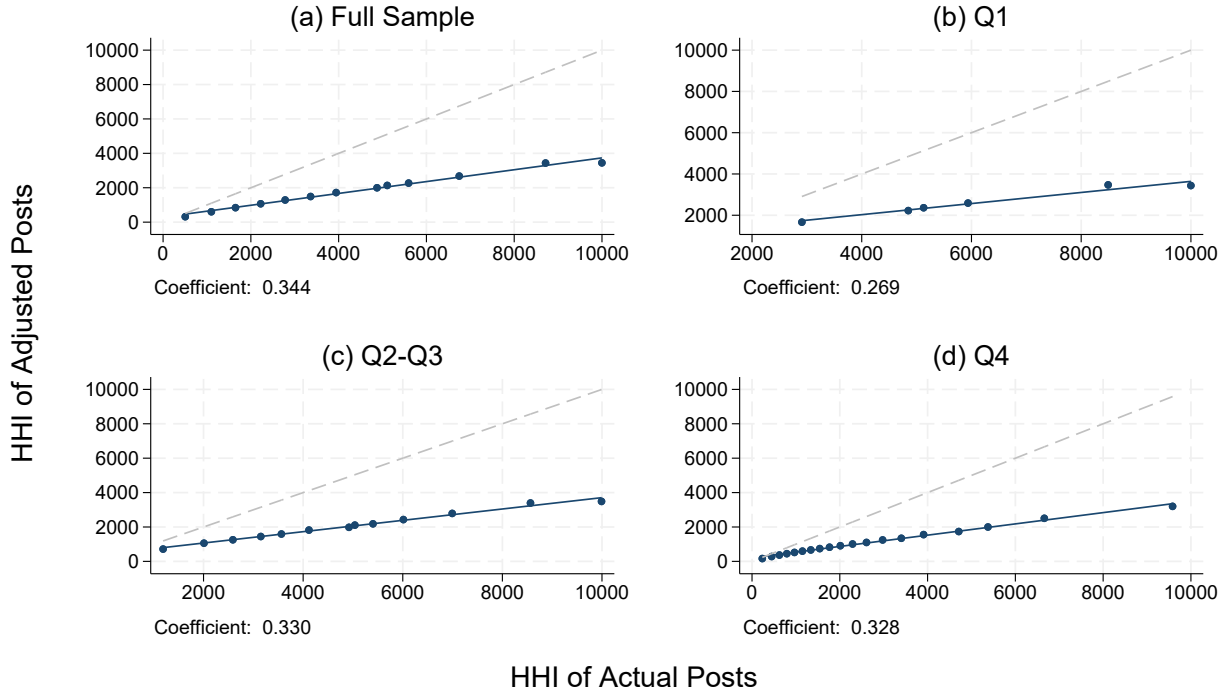
Notes: See figure 4. Here we restrict the sample to establishments with at least 10 employees.

Figure S.7: HHI: Employment Weighted



Notes: See figure 4. Here we aggregate market-month observations into 5-percent bins using employment weights and then plot employment-weighted average HHIs and ranks within each bin.

Figure S.8: HHIs of BG Postings: By Predicted Market Size



Notes: Binscatter of adjusted and actual HHIs in BG postings. We divide markets (defined at the CZ-by-4 digit NAICS-month level) into 5-percent sized bins based on the HHI of BG postings (equation 8) and then plot the average actual HHI and adjusted HHIs (equation 9) within each bin. We average monthly HHIs over 2011-19. Panel a shows the results for the full sample, whereas Panels b, c and d show the bottom quartile, the middle two quartiles combined, and the top quartile of adjusted predicted market size, respectively ($\hat{N}$).