# Employer Concentration and Wages: Evidence from Large Firms' Hiring Shocks

Gregor Schubert, Anna Stansbury, Bledi Taska, and Matthew Dey

July 31, 2025

We estimate the effect of increases in employer concentration on wages, using a new instrument for employer concentration based on changes in large firms' national hiring patterns. We measure employer concentration over mobility-adjusted labor markets: clusters of local occupations identified through asymmetric mobility patterns (using new, highly granular data on occupational mobility from 16 million resumes). We find that increased employer concentration causally reduces wages: moving from the median to the 95th percentile of employer concentration as experienced by workers lowers wages by 4 log points. Overall, we estimate that more than one in six U.S. workers face wage suppression of 2.5% or more as a result of above-median employer concentration. The effects of employer concentration are particularly pronounced in healthcare occupations.

**JEL Codes:** J31, J42, J62
**Keywords:** monopsony, labor markets, concentration, market definition

*G. Schubert: UCLA Anderson School of Management, gregor.schubert@anderson.ucla.edu; A. Stansbury: MIT Sloan School of Management, amms@mit.edu; B. Taska: SkyHive (formerly at Burning Glass Technologies/Lightcast); M. Dey: Bureau of Labor Statistics. This paper draws heavily on analyses from prior versions circulated as "Employer Concentration and Outside Options" and "Monopsony and Outside Options" The outside option analyses from these paper have been moved to a separate paper titled "Occupational Mobility and Outside Options". For their detailed comments and advice, we thank Justin Bloesch, Gabriel Chodorow-Reich, Karen Dynan, Richard Freeman, Xavier Gabaix, Ed Glaeser, Claudia Goldin, Benny Goldman, Larry Katz, Isaac Sorkin, Larry Summers, Maya Sen, and Betsey Stevenson. For helpful discussions, we thank David Balan, John Coglianese, Oren Danieli, David Deming, Martin Feldstein, Emma Harrington, Peter Hull, Simon Jaeger, Mike Lipsitz, Ioana Marinescu, Jeff Miron, Simon Mongey, Suresh Naidu, Nancy Rose, Elizabeth Weber Handwerker, Ron Yang, seminar participants at Harvard, MIT Sloan, Chicago Booth, Oxford, the Federal Trade Commission, the OECD, and the Peterson Institute for International Economics, the NBER Wage Dynamics in the 21st Century Conference, the briq Workshop on Firms, Jobs and Inequality, the Federal Reserve System Community Development Research Conference, the IZA Summer School on Labor Economics, the IZA/CAIS Workshop on Matching in Online Labor Markets, the Urban Economic Association Meetings, the Wharton People and Organizations Conference, the Asian and Australasian Society of Labour Economics Conference, the Oxford NuCamp Virtual PhD Workshop, the ASSA meetings, the Southern Economic Association meetings, the Equitable Growth Grantee Conference, UT Austin, University of Copenhagen, Copenhagen Business School, Wayne State University, Georgetown Law School, George Mason Law School, University of Virginia, Queen Mary University London, U Mass Boston, Brown, UC Santa Barbara, and the Chicago Fed. This research was supported by the Washington Center for Equitable Growth (Schubert & Stansbury) and the James M. and Kathleen D. Stone PhD Scholarship in Inequality and Wealth Concentration (Stansbury). 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

In recent years, concerns about employer concentration have increased. As a source of monopsony power, employer concentration has been posited as a possible explanation for inequality, low pay, and stagnant pay growth. Concerns around high employer concentration have bolstered calls to raise minimum wages and strengthen collective bargaining, and antitrust authorities have been called on to consider employer concentration in merger and acquisition reviews.[1] Yet the degree to which employer concentration suppresses wages—and for which workers—is still a subject of intense debate, as researchers confront methodological challenges in estimating the causal effects of employer concentration on wages.

In this paper, we estimate the causal effect of employer concentration on average hourly wages across a broad set of U.S. labor markets, contributing two important innovations. First, estimates of the relationship between employer concentration and wages are typically beset by endogeneity concerns: we make progress by developing a new plausibly exogenous instrument for employer concentration, based on changes in large firms' national hiring patterns. Second, measuring employer concentration requires defining the appropriate market: we propose a new asymmetric mobility-adjusted market definition, using novel and highly granular data on occupation-to-occupation flows, and show that this approach can be used for empirical practice in academia, antitrust, and policy. Overall, our estimates suggest around one in six U.S. workers experience non-trivial wage suppression as a result of above-median employer concentration, and that many of the most affected workers are in healthcare occupations. We outline our approach in more detail below.

Our core dataset is the near-universe of online vacancy postings, sourced from Burning Glass Technologies. We use these data to measure employer concentration as the Herfindhal-Hirschmann Index ("HHI") across vacancies for occupation-metro area labor markets covering over three-quarters of the US workforce for each year from 2011–2019. Our key outcome variable is hourly wages, obtained from the BLS Occupational Employment and Wage Statistics Program ("OEWS"). The question we seek to answer is: what is the causal effect of employer concentration on wages in US labor markets? Our approach brings two major innovations relative to prior work, which we discuss below: causal identification, and market definition.

To make progress on causal identification, we develop a new identification approach for the effects of employer concentration on wages. This approach is designed to address the concern that employer concentration is endogenously determined by firm fundamentals and

[1] See, variously, Bahn (2018); Shambaugh, Nunn, Breitwieser, and Liu (2018); Krueger and Posner (2018); Naidu, Posner, and Weyl (2018); Marinescu and Hovenkamp (2019); Marinescu and Posner (2020).

local competition, leading to bias in simple regressions of wages on measures of concentration. We draw on shift-share and granular IV methodology (Borusyak, Hull, and Jaravel, 2022; Gabaix and Koijen, 2024) to develop an instrument that captures plausibly exogenous shifts in employer concentration within a particular local occupation. This instrument can be applied to a broad set of occupationally and geographically diverse labor markets.

Our instrument is based on differential exposure to large firms' national hiring patterns. Specifically, we use the fact that large companies often expand their hiring in national waves – driven, for example, by national demand shocks or strategic shifts that are unrelated to local characteristics in any individual market where they have an existing presence. These national hiring waves lead to an increase in the company's employment share in its individual locations. Different local labor markets are differently exposed to these hiring changes, depending on the employer's pre-existing relative size in each of those markets: whether the expanding employer initially had a small or large presence in a market determines whether its national hiring push decreases or increases local employer concentration. These shocks from large firms' national hiring waves are plausibly orthogonal to local occupation-specific productivity, with the core identifying assumption being that each large firm's decision to increase its hiring nationwide is exogenous with respect to the economic conditions in any specific local occupation in question. Our baseline estimates also control for a rich set of potentially confounding trends by including occupation-by-year fixed effects and either metro area-by-year or occupation-by-metro area fixed effects. Moreover, we also control for proxies for changes in local labor demand that might arise as a result of the national hiring waves.

To make progress on market definition, we propose a new "mobility-adjusted labor market" definition (MALMA). We start by using new, highly-granular occupational mobility data constructed from 16 million US workers' resumes,[2] to show that occupational mobility is high, and highly heterogeneous across occupations, making consideration of worker mobility a first order issue. We then use this data to define the "mobility-adjusted labor market" (MALMA) for workers in each occupation. This is a market definition which includes all other occupations that are likely targets for worker moves. These MALMAs differ from other cluster-based approaches to labor market definition because they are asymmetric: occupation $p$ may be a relevant option for workers in occupation $o$, but not vice versa. This is an important distinction, informed by the asymmetry of actual occupation flows we observe in our new mobility data. All our estimates are based on measures of employer concentration calculated over these MALMAs.

[2]The large sample size – an order of magnitude more than other data sources – enables us to estimate occupational transitions reliably between a large share of US occupations. This new occupational mobility dataset is publicly available on our websites.

How much does employer concentration matter for wages? Our baseline IV results show a large negative causal effect of employer concentration on wages: our estimates suggest that going from the HHI faced by the median worker to the HHI faced by the worker at the 95th percentile in 2019 would lead to 4 log points lower hourly wages. We show that these estimates are robust to variations in the empirical specification, and control variables used; how units are weighted in the estimation; excluding non-tradable occupations; excluding labor markets that start out at low levels of concentration; and omitting occupations that are under-represented in our vacancy data. Moreover, we show that our findings are not merely an artefact of using online vacancies to measure employer concentration: using restricted-access micro data from the BLS' OEWS program, we show that we obtain qualitatively similar estimates when we use HHI measures based on the stock of employment as measured by the OEWS.

We use our estimates to ask how much employer concentration matters in the U.S. labor market, and for whom. Our approach is uniquely suited to do so, given (i) our causal identification, using variation from a large share of the U.S.' labor markets, and (ii) our mobility-adjusted approach to labor market definition, enabling us to identify which labor markets are more or less concentrated based on a realistic definition of workers' set of feasible job options. Our back-of-the-envelope calculation suggests that over one sixth of the 117 million workers covered by our data in 2019 experienced a non-trivial degree of wage suppression – of 2.5% or more – as a result of being in a labor market with higher-than-median employer concentration.[3] Many of the most-affected workers are healthcare workers, reflecting both high healthcare employment concentration and low mobility into other non-healthcare occupations.

Our estimates of the causal impact of employer concentration for a large majority of US workers provide important evidence to the academic and policy debate on this topic. They point to a middle ground between two common views: when estimated over well-defined labor markets and with credible causal identification, the effects of employer concentration are not irrelevant—e.g. confined to a few factory towns—nor are they a major factor in wage suppression for the majority of US workers. Moreover, our methodologies for mobility-based labor market definition represent pragmatic improvements for future analyses of labor market power and antitrust policy – as well as more broadly in studies of labor market outcomes wherever market definition is important, such as studying the transmission of economic shocks or the effects of labor market regulations.

**Related literature.** We build on a growing body of work on employer concentration, which began in recent years with Azar, Marinescu, Steinbaum, and Taska (2020b), Azar,

[3]Measured as the mobility-adjusted HHI faced by the median worker in our data, which was around 74.

Marinescu, and Steinbaum (2020a), Benmelech, Bergman, and Kim (2022), and Rinz (2022). We make two key contributions to this literature. First, we develop a new instrument to estimate plausibly causal negative effects of employer concentration on wages. This is an important innovation in a context where most evidence is subject to concerns about omitted variable bias arising from local occupation-specific productivity or other economic conditions. In this way, our work complements a small recent literature using M&A activity or deregulation as shocks to labor market concentration (Arnold, 2020; Prager and Schmitt, 2021; Guanziroli, 2022; Thoresson, 2024; Farag, Compton, Steinbaum, Abdelfattah, and Stansbury, 2024).[4] Second, in using a new mobility-adjusted labor market definition (MALMA), we address the market definition challenge, tackling the misspecification problem that arises when administrative labor market boundaries do not capture workers' true labor markets (and therefore do not capture the true degree of employer concentration they face). Finally, we use our causal estimates to quantify the rough magnitude of the employer concentration problem across most US occupations. As we discuss in more detail in section 2, our approach addresses the key critiques of regressions of wages on local employer concentration (e.g. Berry, Gaynor, and Scott Morton, 2019; Rose, 2019), related to concerns that employer concentration is endogenously determined and that there is no appropriate definition of a market on which a meaningful concentration index can be calculated.

Our use of occupation flows to define labor markets also relates to a growing literature which uses worker flows to identify the scope of workers' labor markets (e.g. Shaw, 1987; Manning and Petrongolo, 2017; Nimczik, 2018). Our key innovation relative to these papers is to define labor markets asymmetrically.

Finally, we contribute to a literature on the range of factors which may impede competition in labor markets, including non-compete agreements (Johnson, Lavetti, and Lipsitz, 2025) and occupational licensing (Johnson and Kleiner, 2020), as well as the broader literature on monopsony power (e.g. Manning, 2003).

**Roadmap.** We lay out our empirical approach in section 2, explaining our data and independent variables of interest. In section 3, we discuss our instrument for employer concentration. In section 4, we show our baseline results and explore their robustness to different specifications and estimation strategies. We discuss the implications of our results for the aggregate economy and for policy in section 5, and conclude in section 6.

[4] Note that these estimates from M&A activity leverage a very different source of variation to that in our instrument – only 2% of changes in employer concentration come from M&A activity (Arnold, 2020) – and are subject to concerns about other simultaneous labor market effects of the M&A.

## 2 Empirical Approach

Why might employer concentration matter for wages? A growing theoretical literature outlines a clear causal relationship between higher employer concentration and lower wages across different types of labor market models. In a Cournot oligopsony model of the labor market, firms' market shares determine the slope of their labor supply curves: larger market shares increase firms' market power, enabling them to pay wages which are marked down below the marginal product (e.g. Arnold, 2020; Berger, Herkenhoff, and Mongey, 2022). Similarly, in a search-and-bargaining framework, employer concentration puts downward pressure on wages because it reduces the average value of workers' outside option in the wage bargain (Jarosch, Nimczik, and Sorkin, 2024).[5] In these models, higher levels of employer concentration causally reduce average wages ceteris paribus, and simple indexes of employer concentration are therefore important determinants of the wage.[6]

These theoretical results have, in turn, motivated a rapidly growing empirical literature on the relationship between wages and employer concentration. Specifically, a growing literature documents negative cross-sectional wage-HHI correlations, and negative associations between wages and HHIs over time (e.g. Azar et al., 2020a; Benmelech et al., 2022; Azar et al., 2020b; Rinz, 2022). Policymakers over recent years in the US and other countries have increased their focus on labor market competition in general, and on employer concentration as one axis of labor market competition, calculating HHI indices and using these to extrapolate potential wage-suppressive effects of diminished competition in concentrated markets (e.g. U.S. Department of Justice and Federal Trade Commission, 2023; UK Competition Authority, 2024)

Yet, while a negative wage-HHI correlation may be consistent with a causal effect of employer concentration on wages, there are important concerns with running simple cross-sectional regressions of outcomes like prices or wages on measures of market concentration – as raised in older critiques of the structure-conduct-performance paradigm (e.g. Schmalensee, 1989) as well as in more recent work. For example, Rose (2019) argues that empirical strate-

[5]Notably, these papers address the theoretical critique of the structure-conduct-performance paradigm that there may not always be a clear mechanism linking changes in labor market concentration to changes in wages. For example, Berry et al. (2019) emphasize that concentration measures like the HHI are an outcome, not a structural primitive: concentration and wages can be simultaneously determined by unobserved productivity, search frictions or labor supply shocks. Jarosch et al. (2024) tackle this directly, showing in a search-and-bargaining model that there is a clear conceptual channel by which labor market concentration will always ceteris paribus exert downward pressure on wages, as it reduces the value of workers' outside option set.

[6]For example in Arnold (2020), the relevant index is an HHI across firm employment shares; in Berger et al. (2022), the relevant index is an HHI across firms' wage bills; in Jarosch et al. (2024) the relevant concentration index is a higher-order concentration index across firms' employment shares, which becomes the HHI to a second-order approximation.

gies attempting to identify a causal effect of employer concentration on wages must isolate the effect of employer concentration from changes in labor demand and labor supply for other reasons. Syverson (2019) suggests one solution to this issue, noting that "concentration might be instrumented using [direct measures of plausibly exogenous differences in competition]." One core contribution of our paper is in addressing this critique: our identification strategy, outlined in Section 3, uses national hiring shocks to construct an instrument that captures plausibly exogenous variation in local labor market concentration.

Moreover, the large literature (and body of policy practice) on concentration in product markets must contend with concerns over the appropriate definition of a market for which a meaningful concentration index can be calculated (Syverson, 2019). The second core contribution of our paper is to directly address the analogue of this concern for labor markets: we explicitly take into account that workers' empirical labor markets are broader than their current occupation, using empirical worker mobility flows to dervelop an asymmetric concept of a "mobility-adjusted labor market" (MALMA), as outlined in detail in Section 2.2.

## 2.1 Baseline specification

Our baseline empirical specification regresses the log of the average hourly wage $\bar{w}_{o,m,t}$ on the log of employer concentration $HHI_{o,m,t}$:

$$\ln \bar{w}_{o,m,t} = \alpha_{o,m,t} + \gamma_1 \ln HHI_{o,m,t} + \xi_{o,m,t}, \tag{1}$$

where subscripts refers to the metro area ($m$), occupation ($o$), and year ($t$). $\alpha_{o,m,t}$ are a vector of fixed effects. In our baseline specification, we include occupation-by-year and metro area-by-year fixed effects, which control for any national occupation-level economic shocks and any metro-area-level economic shocks which might affect both concentration and wages. In an alternative specification, we instead include metro area-by-occupation fixed effects and occupation-by-year fixed effects, controlling for market-level factors and identifying from changes within a given occupation-by-metro area labor market over time.

Our baseline labor market definition is an occupation $o$ by metro area $m$. For occupation $o$, we use a mobility-adjusted labor market ("MALMA"), which we construct using occupational mobility flows to cluster together groups of similar SOC 6-digit occupations (explained in section 2.2). Throughout this paper, when we use the term "occupation", we are referring to a MALMA occupation cluster unless otherwise noted. Our wage measure $\bar{w}_{o,m,t}$ is the average hourly wage for all workers in a given occupation, metro area, and year, from BLS OEWS data.[7] Our employer concentration measure, the Herfindahl-Hirschmann Index or

[7]Note that, in keeping with much of the literature on labor market impacts of monopsony power, we focus

HHI, is constructed as the sum of the squared shares of each employer's vacancy postings within a given occupation, metro area, and year, using Burning Glass Technologies' vacancy posting data (discussed in section 2.3).[8] Our baseline time period is 2011–2019, which is the period for which we have the BGT vacancy posting data.

## 2.2 Labor market definition

In this paper, we define the boundaries of a labor market by occupation and geography. For our geographical boundaries, we use metro areas.[9] For our occupational boundaries, we construct a mobility-adjusted labor market ("MALMA") using occupation flows data.

**Occupation flows data.** We measure occupation-to-occupation flows at the SOC 6-digit level in a new data set of occupational transitions, which we constructed using 16 million unique US resumes from Burning Glass Technologies ("BGT").[10] These resumes enable us to observe longitudinal snapshots of workers' job histories over 2002–2018. Using these data, we construct the occupation transition share $\pi_{o \to p}$, defined as the share of consecutive year-occupation pairs where the worker is observed in occupation $o$ in year $t$ and in occupation $p$ in year $t+1$ (as a share of all workers observed in occupation $o$ in year $t$):[11]

$$\begin{aligned} \pi_{o \to p} &= \frac{\#\text{ in occ } o \text{ in year } t \text{ observed in occ } p \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t} \\ &\approx Prob(\text{move from occ } o \text{ to occ } p). \end{aligned} \tag{2}$$

Using this transition share, we document five stylized facts about US occupational mo-

on wages as our key outcome for workers. Firms may change other components of a worker's overall compensation package in response to changes in competition (Lamadon, Mogstad, and Setzler, 2022). We have no reason to believe that increased concentration would cause firms to increase amenities while decreasing wages; instead, it seems most likely from theory that amenities would decline alongside wages. Our wage effects can be thought of as proxying more broadly for the effects of employer concentration on overall compensation to the extent that firms adjust amenities proportional to their adjustments on wages; future work with data on non-wage amenity margins of adjustment would be valuable.

[8] Burning Glass Technologies is now known as Lightcast.

[9] We use "metro areas" to refer to CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas).

[10] BGT sourced the resumes from a variety of partners, including recruitment and staffing agencies, workforce agencies, and job boards. We construct our own data rather than using existing data because the largest publicly available dataset with occupation flow information – the Current Population Survey – has at least an order of magnitude fewer occupational transition observations than our resume data. This matters: with 705,600 possible transition pairs between SOC 6-digit occupations, data sets with even a few million observations are not big enough to capture many transition paths.

[11] We exclude from the denominator any observations in occ $o$ in year $t$ where the worker is not observed at all in year $t+1$. To correct for the over-representation of younger workers in the BGT resume data, we re-weight our observed occupational transitions to match the age distribution of employment within each occupation. We discuss the BGT resume data and the construction of the transition share in more detail in Appendix E.

bility which support our MALMA approach: (i) Occupational mobility across SOC 6-digit boundaries is high;[12] (ii) Occupational mobility is highly heterogeneous across occupations;[13] (iii) Aggregating up the hierarchy to SOC 3-digit occupations does little to capture additional transitions, while at the same time incorporating many irrelevant job options;[14] (iv) The occupation transition matrix is sparse, suggesting that workers' relevant labor markets are typically comprised of only a few SOC 6-digit occupations; and (v) Occupation flows are highly asymmetric. Together, this set of stylized facts suggests that the use of administrative occupational boundaries is inappropriate for labor market analysis, but that occupational transitions can be used to create small asymmetric clusters of similar occupations which are a better reflection of workers' true labor markets. We use these insights in constructing our MALMA labor markets.

**Constructing MALMAs (Mobility-Adjusted Labor Markets).** For each SOC 6-digit occupation $o$, we define the MALMA as a cluster of occupations which includes occupation $o$ and any other occupations $p$ which are relevant outside options for workers in occupation $o$. These occupations $p$ are identified as the occupations which are frequently transitioned-to by workers in occupation $o$. Specifically, the MALMA for a worker in SOC 6-digit occupation $o$ is defined as the set of all job vacancies in metro area $m$, in the occupations $p$ that are transitioned-to by workers in occupation $o$ at a rate above a chosen threshold $x$:

$$MALMA_{o,m,t} = \bigcup_{p:\ \pi_{o\to p} > x} v_{i,p,m,t}, \tag{3}$$

where $v_{i,p,m,t}$ is the set of all vacancies of firms $i$ in occupation $p$ in metro $m$ at time $t$, and $\pi_{o\to p}$ is our empirical transition share, which proxies the share of workers in occupation $o$ who transition to occupation $p$. For our analyses, we set a threshold of $x = 2\%$. This means that any SOC 6-digit occupation $p$ is included in occupation $o$'s MALMA if the annual transition share from $o$ to $p$ is greater than 2%.[15] For ease of exposition, we will simply refer to these

[12] The average $\pi_{o\to o}$ is 67%, meaning that 33% of consecutive occupation-year observations in our resume data involve an occupation transition between 6-digit SOCs. Note that this measure is *not* directly analogous to other annual occupational mobility measures – it will provide a higher estimate of mobility – because if a worker switches occupation at *any point* during year $t$ or $t + 1$ she will be counted in the numerator. We compare our measures of occupational mobility to those in the CPS in Appendix E.

[13] The share leaving their occupation each year is almost twice as high for high-mobility occupations like fast food cooks, retail salespersons, or personal care aides, than it is for low-outward-mobility occupations with occupation-specific skills like pharmacists, lawyers, or electricians. Appendix Figure B1 illustrates this heterogeneity.

[14] The average $\pi_{o\to o}$ at the SOC 3-digit level is less than 70% – only very slightly higher than the average $\pi_{o\to o}$ within 6-digit occupations (Appendix Figure B1).

[15] Intuitively, a $\pi_{o\to p}$ of 2% means that, of all the workers in occupation $o$ in year $t$, around 2% of them are observed in a job in occupation $p$ at some point in year $t$. (Note that most workers, even when they change job, stay in the same 6-digit occupation, and $\pi_{o\to p}$ captures this for $p = o$).

MALMAs from the perspective of a specific occupation $o$ as "occupations" throughout the paper and will note explicitly when we are talking about occupations defined by 6-digit or 3-digit administrative boundaries instead.

What do these MALMAs look like? For the typical SOC 6-digit occupation, the MALMA incorporates that SOC 6-digit occupation itself plus one to three others.[16] The MALMA for Cashiers, for example, also includes Retail Salespersons and Customer Service Representatives; the MALMA for Accountants and Auditors also includes Financial Managers; and the MALMA for Software Developers (Systems) also includes Software Developers (Applications), Computer Programmers, and Computer Occupations (All Other).

**MALMAs vs. administrative labor markets.** The ideal labor market definition captures a large share of workers' true outside options ("completeness"), while also ensuring that few of the jobs within the labor market are irrelevant for workers ("relevance"). Typically, there is a trade-off between the two: expanding the scope of a labor market may improve completeness at the cost of relevance. In Appendix B, we show that MALMAs improve on SOC 6-digit labor markets in terms of their completeness, but at the cost of lower average relevance. In contrast, when compared to SOC 3-digit labor markets MALMAs weakly dominate across these criteria: MALMAs are on average more relevant than their analogous SOC 3-digit occupations (containing fewer irrelevant job options), while being no less complete. Thus, we see MALMAs as an improvement upon existing administrative labor market definitions. Nonetheless, we also replicate our results for SOC 6-digit and SOC 3-digit labor markets in Appendix B.

**Asymmetric vs. symmetric labor market definition.** Some recent papers also construct labor markets based on mobility data (e.g Schmutte (2015); Nimczik (2018); Arnold (2020)). A key difference in our approach, which we believe is essential for antitrust applications and the study of labor market power, is the asymmetry of our MALMA occupation clusters: the MALMA for occupation $o$ may not contain occupation $p$, even if the MALMA for occupation $p$ contains occupation $o$. Why is this asymmetry important? Flows from occupation $o$ to another occupation $p$ can be thought of as representing some combination of the feasibility and desirability of that transition. Often, either the feasibility or desirability of a move is much greater in one direction than in the other: for example, in cases where a move requires more specialized skills or training in one direction, or in cases where one of the two occupa-

[16]For 65 SOC 6-digit occupations, the MALMA incorporates five or more occupations. Meanwhile, for 210 of the 789 SOC 6-digit occupations in our data, the MALMA is simply the focal SOC 6-digit occupation $o$ itself: these occupations either have very low degrees of outward mobility and/or workers who leave these occupations move to a very dispersed set of other jobs, meaning that no other occupations are important enough outside options to meet the 2% mobility threshold.

tions pays better or reflects a career progression.[17] Defining a labor market as a symmetric cluster of occupations – failing to allow for this asymmetry – means frequently incorporating irrelevant job options into the labor market definition, which will affect conclusions about the scope of workers' labor market and therefore the degree of concentration they face.

**Geographic mobility.** Our market definition approach in this paper allows for mobility across occupation boundaries but not across geographical boundaries: we simply use existing metropolitan areas. This is because geographic mobility is substantially lower than occupational mobility: using the 2013-17 ACS we estimate that the share of people who leave their metro area annually is less than 4%. Indeed, other researchers have frequently argued that metro areas are too big, rather than too small, as measures of workers' relevant labor markets, particularly for low wage workers (Bartik, 2024), and so ideally we would use a smaller geographic definition – but we cannot, because the metro area is the smallest unit for which we can obtain occupation-level wage data.

## 2.3 Measuring employer concentration

To measure employer concentration, we use Burning Glass Technologies' database of online vacancy postings, following Azar et al. (2020b). The BGT vacancy posting data covers the near-universe of online job postings, drawn from over 40,000 sources including company websites and online job boards, with no more than 5% of vacancies from any one source (Hazell and Taska, 2019). We calculate the Herfindahl-Hirschman Index (HHI) of each employer's share of vacancy postings within each occupation-by-metro labor market, in each year 2011–2019:

$$HHI_{o,m,t} = \sum_{i=1}^{N} \left( \frac{v_{i,o,m,t}}{\sum_{i=1}^{N} v_{i,o,m,t}} \right)^2 \tag{4}$$

where $v_{i,o,m,t}$ denotes the number of vacancy postings from employer $i$ in occupation $o$ and metro area $m$ in year $t$. The occupation $o$ is the MALMA, as defined above (such that the HHI is calculated across the cluster of SOC 6-digit occupations which make up the MALMA for the SOC 6-digit occupation $o$ in question).[18]

[17] This is quantitatively important for many occupations. For example, Appendix Table B2 shows that licensed vocational nurses are likely to become registered nurses, HR assistants often become HR specialists, and legal secretaries become paralegals—but moves in the reverse direction are rare, as these moves represent career progressions that normally do not operate in reverse. Specialized skills or training also affect the feasibility of a move: for example, instances where carpenters become truck drivers are more likely than the reverse. And some asymmetric flows can be explained by differences in desirability. For instance, cashiers are more likely to become customer service representatives than the other way around, likely in part due to the fact that the latter occupation earned 63% more per hour in 2012 than the former.

[18] Formally: $HHI^{\text{MALMA}}_{o,m,t} = \sum_{i=1}^{N} \left( \frac{v_{i,p,m,t} \times \mathbb{1}[\pi_{o \to p} > x]}{MALMA_{o,m,t}} \right)^2$.

There are two possible concerns in terms of the representativeness of online job vacancies. First, not all vacancies are posted online. Azar et al. (2020b) estimate that in 2016, the BGT vacancy database captured around 85% of all job vacancies both online and offline (as measured in the Help Wanted Online database), but this is lower for occupations where a large share of jobs are advertised offline or informally. Second, in occupations where firms tend to hire many workers for each posted vacancy, our estimates of employer concentration will be biased to the degree that larger firms may hire more people per vacancy posting.[19] A recent comparison of BGT vacancy postings to JOLTS job openings micro data by Dalton, Kahn, and Mueller (2025) assuages these concerns to a large degree: they find that while measurement error in BGT data affects the level of estimated HHIs, relative HHIs across labor markets are well-measured within the BGT vacancy data. Nonetheless, to understand the degree to which these measurement error concerns may be an issue in our setting, we calculate a measure of 'represented-ness' of each occupation in the BGT data: the occupation's share of vacancy postings in the BGT database relative to the occupation's share of total employment (from BLS OEWS). Occupations which are particularly underrepresented include low-wage food service, cleaners, home health aides, laborers, and cashiers. To ensure that our estimates of the effect of employer concentration on wages are not driven by this, we include occupation-by-year fixed effects in our baseline specification, and also show in robustness checks that our results are robust to excluding underrepresented occupations. For further discussion of the BGT vacancy data, see Appendix D.

## 2.4 Summary statistics

Our main analysis sample – the set of occupation-metro area-year observations with information on wages and employment (from BLS OEWS), and the HHI and HHI instrument (from BGT vacancy data) – contains 652 occupations and 387 metro areas, and a total of 87,716 distinct occupation-metro area cells. Not all occupation-metro area cells have observations for all 9 years in our sample. In total, the occupation-metro area cells in our sample represent a little over a third of the US workforce (58.8 million workers as of 2019).[20]

[19] If small firms or households are disproportionately less likely to post vacancies online, which seems likely, we will overestimate employer concentration for underrepresented occupations. If large firms hire more workers per job posting than small firms, we will underestimate concentration in labor markets with skewed employer size distributions, relative to those with more symmetric employer size distributions. Our measures of employer concentration are therefore less reliable for occupations for which households are a large employer, and for occupations for which there are many large employers who hire a lot of workers for undifferentiated job roles.

[20] We have HHIs, wage, and employment information for over three-quarters of the US workforce, but can only calculate our HHI instrument for this smaller sample.

Table 1 shows the distribution of wages, employment, and HHIs in our data.[21]

Employer concentration varies widely across occupation-metro area labor markets (Panel A), with the 5th percentile having an HHI of only 21 (equivalent to almost 500 equal-sized employers in the labor market), while the 95th percentile has an HHI of 2,448 (equivalent to 4 equal-sized employers in the labor market). There is less variation in the employer concentration experienced by workers (Panel B), because the smallest labor markets tend to be the most concentrated. As a result, the 95th percentile worker is in a labor market with an HHI of 830 (equivalent to 12 equal-sized employers in the labor market).

# 3 Identification: shocks to employer concentration

As discussed in Section 2, employer concentration and wages are jointly determined: supply, demand, and other factors affect both the presence and size of the employers in any given labor market, as well as wages. To isolate the causal arrow from employer concentration to wages, we therefore need plausibly exogenous variation in local employer concentration.

We develop a novel instrument that induces exogenous variation in local labor market concentration by leveraging differential local occupation-level exposure to large national firms' hiring growth. Our strategy can be interpreted as a shift-share 'Bartik' instrument or through the lens of the "granular" instrumental variable approach (GIV) of Gabaix and Koijen (2024). Our instrument relates most closely to a small recent literature which uses national company mergers as shocks to local concentration (Arnold, 2020; Prager and Schmitt, 2021; Guanziroli, 2022; Farag et al., 2024). Conceptually, one can think of our approach as a continuous analog of these discrete merger IV designs, identifying exogenous changes in concentration from changes in local employment shares driven by national hiring campaigns, rather than from changes in local employment shares driven by national mergers between firms. In the rest of this section, we explain the logic and construction of our instrument, and outline the identification conditions.

## 3.1 Instrument construction: Large national hiring shocks

The logic of our instrument comes from four facts: (a) increases in local employer concentration are often driven by individual large firms growing, (b) these firms usually operate across many labor markets, (c) local labor markets are differentially exposed to different large firms, and (d) the employment growth of these large firms nationally is likely orthogonal to economic conditions in a specific local occupation.

[21] The table also shows that our sample is closely representative of the full US workforce in terms of the wage distribution (Panel E).

Specifically, we note that the growth in local employer concentration in occupation $o$ is a function of the growth in local occupational employment for each employer $j$, $g_{j,o,m,t}$ (leaving aside firm entry): $\Delta HHI_{o,m,t} = \sum_j \sigma^2_{j,o,m,t} - \sum_j \sigma^2_{j,o,m,t-1} = \sum_j \sigma^2_{j,o,m,t-1}\left(\frac{(1+g_{j,o,m,t})^2}{(1+g_{o,m,t})^2} - 1\right)$. The increase in local occupational employer concentration is a function both of initial concentration and of the growth rates of firm-level vacancies $g_{j,o,m,t}$ relative to overall vacancy growth in the labor market $g_{o,m,t}$. For firms with a presence across multiple metropolitan areas, we instrument for the vacancy growth for each firm $j$ in occupation $o$ and metro area $m$ with the national vacancy growth of that firm $j$ in occupation $o$, leaving out the metro area in question $m$, (which we denote $\tilde{g}_{j,o,t}$).[22] Our instrument for the log HHI, $Z^{HHI}_{o,m,t}$, is therefore:

$$Z^{HHI}_{o,m,t} = \ln\left(\sum_j \sigma^2_{j,o,m,t-1}\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,m,t})^2} - 1\right)\right) \tag{5}$$

where $\tilde{g}_{o,m,t} = \sum_j \sigma_{j,o,m,t-1}\tilde{g}_{j,o,t}$ is the predicted local growth rate in vacancies, as predicted from the national (leave-one-out) growth of hiring in occupation $o$ by each large national firm $j$.

Through the lens of shift-share instruments (Borusyak et al., 2022), our instrument features plausibly exogenous 'shifts' (a function of firms' national hiring growth), and possibly endogenous exposure 'shares' (the last-period local occupational vacancy shares of each of those firms). Through the lens of granular instrumental variables (Gabaix and Koijen, 2024), our strategy uses plausibly exogenous idiosyncratic firm-level variation to instrument for changes in market-level aggregate concentration.

Two key assumptions for our instrument to be valid are that the firm's national leave-one-out vacancy growth is (i) correlated with its local vacancy growth (the relevance assumption), but (ii) uncorrelated with the other determinants of occupation-specific wage growth in any given metro area $m$, conditional on our occupation-year and metro area-year fixed effects (the exclusion restriction)—or the occupation-metro and occupation-year fixed effects we use in an alternative specification. We examine these assumptions in the following sections.

[22] Specifically, we only use firms $j$ with presence in at least 5 metro areas in that occupation nationwide, on the premise that these firms are sufficiently geographically dispersed that their overall nationwide hiring decisions are unlikely to be driven by local economic trends in metro area $m$ and occupation $o$. We also only use positive shocks (hiring growth), and only use intensive margin shocks (excluding changes to or from zero). Note that we are instrumenting for the local level of the HHI with an instrument derived from an expression for the change in the HHI, and by taking the log we exclude observations where the predicted change in HHI is negative.

## 3.2 Motivation: large firm hiring shocks

We start by considering the theoretical context in which our instrument would be valid. The motivation behind our instrument is the existence of national firm-level labor demand shocks which are unrelated to the occupation-specific economic conditions in a particular local area. More specifically, assume firm $i$'s growth in labor demand (growth in posted vacancies $g_{i,o,m,t}$) in occupation $o$ and local area $m$ is a function of national occupation-level factors $\nu_{o,t}$, local metro area-level factors $\nu_{m,t}$, national firm-level factors $\nu_{i,t}$, and local firm-occupation-specific idiosyncratic factors $\nu_{i,o,m,t}$:

$$g_{i,o,m,t} = \nu_{o,t} + \nu_{m,t} + \nu_{i,t} + \nu_{i,o,m,t}. \tag{6}$$

Our baseline fixed effects hold constant occupation-level and metro area-level components of labor demand which are common to all firms, meaning that we exclude components $\nu_{o,t}$ and $\nu_{m,t}$. Our leave-one-out identification strategy, where we instrument for the growth of local labor demand $g_{i,o,m,t}$ with the average of the firm's national labor demand in other metro areas, leaving out metro area $m$ ($\tilde{g}_{o,m,t}$), means that we also exclude any local firm-occupation-specific idiosyncratic shocks $\nu_{i,o,m,t}$. Thus, our identification isolates the effect of national firm-level shocks to labor demand $\nu_{i,t}$. This could include, for example, shocks to productivity (determined by national technology investments, IP, or production processes and systems), or shocks to national product demand (perhaps a function of national brand recognition or advertising campaigns). These national firm-level shocks would be expected to increase firm $i$'s hiring nationwide relative to other firms in the same occupation, and therefore to also increase firm $i$'s hiring in any given local area $m$ relative to other firms in the same occupation, for reasons that are orthogonal to specific labor supply, demand, or productivity changes in occupation $o$ and metro area $m$.

For intuition, consider a hypothetical example: assume that in Bloomington, IL, State Farm is the biggest employer of insurance sales agents, while in Amarillo, TX, employment is more concentrated in other insurance companies. If State Farm's brand improves nationwide – for example, through a successful advertising campaign – such that demand for its products increases, it will grow faster than other insurance companies. Under most assumptions about how that growth is allocated geographically, employer concentration of insurance sales agents will grow by more in Bloomington, IL, than in Amarillo, TX, for reasons entirely orthogonal to specific local labor supply or demand conditions for insurance sales agents in Bloomington, IL.

**Large hiring shock examples.** What are some specific examples from our data of these national firm-level growth shocks? In Table 2 we show examples of large employer-occupation pairs in our BGT vacancy data, and their growth over 2018-19. The table illustrates that the large hiring shocks in our data are reflective of both broader economic trends and idiosyncratic firm-level factors (although, note that our occupation-year fixed effects will remove the effect of any common economic trends from our main results). For example, Panel A (focusing on tradable occupations) shows that IBM, Deloitte, and Amazon all hired for a large number of new positions in computer occupations and software development, likely due to national growth in these companies; this national growth would be more likely to increase concentration in locations where these large companies usually hire their technical staff. Panel B of the table (focusing on non-tradable occupations) shows that large hiring shocks for specific occupations can also arise in association with broader secular trends, such as the expansion of coffee chains like Starbucks and franchise restaurants like McDonald's, and the rise of dollar stores like Dollar General. As a result, the local areas that are ex ante more exposed to the particular companies driving those trends will see exogenously larger changes in local employment concentration in the affected occupations.

## 3.3 Identification: relevance condition

Relevance of the instrument requires that these national firm demand shocks predict firms' local hiring decisions. This is testable in our data: the vacancy growth rate of firm $j$ in occupation $o$ in all metro areas not-$m$ strongly positively predicts the vacancy growth rate of firm $j$ in occupation $o$ and metro area $m$.[23]

This relationship between a firm's national hiring and its local hiring may be most clearly justifiable in occupations which produce tradable products or services, where demand is primarily driven by national and not local factors, and therefore where a national demand shock for the firms' products will induce increased production, increasing employment across their facilities. In a robustness check we therefore show that our main wage-HHI results hold when limiting our sample to tradable occupations only. But even in non-tradables, there are reasons to expect the relevance condition to hold: changes in company productivity or strategy may affect hiring decisions even if demand is primarily locally determined. For instance, Giroud and Mueller (2019) find that in non-tradable industries, consumer demand shocks in the headquarters region of a given company affect establishment-level employment in even distant locations of that company. Among non-tradable occupations, a prominent example of a large national firm growing by increasing its employment in its existing markets is the diffusion of Wal-Mart stores across the U.S.: Holmes (2011) documents that new Wal-

[23]Specifically, a regression of the latter on the former gives a coefficient of 0.35 (standard error 0.0006), for the sample of firms on which we construct the instrument (firms with vacancies in 5 or more metro areas in the occupation $o$ in question).

Mart stores diffused slowly, radiating outward from Bentonville, AK, because economies of density meant that when Wal-Mart expanded it did so by opening new stores and increasing employment at nearby existing stores. This "saturation strategy" for expansion meant that Wal-Mart increased its share of the local labor market near its existing locations over time (Neumark, Zhang, and Ciccarella, 2008), which then would have increased employer concentration if Wal-Mart already had a substantial share of local employment.

Relevance of the instrument also requires that greater exposure to these large firms' hiring shocks actually does increase employer concentration: that the first stage holds. A sufficient condition for this, under *most* initial employer share distributions, is that firm $j$'s new vacancies are allocated proportionally across occupation-metro area labor markets, such that each occupation-metro area labor market sees the same growth rate in its firm $j$ vacancies as the national average.[24] However, this condition is not necessary: in fact, the first stage can be valid even if the growth rate of firm $j$'s new vacancies in low-initial-employment-share occupation-metro area labor markets is higher than in high-initial-employment-share labor markets, as long as this relationship is not too strong. Empirically, our first stage holds strongly, as we discuss in section 4.1.

## 3.4 Identification: exclusion restriction

The exclusion restriction requires that local (size-squared-weighted) exposure to these large firm hiring shocks is uncorrelated with the other determinants of occupation-specific wage growth in any given metro area $m$, conditional on our occupation-year and metro area-year fixed effects (or the occupation-metro and occupation-year fixed effects we use in an alternative specification) and any control variables.

**Controlling for labor demand.** The primary concern in this vein is that greater local exposure to fast-growing national firms may directly affect total local labor demand, as well as increasing employer concentration. Since increasing labor demand would be expected to increase wages, this would be expected to bias our estimated coefficient toward zero, which means that our (negative) estimates of the wage effect of labor market concentration would

[24]Note that the first stage might not hold for *all* possible combinations of the distribution of employment growth and initial employer shares. For example, consider a world in which there is a labor market where Employer X has 80% of the market in one metro area, and the rest of the market is comprised of atomistic firms (HHI=6,400); and Employer X has 65% of the market in another metro area, with the rest of the market comprised of atomistic firms (HHI=4,200). If Employer X grows by 10% in both locations in a given year, and the other firms do not grow at all, employer concentration will actually increase by more in the latter than the former market. This circumstance, however, only occurs when comparing two labor markets which both have extremely high levels of employer concentration already, and so is not relevant for the vast majority of the labor markets in our data (in our analysis sample, only 1% of workers had an HHI greater than 2,000 in our data in 2019).

be, if anything, too small in magnitude. To address this, we control for two correlates of the change in local occupation-specific labor demand: (1) the growth rate of local vacancies in the occupation-metro area ($g_{o,m,t} = \sum_j \sigma_{j,o,m,t-1} g_{j,o,m,t}$), and (2) the predicted growth rate of local vacancies based on large firms' national growth (i.e. the direct linear analog to our concentration index: $\tilde{g}_{o,m,t} = \sum_j \sigma_{j,o,m,t-1} \tilde{g}_{j,o,t}$ as defined above). With these controls, we are estimating the effect of a change in local labor market concentration due to changes in large firms' employment, holding constant any direct linear effect on local labor demand or productivity.[25] In a robustness check, we also include a control for average vacancy growth across firms within a local occupation, with each firm weighted equally, $\frac{1}{N} \sum_j g_{j,o,m,t}$. This is suggested by Gabaix and Koijen (2024) as an appropriate control for local demand effects in a granular IV setting, as it controls for the increase in vacancies experienced commonly across all firms in the local labor market. The identification assumptions in the specification with this control would require that local squared exposure to excess national firm-level growth is random in expectation conditional on this proxy for local labor demand (alongside the other controls and fixed effects). We find that including these labor demand controls has little effect on our baseline coefficient estimates.

Note that distinguishing labor demand and concentration effects in this way is not simply "identification by functional form" – the insight of our IV approach is precisely that the same average positive change in labor demand can affect concentration positively *or* negatively, depending on how the labor demand shocks are distributed across firms that start out with a small or a large share of local employment. It is this interaction of pre-existing employer market shares with national shocks that drives the local market variation in the instrument (and even nonlinear functions of mean labor demand changes would therefore be unable to eliminate market exposure to the instrument). Specifically, our approach allows for different local occupations to have different expected average growth rates based on national firm growth in a way that correlates with local shocks to wages. It only requires local wage residuals across local occupations to be uncorrelated with whether this growth is driven by the *national* growth of locally *large* firms vs. small firms. To illustrate this intuition, return to the hypothetical example of insurance sales agents in Bloomington, IL and in Amarillo, TX. Our granular IV identification approach controls for local growth rates of overall insurance sales agent employment in both metro areas, allowing for each metro area to be exposed differently to overall trends in the demand for insurance sales agents. The identification only requires that once we account for overall metro area exposure to insurance sales agent demand, whether that demand was driven by the metro area's major employer or

[25] Controlling for national trend exposure directly to prevent it from confounding a nonlinear IV is similar to the "double Bartik" approach in Chodorow-Reich and Wieland (2020).

smaller employers is not correlated with local idiosyncratic wage shocks for insurance sales agents.

**Controlling for labor market trends.** A second possible exclusion restriction concern is that firms which expand more nationally may be located in weaker labor markets where wage growth is lower: for firms' national hiring growth shocks to be "as good as randomly assigned", we must assume that the presence of a firm which is nationally fast-growing vs. nationally slower-growing in a local occupation is uncorrelated with local occupation-specific economic conditions, conditional on our fixed effects. Our baseline analyses include metro area-by-year fixed effects to capture any aggregate weakness in the local labor market, and we also show analyses with occupation-by-metro area fixed effects to capture any occupation-specific labor market weakness in a given metro area. Robustness to both of these indicates that this is likely not the variation driving our results.

**Additional identification assumptions.** The non-linear structure of our IV can introduce bias: different local labor markets may be systematically exposed to the "shocks" of the growth of large national firms, and this pattern may be correlated with outcomes related to wages. This is a general concern in non-linear shift share instrumental variables, as identified by Borusyak and Hull (2023). We follow their recommendation in incorporating a control for the expected value of the HHI instrument in all our baseline regressions.[26] When we control for the expected HHI instrument in our regressions, we identify only off the difference between the actual firm-level vacancy growth shocks experienced, as compared to a counterfactual where the firm distribution was the same but vacancy growth shocks were different. This removes any omitted variable bias arising from non-random exposure to the national firm vacancy growth shocks.

We also note that our instrument is unlikely to be strong for initially unconcentrated labor markets – if each firm has only a trivial share of local employment, even substantial hiring growth will not much change local employer concentration. Our estimation therefore includes a robustness check that omits the least concentrated labor markets from the analysis. Moreover, in our discussion of the implications of our estimates, we apply our estimates of the effect of employer concentration on wages only to local labor markets with above-median employer concentration.

[26] To construct this expected HHI instrument control, we first construct 100 counterfactual HHI instruments by randomly re-assigning the actual observed firm-occupation vacancy growth rates to the other firms in the data, and using these counterfactual firm-occupation vacancy growth rates $\hat{g}_{j,o,t}$ to construct a counterfactual HHI instrument $e_{o,m,t} = \left(\sum_j \sigma^2_{j,o,m,t-1} \left(\frac{(1+\hat{g}_{j,o,t})^2}{(1+\hat{g}_{o,m,t})^2} - 1\right)\right)$. We then construct our expected HHI instrument by taking 100 of these counterfactual HHI instruments, averaging over them, and taking the log.

### 3.5 Comparison to other approaches

We see our approach as a novel contribution with regard to the problem of estimating the causal effect of employer concentration on wages. Much of the recent empirical work instruments for changes in employer concentration in a given local occupation with changes in (the inverse of) the number of employers in the same occupation in other local areas (e.g. Azar et al. (2020b;a); Rinz (2022); Qiu and Sojourner (2019); Marinescu, Ouss, and Pape (2021); Gibbons, Greenman, Norlander, and Sørensen (2019)). This circumvents some endogeneity issues, but a major concern remains that national occupation trends in concentration may be correlated with unobservable national trends in occupational productivity, demand, or supply, which could confound estimated wage effects.[27] Our strategy based on local variation in concentration allows us to use occupation-year fixed effects to control for national occupation-level factors which affect wages.

Other recent empirical work uses M&A activity to generate plausibly exogenous variation in local labor market concentration: Arnold (2020) for all industries, Prager and Schmitt (2021) for hospitals, and Guanziroli (2022) and Farag et al. (2024) for pharmacies. This avoids endogeneity concerns about the cause of the change in concentration, but reflects one specific source of concentration (M&A activity accounts for less than 2% of changes in local employer concentration (Arnold, 2020)) and cannot fully isolate the effects of employer concentration from other local economic effects of the M&A activity.[28] Our approach examines the effects of a different source of variation: increases in concentration as a result of large employers' national hiring decisions. This allows us to control for effects on local labor demand. Ultimately, we believe that this set of complementary identification approaches – based off different variation, and with different strengths – can together provide a useful picture of the effects of employer concentration on wages.

# 4 Results

Our results show that higher employer concentration (as a result of large employers' national hiring shocks) causally reduces wages. These results are robust to a range of alternate specifications and control variables. We discuss these findings further below.

[27] The authors control for variables like labor market tightness to address this.

[28] Dodini, Lovenheim, Salvanes, and Willén (2024) adopt an additional different strategy, demonstrating that workers laid-off in mass layoffs see larger wage losses in more concentrated labor markets in Norway.

## 4.1 First-stage results

To show the relevance of our national hiring shock instrument for local labor market concentration, we first show the relationship between our instrument and the HHI. The first stage regression estimates are shown in Table 3, column 4, and Figure 1 shows binscatters representing the first stage relationship between the (residualized) instrument and HHIs. The instrument is strong (with a high F-statistic) and positively and monotonically predicts the HHI even when including the baseline controls and detailed fixed effects in the regression. That is, local labor markets that are more exposed to firms that are increasing their hiring nationally see an increase in the concentration of local hiring. This finding is consistent with results in the literature on product market concentration: Rossi-Hansberg, Sarte, and Trachter (2021) show that when top firms in an industry enter a new location, concentration in the local product market falls, but as these top firms expand, local product market concentration rises again (for a subset of industries). Our results are consistent with this latter pattern: we show that, when large firms expand nationally, local labor markets where those firms already have a meaningful presence see increasing concentration.

## 4.2 Main results

In our data, there is a robust negative correlation between log vacancy HHIs and log wages at the occupation-metro area level, as shown in Figure 1 (and as found in other work, e.g. Azar et al. (2020a); Benmelech et al. (2022); Rinz (2022)).

We show our main regression results in Table 3. Our baseline regressions with occupation-year and metro area-year fixed effects are shown in Panel A. We see a large negative OLS relationship with a coefficient of -0.016 (column 1). When instrumenting for the HHI, the coefficient is larger at -0.029 (suggesting omitted variable bias or measurement error in the OLS specification), and statistically significant at the 1% level.[29]

In Panel B, we show these regressions with alternative fixed effects: occupation-metro area and occupation-year fixed effects. Here, identifying variation comes from differential year-to-year national growth rates of the large employers present in a given occupation-metro area labor market over the period 2011–2019, rather than from differential exposure to the fast-growing national employers in a local occupation relative to other localities and occupations. Again, we see a negative relationship between wages and concentration in both the OLS and IV specifications. The IV coefficient is substantially smaller with this set of fixed effects than with our baseline fixed effects, at -0.008, but remains statistically significant

[29] We limit the sample for all our regressions to the set of MALMAs where we can calculate a wage, employment, HHI, and HHI instrument for the focal SOC 6-digit occupation.

at the 1% level. In Figure 1 we visualize the reduced form relationship between the log wage and log HHI instrument with both sets of fixed effects.

How big is this effect? Our IV coefficient estimates suggest that going from the HHI faced by the median worker to the HHI faced by the worker at the 95th percentile (from an $HHI$ of 68 to 607) would lead to 6.3 log points lower hourly wages in our baseline specification, or 1.8 log points lower hourly wages in the specification using occupation-by-metro area fixed effects.[30] Since both of these are reasonable specifications which each control for different potential sources of omitted variable bias, we can simply take the average of the two, suggesting that moving from the HHI faced by the median to the 95th percentile worker results in around 4 log points lower wages.

### 4.3 Robustness

We explore a number of variations on our baseline analyses, illustrated in Figure 2 (coefficient estimates and standard errors shown in Appendix Tables A1 and A2).

First, we run our regressions with different specifications. Without employment weighting, our coefficient estimates remain significant but are smaller, suggesting that our large negative estimates of the effect of employer concentration on wages are not driven by small (unrepresentative) labor markets. Our coefficient estimates remain similar if we remove the vacancy growth controls which proxy for labor demand, or if we follow Gabaix and Koijen (2024) in adding an additional labor demand control for the equal-weighted vacancy growth of local firms ($g^e_{o,m,t} = \frac{1}{N}\sum_j^N g_{j,o,m,t}$) to reflect *local* occupation-specific shocks that are common to all firms – suggesting that correlated labor demand shocks are unlikely to be playing a major role. Similarly, our coefficient estimates remain similar if we remove the control for the expected HHI instrument.

Second, we run our regressions with different samples. Since the logic of our HHI instrument holds less well for very low HHI cells, we drop occ-metro area cells with very low HHIs (with $HHI$ less than 15). To address concerns about representativeness in our BGT vacancy data, we exclude any occupations which are substantially underrepresented in the vacancy data (with 'represented-ness' less than 0.2). To address the concern that the logic of our instrument may work less well for occupations producing non-tradable goods or services (since local hiring decisions are more likely to be driven by local occupation-specific economic conditions, making our first stage less likely to hold), we run a specification with only occupations which produce tradable goods or services.[31] In all these cases, our coefficient

[30] Calculated as $(\ln(607) - \ln(67)) \cdot -0.008 = -0.018; (\ln(607) - \ln(67)) \cdot -0.029 = -0.063$. This is at the low end of the range of existing estimates presented in Marinescu and Hovenkamp (2019).

[31] We define tradable occupations as all occupations in manufacturing or production, extraction, or farming,

estimates are similar to the baseline.

Third, we run our regressions with an alternate version of our HHI instrument. In our baseline regressions, we instrument for the local vacancy growth of firm $j$ in occupation $o$ and metro area $m$ with that firm's nationwide vacancy growth in occupation $o$, leaving out metro area $m$; to address endogeneity concerns, we only use firms with presence in at least 5 metro areas in that occupation $o$ to construct the instrument. In our robustness check, we construct an alternate version of the instrument that uses all firms with positive vacancy growth in occupation $o$ in any other metro area $m$ to construct the instrument, removing the limit that they must be present in at least 5 metro areas.

Finally, for our regressions with occupation-metro area fixed effects, we estimate with metro-year or year fixed effects (rather than occupation-year fixed effects). The coefficient estimates remain similar to the specification with occupation-metro area and occupation-year fixed effects, and statistically significant at the 10% level.

## 4.4 Alternate labor market definitions

In Appendix B we replicate our results measuring the occupation at the 6-digit SOC or 3-digit SOC level, rather than using our MALMA occupation clusters. Our findings are very similar to those in the main results, with large, negative, and statistically significant effects of employer concentration on wages in our baseline specifications and across robustness checks. With our baseline fixed effects, we estimate a wage-concentration elasticity of -0.020 at the SOC 6-digit level and -0.030 at the SOC 3-digit level; these are very similar to the elasticity of -0.029 we estimate with MALMA labor markets in Table 3. With occupation-metro fixed effects, we estimate smaller wage-concentration elasticities of -0.005 and -0.006 for SOC 6-digit and SOC 3-digit occupations respectively, with the point estimate again very similar to the -0.008 estimated in our main MALMA specification, although the estimates for the SOC 6-digit and SOC 3-digit specifications in this case are noisy and not statistically significantly different from zero.[32] Note that while the estimated elasticities are very similar across these market definitions, the market definition matters for the implication drawn for specific occupations: for example, a specific local labor market for bank tellers might appear highly concentrated at the SOC 6-digit level, but the MALMA labor market is likely to be substantially less concentrated, meaning that one would predict a large wage suppressive

fishing, and forestry (SOC codes starting in "45", "47-5", and "51"), as well as all occupations defined as at least somewhat teleworkable by Dingel and Neiman (2020) (i.e. those with a teleworkability index > 0).

[32] Note that we believe SOC 6-digit 3-digit occupations are worse at capturing workers' true labor market than our MALMAs: SOC 6-digit occupations are less complete, including a smaller share of the relevant labor market, and SOC 3-digit occupations are more relevant, since they include a much larger share of irrelevant outside options.

effect of concentration if using a SOC 6-digit market definition but not if using a MALMA. We discuss these results in detail in Appendix B, and compare MALMA labor markets to SOC 6-digit and SOC 3-digit labor markets in Figures B1 and B2.

## 4.5 Replication in OEWS micro data

Our HHIs are measured in the BGT vacancy data. We choose to use vacancies rather than employment to construct HHIs because we see vacancies as the most appropriate reflection of the job options that are feasibly available to workers. Moreover, there is no US dataset which accurately captures occupation-level employment by employer. However, there are some potential concerns with using this vacancy data as a proxy for job openings: on the one hand, a single job posting can reflect multiple job openings with the same description and this mismatch may be skewed across different types of employers or occupations. On the other hand, the BGT online vacancy posting data may not capture the totality of all job postings in a way that skews our estimates of employer concentration (see Appendix section D.4 for further discussion of these concerns).

As a result, it may be of interest to understand if our results also hold with HHIs measured over employment. We therefore replicate our main results using estimates of employment HHIS constructed from OEWS restricted-access micro data. Since these data only provide occupation-level employment figures for a sample of surveyed establishments, we follow Handwerker and Dey (2024) and impute occupation-level employment to the near universe of employers in the Quarterly Census of Employment and Wages (QCEW). We then estimate occupation-metro area level HHIs from this combined pseudo-universe dataset. Interestingly, despite the differences between vacancy postings and employment, and the (likely uncorrelated) potential sources of measurement error in each dataset, the HHIs calculated with this method are highly correlated with our HHIs from the BGT vacancy posting data, with a correlation coefficient of 0.61 (Appendix Table C1, Appendix Figure C1). When replicating our main analyses with the OEWS HHIs (and our original instrument using large firm hiring shocks from the BGT vacancy posting data), we find similarly large, negative, and statistically significant effects of employer concentration on wages.[33] This analysis serves both to assuage concerns that our results are an artefact of using vacancies rather than employment to measure HHIs, as well as to assuage concerns that our results were induced by measurement error in our BGT vacancy data. We describe this replication exercise in full in Appendix C.

[33]We are able to fully replicate our baseline analyses with occupation-year and metro area-year fixed effects. Our hiring shock instrument has no first stage in the OEWS employment data when we include occupation-metro area fixed effects, so we cannot replicate those analyses at all in the OEWS data.

## 5 Discussion and Implications

**Who is most affected by employer concentration?** We use our coefficient estimates in a back-of-the-envelope quantification of the "wage effect" of employer concentration in each above-median-concentration occupation-by-metro area labor market in 2019, relative to a scenario where their HHI is reduced to 74 (which is the degree of concentration across occupation (MALMA)-metro area labor markets faced by the median worker in our sample).[34] Specifically, we calculate a counterfactual wage effect for each occupation-metro labor market in 2019:

$$\text{wage effect}_{o,m,t} = (log(HHI_{o,m,t}) - log(74)) \cdot \gamma \tag{7}$$

where $\gamma$ denotes the estimated coefficient on the log(HHI) in our IV regression specification in Table 3 column (2). We set $\gamma = -0.0185$ as the average between the coefficient estimates in Panel A and Panel B (using different sets of fixed effects). Note that this exercise considers the effect of changes in employer concentration *holding all else constant*, including local productivity. It can illustrate the degree to which wages may be marked down from local occupational productivity as a result of employer concentration, but cannot necessarily illustrate what would happen if a specific policy or business decision were to change local employer concentration (as it might also change local productivity). It also rests on the assumption that we can apply our estimated coefficients linearly.

Of the 117 million workers in labor markets covered by our wage, employment, and concentration data in 2019, our counterfactual wage exercise suggests that roughly 19% had wages which were at least 2.5% lower as a result of above-median employer concentration.[35] Our estimates suggest that employer concentration slightly widens inequality: the share of workers affected by employer concentration is larger in low-wage occupations and in low-wage cities (Appendix Figures A1 and A2).

Which occupations are most affected by employer concentration? In Figure 3, we show the average estimated wage effect of above-median employer concentration across occupations, aggregated for ease of interpretation into SOC 2-digit occupation groups. Healthcare occupations stand out as being the most affected. To focus on more granular occupations,

[34] An HHI of 74 is a very unconcentrated market: it is equivalent to a labor market with roughly 135 equal-sized employers, for example, or a labor market where two large firms each employ around 6% of the workforce and an atomistic "fringe" of firms employ the rest.

[35] With a coefficient of -0.0185, this means any cells with an $HHI > 218$. We estimate a very similar share excluding occupations which are highly underrepresented in the BGT vacancy data (a 'represented-ness' of less than 0.5). Our estimate of 19% is likely an underestimate of the effect of employer concentration across the full U.S. labor market since our data lacks coverage of non-metropolitan areas and of some small occupations, where one would expect employer concentration to be higher than average.

in Table 4, we show the twenty largest SOC 6-digit occupations which are "highly affected" by employer concentration, which we define as having at least one third of their workforce in labor markets with an estimated wage effect of concentration of 2.5% or more. Again, healthcare occupations stand out: an estimated 39% of registered nurses, 44% of nursing assistants, 58% of medical assistants, 40% of licensed practical and vocational nurses, 92% of pharmacy technicians, 81% of pharmacists, 79% of radiologic technologists, 41% of nurse practitioners, 46% of phlebotomists, and 64% of physician assistants experience wage suppression of greater than 2.5% as a result of above-median employer concentration. Healthcare workers are particularly affected by employer concentration because they often have (i) high employer concentration within their narrow SOC 6-digit occupations, (ii) low occupational mobility, and (iii) what mobility they do have is often to other healthcare occupations, which often share the same set of employers (such as large hospitals).[36]

Importantly, the list of most-affected occupations is quite different using our HHIs measured over MALMA mobility-adjusted labor markets, than it would be if we had carried out this analysis using single SOC 6-digit occupations. Specifically, using SOC 6-digit occupations leads to overestimation of the effect of employer concentration for high-mobility occupations with a diverse range of possible employers, like bank tellers or short order cooks, and underestimation of the effects of employer concentration for low-mobility occupations with a limited range of possible employers, like registered nurses or pharmacists.

While our estimates above suggest that employer concentration suppresses wages for several million workers, we underscore also that the majority of American workers do not work in labor markets with even moderate degrees of employer concentration – and therefore likely do not experience significant wage suppression as a result of employer concentration. Thus, policymakers should focus attention on the subset of workers who face concentrated labor markets, defined appropriately to reflect their own occupation and any other relevant outside options.

Note, moreover, that while our estimates show that increases in employer concentration reduce local wages, they cannot tell us whether the ultimate incidence of these wage reductions falls on firms in the form of higher profits, or consumers in the form of lower prices. The balance likely depends on the nature of product market competition.[37]

**Antitrust implications.** One area where our analysis can be applied is antitrust.[38] The

[36]This is in keeping with recent work that has found large effects of hospital mergers on wages of nursing and pharmacy workers (Prager and Schmitt, 2021), and a low elasticity of the labor supply of registered nurses to individual hospitals (Staiger, Spetz, and Phibbs, 2010).

[37]Kahn and Tracy (2024) argue that the ultimate incidence of local labor market concentration falls to a large extent on local landowners as lower local wages reduce local rents and house prices. Understanding the ultimate incidence of these effects is important to determine the appropriate policy response.

[38]Several scholars have called for antitrust authorities to pay attention to employer concentration (Marinescu

2023 revision to the DoJ-FTC Horizontal Merger Guidelines now explicitly incorporates a focus on labor market competition, and several recent merger reviews have used measures of employer concentration – in particular, HHIs calculated on 6-digit SOC labor markets – as a preliminary screen for anticompetitive effects of mergers in labor markets (e.g. Federal Trade Commission, 2020; 2022). Our analysis using mobility-adjusted labor markets suggests that it is crucial that labor markets are appropriately defined in this screening process: using simple official labor market definitions like the SOC 6-digit occupation would likely lead to some mergers being scrutinized even if they would have little effect on wages, while others which may have anti-competitive effects may go unnoticed.

It is important to note that our findings do not tell us that *all* increases in employer concentration reduce wages. If higher employer concentration comes alongside higher productivity, workers' wages may be higher in the high-concentration high-productivity scenario than a low-concentration lower-productivity scenario, so seeking to reduce employer concentration may not be the best response: close scrutiny of individual cases, and industry- and occupation-specific studies, are necessary to understand whether antitrust action would be appropriate in any specific circumstance.[39] In addition, while increased antitrust scrutiny of labor markets is important, it is unlikely to affect the majority of workers impacted by employer concentration (Naidu and Posner, 2021), since most changes in employer concentration are not caused by mergers and acquisitions and many concentrated labor markets do not feature illegal anti-competitive practices.

**Labor market analysis.** Our analysis of occupational mobility revealed large and heterogeneous rates of workers moving out of their current occupation every year, showing that their effective labor market transcends the boundaries of their current (narrow) occupation, and that aggregating up to larger administrative occupation definitions like the SOC 3-digit labor market does not solve this problem. Moreover, our results suggest that ignoring the effective boundaries of dynamic labor markets may lead to biased estimates of labor market impacts, for instance by neglecting the role that simultaneous changes in outside options play in worker outcomes. This means that academic or policy-oriented analyses of the impact of shocks on workers should take into account both the larger scope of which jobs are relevant for workers, and the potential for spillovers of labor market changes between occupations.

---

and Hovenkamp, 2019; Naidu et al., 2018; Hemphill and Rose, 2017; Steinbaum and Stucke, 2020; Hovenkamp, 2018; Krueger and Posner, 2018). Historically antitrust authorities paid little attention to employer concentration (though monopsony is referred to in the 1992 DoJ-FTC Horizontal Merger Guidelines (Phillips, 2019)), but this has changed in recent years.

[39] As also emphasized by Hovenkamp (2018), Berger et al. (2022), and Arnold (2020). Naidu et al. (2018) argue that antitrust authorities should permit mergers where the incremental increase in workers' wages because of increased productivity would *outweigh* any incremental decrease in workers' wages induced by the increase in employer concentration.

Concretely, research on labor markets should explicitly grapple with, and justify, why a particular administrative definition (like SOC 6-digit occupations, or NAICS 4-digit industries) captures the relevant labor market for the issue at hand, and should consider probabilistically incorporating other occupations or industries that may affect a worker's outcomes.[40] While studies of product markets have long recognized the need to define product markets well, the same care is needed when discussing labor markets.

# 6 Conclusion

Our findings point to a middle ground between two prominent views about the effects of employer concentration in the US labor market. On the one hand, employer concentration is *not* a niche issue confined to a few factory towns: we find large, negative, and significant effects of employer concentration on wages when estimated using nuanced market definitions and plausibly exogenous variation across the majority of the US labor market, and our back-of-the-envelope calculations suggest that one in six U.S. private sector workers experience non-trivial wage effects of employer concentration. On the other hand, most workers are not in highly concentrated labor markets, and the effects of employer concentration therefore do not seem big enough to have a substantial effect on the aggregate wage level or degree of income inequality in the U.S. economy (though other sources of monopsony power may still be important).[41] The fact that employer concentration affects wages for several million American workers suggests that increased policy attention to this issue is appropriate, in terms of antitrust, policies to raise wages, and policies to increase worker mobility. For these policy decisions, our work underscores that the definition of the labor market is vitally important, that the local occupation is a good definition of the local labor market for occupations with low outward mobility but broader mobility-adjusted concepts of the labor market can and should be used for more outwardly mobile occupations, and that when estimated within well-defined labor markets, employer concentration can have large effects on workers' wages.

[40] We make our matrix of occupational mobility across SOC 6-digit occupations freely available for download from our websites, to be used in other studies of occupational labor markets.

[41] Similarly, Rinz (2022), Berger et al. (2022), and Lipsius (2018) show that employer concentration has fallen over recent decades in most local industries, casting doubt on the argument that changing employer concentration can explain median pay stagnation or rising income inequality. It is possible, however, that the decline in countervailing worker power has exposed firms' latent monopsony power, meaning that employer concentration (and other sources of monopsony power) have greater wage effects than in the past (Erickson and Mitchell, 2007; Naidu et al., 2018; Stansbury and Summers, 2020).

# References

David Arnold. Mergers and acquisitions, local labor market concentration, and worker outcomes. Working paper, 2020.

José Azar, Ioana Marinescu, and Marshall Steinbaum. Labor market concentration. *Journal of Human Resources*, pages 1218–9914R1, 2020a.

José A Azar, Ioana Marinescu, Marshall I Steinbaum, and Bledi Taska. Concentration in us labor markets: Evidence from online vacancy data. *Labour Economics*, 66:101886, 2020b.

Kate Bahn. Understanding the importance of monopsony power in the u.s. labor market. Technical report, Washington Center for Equitable Growth, 2018.

Timothy J Bartik. Local labor markets should be redefined: New definitions based on estimated demand-shock spillovers. Technical report, Upjohn Institute Working Paper, 2024.

Efraim Benmelech, Nittai K Bergman, and Hyunseob Kim. Strong employers and weak employees how does employer concentration affect wages? *Journal of Human Resources*, 57(S):S200–S250, 2022.

David Berger, Kyle Herkenhoff, and Simon Mongey. Labor market power. *American Economic Review*, 112(4):1147–93, 2022.

Steven Berry, Martin Gaynor, and Fiona Scott Morton. Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives*, 33(3): 44–68, 2019.

Kirill Borusyak and Peter Hull. Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185, 2023.

Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213, 2022.

Anthony P. Carnevale, Tamara Jayasundera, and Dmitri Repnikov. Understanding online job ads data: A technical report. 2014.

Gabriel Chodorow-Reich and Johannes Wieland. Secular labor reallocation and business cycles. *Journal of Political Economy*, 2020.

Lauren Cohen, Umit Gurun, and N Bugra Ozel. Too many managers: The strategic use of titles to avoid overtime payments. Technical report, National Bureau of Economic Research, 2023.

Michael R Dalton, Lisa B Kahn, and Andreas I Mueller. Do online job postings capture job vacancies? In *Mimeo*. 2025.

Jonathan I Dingel and Brent Neiman. How many jobs can be done at home? *Journal of Public Economics*, 189:104235, 2020.

Samuel Dodini, Michael Lovenheim, Kjell Salvanes, and Alexander Willén. Monopsony, job tasks and labour market concentration. *The Economic Journal*, 134(661):1914–1949, 2024.

Christopher L Erickson and Daniel JB Mitchell. Monopsony as a metaphor for the emerging post-union labour market. *International Labour Review*, 146(3-4):163–187, 2007.

Enas Farag, Chris Compton, Marshall Steinbaum, Alaa Abdelfattah, and Anna Stansbury. A retrospective analysis of the acquisition of target's pharmacy business by cvs health: Labor market perspective. *Available at SSRN 4644895*, 2024.

Federal Trade Commission. Staff comment to the texas health and human services commission on certificate of public advantage applications by hendrick health system and shannon health system. Staff comment, Federal Trade Commission, Bureau of Competition & Bureau of Economics, September 2020. URL https://www.ftc.gov/system/files/documents/advocacy_documents/ftc-staff-comment-texas-health-human-services-commission-regarding-certificate-public-advantage/20100902010119texashhsccopacomment.pdf.

Federal Trade Commission. Staff submission to the new york state department of health regarding the certificate of public advantage application of suny upstate medical university and crouse health system. Staff Comment FTC File No. 221-0126, Federal Trade Commission, Bureau of Competition & Bureau of Economics, October 2022. URL https://www.ftc.gov/system/files/ftc_gov/pdf/FTC-NY-COPA-Comment-10-7-2022-PUBLIC.pdf.

Xavier Gabaix and Ralph SJ Koijen. Granular instrumental variables. *Journal of Political Economy*, 132(7):2274–2303, 2024.

Eric M Gibbons, Allie Greenman, Peter Norlander, and Todd Sørensen. Monopsony power and guest worker programs. *The Antitrust Bulletin*, 64(4):540–565, 2019.

Xavier Giroud and Holger M Mueller. Firms' internal networks and local economic shocks. *American Economic Review*, 109(10):3617–49, 2019.

Tomas Guanziroli. Does labor market concentration decrease wages? evidence from a retail pharmacy merger.", 2022.

Elizabeth Weber Handwerker and Matthew Dey. Some facts about concentrated labor markets in the united states. *Industrial Relations: A Journal of Economy and Society*, 63(2): 132–151, 2024.

Jonathon Hazell and Bledi Taska. Downward rigidity in the wage for new hires. 2019.

C Scott Hemphill and Nancy L Rose. Mergers that harm sellers. *Yale LJ*, 127:2078, 2017.

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

Thomas J Holmes. The diffusion of wal-mart and economies of density. *Econometrica*, 79

(1):253–302, 2011.

Herbert Hovenkamp. Whatever did happen to the antitrust movement. *Notre Dame L. Rev.*, 94:583, 2018.

Gregor Jarosch, Jan Sebastian Nimczik, and Isaac Sorkin. Granular search, market structure, and wages. *Review of Economic Studies*, 91(6):3569–3607, 2024.

Janna E Johnson and Morris M Kleiner. Is occupational licensing a barrier to interstate migration? *American Economic Journal: Economic Policy*, 2020.

Matthew Johnson, Kurt Lavetti, and Michael Lipsitz. The labor market effects of legal restrictions on worker mobility. *Journal of Political Economy*, 133(9), September 2025.

Matthew E Kahn and Joseph Tracy. Monopsony in spatial equilibrium. *Regional Science and Urban Economics*, 104:103956, 2024.

Gueorgui Kambourov and Iourii Manovskii. Rising occupational and industry mobility in the united states: 1968–97. *International Economic Review*, 49(1):41–79, 2008.

Gueorgui Kambourov and Iourii Manovskii. A cautionary note on using (march) current population survey and panel study of income dynamics data to study worker mobility. *Macroeconomic Dynamics*, 17:172–194, 2013.

Daniel Kreisman, Jonathan Smith, and Bondi Arifin. Labor market signaling and the value of college: Evidence from resumes and the truth. *Journal of Human Resources*, 58(6): 1820–1849, 2023.

Alan B Krueger and Orley Ashenfelter. Theory and evidence on employer collusion in the franchise sector. *Journal of Human Resources*, 57(S):S324–S348, 2022.

Alan B. Krueger and Eric A. Posner. A proposal for protecting low-income workers from monopsony and collusion. 2018.

Thibaut Lamadon, Magne Mogstad, and Bradley Setzler. Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, 112(1):169–212, 2022.

Ben Lipsius. Labor market concentration does not explain the falling labor share. Working paper, 2018.

Alan Manning. *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press, 2003.

Alan Manning and Barbara Petrongolo. How local are labor markets? evidence from a spatial job search model. *American Economic Review*, 107(10):2877–2907, 2017.

Ioana Marinescu and Herbert Hovenkamp. Anticompetitive mergers in labor markets. *Indiana Law Journal*, 2019.

Ioana Elena Marinescu and Eric A. Posner. Why has antitrust law failed workers? *Cornell Law Review*, 2020.

Ioana Elena Marinescu, Ivan Ouss, and Louis-Daniel Pape. Wages, hires, and labor market concentration. *Journal of Economic Behavior and Organization*, 2021.

Suresh Naidu and Eric A. Posner. Labor monopsony and the limits of the law. *Journal of Human Resources*, 2021.

Suresh Naidu, Eric A. Posner, and E. Glen Weyl. Antitrust remedies for labor market power. *Harvard Law Review*, 2018.

David Neumark, Junfu Zhang, and Stephen Ciccarella. The effects of wal-mart on local labor markets. *Journal of Urban Economics*, 63(2):405–430, 2008.

Jan Sebastian Nimczik. Job mobility networks and endogenous labor markets. 2018.

Israel P Nosnik, Patricia Friedmann, Harris M Nagler, and Caner Z Dinlenc. Resume fraud: unverifiable publications of urology training program applicants. *The Journal of urology*, 183(4):1520–1523, 2010.

Noah Joshua Phillips. Prepared statement before the u.s. house of representatives on antitrust and economic opportunity: Competition in labor markets. 2019.

Elena Prager and Matthew Schmitt. Employer consolidation and wages: Evidence from hospitals. *American Economic Review*, 2021.

Yue Qiu and Aaron Sojourner. Labor-market concentration and labor compensation. 2019.

Kevin Rinz. Labor market concentration, earnings, and inequality. *Journal of Human Resources*, 57(S):S251–S283, 2022.

Nancy Rose. Thinking through anticompetitive effects of mergers on workers. Technical report, American Bar Association 2019 Antitrust Spring Meeting, 2019.

Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter. Diverging trends in national and local concentration. *NBER Macroeconomics Annual*, 35(1):115–150, 2021.

Richard Schmalensee. Inter-industry studies of structure and performance. *Handbook of industrial organization*, 2:951–1009, 1989.

Ian M Schmutte. Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics*, 33(1):1–32, 2015.

Jay Shambaugh, Ryan Nunn, Audrey Breitwieser, and Patrick Liu. The state of competition and dynamism: Facts about concentration, start-ups, and related policies. 2018.

Kathryn L Shaw. Occupational change, employer change, and the transferability of skills. *Southern Economic Journal*, pages 702–719, 1987.

Arthur A Sloane. Countering resume fraud within and beyond banking: no excuse for not doing more. *Labor Law Journal*, 42(5):303, 1991.

Douglas O. Staiger, Joanna Spetz, and Ciaran S. Phibbs. Is there monopsony in the labor market? evidence from a natural experiment. *Journal of Labor Economics*, 28(2):211–236, 2010.

Anna Stansbury and Lawrence H Summers. The declining worker power hypothesis: An explanation for the recent evolution of the american economy. *Brookings Papers on Economic Activity*, 2020.

Marshall Steinbaum and Maurice E Stucke. The effective competition standard: A new standard for antitrust. *University of Chicago Law Review*, 2020.

Chad Syverson. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3):23–43, 2019.

Anna Thoresson. Employer concentration and wages for specialized workers. *American Economic Journal: Applied Economics*, 16(1):447–479, 2024.

UK Competition Authority. Competition and market power in uk labour markets, 2024.

U.S. Department of Justice and Federal Trade Commission. Horizontal merger guidelines. 2023.

David Weil. *The Fissured Workplace*. 2014.

Ming Xu. Understanding the decline in occupational mobility. *Working Paper*, 2018.

# Tables and Figures

Table 1: Summary Statistics, 2019

| | p1 | p5 | p10 | p25 | p50 | p75 | p90 | p95 | p99 |
|---|---|---|---|---|---|---|---|---|---|
| *Panel A: Our analysis sample, unweighted* | | | | | | | | | |
| Hourly wage | 9.93 | 12.06 | 13.53 | 17.64 | 25.85 | 39.73 | 50.97 | 57.19 | 70.15 |
| Employment | 40 | 80 | 130 | 330 | 940 | 2,960 | 8,990 | 18,650 | 57,550 |
| HHI | 12 | 25 | 42 | 99 | 267 | 633 | 1,339 | 2,015 | 3,817 |
| *Panel B: Our analysis sample, employment-weighted* | | | | | | | | | |
| Hourly wage | 9.55 | 11.24 | 12.59 | 14.69 | 20.00 | 33.43 | 49.73 | 58.61 | 72.06 |
| Employment | 180 | 544 | 980 | 2,810 | 10,350 | 33,660 | 80,790 | 123,580 | 290,640 |
| HHI | 6 | 12 | 17 | 33 | 84 | 216 | 537 | 869 | 2,084 |
| *Panel C: Full sample, unweighted* | | | | | | | | | |
| Hourly wage | 10.20 | 12.55 | 14.21 | 18.23 | 26.55 | 40.71 | 51.51 | 57.95 | 72.12 |
| Employment | 30 | 60 | 90 | 240 | 820 | 3,120 | 10,610 | 22,960 | 74,049 |
| HHI | 11 | 22 | 37 | 98 | 292 | 824 | 2,188 | 3,750 | 10,000 |
| *Panel D: Full sample, employment-weighted* | | | | | | | | | |
| Hourly wage | 9.69 | 11.58 | 12.99 | 15.53 | 21.64 | 38.59 | 54.13 | 62.34 | 78.09 |
| Employment | 140 | 470 | 930 | 3,150 | 12,260 | 40,550 | 96,730 | 155,420 | 290,640 |
| HHI | 5 | 9 | 13 | 26 | 74 | 205 | 560 | 979 | 2,959 |
| *Panel E: US national wage distribution* | | | | | | | | | |
| Hourly wage | – | – | 10.35 | 13.02 | 19.14 | 30.88 | 48.57 | – | – |

*Source:* Wage and employment from BLS OEWS, calculated at MALMA occupation by metro area level; HHIs from BGT vacancy data, calculated at MALMA occupation by metro area level; US national wage distribution data from BLS OEWS national statistics. *Notes:* Summary statistics across occupation (MALMA)-by-metro area cells in our data, 2019. Panels A and B show summary statistics for our analysis sample used for our main regressions (the sample on which we can calculate the HHI instrument in the BGT vacancy data, N=43,231); Panels C and D show summary statistics for our full sample of occ-metro area cells for which we have wage, employment, and HHIs (N=107,837). Panels B and D weight cells by employment of the focal 6-digit occupation within the MALMA, in the metro area in question. Thus, the interpretation of $p50$, for example, is that the median occupation-metro area cell has an average hourly wage of \$25.85 but the median worker is in an occupation-metro area cell with an average hourly wage of \$20. Panel E shows the 10th, 25th, 50th, 75th, and 90th percentiles of the full national hourly wage distribution in 2019, showing that our analysis sample is slightly higher-wage than the overall national wage distribution.

Table 2: Examples of employer-occupation pairs with large national job posting growth

| Employer | Occupation | Total jobs (2018) | Δ jobs 2018-19 | YoY growth 2018-19, % |
|---|---|---|---|---|
| *Panel A: Tradable Occupations* | | | | |
| Lowe's Companies, Inc. | Stock clerks and order fillers | 18,576 | 21,456 | 116 |
| IBM | Software developers, applications | 11,155 | 17,497 | 157 |
| State Farm | Insurance sales agents | 8,656 | 15,420 | 178 |
| JPMorgan Chase | Secur., commod., and fin. svcs sales agents | 10,043 | 14,465 | 144 |
| Walgreens | Customer service representatives | 12,280 | 13,556 | 110 |
| Deloitte | Computer occupations, all other | 22,366 | 12,782 | 57 |
| Dollar Tree | General and operations managers | 8,970 | 12,208 | 136 |
| Amazon | Software developers, applications | 9,095 | 11,967 | 132 |
| Anthem BCBS | Medical and health services managers | 30,462 | 11,318 | 37 |
| KinderCare | Preschool teachers, except special education | 7,173 | 10,561 | 147 |
| Bank of America | Personal financial advisors | 10,311 | 9,623 | 93 |
| Lowe's Companies, Inc. | Customer service representatives | 7,445 | 9,297 | 125 |
| Amazon | Computer occupations, all other | 5,180 | 8,890 | 172 |
| KPMG | Accountants and auditors | 6,162 | 7,062 | 115 |
| Home Depot | Cargo and freight agents | 5,901 | 5,585 | 95 |
| O'Reilly Automotive | Sales rep., wholes. & mfg., exc. tech. & scient. | 5,989 | 5,217 | 87 |
| Rent-A-Center | Sales rep., wholes. & mfg., exc. tech. & scient. | 5,575 | 5,031 | 90 |
| Home Depot | Stock clerks and order fillers | 7,199 | 4,969 | 69 |
| Schwan's | Sales representatives, services, all other | 5,270 | 4,956 | 94 |
| HCA Healthcare | Medical and health services managers | 5,285 | 4,897 | 93 |
| *Panel B: Non-tradable Occupations* | | | | |
| Lowe's Companies, Inc. | Retail salespersons | 60,360 | 88,270 | 146 |
| Dollar General | Retail salespersons | 33,357 | 59,959 | 180 |
| Dollar General | First-line supervisors of retail sales workers | 28,941 | 49,589 | 171 |
| Allied Universal | Security guards | 46,662 | 45,208 | 97 |
| Roehl Transport | Heavy and tractor-trailer truck drivers | 20,621 | 31,969 | 155 |
| AlliedBarton Sec. Svcs | Security guards | 15,148 | 29,586 | 195 |
| Starbucks Coffee | Counter attend., cafet., food, & coffee shop | 8,917 | 28,329 | 318 |
| Macy's | Retail salespersons | 33,832 | 27,724 | 82 |
| H& R Block | Tax preparers | 20,852 | 26,502 | 127 |
| Compass Group | Comb. food prep. & serving, incl. fast food | 21,887 | 25,389 | 116 |
| McDonald's | Food service managers | 13,890 | 25,132 | 181 |
| Starbucks Coffee | First-line supervisors, food prep. & serving | 6,287 | 24,897 | 396 |
| HCA Healthcare | Registered nurses | 33,803 | 23,781 | 70 |
| Marriott International | Hotel, motel, and resort desk clerks | 17,529 | 22,677 | 129 |
| CVS Health | Pharmacy technicians | 18,041 | 22,581 | 125 |
| Securitas | Security guards | 20,424 | 20,796 | 102 |
| Dick's Sporting Goods | Retail salespersons | 21,727 | 20,217 | 93 |
| Sprint | Retail salespersons | 15,523 | 20,013 | 129 |
| Walgreens | First-line supervisors of retail sales workers | 13,697 | 18,187 | 133 |
| Lowe's Companies, Inc. | Cashiers | 16,828 | 18,004 | 107 |

*Source:* BGT vacancy data. *Notes:* Table shows examples of the 20 employer-occupation pairs (> 5K job postings in 2018) with the largest change 2018-2019 in job postings in the BGT vacancy data, separately for tradable occupations (panel A) and non-tradable occupations (panel B). Tradable occupations are defined as either being teleworkable to some degree based on the measure by Dingel and Neiman (2020); being a production occupation; a farming, fishing, and forestry occupation; or an extraction worker. Where necessary, official occupation names have been abbreviated to retain table formatting.

Table 3: Main results

| *Dep var:* | *Log wage* | | | *Log HHI* |
|---|---|---|---|---|
| | OLS | IV | Reduced Form | First Stage |
| | (1) | (2) | (3) | (4) |
| *Panel A: Occ-Year and Metro Area-Year Fixed Effects* | | | | |
| Log HHI | -0.016*** | -0.029*** | | |
| | (0.001) | (0.003) | | |
| Log HHI instrument | | | -0.003*** | 0.119*** |
| | | | (0.000) | (0.003) |
| Vacancy growth | -0.003*** | -0.004*** | -0.004*** | 0.009 |
| | (0.001) | (0.001) | (0.001) | (0.011) |
| Predicted vacancy growth | 0.001 | 0.005*** | 0.003** | 0.079*** |
| | (0.001) | (0.002) | (0.001) | (0.010) |
| Expected HHI instrument | -0.003*** | 0.002 | -0.007*** | 0.331*** |
| | (0.001) | (0.002) | (0.001) | (0.009) |
| Observations | 386,479 | 366,250 | 366,250 | 366,250 |
| F-stat | | 1,326 | | |
| *Panel B: Occ-Metro Area and Occ-Year Fixed Effects* | | | | |
| Log HHI | 0.000 | -0.008*** | | |
| | (0.001) | (0.002) | | |
| Log HHI instrument | | | -0.001*** | 0.079*** |
| | | | (0.000) | (0.002) |
| Vacancy growth | -0.001*** | -0.002*** | -0.001** | -0.188*** |
| | (0.000) | (0.001) | (0.000) | (0.008) |
| Predicted vacancy growth | 0.000 | 0.003*** | 0.001* | 0.175*** |
| | (0.001) | (0.001) | (0.001) | (0.008) |
| Expected HHI instrument | -0.002*** | -0.000 | -0.001*** | 0.129*** |
| | (0.000) | (0.001) | (0.000) | (0.005) |
| Observations | 370,199 | 349,826 | 349,826 | 349,826 |
| F-stat | | 1,218 | | |

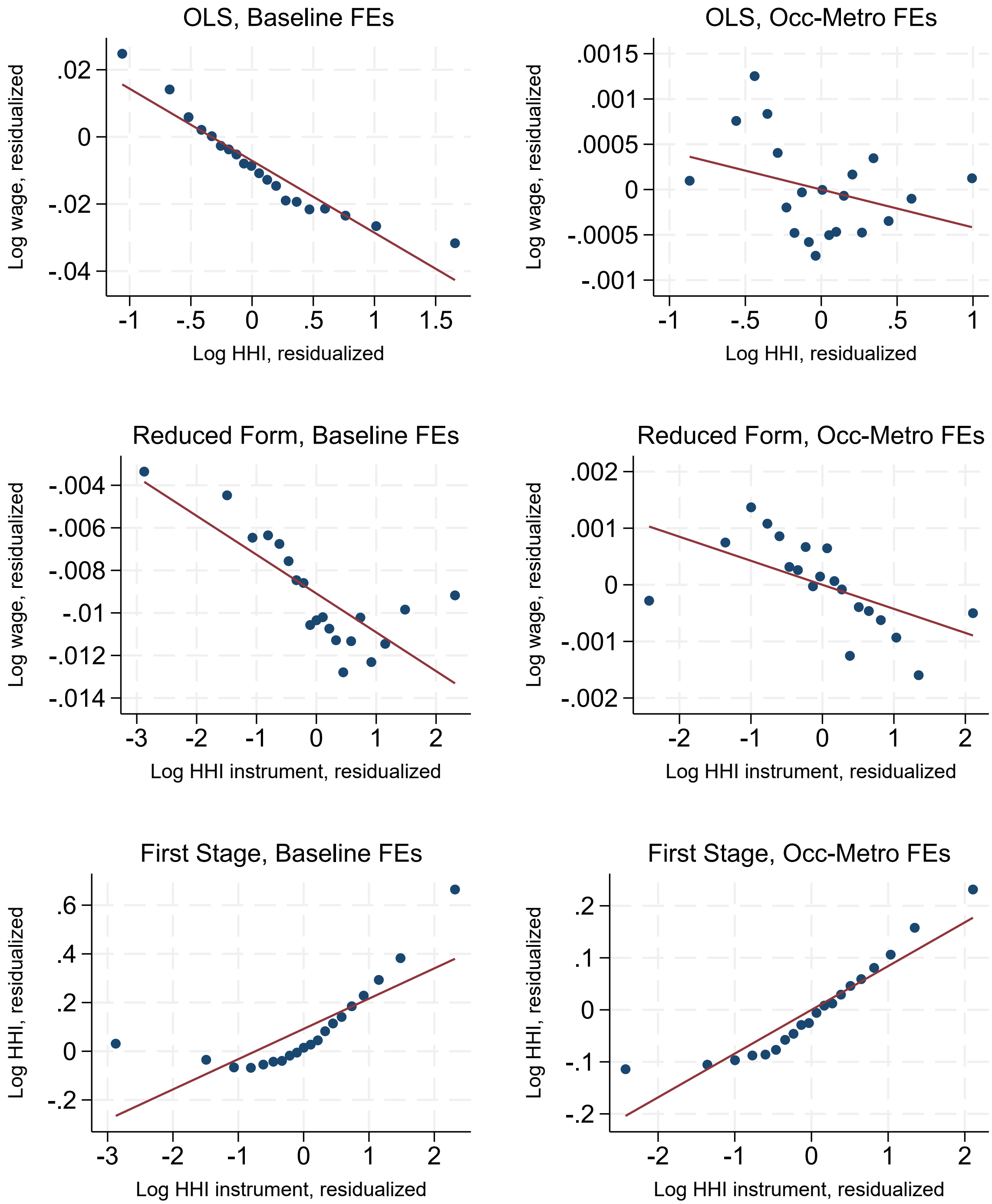
*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at occupation-metro area level in parentheses. Fixed effects noted in Panel titles. Weighted by average employment of occ-metro area cell. Unit of analysis: MALMA occupation X metro area X year, 2011-2019.

Table 4: Large occupations highly affected by employer concentration, SOC 6-digit level

| Occupation (SOC 6-digit) | National Employment (2019) | Share with Wage Effect >2.5% |
|---|---|---|
| Registered nurses | 2,622,400 | .39 |
| Nursing assistants | 1,191,610 | .44 |
| Medical assistants | 646,590 | .58 |
| Licensed practical and licensed vocational nurses | 578,130 | .4 |
| Securities, commodities, and financial services sales agents | 410,300 | .99 |
| Pharmacy technicians | 364,930 | .92 |
| Hairdressers, hairstylists, and cosmetologists | 353,710 | 1 |
| Industrial machinery mechanics | 300,110 | .34 |
| Pharmacists | 272,510 | .81 |
| Radiologic technologists | 173,790 | .79 |
| Nurse practitioners | 172,320 | .41 |
| Healthcare social workers | 153,210 | .35 |
| Medical and clinical laboratory technicians | 141,325 | .35 |
| Parking lot attendants | 137,180 | .41 |
| Automotive body and related repairers | 118,670 | .5 |
| Phlebotomists | 111,540 | .46 |
| Electronics engineers, except computer | 110,320 | .38 |
| Butchers and meat cutters | 107,360 | 1 |
| Aircraft mechanics and service technicians | 105,860 | .89 |
| Physician assistants | 104,870 | .64 |

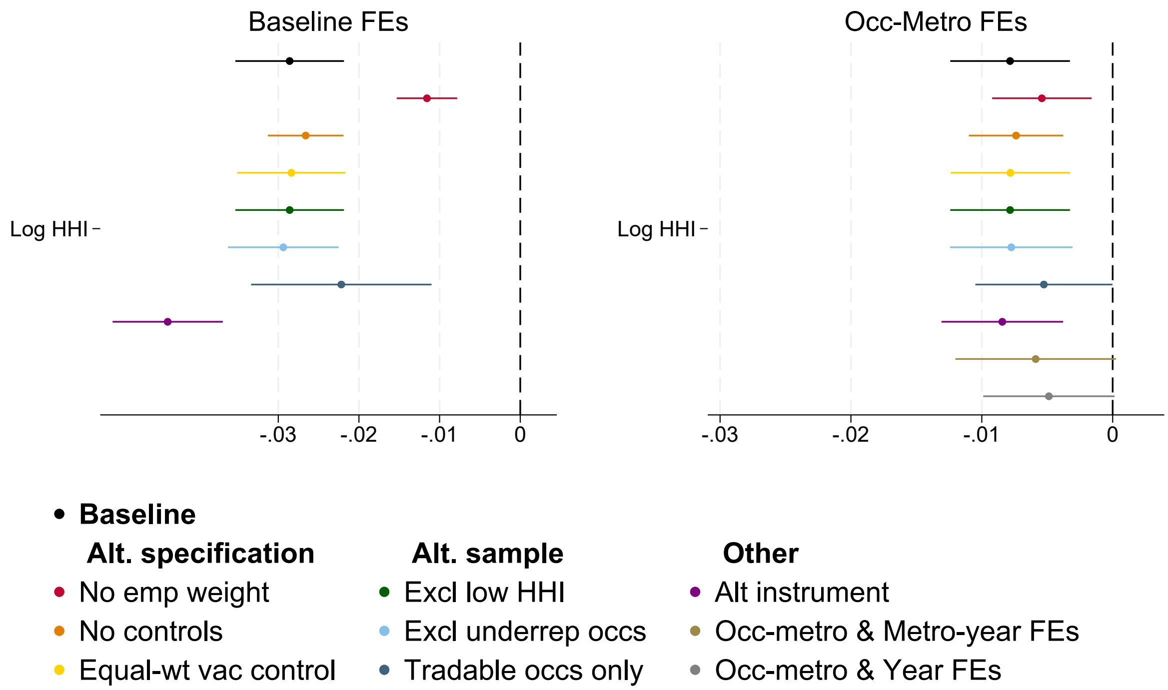
*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This table shows the twenty largest SOC 6-digit occupations which are highly affected by employer concentration, which we define as occupations where at least one third of the national workforce are in labor markets with estimated wage effect of employer concentration greater than 2.5%. The average wage effect is calculated as described in Section 5.

Figure 1: Binned scatter plots of relationship between HHI instrument, HHI, and wage



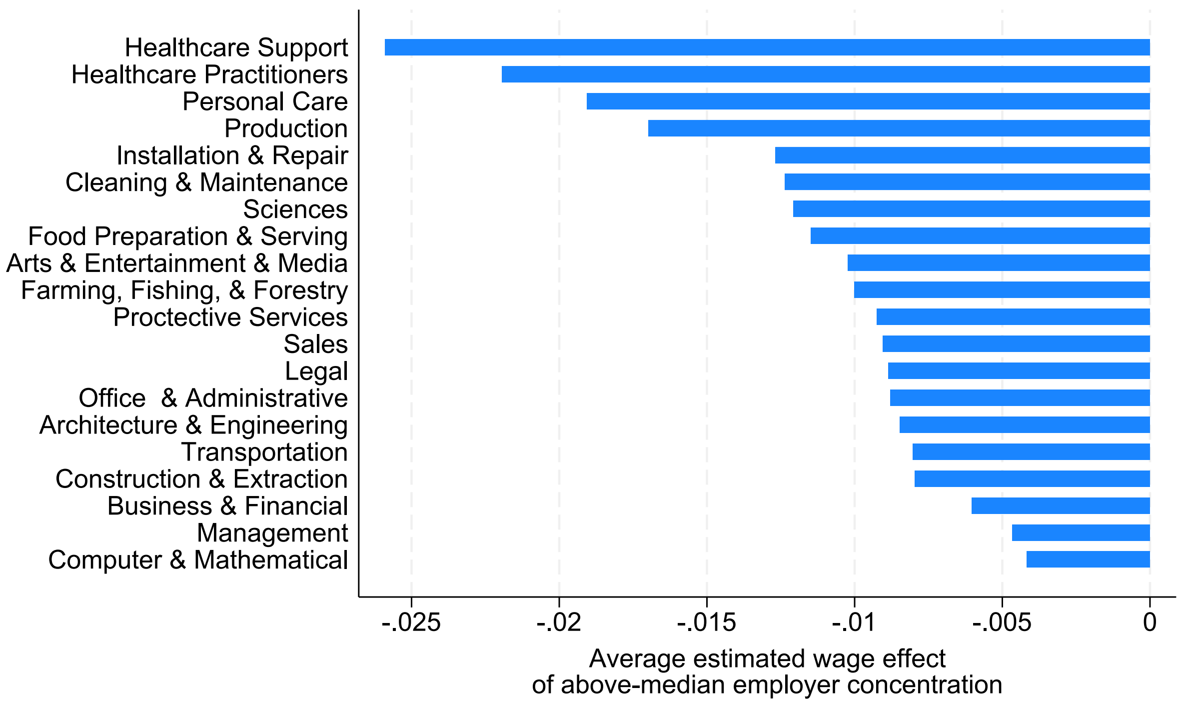

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Each figure is a binned scatter plots showing the relationship between $y$ and $x$ residualized on fixed effects and on our baseline controls. Left panel with "Baseline FEs" residualizes on occ-year and metro-year fixed effects. Right panel with "Occ-Metro FEs" residualizes on occ-metro and occ-year fixed effects. $y$ and $x$ variables are, respectively (i) "OLS": log wage on log HHI; (ii) "Reduced Form": log wage on log HHI instrument; (iii) "First Stage": log HHI on log HHI instrument.

Figure 2: Robustness checks: wage-HHI IV regressions




*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Coefficient plots show alternate specifications for our baseline regressions shown in Table 3. The left panel "Baseline FEs" has the same fixed effects as in Panel A of Table 3 (occ-year and metro area-year); the right panel "Occ-Metro FEs" has the same fixed effects as in Panel B of Table 3 (occ-metro area and occ-year). The black coefficient reproduces the main coefficient from Table 3. Each other coefficient shows a regression with one variation relative to the baseline. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms). "Occ-metro & Metro-year FEs" and "Occ-metro & Year FEs" show, respectively, regressions with each of these sets of fixed effects.

Figure 3: Most-affected occupations by employer concentration, by SOC 2-digit occupation group

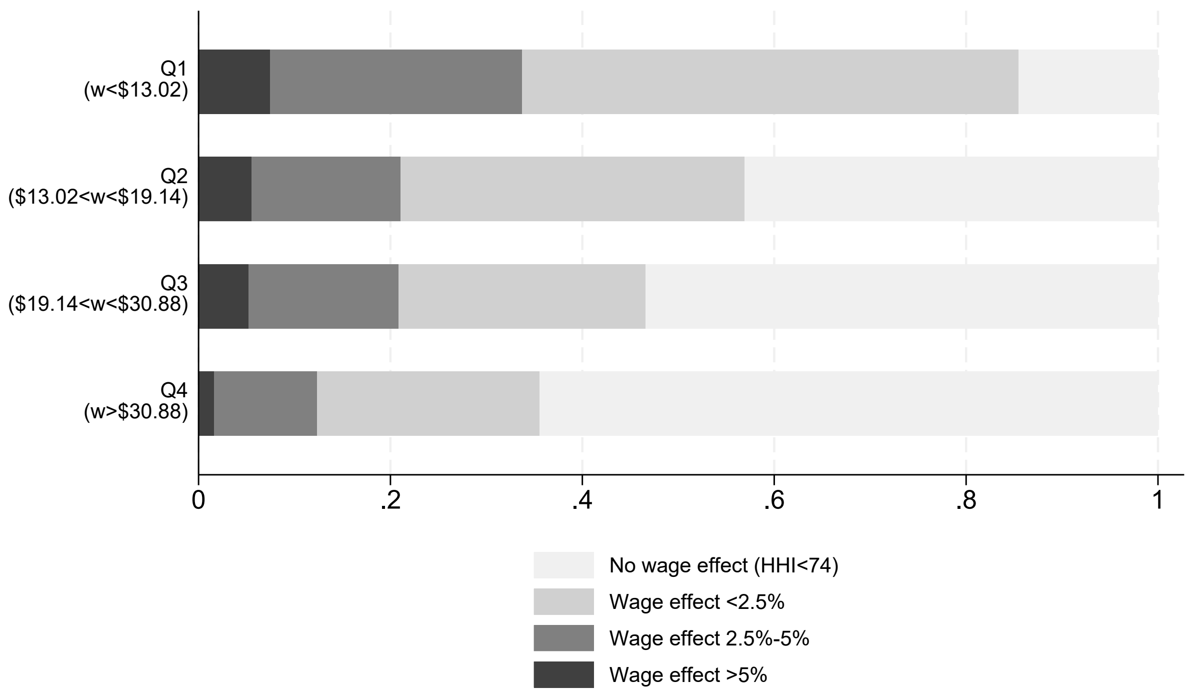



*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This figure shows the average estimated wage effect of above-median employer concentration by SOC 2-digit occupation group, where the average wage effect is calculated as described in Section 5.

# Online Appendix

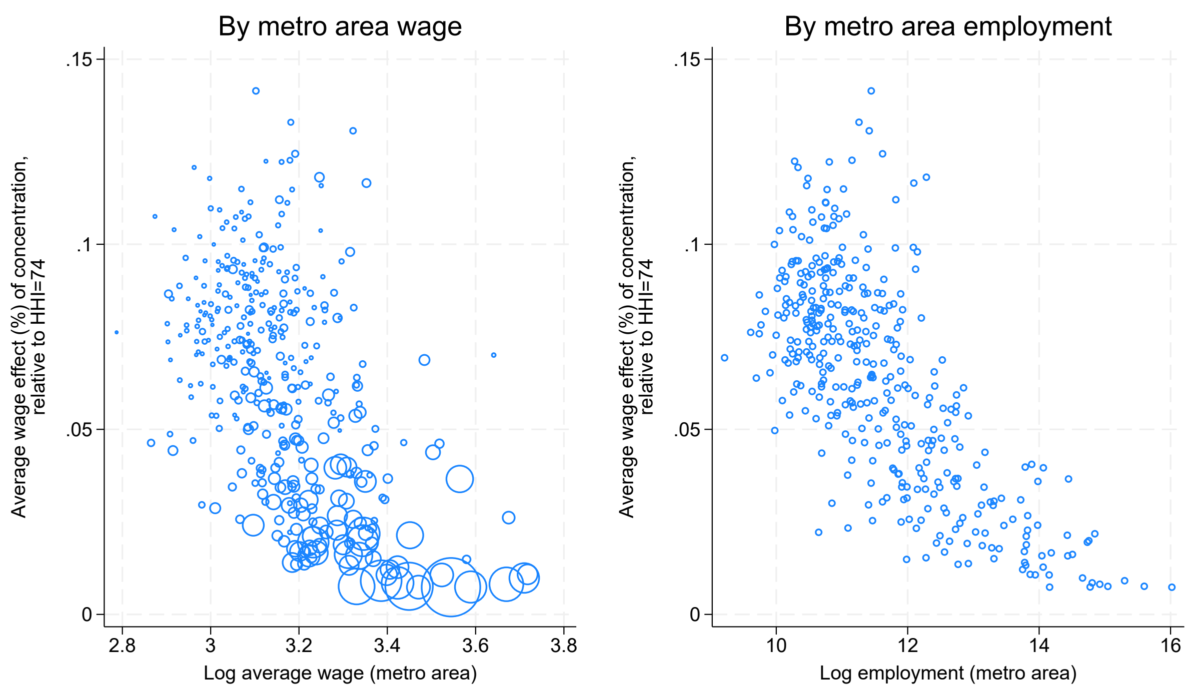
# A Appendix: Additional Tables and Figures

Figure A1: Estimated wage effect of above-median employer concentration, by quartile of the wage distribution




*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This figure shows the average estimated wage effect of above-median employer concentration, by quartile of the wage distribution. 6-digit SOC occupation-metro area cells are grouped into quartiles by their average hourly wage in the BLS OEWS. The average wage effect is calculated as described in Section 5.

Figure A2: Estimated wage effect of above-median employer concentration, by metro area




*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This figure shows the average estimated wage effect of above-median employer concentration, by metro area. For the left panel, bubble size reflects metro area employment. The average wage effect is calculated as described in Section 5.

Table A1: Robustness regressions, baseline fixed effects

| | No emp. weight (1) | No controls (2) | Equal-wt. vac. control (3) | Excl. low HHI (4) | Excl. underrep. occs (5) | Tradable occs only (6) | Alt. instrument (7) |
|---|---|---|---|---|---|---|---|
| Log HHI | -0.012*** | -0.027*** | -0.028*** | -0.029*** | -0.029*** | -0.022*** | -0.044*** |
| | (0.002) | (0.002) | (0.003) | (0.003) | (0.004) | (0.006) | (0.003) |
| Vacancy growth | -0.003*** | | -0.005*** | -0.004*** | -0.004*** | -0.004** | -0.004*** |
| | (0.001) | | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) |
| Predicted vacancy growth | 0.003*** | | 0.005*** | 0.005*** | 0.005*** | 0.001 | 0.008*** |
| | (0.001) | | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Expected HHI instrument | -0.001* | | 0.002 | 0.002 | 0.002 | 0.011*** | 0.008*** |
| | (0.000) | | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Equal-wt. vacancy growth | | | 0.002*** | | | | |
| | | | (0.001) | | | | |
| Observations | 366,250 | 395,450 | 366,250 | 366,250 | 355,342 | 172,425 | 364,435 |
| F-stat | 9,485 | 2,494 | 1,322 | 1,326 | 1,285 | 799 | 1,503 |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This table shows the coefficients corresponding to the robustness exercises in Figure 2, left panel. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at occupation-metro area level in parentheses. Occupation-year and metro area-year fixed effects. Weighted by average employment of occ-metro area cell unless otherwise noted. Unit of analysis: MALMA occupation X metro area X year, 2011-2019. Each column replicates the main IV specification in Table 3, Panel A, column 2, but with a change as denoted in the row header. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms).

Table A2: Robustness regressions, occ-metro fixed effects

| | No emp weight | No controls | Equal-wt. vac. control | Excl. low HHI | Excl. underrep. occs | Tradable occs only | Alt. inst-rument | Metro-year FEs | Year FEs |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log HHI | -0.005*** | -0.007*** | -0.008*** | -0.008*** | -0.008*** | -0.005** | -0.008*** | -0.006* | -0.005* |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) | (0.003) | (0.003) |
| Vacancy growth | -0.002*** | | -0.003*** | -0.002*** | -0.003*** | -0.004*** | -0.003*** | -0.001 | -0.003*** |
| | (0.000) | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Predicted vacancy | 0.001 | | 0.003*** | 0.003*** | 0.003*** | 0.002 | 0.003*** | -0.004*** | -0.003** |
| growth | (0.001) | | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) | | |
| Expected HHI | 0.000 | | -0.000 | -0.000 | 0.000 | -0.000 | -0.000 | 0.000 | -0.000 |
| instrument | (0.000) | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Equal wt. | | | 0.000 | | | | | | |
| vac. growth | | | (0.000) | | | | | | |
| Observations | 349,826 | 379,300 | 349,826 | 349,826 | 340,411 | 164,531 | 348,022 | 350,219 | 350,224 |
| F-stat | 5,736 | 1,730 | 1,214 | 1,218 | 1,164 | 884 | 1,300 | 843 | 1,089 |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This table shows the coefficients corresponding to the robustness exercises in Figure 2, right panel. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at occupation-metro area level in parentheses. Occupation-metro area and occupation-year fixed effects unless otherwise noted. Weighted by average employment of occ-metro area cell unless otherwise noted. Unit of analysis: MALMA occupation X metro area X year, 2011-2019. Each column replicates the main IV specification in Table 3, Panel B, column 2, but with a change as denoted in the row header. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms). "Metro-year FEs" shows results with metro area-year and occ-metro fixed effecs. "Year FEs" shows results with year and occ-metro fixed effects.

# B Appendix: Alternate Labor Market Definitions

In our paper, we use our MALMAs (mobility-adjusted labor markets) for our labor market definition rather than official occupation categories. In this appendix, we discuss the advantages of using MALMAs as opposed to SOC 6-digit or 3-digit occupations in section B.1, and we replicate our main wage-concentration analyses in the paper at the SOC 6-digit and 3-digit level in section B.2.

## B.1 Advantages of MALMAs

When defining a labor market, we face a trade-off between two features:

1. *Market completeness:* A market definition for worker $i$ is more complete when more of worker $i$'s set of possible job options are captured within that market.
2. *Market relevance:* A market definition for worker $i$ has higher relevance when more of the jobs within the market are actual possible job options for worker $i$.

These two factors trade off against each other because as the market definition gets broader, market completeness will increase (as the market captures a greater share of all possible job options), but market relevance will decrease (as the jobs within the market, on average, are less relevant for the worker $i$ in question).

We can construct empirical analogs for these two concepts as follows:

1. *Market completeness:* The share of transitions which remain *within* the market. Specifically, we calculate $\sum_p \pi_{o\to p}$ across all 6-digit occupations $p$ which are within the broader market definition. For 6-digit occupations, this is simply $\pi_{o\to o}$; for the 3-digit occupation, this is the sum across all 6-digit occupations within the broader 3-digit occupation, and for the MALMA this is the sum across all 6-digit occupations within the MALMA.
2. *Market relevance:* The average transition share *to* jobs within the market (weighted by each destination occupation's employment share within the market). Specifically, we calculate $\sum_p \frac{emp_p}{\sum_p emp_p}\pi_{o\to p}$. For 6-digit occupations this is again, simply $\pi_{o\to o}$. For 3-digit occupations, the set of occupations $p$ are the 6-digit occupations within the 3-digit occupation. For MALMAs, the set of occupations $p$ are the 6-digit occupations within the MALMA.

Figure B1 shows kernel density plots of market completeness and market relevance for each of our market definitions. The figure shows the expected tradeoffs between relevance and completeness: for the most narrowly defined labor market (SOC 6-digit), relevance is highest but completeness is lowest. However, it also shows that not all broader labor market definitions are created equal: specifically, both the MALMA and the 3-digit SOC perform roughly the same on completeness (in fact, the MALMA labor markets are slightly more complete on average), *but* the MALMA labor markets perform significantly better on relevance. Thus, the SOC 3-digit labor market is strictly worse as a market definition than the MALMA, since it performs worse on relevance and no better on completeness.

This can also be shown in Figure B2. For each SOC 6-digit starting occupation, this plots the difference in market completeness and relevance, respectively, between its analogous 3-digit or MALMA market definition. The red lines at zero split the figure into quadrants. In the top right quadrant, the 3-digit market definition strictly dominates the MALMA market definition: the market is both more complete and more relevant. There are almost no occupations in this category. In the bottom left quadrant, the MALMA market definition strictly dominates the 3-digit definition. There are many occupations in this category. In the top left and bottom right quadrants, there is a tradeoff between relevance and completeness when comparing the two definitions. The overall pattern shows that in the cases where the 3-digit occupation is a little more complete (bottom right quadrant), it is often substantially less relevant than the MALMA; and in the cases where the 3-digit occupation is more relevant (top left quadrant), it is often substantially less complete.

Why does the MALMA tend to perform better on completeness and relevance than the 3-digit SOC? First, MALMAs are narrower: the SOC 3-digit labor markets simply incorporate many additional irrelevant SOC 6-digit occupations. Moreover, the SOC 3-digit occupations often do not capture important outside options, whereas MALMAs do. This is because many workers have relevant outside options which are not captured by the SOC 3-digit aggregation logic (which is not necessarily based on skill similarity). We give a few examples from large occupations in Table B1, and discuss each of them below.

Registered nurses rarely leave their occupation, and if they do, their only frequent transition path is to become medical and health services managers ($\pi_{o\to p} = 4.2\%$). This occupation is not in their SOC 3-digit occupation. But, there are 31 other SOC 6-digit occupations within registered nurses' SOC 3-digit occupation, almost all of which are entirely irrelevant as outside options for registered nurses ($\pi_{o\to p} < 0.1\%$), including surgeons, veterinarians, optometrists, dentists, pharmacists. Thus, the 3-digit occupation for nurses is far less relevant than the MALMA, and also is less complete.

Accountants and auditors' MALMA includes financial managers, which is a frequent transition path ($\pi_{o\to p} = 4.0\%$) but is not in their 3-digit occupation. There are many occupations within accountants and auditors' 3-digit occupation which are effectively irrelevant to them ($\pi_{o\to p} < 0.1\%$), including insurance underwriters, loan officers, and appraisers of real estate.

Restaurant cooks' MALMA includes two other SOC 6-digit occupations: Chefs and head cooks ($\pi_{o\to p} = 3.1\%$), and Food service managers ($\pi_{o\to p} = 2.0\%$). Neither of these are within their 3-digit SOC occupation. In contrast, their 3-digit occupation comprises 4 other SOC 6-digit occupations which are mostly not relevant to restaurant cooks: Food preparation workers, Institution and cafeteria cooks, Short order cooks, and Fast food cooks. These are not relevant transition paths for restaurant cooks in large part because restaurant cooks tend to be higher-paying jobs requiring substantially more advanced and different skills. But, these are also very large occupations, meaning that the average relevance of the SOC 3-digit occupation is very low (since there are many jobs included which are not relevant for restaurant cooks).

Retail salespersons' MALMA includes only that single SOC 6-digit occupation. This is because while retail salespersons frequently leave their job, they move in a very dispersed fashion to a very wide range of other jobs. In contrast, their 3-digit occupation also includes cashiers, counter and rental clerks, parts salespersons, and gaming change persons. Retail salespersons almost never move to any of these jobs (the thickest transition path, to cashiers, is $\pi_{o \to p} = 0.8\%$), but because cashiers in particular is such a large occupation, the inclusion of cashiers in retail salespersons' labor market makes the average job in this labor market much less relevant.

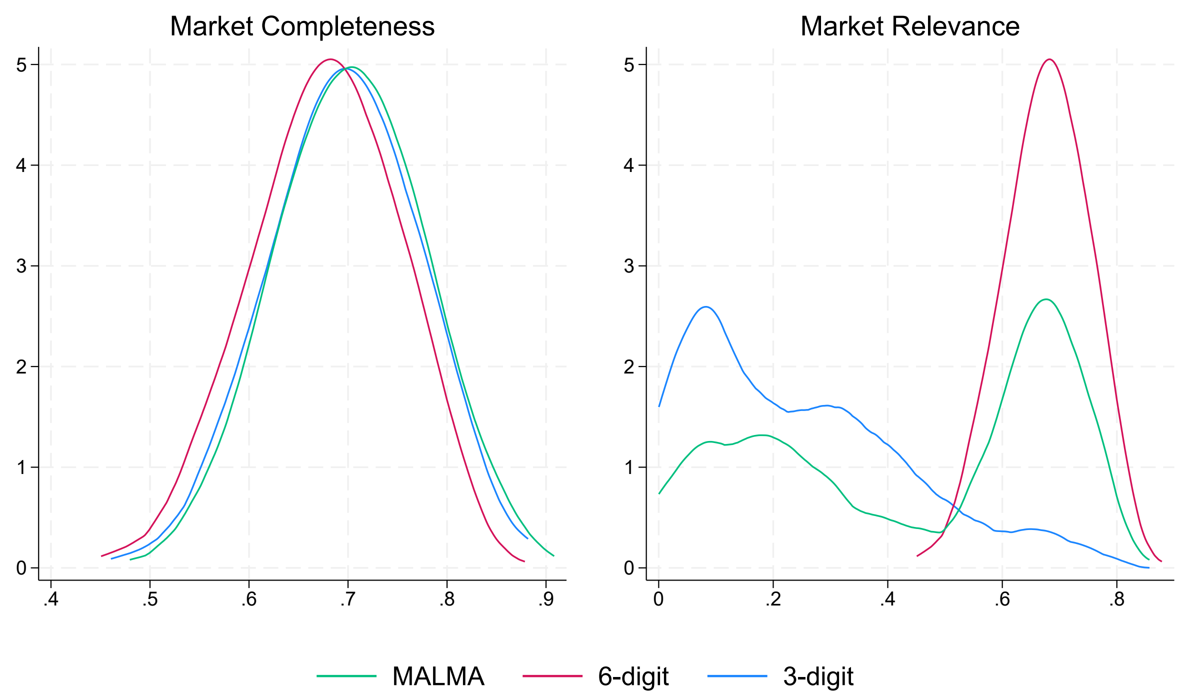
As noted in the main text, MALMAs also are able to capture the asymmetry of occupational mobility. The importance of this is highlighted by the examples in Appendix Table B2, which shows that there are many large occupation pairs where flows are frequent in one direction, but not likely in the opposite direction.

Table B1: Example large occupations where MALMA has substantially greater relevance than 3-digit

| Occupation (SOC 6-digit) | Employment (2012, national) | MALMA completeness | 3-digit completeness | MALMA relevance | 3-digit relevance |
|---|---|---|---|---|---|
| Registered Nurses | 2,633,980 | 77% | 74% | 69% | 42% |
| Accountants and auditors | 1,129,340 | 80% | 78% | 56% | 36% |
| Cooks, restaurant | 1,000,710 | 70% | 66% | 54% | 23% |
| Retail salespersons | 4,340,000 | 64% | 65% | 64% | 34% |

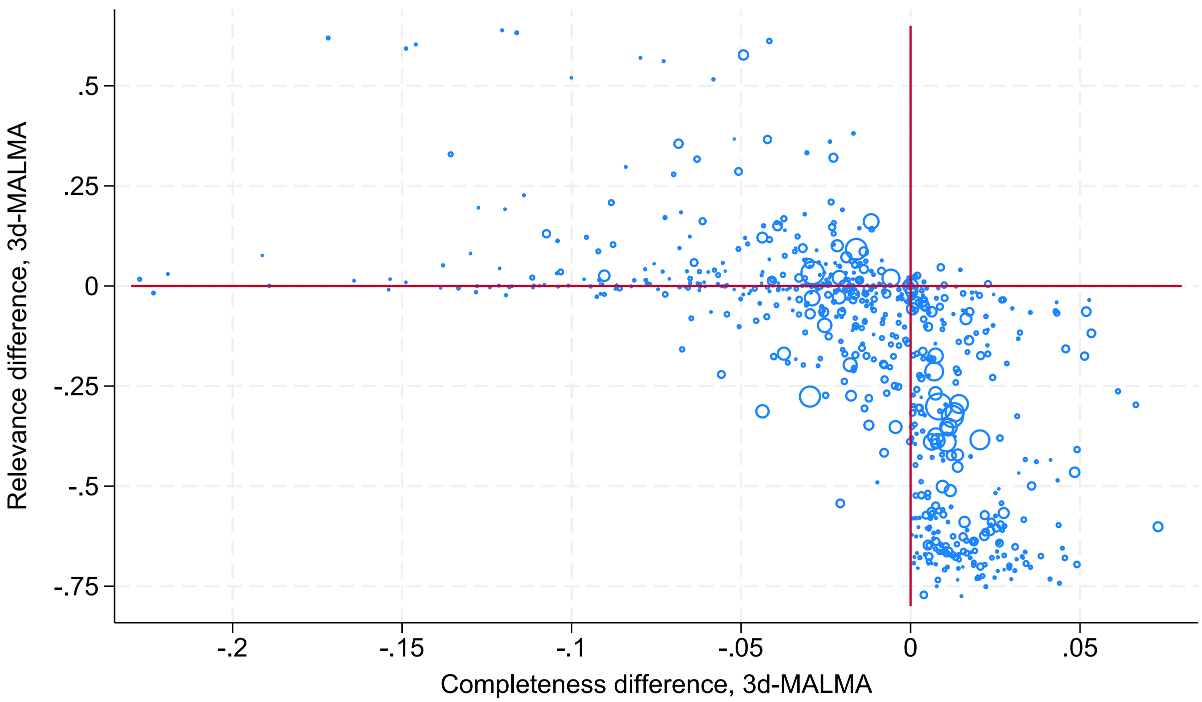
*Source:* Relevance and completeness from BGT resume data; wage from BLS OEWS. *Notes:* Table shows some example large occupations where MALMA market relevance is substantially greater than 3-digit market relevance, with little or no tradeoff on market completeness.

Figure B1: Market relevance and completeness for three labor market definitions




*Source:* BGT resume data. *Notes:* Figure shows completeness and relevance calculated from SOC 6-digit level occupation transition data extracted from BGT resumes. Figure is at occupation level, weighted by national occupation employment in 2012 from BLS OEWS data. Occupations with less than 500 job-year observations in the BGT resume data are excluded.

Figure B2: Difference in market relevance and completeness for 3-digit vs MALMA labor markets




*Source:* BGT resume data. *Notes:* Figure shows difference in completeness and relevance for each starting SOC 6-digit occupation, calculating the difference as the 3-digit completeness (or relevance) minus the MALMA completeness (or relevance). Figure is at 6-digit starting occupation level, with bubbles weighted by national occupation employment in 2012 from BLS OEWS data. Occupations with less than 500 job-year observations in the BGT resume data, and occupations with less than 10,000 employees nationwide in 2012, are excluded.

Table B2: Examples of large occupation pairs with asymmetric worker mobility

| Origin Occupation Title | Origin Wage ($) | Destination Occupation Title | Dest. Wage ($) | $\pi_{1\to2}$ (%) | $\pi_{2\to1}$ (%) |
|---|---|---|---|---|---|
| Licensed practical and licensed vocational nurses | 20.39 | Registered nurses | 32.66 | 7.06 | 0.53 |
| HR assistants, exc. payroll & timekeeping | 18.43 | Human resources specialists | 29.16 | 4.80 | 0.29 |
| Legal secretaries | 21.34 | Paralegals and legal assistants | 24.15 | 3.49 | 0.57 |
| Tellers | 12.40 | Customer service representatives | 15.92 | 3.10 | 0.21 |
| Receptionists and information clerks | 13.00 | Secretaries & admin. assist. exc. legal, med., & exec. | 16.13 | 2.98 | 0.42 |
| Lifeguards, ski patrol, & other recr. protective svcs | 9.96 | Fitness trainers and aerobics instructors | 17.74 | 2.91 | 0.59 |
| Cashiers | 9.79 | Customer service representatives | 15.92 | 2.93 | 0.61 |
| Sales representatives, services, all other | 29.22 | Sales rep., whsle & mfg., exc. tech. & scient. prod. | 30.91 | 2.48 | 0.29 |
| Computer network support specialists | 30.27 | Network and computer systems administrators | 36.69 | 2.68 | 0.53 |
| Phlebotomists | 14.86 | Medical and clinical laboratory technicians | 18.91 | 2.51 | 0.40 |
| Carpenters | 21.41 | Heavy and tractor-trailer truck drivers | 19.40 | 2.14 | 0.08 |
| Bakers | 12.05 | Chefs and head cooks | 22.39 | 2.31 | 0.28 |
| Loan interviewers and clerks | 17.40 | Loan officers | 33.82 | 2.77 | 0.76 |
| Dishwashers | 9.10 | Laborers; freight, stock, & material hand-movers | 12.70 | 2.11 | 0.11 |
| Telecomm. equip. (exc. line) installers/repairers | 25.82 | Computer user support specialists | 24.10 | 2.10 | 0.15 |
| Dishwashers | 9.10 | Cooks, restaurant | 11.20 | 2.18 | 0.26 |
| Industrial truck and tractor operators | 15.43 | Laborers; freight, stock, & material hand-movers | 12.70 | 2.87 | 0.97 |
| Bartenders | 10.40 | Waiters and waitresses | 9.95 | 3.31 | 1.43 |
| Insurance sales agents | 30.48 | Sales rep., whsle & mfg., exc. tech. & scient. prod. | 30.91 | 2.19 | 0.33 |
| Food batchmakers | 13.63 | Chefs and head cooks | 22.39 | 1.89 | 0.04 |

*Source:* BGT resume data; BLS OEWS. *Notes:* Table shows examples of the 20 occupation pairs with the largest asymmetry in the probability of origin workers moving to the destination occupation compared to the probability of destination workers moving to the origin occupation. The ranking is computed after filtering for occupation pairs where the origin occupation has at least 100,000 workers nationally in 2012, where origin and destination occupation differ in national employment by a multiple of less than 5, where wage data is available for the origin and destination occupations, and where the destination occupation is not a management occupation (major SOC group 11), not in the same broad SOC group as the origin occupation, and not a residual "…,all other" occupation; and where the origin occupation has a "represented-ness" in the vacancy data of at least 0.2. Wage data in the table represents average hourly wages in 2012 from the BLS OEWS. Where necessary, official occupation names have been abbreviated to retain table formatting.

## B.2 Replication with SOC 6-digit and SOC 3-digit labor markets

In this section, we replicate our baseline analyses in section 4, but using SOC 6-digit or SOC 3-digit occupation boundaries instead of our MALMA occupation labor markets.

First, in Table B3, we show a comparison of the HHI for MALMA vs 6-digit and 3-digit SOC occupations. Panel A shows the distribution of the HHIs across occupation X metro area labor markets in 2019, for each occupation definition, and Panel B shows the employment-weighted distribution (such that, for example, *p50* represents the HHI of the labor market of the median worker). As would be expected, concentration is highest overall for the narrowest labor market (6-digit SOC) and lowest for the broadest labor market (3-digit SOC). The HHI within MALMA labor markets is typically lower than the HHI for the analogous 6-digit labor market, but the ratio varies substantially: indeed, for about 8% of SOC 6-digit labor markets, the HHI calculated over the MALMA is actually higher than the HHI calculated over the SOC 6-digit occupation alone (because the other occupations included in the MALMA are highly concentrated, and/or because they share the same employers as the focal SOC 6-digit occupation).

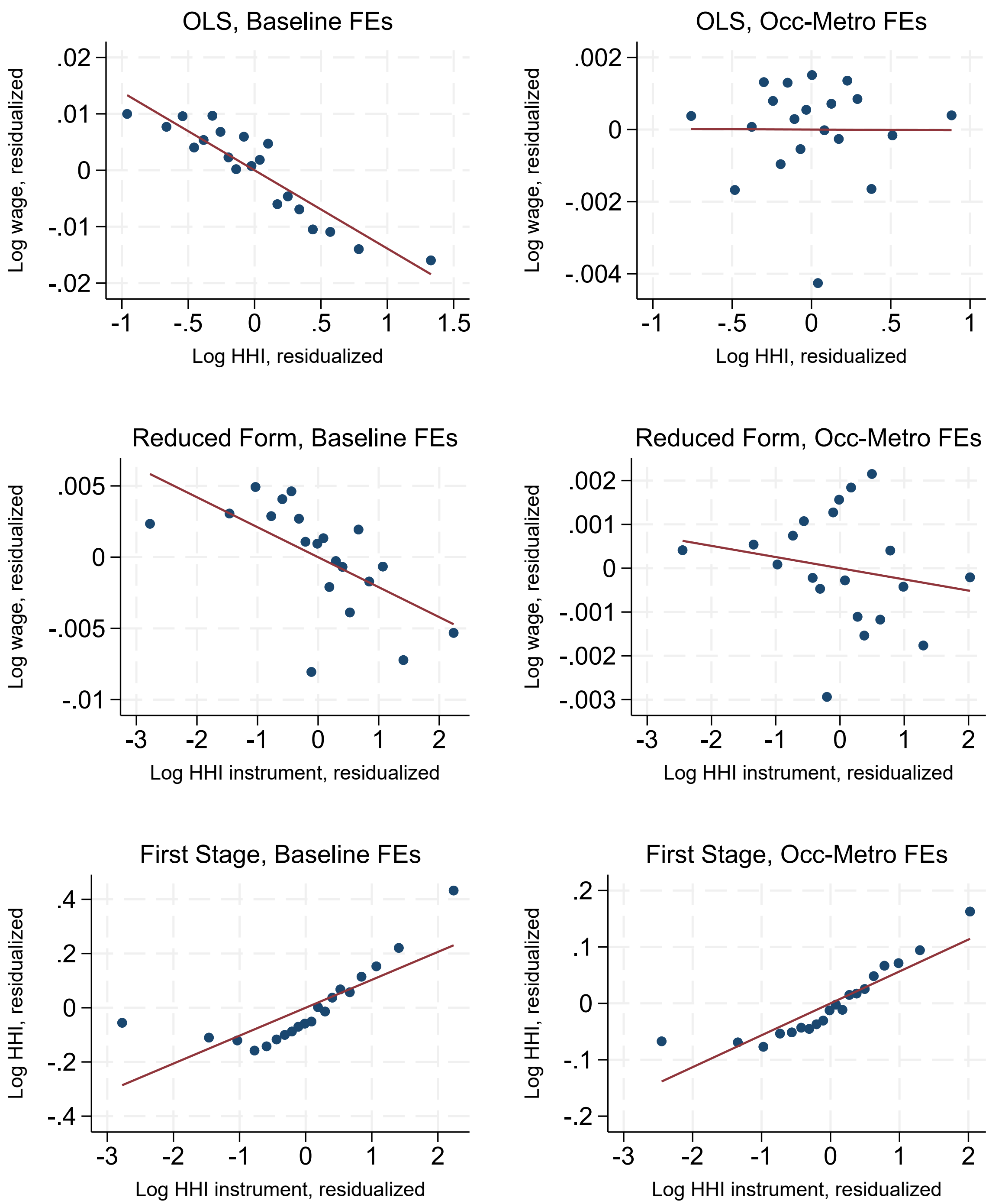
Next in Tables B4 and B5, we show our main regression results for the 6-digit and 3-digit occupation definitions respectively, with binned scatter plots showing the first stage and reduced form relationships in Figures B3 and B4, and robustness checks in Figures B5 and B6. As these set of results show, consistently we find large, negative, and statistically significant effects of employer concentration on wages across almost all specifications for both of these alternate labor market definitions.

Table B3: Comparing HHI MALMA and HHI for SOC 6-digit occupations, 2019

| | p1 | p5 | p10 | p25 | p50 | p75 | p90 | p95 | p99 |
|---|---|---|---|---|---|---|---|---|---|
| *Panel A: Our analysis sample, unweighted* | | | | | | | | | |
| HHI (MALMA) | 12 | 25 | 42 | 99 | 267 | 633 | 1,339 | 2,015 | 3,817 |
| HHI (6-digit SOC) | 24 | 66 | 112 | 251 | 577 | 1,215 | 2,272 | 3,278 | 5,944 |
| HHI (3-digit SOC) | 16 | 41 | 71 | 172 | 433 | 1,046 | 2,273 | 3,600 | 10,000 |
| *Panel B: Our analysis sample, employment-weighted* | | | | | | | | | |
| HHI (MALMA) | 6 | 12 | 17 | 33 | 84 | 216 | 537 | 869 | 2,084 |
| HHI (6-digit SOC) | 8 | 17 | 26 | 59 | 140 | 355 | 808 | 1,289 | 2,959 |
| HHI (3-digit SOC) | 6 | 10 | 13 | 27 | 69 | 168 | 421 | 716 | 1,980 |

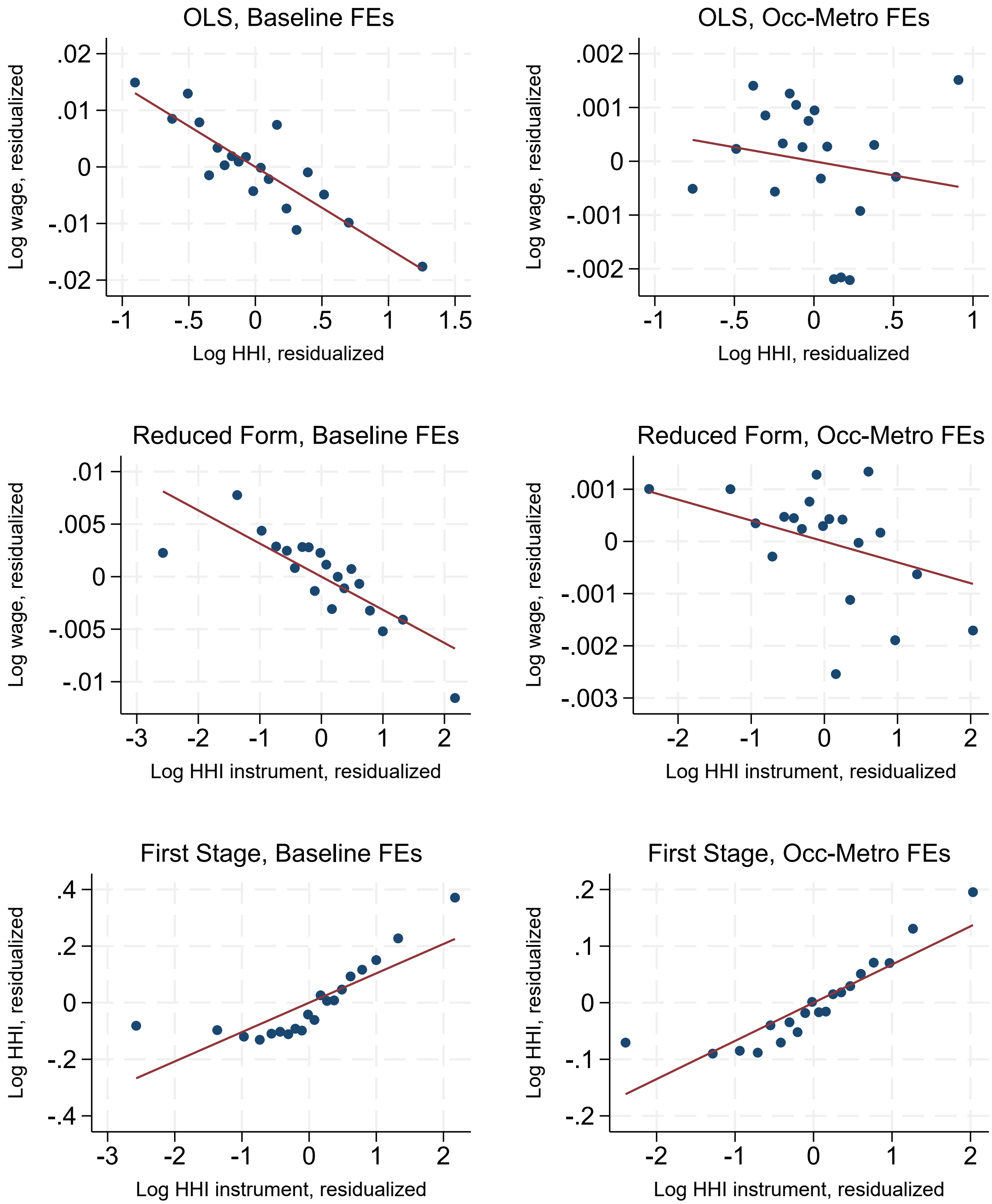
*Source:* BGT vacancy data and BGT resume data; authors' calculations. *Notes:* This table shows the distribution of HHIs across occupation X metro area labor markets in our analysis sample in 2019. Unit of observation is an occupation X metro area cell, with occupation defined as MALMA, SOC 6-digit, or SOC 3-digit respectively. Panel A is unweighted; Panel B weights by cell employment in 2019.

Figure B3: Binned scatter plots: SOC 6-digit



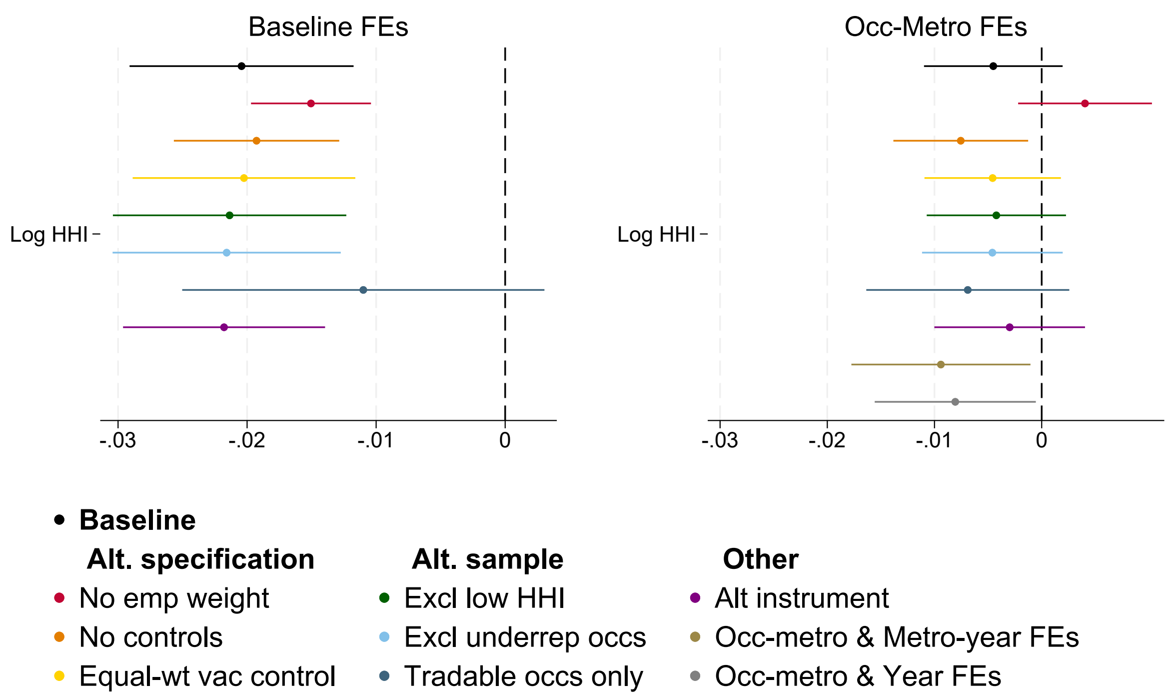

*Source:* BGT vacancy data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Each figure is a binned scatter plots showing the relationship between $y$ and $x$ residualized on fixed effects. Left panel with "Baseline FEs" residualizes on occ-year and metro-year fixed effects. Right panel with "Occ-Metro FEs" residualizes on occ-metro and occ-year fixed effects. $y$ and $x$ variables are, respectively (i) "OLS": log wage on log HHI; (ii) "Reduced Form": log wage on log HHI instrument; (iii) "First Stage": log HHI on log HHI instrument.

Figure B4: Binned scatter plots: SOC 3-digit



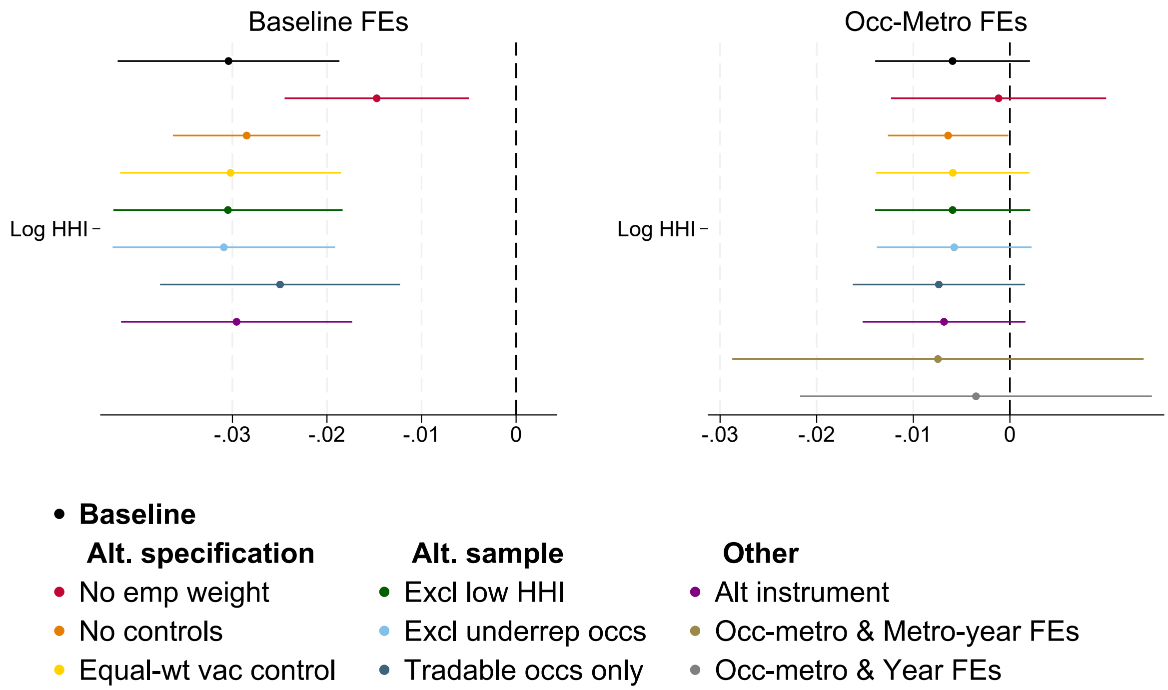

*Source:* BGT vacancy data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Each figure is a binned scatter plots showing the relationship between $y$ and $x$ residualized on fixed effects. Left panel with "Baseline FEs" residualizes on occ-year and metro-year fixed effects. Right panel with "Occ-Metro FEs" residualizes on occ-metro and occ-year fixed effects. $y$ and $x$ variables are, respectively (i) "OLS": log wage on log HHI; (ii) "Reduced Form": log wage on log HHI instrument; (iii) "First Stage": log HHI on log HHI instrument.

Figure B5: Robustness: SOC 6-digit




*Source:* BGT vacancy data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Coefficient plots show alternate specifications for our regressions at SOC 6-digit level shown in Table B4. The left panel "Baseline FEs" has the same fixed effects as in Panel A of Table B4 (occ-year and metro area-year); the right panel "Occ-Metro FEs" has the same fixed effects as in Panel B of Table B4 (occ-metro area and occ-year). The black coefficient reproduces the main coefficient from Table B4. Each other coefficient shows a regression with one variation relative to the baseline. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms). "Occ-metro & Metro-year FEs" and "Occ-metro & Year FEs" show, respectively, regressions with each of these sets of fixed effects.

Figure B6: Robustness: SOC 3-digit




*Source:* BGT vacancy data; BLS OEWS wage and employment data; authors' calculations. *Notes:* Coefficient plots show alternate specifications for our regressions at SOC 3-digit level shown in Table B5. The left panel "Baseline FEs" has the same fixed effects as in Panel A of Table B5 (occ-year and metro area-year); the right panel "Occ-Metro FEs" has the same fixed effects as in Panel B of Table B5 (occ-metro area and occ-year). The black coefficient reproduces the main coefficient from Table B5. Each other coefficient shows a regression with one variation relative to the baseline. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms). "Occ-metro & Metro-year FEs" and "Occ-metro & Year FEs" show, respectively, regressions with each of these sets of fixed effects.

Table B4: Main results: SOC 6-digit

| *Dep var:* | *Log wage* | | | *Log HHI* |
|---|---|---|---|---|
| | OLS | IV | Reduced Form | First Stage |
| | (1) | (2) | (3) | (4) |
| *Panel A: Occ-Year and Metro Area-Year Fixed Effects* | | | | |
| Log HHI | -0.014*** | -0.020*** | | |
| | (0.002) | (0.004) | | |
| Log HHI instrument | | | -0.002*** | 0.103*** |
| | | | (0.000) | (0.004) |
| Vacancy growth | -0.342*** | -0.384*** | -0.243*** | -6.931*** |
| | (0.079) | (0.079) | (0.081) | (0.739) |
| Predicted vacancy growth | 0.041 | 0.128 | -0.067 | 9.550*** |
| | (0.112) | (0.121) | (0.111) | (0.638) |
| Expected HHI instrument | -0.001 | 0.001 | -0.003*** | 0.199*** |
| | (0.001) | (0.001) | (0.001) | (0.008) |
| Observations | 385,249 | 385,249 | 385,249 | 385,249 |
| F-stat | | 809 | | |
| *Panel B: Occ-Metro Area and Occ-Year Fixed Effects* | | | | |
| Log HHI | -0.000 | -0.005 | | |
| | (0.001) | (0.003) | | |
| Log HHI instrument | | | -0.000 | 0.057*** |
| | | | (0.000) | (0.002) |
| Vacancy growth | -0.066 | -0.137** | -0.064 | -16.048*** |
| | (0.041) | (0.057) | (0.040) | (0.561) |
| Predicted vacancy growth | -0.072 | -0.006 | -0.063 | 12.668*** |
| | (0.064) | (0.072) | (0.064) | (0.582) |
| Expected HHI instrument | -0.000 | 0.000 | -0.000 | 0.067*** |
| | (0.000) | (0.000) | (0.000) | (0.004) |
| Observations | 368,449 | 368,449 | 368,449 | 368,449 |
| F-stat | | 600 | | |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at occupation-metro area level in parentheses. Fixed effects noted in Panel titles. Weighted by average employment of occ-metro area cell. Unit of analysis: MALMA occupation X metro area X year, 2011-2019. This table replications Table 3, but using 6-digit occupation X metro area instead of MALMA X metro area as the labor market definition.

Table B5: Main results: SOC 3-digit

| *Dep var:* | *Log wage* | | | *Log HHI* |
|---|---|---|---|---|
| | OLS | IV | Reduced Form | First Stage |
| | (1) | (2) | (3) | (4) |
| *Panel A: Occ-Year and Metro Area-Year Fixed Effects* | | | | |
| Log HHI | -0.014*** | -0.030*** | | |
| | (0.002) | (0.006) | | |
| Log HHI instrument | | | -0.003*** | 0.104*** |
| | | | (0.001) | (0.005) |
| Vacancy growth | -0.008*** | -0.009*** | -0.006*** | -0.093*** |
| | (0.001) | (0.002) | (0.001) | (0.013) |
| Predicted vacancy growth | 0.003 | 0.006*** | 0.002 | 0.116*** |
| | (0.002) | (0.002) | (0.002) | (0.012) |
| Expected HHI instrument | -0.003** | 0.002 | -0.006*** | 0.283*** |
| | (0.002) | (0.003) | (0.002) | (0.010) |
| Observations | 166,241 | 166,241 | 166,241 | 166,241 |
| F-stat | | 491 | | |
| *Panel B: Occ-Metro Area and Occ-Year Fixed Effects* | | | | |
| Log HHI | -0.001 | -0.006 | | |
| | (0.001) | (0.004) | | |
| Log HHI instrument | | | -0.000 | 0.067*** |
| | | | (0.000) | (0.003) |
| Vacancy growth | -0.000 | -0.001 | 0.000 | -0.250*** |
| | (0.001) | (0.001) | (0.001) | (0.008) |
| Predicted vacancy growth | -0.001 | 0.000 | -0.001 | 0.179*** |
| | (0.001) | (0.001) | (0.001) | (0.011) |
| Expected HHI instrument | 0.000 | 0.001 | 0.000 | 0.135*** |
| | (0.001) | (0.001) | (0.001) | (0.007) |
| Observations | 164,135 | 164,135 | 164,135 | 164,135 |
| F-stat | | 423 | | |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at occupation-metro area level in parentheses. Fixed effects noted in Panel titles. Weighted by average employment of occ-metro area cell. Unit of analysis: MALMA occupation X metro area X year, 2011-2019. This table replications Table 3, but using 3-digit occupation X metro area instead of MALMA X metro area as the labor market definition.

# C Appendix: OEWS Replication

In this section, we replicate our main results using the Bureau of Labor Statistics' Occupational Employment and Wage Statistics program ("OEWS") restricted access micro data. The OEWS program collects establishment-level employment and wages, enabling us to construct an occupation-by-metro area level HHI over establishment-level employment (rather than over firm-level vacancy postings, as in our BGT data which we use in the main paper). The rationale for using the OEWS data to construct HHIs is twofold: (1) one may be concerned that vacancy postings do not map one-to-one to true vacancies, which would bias our results if this measurement error were correlated with both our instrument for employer concentration and wages; and (2) different theoretical frameworks predict differently as to whether the employment HHI or vacancy HHI is the best proxy for labor market competition across the set of jobs available to workers.

The OEWS data are constructed from a sample of roughly 200,000 establishments surveyed each May and November. Respondents report employment counts by detailed occupation code and coarse wage bands. Over each three-year period, the OEWS uses employment and wages from these establishment surveys to create estimates of employment and wages for each individual SOC 6-digit occupation within geographic areas. (These average employment and wage estimates, by SOC 6-digit occupation and metro area, are what we use for our wage and employment data in our main analyses). Handwerker and Dey (2024) use the OEWS micro data to estimate employer concentration for each US occupation-by-metro area labor market. Since the OEWS is a survey of establishments, estimating concentration measures requires that they impute occupational employment levels for the non-sampled establishments in the BLS' census of employers. See Handwerker and Dey (2024) for full details on this imputation procedure.

We follow Handwerker and Dey (2024) in estimating HHIs at the level of occupation-by-metro area labor markets. To be consistent with our labor market definitions in this paper, we estimate employment HHIs at the level of (i) MALMA-by-metro area, (ii) 6-digit occupation-by-metro area, and (iii) 3-digit occupation-by-metro area labor markets.[42] We

[42] The 6-digit and 3-digit occupation codes in some cases are slightly differently aggregated in these data than in our main analysis sample, due to data limitations. Note also that the definition of an employer is slightly different than our definition in the BGT data due to data differences: in the OEWS data, a single employer is a collection of establishments which share a common EIN, while in the BGT data, we identify a single employer by employer name as listed on the vacancy posting. As Handwerker and Dey (2024) note, some large firms use multiple EINs, meaning that the OEWS approach to defining an employer may understate employer concentration. Meanwhile, some firms (particularly franchised firms) have the same name without being the same legal employment entity, meaning that our approach to defining an employer in the BGT data may overstate employer concentration in some settings.

show summary statistics of the employment HHI (and average hourly wage) in the OEWS data for 2019 in Table C2.

We find that the employment HHIs constructed from the OEWS data are very highly correlated with the vacancy posting HHIs constructed from the BGT data. Table C1 shows an (employment-weighted) correlation coefficient of 0.61. Moreover, as shown in the binned scatter plots in Figure C1, this correlation holds both in terms of the raw HHIs and in terms of the residualized HHIs (residualized on our baseline occupation-year and metro area-year fixed effects), and holds across all three of our labor market definitions. Thus, to a first approximation we are unlikely to expect measurement error in the BGT vacancy postings data to bias our results.

Next, we use these data to replicate our baseline regressions, using our original instrument for the HHI (constructed from national hiring shocks from the BGT vacancy postings), but using the OEWS employment HHIs, employment, and wage data. The dependent variable, as before, is the log average hourly wage (constructed from the OEWS).

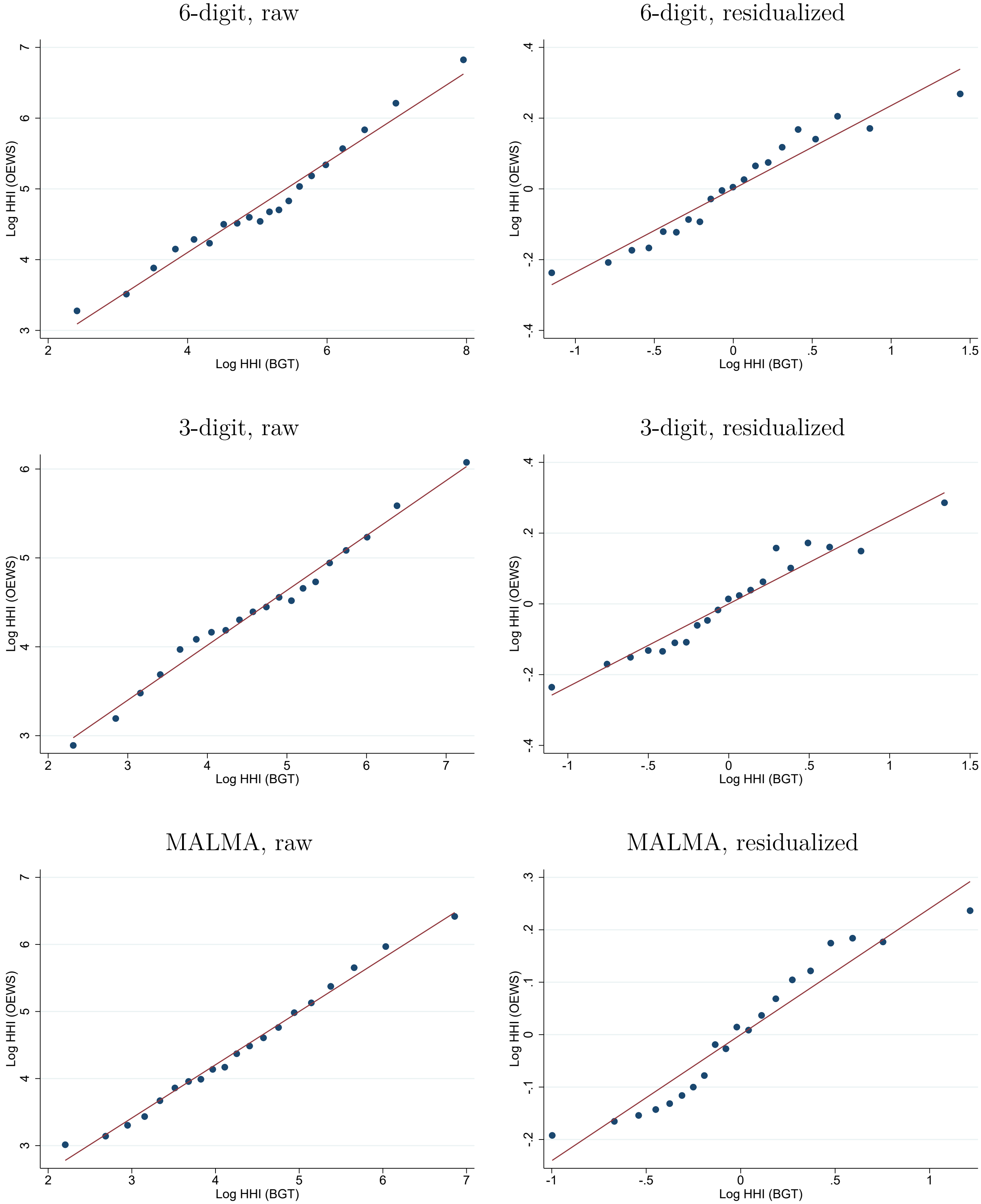
Our results replicate well in the OEWS data. As shown in Table C3, we find large, negative, and statistically significant effects of employer concentration on wages in our baseline specification with occupation-year and metro area-year, using our HHI instrument (constructed from the BGT postings data using shifts in large employer hiring behavior), but using this to instrument for the OEWS HHI. With our baseline fixed effects, our hiring instrument is a substantially weaker instrument for the OEWS establishment employment HHI than it is for the BGT firm vacancy posting HHI (as would be expected), but the Kleibergen-Paap F-stat is still sufficiently strong at 38.

We similarly find large, negative, and statistically significant effects of employer concentration on wages at the 6-digit or 3-digit occupation level (Table C4). There is a consistently negative causal effect of HHI on wages across our suite of robustness checks (Table C5).[43]

Unfortunately, we cannot replicate our main regressions with occupation-metro area fixed effects (the results in Table 3, Panel B). This is because our hiring instrument from the BGT data is no longer an effective instrument for the OEWS employment HHI when we include occupation-metro area fixed effects: depending on the specification, we either find a serious weak instrument problem (with a Kleibergen-Paap F-stat of less than 2), and/or a weakly negative first stage. Thus, we cannot use the BGT hiring instrument to instrument for OEWS employment HHI with occ-metro fixed effects, and so we do not present any regression results from the OEWS data for specifications with occupation-metro area fixed effects.

[43] As in our main regressions, we limit the sample in the MALMA regressions to only those occupation-metro area cells for which we could construct an HHI, wage, and HHI instrument for the focal 6-digit occupation.

Figure C1: Comparing OEWS and BGT HHIs




*Source:* Log Employment HHI from OEWS data at occupation by metro area level; Log Vacancy HHI from BGT data at occupation by metro area level. *Notes:* "Raw" shows the raw binscatter of the OEWS HHI on BGT HHI. "Residualized" means residualized on our baseline fixed effects: occ-year and metro area-year.

Table C1: Correlation between BGT and OEWS HHIs

| | MALMA | 6-digit | 3-digit |
|---|---|---|---|
| Weighted | 0.61 | 0.61 | 0.61 |
| Unweighted | 0.49 | 0.58 | 0.57 |

*Source:* BGT vacancy data; OEWS data. *Notes:* Table shows correlation between HHIs calculated in BGT vacancy data and OEWS employment data. HHIs are calculated over markets defined as, respectively for each column, MALMA x metro area, SOC 6-digit x metro area, or SOC 3-digit x metro area. Rows report correlations weighted by cell employment or unweighted.

Table C2: Summary statistics across occ-metro labor markets, 2019, OEWS data

| Variable | p1 | p5 | p10 | p25 | p50 | p75 | p90 | p95 | p99 |
|---|---|---|---|---|---|---|---|---|---|
| *MALMA, unweighted* | | | | | | | | | |
| OEWS HHI | 13 | 37 | 67 | 192 | 549 | 1,505 | 3,580 | 5,556 | 10,000 |
| OEWS Wage | 10.81 | 13.53 | 17.41 | 25.73 | 36.94 | 47.18 | 56.87 | 64.41 | 87.54 |
| *MALMA, employment-weighted* | | | | | | | | | |
| OEWS HHI | 4 | 8 | 12 | 29 | 67 | 199 | 704 | 1,301 | 3,161 |
| OEWS Wage | 10.64 | 11.54 | 12.42 | 17.03 | 38.13 | 52.02 | 63.13 | 70.21 | 82.14 |
| *6-digit, unweighted* | | | | | | | | | |
| OEWS HHI | 34 | 101 | 178 | 467 | 1,372 | 3,635 | 8,347 | 10,000 | 10,000 |
| OEWS Wage | 10.04 | 12.75 | 15.34 | 21.10 | 29.40 | 41.74 | 58.32 | 74.83 | 132.30 |
| *6-digit, employment-weighted* | | | | | | | | | |
| OEWS HHI | 6 | 14 | 22 | 41 | 97 | 284 | 964 | 1,923 | 5,000 |
| OEWS Wage | 9.44 | 10.68 | 11.85 | 14.18 | 23.46 | 38.84 | 55.60 | 68.00 | 98.62 |
| *3-digit, unweighted* | | | | | | | | | |
| OEWS HHI | 20 | 51 | 84 | 186 | 512 | 1,479 | 3,527 | 5,556 | 10,000 |
| OEWS Wage | 9.46 | 11.13 | 12.38 | 15.53 | 20.78 | 29.83 | 41.29 | 50.02 | 65.08 |
| *3-digit, employment-weighted* | | | | | | | | | |
| OEWS HHI | 5 | 11 | 16 | 33 | 74 | 180 | 482 | 910 | 3,134 |
| OEWS Wage | 9.80 | 11.20 | 12.36 | 15.01 | 19.35 | 31.83 | 48.43 | 58.31 | 74.76 |

*Source:* OEWS. *Notes:* Table shows percentiles of HHI and wage across occupation-metro area labor markets in the 2019 OEWS. Occupations are defined, respectively, at MALMA, SOC 6-digit, or SOC 3-digit level. Employment-weighted statistics are weighted by cell employment in 2019.

Table C3: OEWS replication: main results (MALMA labor markets)

| *Dep var:* | | *Log wage* | | *Log HHI* |
|---|---|---|---|---|
| | OLS | IV | Reduced Form | First Stage |
| | (1) | (2) | (3) | (4) |
| Log HHI | 0.003 | -0.109*** | | |
| | (0.002) | (0.028) | | |
| Log HHI instrument | | | -0.002*** | 0.022*** |
| | | | (0.000) | (0.004) |
| Vacancy growth | -0.012*** | -0.012*** | -0.012*** | 0.003 |
| | (0.002) | (0.003) | (0.002) | (0.014) |
| Predicted vacancy growth 0.005* | 0.002 | 0.008*** | -0.051*** | |
| | (0.003) | (0.003) | (0.003) | (0.018) |
| Expected HHI instrument | -0.009*** | 0.007 | -0.007*** | 0.131*** |
| | (0.002) | (0.004) | (0.002) | (0.011) |
| Observations | 268,192 | 268,192 | 268,192 | 268,192 |
| F-stat | | 38 | | |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* * p<0.10, ** p<0.05, *** p<0.01. This table replications our main Table 3, Panel A, but using the OEWS data to calculate HHIs (rather than the BGT vacancy data). Standard errors clustered at occupation-metro area level in parentheses. Occupation-year and metro area-year fixed effects. Weighted by average employment of occ-metro area cell. Unit of analysis: MALMA occupation X metro area X year, 2011-2019.

Table C4: OEWS replication: main results for 6-digit and 3-digit occupation labor markets

| *Dep var:* | *Log wage* | | | *Log HHI* |
|---|---|---|---|---|
| | OLS | IV | Reduced Form | First Stage |
| | (1) | (2) | (3) | (4) |
| *Panel A: 6-digit occupations* | | | | |
| Log HHI | 0.006** | -0.130*** | | |
| | (0.003) | (0.042) | | |
| Log HHI instrument | | | -0.003*** | 0.020*** |
| | | | (0.001) | (0.004) |
| Vacancy growth | -0.007*** | -0.007*** | -0.007*** | 0.002 |
| | (0.001) | (0.002) | (0.001) | (0.008) |
| Predicted vacancy growth | -0.000 | -0.000 | 0.001 | -0.012 |
| | (0.002) | (0.002) | (0.002) | (0.012) |
| Expected HHI instrument | -0.012*** | 0.004 | -0.009*** | 0.104*** |
| | (0.002) | (0.005) | (0.002) | (0.010) |
| Observations | 186,580 | 186,580 | 186,580 | 186,580 |
| F-stat | | 24 | | |
| *Panel B: 3-digit occupations* | | | | |
| Log HHI | 0.004 | -0.078*** | | |
| | (0.003) | (0.030) | | |
| Log HHI instrument | | | -0.002*** | 0.024*** |
| | | | (0.001) | (0.005) |
| Vacancy growth | -0.004*** | -0.005*** | -0.004*** | -0.005 |
| | (0.001) | (0.001) | (0.001) | (0.008) |
| Predicted vacancy growth | 0.001 | 0.001 | 0.002 | -0.013 |
| | (0.002) | (0.002) | (0.002) | (0.010) |
| Expected HHI instrument | -0.007*** | 0.002 | -0.005*** | 0.085*** |
| | (0.002) | (0.003) | (0.002) | (0.011) |
| Observations | 283,750 | 283,750 | 283,750 | 283,750 |
| F-stat | | 23 | | |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* * p<0.10, ** p<0.05, *** p<0.01. Panel A of this table replicates Panel A in Table B4, defining labor markets at the SOC 6-digit by metro area level, but using the OEWS data to calculate HHIs (rather than the BGT vacancy data). Panel B of this table replicates Panel A in Table B5, defining labor markets at the SOC 3-digit by metro area level, but using the OEWS data to calculate HHIs. Standard errors clustered at occupation-metro area level in parentheses. Occupation-year and metro area-year fixed effects. Weighted by average employment of occ-metro area cell. Unit of analysis: MALMA occupation X metro area X year, 2011-2019.

Table C5: OEWS replication: robustness

| | No emp. weight (1) | No controls (2) | Equal-wt. vac. control (3) | Excl. low HHI (4) | Excl. underrep rep. occs (5) | Tradable occs only (6) | Alt. instrument (7) |
|---|---|---|---|---|---|---|---|
| *Panel A: MALMA labor markets* | | | | | | | |
| Log HHI | -0.017 | -0.086*** | -0.108*** | -0.015 | -0.108*** | -0.022 | -0.523*** |
| | (0.013) | (0.016) | (0.028) | (0.015) | (0.028) | (0.034) | (0.170) |
| Observations | 268,192 | 287,335 | 268,192 | 240,345 | 268,085 | 124,846 | 267,003 |
| F-stat | 289 | 100 | 38 | 53 | 38 | 14 | 10 |
| *Panel B: 6-digit occupation labor markets* | | | | | | | |
| Log HHI | -0.114*** | -0.068*** | -0.078*** | -0.050*** | -0.078*** | -0.146 | -0.125*** |
| | (0.018) | (0.017) | (0.029) | (0.019) | (0.030) | (0.115) | (0.047) |
| Observations | 283,750 | 312,995 | 283,750 | 274,841 | 281,593 | 151,707 | 281,567 |
| F-stat | 256 | 67 | 24 | 31 | 22 | 3.7 | 13 |
| *Panel C: 3-digit occupation labor markets* | | | | | | | |
| Log HHI | -0.098*** | -0.113*** | -0.128*** | -0.063** | -0.131*** | 0.452 | -0.140*** |
| | (0.022) | (0.020) | (0.041) | (0.026) | (0.042) | (0.360) | (0.040) |
| Observations | 186,580 | 199,619 | 186,580 | 176,123 | 184,486 | 39,000 | 185,553 |
| F-stat | 136 | 83 | 25 | 36 | 24 | 1.3 | 28 |

*Source:* BGT vacancy data; BGT resume data; BLS OEWS wage and employment data; authors' calculations. *Notes:* This table replicates robustness table A1, but using OEWS data to construct HHIs instead of BGT vacancy data. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors clustered at occupation-metro area level in parentheses. Occupation-year and metro area-year fixed effects. Weighted by average employment of occ-metro area cell unless otherwise noted. Unit of analysis: MALMA occupation X metro area X year, 2011-2019. Each column replicates the main IV specification in Table C3, column 2, but with a change as denoted in the row header. "No emp weight" shows unweighted regressions. "No controls" removes the controls for demand (vacancy growth and the expected HHI instrument). "Equal-wt vac control" adds a control for the (unweighted) average vacancy growth across firms in the labor market. "Excl low HHI" excludes cells with low HHIs. "Excl underrep occs" excludes occupations which are heavily underrepresented in the BGT vacancy data. "Tradable occs only" shows results only for occupations which are tradable (defined in notes to Table 2). "Alt instrument" shows results using an instrument constructed from all firms' hiring growth (rather than only large firms).

# D Appendix: Burning Glass Technologies Vacancy Data

This section contains further information about the vacancy posting data set from Burning Glass Technologies ("BGT“), which we use to construct our employer concentration index (as discussed briefly in Section 2.3). Burning Glass Technologies – now known as Lightcast – is an analytics software company that provides real-time data on job growth, skills in demand, and labor market trends. They frequently collaborate with academic researchers by providing data. The BGT vacancy data on online job postings has been used in several other academic papers, including Azar et al. (2020b) and Hazell and Taska (2019). (We also use a different data set from BGT – the resume data set – to construct our measures of occupational mobility. We discuss this further in Appendix E.)

## D.1 Vacancy posting data overview

Burning Glass Technologies constructs its vacancy database by collecting online job postings from about 40,000 websites, capturing the near-universe of online US job vacancies. They only measure *new* vacancy postings. To capture vacancies which firms keep online to hire workers continually for a given job, BGT consider a vacancy to be "new" if the identical vacancy is still online after 60 days (Carnevale, Jayasundera, and Repnikov, 2014). BGT use proprietary algorithms to de-duplicate vacancies (for example if the same vacancy is posted on different websites). Each vacancy posting contains the job title, company name, location, date, and job description. Using proprietary parsing technology, BGT imputes a SOC 6-digit occupation code. More details on the process by which BGT obtains, parses, and deduplicates this data can be found in Carnevale et al. (2014).

We construct HHIs using BGT's vacancy data for the years 2011–2019. We also use the vacancy data from 2010 to construct our HHI instrument, since we use year-to-year growth rates (so we use 2010-11 data to construct the instrument for 2011, and so on). Over the 2011–2019 period, we have data on 248,751,182 vacancies which have been assigned a SOC 6-digit occupation and metropolitan area by BGT. Of these, a little under one third or 74.1 million have no information about the employer. The remaining vacancies have employer names, with a total of 2,474,182 different employers.

## D.2 Defining the employer and calculating the HHI

A key aspect for our purposes is how an "employer" is defined in the data. BGT's algorithm attempts to group together name variants for employers into a standard set, counting for example "Lowe's" or "Lowes" as the same employer. However, there may be some instances where employers which are in reality the same have not been detected by the algorithm due to large differences in spelling, punctuation, or naming conventions. We therefore carry out an additional layer of grouping by removing punctuation, spacing, and capitalization, and adjusting for common spelling differences or acronyms. We also used the Agency for Healthcare Research Quality's "Compendium of US Health Systems" database for 2016 to link hospitals to the health systems which own them where possible, treating a health system as a single employer rather than a specific hospital.[44] We also manually scanned several thousand of the largest employers in the database to group together different employer names which were evidently part of the same ultimate employer.

This means that we for the most part treat vacancies as offered by the same employer if the *name* listed by the employer on the vacancy is sufficiently similar, or if there is a well-known or easily-identifiable relationship between a parent and subsidiary company with different names (such as "Alphabet" and "Google", or two hospitals which are part of the same health system).

We do not capture relationships where one company owns another company but the names are not similar enough to identify this easily: this means that in some cases we will understate employer concentration by attributing vacancies to different employers. On the other hand, our employer categorization means that individual establishments of an employer – or even franchises of a brand – will be treated as the same employer, which may overstate employer concentration if pay decisions are made at the level of the establishment or franchise rather than the overall firm or brand group. It is not entirely conceptually clear whether employer concentration should be measured at the level of the establishment or the firm. On the one hand, individual establishments often have independent hiring policies; but on the other hand, multi-establishment firms often have common internal pay scales meaning they effectively operate as one employer across establishments. Similarly, it is not entirely conceptually clear whether franchises of the same brand should be considered as separate employers. One the one hand, they are independent businesses; on the other hand, franchisees' human resources policies are often at least partly dictated by the franchisor (Weil, 2014), and there have been a number of prominent cases where franchisors have required franchisees not to 'poach' each others' employees (with Krueger and Ashenfelter (2022) estimating that over half of major franchisors have no-poaching agreements in their franchise contract). We view the question of the appropriate *level* at which to calculate employer concentration – taking into account ownership structures across firms, as well as

[44] This match was not always perfect: there are several cases where we have not necessarily succeeded in matching all hospitals to their owner, because of the presence of multiple hospitals in our database with the same name.

establishment structures within firms – as a fruitful avenue for further research.

How do we treat the one third of vacancies which do not include an employer name? When we calculate our HHI statistics for each occupation-metropolitan area-year cell we assume that each vacancy listing by an employer with no name information in the database is a *separate employer* (as do Azar et al. (2020b)). This will lead us to mechanically underestimate the HHI, as it is likely that at least some of these different vacancy postings where no name information is available come from the same employer in practice.[45]

## D.3 Summary statistics

Here, we provide summary statistics for the roughly 175 million vacancies which contain employer names. As one might expect given the skewed distribution of employment, the large majority of these vacancies are accounted for by a small group of large employers: 2,118 employers each posted more than 10,000 vacancies online over 2011–2019, and these 2,118 employers are responsible for a total of 45.7 million vacancies. On the other hand, the median employer in our dataset posted only 2 vacancies over the entirety of 2011–2019 (Table D1). While many of the small employers in our data are only present in the data for a subset of the 2011–2019 period, many large employers are present for all nine years (as shown in Table D1): as a result more than 50% of all vacancies in our database are listed by employers which are present in all nine years of the sample, and more than 75% are listed by employers which are present in at least eight of the nine years. If employers hire a lot in any one year, they also tend to hire a lot in other years: the correlation of vacancies within a given 6-digit SOC occupation and metropolitan area, by employer, from one year to the next is 0.76.

## D.4 Representativeness

To what extent is the online job *posting* data representative of all job *openings*? The Bureau of Labor Statistics' JOLTS database (Job Openings and Labor Turnover Survey) collects data on job *openings*, where each opening represents a specific position that the firm is actively recruiting to fill. The conceptual difference between a job posting and a job opening is that one job posting (a job advertisement) could be used to fill multiple job openings, if the firm needs to hire several people for a job with the same title, job description, and location at the same time. Carnevale et al. (2014) estimated as of 2014 that between 60 to 70 percent of all job openings could be found in the BGT online vacancy posting data.

[45] Azar et al. (2020b) note that the vacancy postings without employer name information are often due to staffing companies not disclosing on whose behalf they are posting a given job.

Table D1: Summary statistics for BGT vacancy data

| | p5 | p10 | p25 | p50 | p75 | p90 | p95 |
|---|---|---|---|---|---|---|---|
| Total vacancies posted by employer (by employer) | 1 | 1 | 1 | 2 | 7 | 24 | 64 |
| No. of years employer present (by employer) | 1 | 1 | 1 | 1 | 2 | 4 | 6 |
| No. of years employer present (vacancy-weight) | 2 | 4 | 8 | 9 | 9 | 9 | 9 |
| Occ. share relative to BLS OEWS (by occ.) | 0.12 | 0.19 | 0.38 | 0.83 | 1.92 | 4.63 | 7.51 |
| Occ. share relative to BLS OEWS (emp.-weight) | 0.17 | 0.21 | 0.33 | 0.60 | 1.18 | 2.13 | 3.10 |
| Metro area share relative to BLS OEWS (by metro area) | 0.56 | 0.65 | 0.75 | 0.88 | 1.05 | 1.22 | 1.35 |
| Metro area relative to BLS OEWS (emp.-weight) | 0.62 | 0.62 | 0.77 | 0.92 | 1.08 | 1.19 | 1.36 |

Note: This table shows some summary statistics from the BGT vacancy data. 'Total vacancies posted by employer' shows percentiles, across employers, of total vacancies posted by each unique named employer over 2011–2019 (aka, the median employer posted 2 vacancies over 2011–2019). 'No. of years employer present' refers to the number of years in which a given employer posted at least one vacancy, with a maximum of 9 (2011–2019 inclusive). The vacancy-weighted version of this statistic weights each observation by the number of vacancies an employer posted. 'Occ. (or metro area) share relative to BLS OEWS' refers to the share of each SOC 6-digit occupation (/metro area) in our vacancy data, relative to the share of that SOC 6-digit occupation (metro area) in the BLS OEWS data for the entire country (calculated for each year 2011–2019 then averaged across the nine years). The employment-weighted version of this statistic weights each occupation-metro area cell by employment in that cell in 2019.

They do this by comparing the number of new job postings (as measured by BGT) to the number of active job openings as measured by the JOLTS database (inflating the BGT job postings number by the new jobs to active jobs ratio in the Help Wanted Online database to take account of the fact that BGT only captures new postings while JOLTS captures all active job postings). Azar et al. (2020b), using the same methodology, estimate that the share of job openings online as captured by BGT is roughly 85% of total job openings as measured by the JOLTS database in 2016, and the jobs that are not online are usually offered by small businesses and union hiring halls.

The BGT vacancy data has been used in several other academic papers in recent years, which have carried out detailed analyses of its representativeness. We provide a brief summary of the representativeness of the BGT vacancy data here and refer the interested reader to Carnevale et al. (2014), Hershbein and Kahn (2018), and Azar et al. (2020b) for more details. Note in particular that Azar et al. (2020b) use the BGT vacancy data for the same purposes as we do: to calculate local occupation-specific employer concentration HHI indices.

Hershbein and Kahn (2018) compare the distribution of BGT vacancies across major industry groups to the distribution of job vacancies in the Bureau of Labor Statistics' JOLTS database. While BGT is overrepresented in health care and social assistance, finance and insurance, and education, and underrepresented in accommodation and food services, public administration/government, and construction, the differences are mostly small in magnitude.

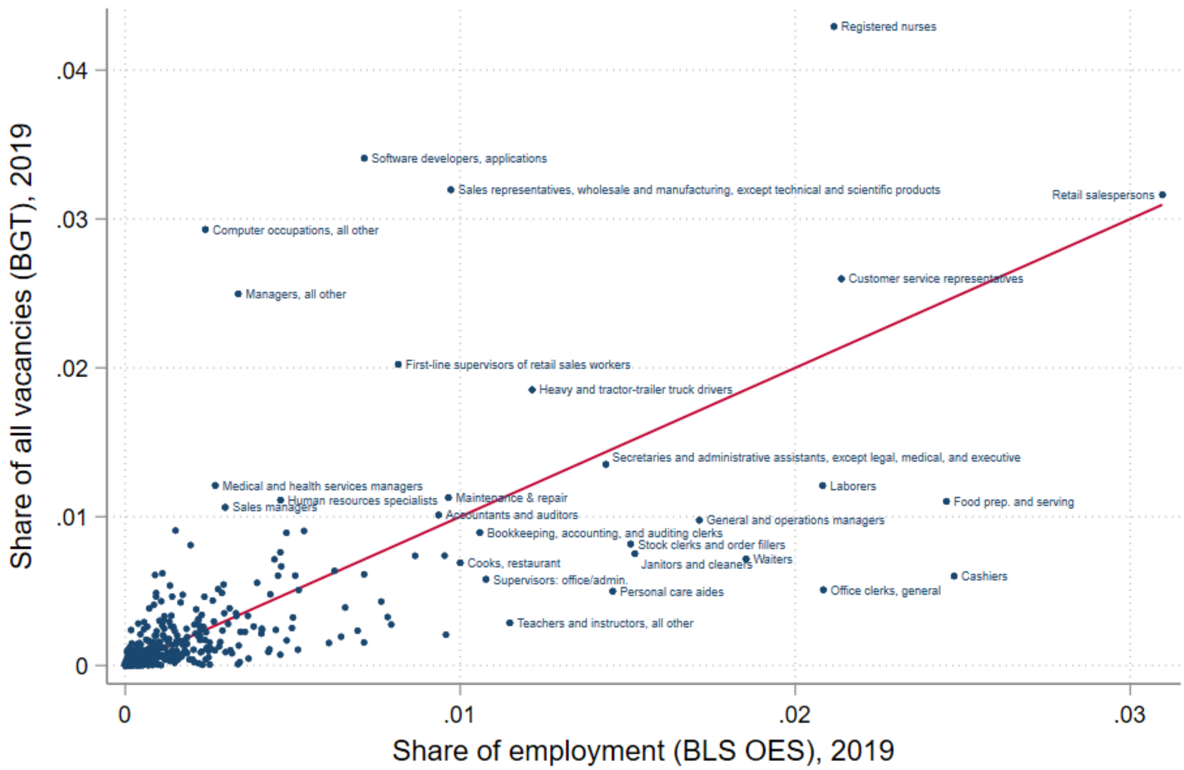
Hershbein and Kahn (2018) also compare the distribution of BGT vacancies by occupation to both the stock and flow of employment in the United States, showing that BGT vacancy data has a much larger than average representation of computer and mathematical occupations, management, healthcare, and business and financial operations, and lower representation in transportation, food preparation and serving, production, and construction. This degree of representativeness does not change much over time in the BGT sample.

To analyze representativeness by occupation systematically, we calculate a measure we call 'represented-ness': the share of all vacancies in our data represented by each SOC 6-digit occupation in a given year, divided by the share of all employment in the BLS occupational employment statistics database which is represented by each SOC 6-digit occupation in that year. Note that our 'represented-ness' measure captures three dimensions: (i) the degree to which the BGT vacancy *posting* data is representative of the totality of vacancy postings in the US; (ii) the degree to which vacancy *postings* are representative of true vacancies (job openings); (iii) is the degree to which individual occupations have high or low turnover (and as a result, a high or low ratio of vacancies to employment). We are interested primarily in the first two of these three, and would ideally compare the representativeness of our BGT vacancy data to a data set of the universe of online *and* offline vacancies by occupation, but this is not available. We show a scatter plot of the share of vacancies each occupation accounts for in our data, relative to the share of employment that occupation accounts for in the BLS OEWS, in Appendix Figure D1.

Of the largest occupations in the data, retail salespersons, customer service representatives, secretaries and executive assistants, and heavy truck drivers are relatively equally represented in BGT data as compared to the BLS OEWS. Registered nurses, software developers and other computer occupations, and sales representatives for wholesale and manufacturing are overrepresented, while laborers, cashiers, waiters, janitors, personal care aides, and food preparation and serving workers are substantially underrepresented in the BGT vacancy data. This pattern of underrepresentedness may not be surprising. These underrepresented occupations are all occupations which tend to have a higher share of their employment accounted for by self-employment, households, or small employers, who may be more likely to advertise through local advertisement channels (posted, for example, on physical job boards, or hired through local agents) or through networks, referrals, or word-of-mouth. In addition, some of these underrepresented occupations may be more likely to have a high ratio of job openings to job postings (a high number of workers hired per job posting).

Similarly, zooming in on the next tier of occupations by size, we see overrepresentation of financial, information, management, and healthcare occupations, relatively even representation of sales, delivery, and mechanical occupations, and underrepresentation of workers

Figure D1: BGT Vacancy Data: representedness of occupations, relative to BLS OEWS, 2019




Note: Comparison of distribution of share of vacancies accounted for by each SOC 6-digit occupation in the BGT vacancy data in 2019, relative to that occupation's share of total employment in 2019 in the BLS occupational employment statistics. Occupations comprising greater than 1% share of either data set are labeled. Red line is the 45 degree line. The vacancy data is discussed in detail in Appendix D.

in occupations with a large share of self-employment (construction, plumbing, landscaping), employment by individual households (maids and housekeeping cleaners, home health aides), or employment where firms may run single job ads for many workers, or which may advertise informally (dishwashers, cooks, food preparation workers, receptionists).

For our purposes, we have two potential representativeness concerns. One concern might be that the representativeness of our data is correlated in some way with factors which would affect both employer concentration and the wage. The underrepresentation of low-wage occupations, for example, might be a concern. This concern is only relevant for the *estimated effect of concentration in our regressions* if our database systematically underrepresents low-wage occupation-metro area labor markets even when controlling for occupation and metro area fixed effects: that is, that within a given occupation, the lower-wage metro areas are underrepresented and within a given metro area, the lower-wage occupations are underrepresented. Moreover, the concern that a single job posting can reflect multiple vacancies may also be a concern here, as a large employer may hire more workers per job posting than a small employer, and so we would systematically underestimate concentration in labor markets with a highly skewed distribution of employer size, relative to labor markets with more symmetric distributions of employer size. For example, when hiring for warehouse laborers, a large warehousing company like Amazon might hire several workers under a job ad for a "Warehouse Associate".[46] On the other hand, for occupations where there is a high degree of granularity of individual job titles and job requirements within an occupation, we may be more likely to observe a one-to-one mapping between job *postings* and job *openings*. One might expect, therefore, that our measures of employer concentration will be less reliable for occupations for which there are many large employers who hire a lot of workers who are not required to be much differentiated in their job tasks, job titles, and qualifications or skills.

For our conclusions in terms of estimating the *aggregate number* of workers who are affected by employer concentration, and creating a ranking of which occupations are more or less affected, underrepresentativeness of the data for either of these reasons is more of a concern: if some occupations are underrepresented in the BGT resume data, they may appear more concentrated when in fact, it is simply the case that online vacancy postings reflect fewer of the true vacancies available in the labor market for that occupation. As

[46] In the extreme case, where each firm only posts one vacancy per occupation that it is hiring for, our measure of the HHI will actually be a measure of 1/N where N is the number of firms hiring for that occupation in that local area. However, in our data there is still substantial variation in the HHI that the inverse number of firms doesn't capture (the R-squared in a cross-sectional regression of log HHI on the inverse number of firms in 2019 is 47%). Nonetheless, the inverse number of firms may be an appropriate measure of employer concentration under some models - rather than the HHI - and indeed in a regression of the log wage on the inverse number of firms in a local occupation (with occupation-year and metro area-year fixed effects) we see a large, negative, and significant coefficient of -0.013 (standard error 0.005).

such, we take care when drawing these conclusions not to isolate specific occupations which appear to be severely underrepresented in our data.

Ideally, to fully address this representativeness concern, we would be able to calculate employer concentration at the level of true job openings/vacancies, rather than vacancy postings. We are not aware of a data set that enables us to observe firm-level local occupational vacancies in the US. We can, however, replicate our main estimates using estimates of employer concentration across *employment* in the BLS OEWS micro data, which we describe in section 4.5 and Appendix section C.

A related concern is that the wages that we use as a dependent variable are measured for the *stock* of employed workers rather than for job vacancies. While wages for new job postings better reflect the impact of changes in employer concentration on marginal new hire wages, firms do not consistently publish wages in job postings for different occupations and often limit the provided information to ranges. Using wages measured for the stock of employed workers means that our estimates represent the effects of changes in employer concentration among new hires on the wages among all workers, which might bias the absolute size of our estimates towards zero if infra-marginal workers do not see their wages adjusted frequently in response to labor market conditions.

# E Appendix: Burning Glass Technologies Resume Data

The Burning Glass Technologies resume data set is a new proprietary data set of 16 million unique US resumes spanning years over 2002–2018. Resumes are sourced by BGT from a variety of BGT partners, including recruitment and staffing agencies, workforce agencies, and job boards. Using the raw resumes, BGT populates a database which contains observations for each individual, denoting their education, jobs, and years in which they worked in each job. BGT's proprietary occupation parser assigns SOC 6-digit occupation codes to each job title listed on each resume. With this data set, we are able to observe 16 million unique workers' job histories and education up until the point where they submit their resume, effectively making it a longitudinal data set (spanning different segments of the 2002–2018 period for different workers). In this paper, we use the resume data to construct occupational transition matrices between SOC 6-digit occupations at a highly granular level. We describe the data set and our methods further below.

## E.1 Construction of occupation transition matrices

Before calculating occupation transition matrices, we apply a number of filters to the raw BGT data:

- Reduce the number of mis-parsed job or resume observations in our data set: eliminate all jobs listed as having lasted more than 70 years, and eliminate any resumes submitted by workers whose imputed age is less than 16 or greater than 100.[47]
- Eliminate all jobs held before 2001.
- Eliminate all resumes with non-US addresses.
- Eliminate any jobs which are listed as having lasted less than 6 months, to ensure that we are only capturing actual jobs rather than short-term internships, workshops etc.

The final number of resumes that contain at least two sequential years of job data under these restrictions is 15.8 million.

From each of these resumes, we extract a separate observation for each job a worker was observed in, in each year they were observed in that job. (We define a 'job' as a unique job title-employer-occupation combination, meaning that a worker can in theory switch job but remain at the same employer and/or in the same occupation.) For each job, we retain information on the SOC 6-digit occupation code. This gives us a data set of 80.2 million worker-job-occupation-year observations, where each worker might be observed in multiple jobs in the same year (either if jobs were held concurrently or the worker switched from one job to another within a given year).

To identify occupational transitions from year to year, we match all sequential pairs of worker-job-occupation-year observations. For instance, we may observe a resume where a person is listed as working as a purchasing manager at Schubert Corp from 2003-2004, as a compliance officer at Stansbury Inc from 2004-2006, and then as a compliance officer at Taska Ltd from 2006-2010: We consider the switch from being a purchasing manager to

Illustrative example of a resume.

| | |
|---|---|
| 2003-2004 | Purchasing Manager, Schubert Corp |
| 2004-2006 | Compliance Officer, Stansbury Inc |
| 2006-2010 | Compliance Officer, Taska Ltd |

being a compliance officier as *both* a change of job and a change of occupation, and the switch from being a compliance officer at Stansbury Inc to being a compliance officer at Taska Ltd as being a change of job but not a change of occupation. For the period 2004-2006, we would therefore record the sequential job holding patterns as follows: This matching of sequential job-year coincidence pairs results in 178.5 million observations (including year-to-year pairs where workers are observed in the same occupation in both years). We use these sequential job-year coincidence pairs to construct our measures of occupational mobility, the transition share $\pi_{o\to p}$, as defined in section 2.2 of the main paper. Specifically, for each pair of

[47] See the next subsection for more details on how we impute ages to the resumes.

Illustrative example of sequential job holding data.

| Year: | 2004 | 2005 | 2006 |
|---|---|---|---|
| *Occ. in year t* | *Occ. in year t+1* | | |
| Purchasing Mgr. (11-3061) | 11-3061<br>13-1040 | | |
| Compliance Off. (13-1040) | | 13-1040 | 13-1040 |

occupations $o$ to $p$, we count the total number of sequential job-year coincidence pairs where the worker is observed in occupation $o$ at any point in year $t$ and is observed in occupation $p$ at any point in year $t+1$. We then divide this by the total number of workers in occupation $o$ in year $t$ who are still observed in the sample in the following year $t+1$.

Since our data is not fully representative on age within occupations, we compute these occupation transition shares separately for different age categories (24 and under, 25 to 34, 35 to 44, 45 to 54, and 55 and over).[48] We then aggregate them, reweighting by the average proportion of employment in each of these age categories in that occupation in the US labor force over 2012–2017 (from the BLS Occupational Employment Statistics). Our aggregate occupational mobility matrix has therefore been reweighted to correspond to the empirical within-occupation age distribution in the labor force, reducing the potential for bias arising from the skewed age distribution of our sample.

## E.2 Summary statistics

Below, we describe the characteristics of the BGT resume data and how it compares to other data sets. All statistics refer to the final set of 15.8 million filtered resumes, or 178.5 million observations of sequential job-year coincidence pairs ('observations') from these resumes, unless otherwise noted.

**Job number and duration:** The median number of jobs on a resume is 4, with the 10th percentile 3 jobs and the 75th percentile 6 jobs. More than 95% of resumes list 10 or fewer jobs (note that a change of job under our definition could include a change of job title or occupation within the same employer). The median length job was 2 years, with the 25th percentile just under 1 year, the 75th percentile 4 years, and the 90th percentile just over 8 years. The median span of years we observe on a resume (from date started first job to date ended last job) is 12 years.

**Gender:** BGT imputes gender to the resumes using a probabilistic algorithm based on the names of those submitting the resumes. Of our observations, 88% are on resumes where

[48] Where we impute age based on the year in which the worker finished either college or high school, as described in the next section.

BGT was able to impute a gender probabilistically. According to this imputation, precisely 50% of our observations are imputed to be more likely to be male, and 50% are imputed to be more likely to be female. This suggests that relative to the employed labor force, women are very slightly over-represented in our data. According to the BLS, 46.9% of employed people were women in 2018.

**Education:** 141.3 million of our observations are on resumes containing some information about education. The breakdown of education in our data for these data points is as follows: the highest educational level is postgraduate for 25%, bachelor's degree for 48%, some college for 19%, high school for 8% and below high school for less than 1%. This substantially overrepresents bachelor's degree-holders and post-college qualifications: only 40% of the labor force in 2017 had a bachelor's degree or higher according to the BLS, compared to 73% in this sample (full comparisons to the labor force are shown in Figure E1). It is, however, to be expected that the sample of the resumes which *provide* educational information are biased towards those with tertiary qualifications, because it is uncommon to put high school on a resume. Imputing high school only education for all resumes which are missing educational information substantially reduces the overrepresentation of those with a BA and higher: by this metric, only 58% of the BGT sample have a bachelor's degree or higher. This remains an overrepresentation, but this is to be expected: a sample drawn from online resume submissions is likely to draw a more highly-educated population than the national labor force average both because many jobs requiring little formal education also do not require online applications, and because we expect online applications to be used more heavily by younger workers, who on average have more formal education. As long as we have enough data to compute mobility patterns for each occupation, and workers of different education levels *within* occupations do not have substantially different mobility patterns, this should not be a reason for too much concern.

**Age:** We impute individuals' birth year from their educational information and from the date they started their first job which was longer than 6 months (to exclude internships and temporary jobs). Specifically, we calculate the imputed birth year as the year when a worker started their first job, minus the number of years the worker's maximum educational qualification requires, minus 6 years. High school is assumed to require 12 years, BA 16 years, etc. For those who do not list any educational qualification on their resume, we impute that they have high school only, i.e. 12 years of education. Since we effectively observe these individuals longitudinally - over the entire period covered in their resume - we impute their age for each year covered in their resume.

As a representativeness check, we compared the imputed age of the people corresponding to our 2002-2018 sample of sequential job observations in the BGT sample to the age distribution of the labor force in 2018, as computed by the BLS. The BGT data of job observations substantially overrepresents workers between 25 and 40 and underrepresents the other groups, particularly workers over 55. 55% of observations in the BGT sample would have been for workers 25-40 in 2017, compared to 33% of the US labor force - see Figure E2 for the full distribution. One would expect a sample drawn from online resume submissions to overweight younger workers for three reasons: (1) because younger workers may be more familiar with and likely to use online application systems, (2) because older workers are less likely to switch jobs than younger workers, and (3) because the method for job search for more experienced (older) workers is more likely to be through direct recruitment or networks rather than online applications. Moreover, by the nature of a longitudinal work history sample, young observations will be overweighted, as older workers will include work experiences when they are young on their resumes, whereas younger workers, of course, will never be able to include work experiences when they are old on their current resumes. Therefore, even if the distribution of resumes was not skewed in its age distribution, the sample of job observations would still skew younger.

As noted above, we directly address this issue by computing occupational mobility only after reweighting observations to adjust the relative prevalence of different ages in our sample relative to the labor force. For instance, this means that we overweight our observations for 45-49 year olds, as this age category is underrepresented in our sample relative to the labor force.

**Occupation:** The BGT automatic resume parser imputes the 6-digit SOC occupation for each job in the dataset, based on the job title. Of 178.5 million useable observations in the data set, 169.6 million could be coded into non-military 6-digit SOC occupations by the BGT parser. 833 of the 840 6-digit SOC occupations are present, some with few observations and some with very many. Ranking occupations by the number of observations, the 10th percentile is 1,226 observations, 25th percentile is 4,173, the median is 20,526, 75th percentile is 117,538, and the 90th percentile is 495,699. We observe 216 occupations with more than 100,000 observations, 83 occupations with more than 500,000 observations, and 19 occupations with more than 2 million observations.[49]

Figure E3 compares the prevalence of occupations at the 2-digit SOC level in our BGT

[49] The occupations with more than 2 million observations are: General and Operations Managers; Sales Managers; Managers, All Other; Human Resources Specialists; Management Analysts; Software Developers, Applications; Computer User Support Specialists; Computer Occupations, All Other; First-Line Supervisors of Retail Sales Workers; Retail Salespersons; Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products; First-Line Supervisors of Office and Administrative Support Workers; Customer Service Representatives; Secretaries and Administrative Assistants, Except Legal, Medical, and Executive; Office Clerks, General; Heavy and Tractor-Trailer Truck Drivers; Financial Managers; Food Service Managers; Medical and Health Services Managers.

data to the share of employment in that occupation group in the labor force according to the BLS in 2017. As the figure shows, at a 2-digit SOC level, management occupations, business and finance, and computer-related occupations are substantially overweight in the BGT data relative to the labor force overall, while manual occupations, healthcare and education are substantially underrepresented.

**Location:** Since not all workers list the location where they work at their current job, we assign workers a location based on the address they list at the top of their resume (if any address is provided). 115.4 million of our observations come from resumes that list an address in the 50 US states or District of Columbia. The broad patterns of the demographic distribution of populations across the US is reflected in our data. By Census region, the Northeast, Midwest, South, and West regions represent 24%, 22%, 38%, and 16% of our BGT sample, while they constitute 18%, 22%, 37%, and 24% of the BLS labor force: that is, our sample is very close to representative for the Midwest and South regions, somewhat overweights the Northeast, and underweights workers from the West region. Zooming in on US states (Figure E4), we see that New Jersey, Maryland and Delaware, for instance, are 1.5-2x as prevalent in our data as they are in the overall US labor force (probably partly because our identification of location is based on residence and the BLS OEWS data is based on workplace), while Nebraska, Montana, South Dakota, Alaska, Idaho and Wyoming are less than half as prevalent in our data as they are in the overall US labor force.

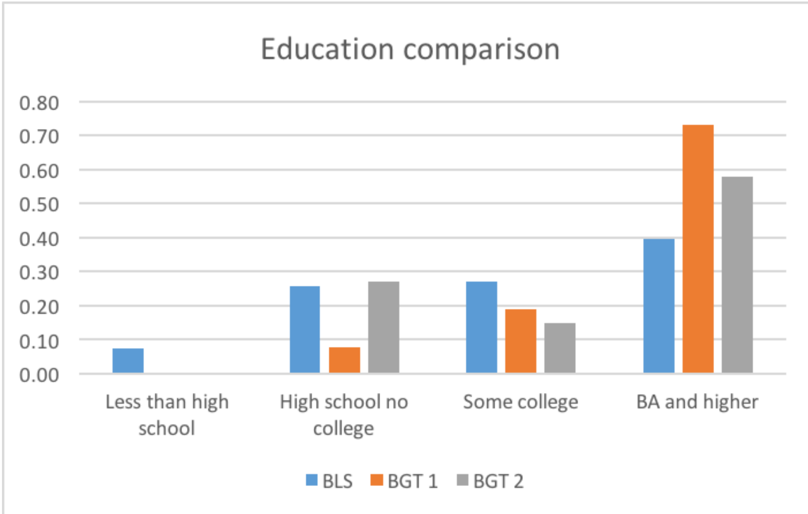
## E.3 Advantages over other datasets

As a large, nationally-representative sample with information about labor market history over the past year, the Current Population Survey is often used to study annual occupational mobility. Kambourov and Manovskii (2013) argue however that the CPS should be used with caution to study occupational mobility. First, the coding is often characterized by substantial measurement error. This is particularly a concern for measuring mobility from one year to the next, as independent coding is often used when there are changes in employers, changes in duties, or proxy responses, and this raises the likelihood of an occupational switch being incorrectly identified when in fact the occupation remained the same.

Due to its structure, the CPS is also only able to identify occupational mobility at an annual or shorter frequency. The PSID is another data source frequently used to study occupational mobility. As a truly longitudinal dataset it is able to capture annual mobility as well as mobility over longer horizons, but its small sample size means that it is unable to provide a more granular picture of mobility between different pairs of occupations.

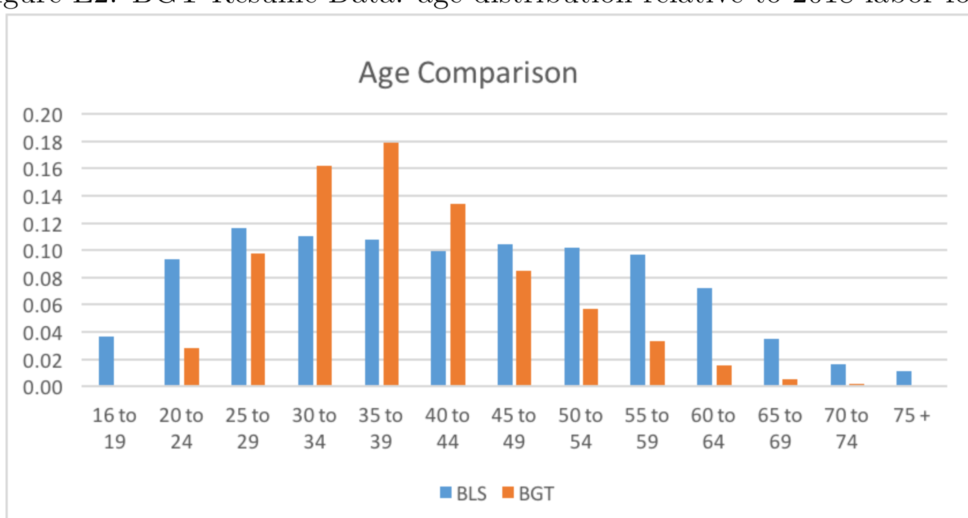
The BGT dataset allows us to circumvent some of these concerns. Its key advantage is its sample size: with 16 million resumes (after our parsing) covering over 80 million job-year

Figure E1: BGT Resume Data: education relative to 2018 labor force



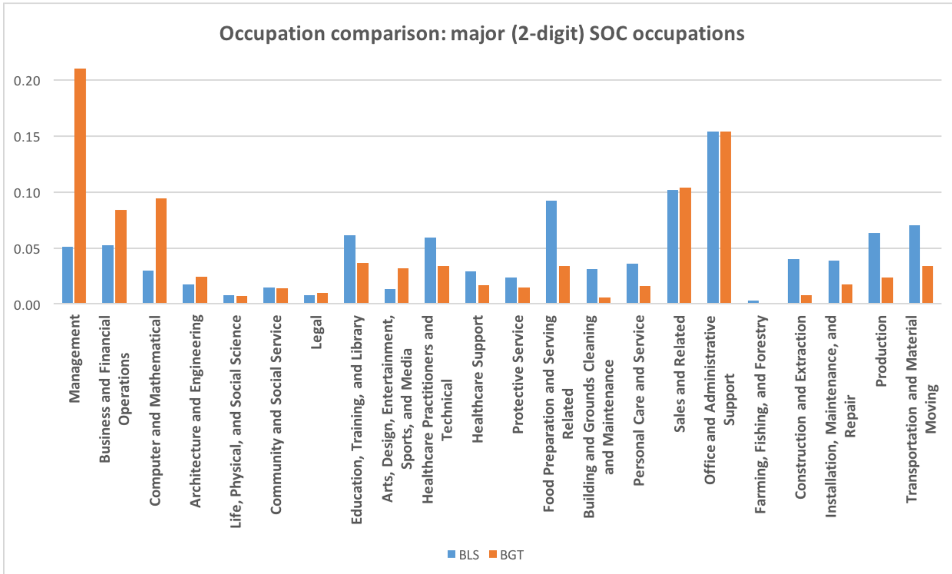

Note: Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT resume data. Two versions are shown: BGT 1 excludes all resumes missing educational information, while BGT 2 assumes all resumes missing educational information have high school education but no college. The resume data is discussed in detail in Appendix E.

Figure E2: BGT Resume Data: age distribution relative to 2018 labor force



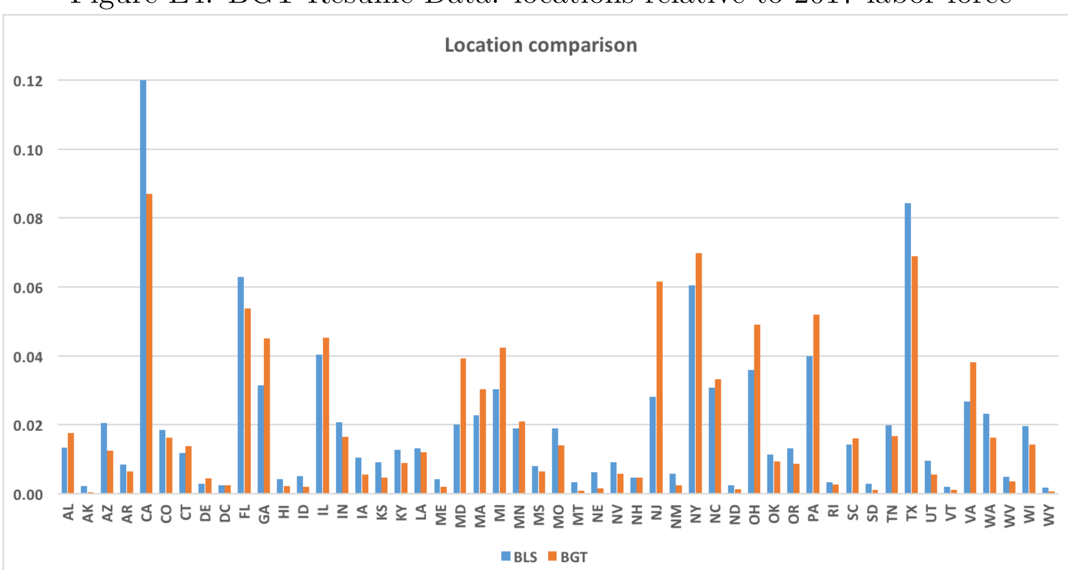

Note: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT resume data. The resume data is discussed in detail in Appendix E.

Figure E3: BGT Resume Data: occupations relative to 2017 labor force




Note: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT resume data. The resume data is discussed in detail in Appendix E.

Figure E4: BGT Resume Data: locations relative to 2017 labor force




Note: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT resume data. Graph shows share of total in each state. The resume data is discussed in detail in Appendix E.

observations, we are able to observe a very large number of job transitions and therefore also to observe a very large number of transitions between different pairs of occupations. Our sample of job-year observations is more than an order of magnitude larger than that which would be available from the CPS when pooling over the same time period we use (2002–2018). And individuals listing their own jobs means that there is less of a risk of independent coding falsely identifying an occupational switch when none occurred.[50]

## E.4 Caveats and concerns

The BGT resume data set does, however, have other features which should be noted as caveats to the analysis.

**1/ Sample selection:** There are three areas of concern over sample selection: first, our data is likely to over-sample people who are more mobile between jobs, as the data is collected only when people apply for jobs; second, our data is likely to over-sample the types of people who are likely to apply for jobs online rather than through other means; and third, our data is likely to over-sample the types of people who apply for the types of jobs which are listed through online applications.

**2/ Individuals choose what to put on their resume:** We only observe whatever individuals have chosen to put on their resume. To the extent that people try to present the best possible picture of their education and employment history, and even sometimes lie, we may not observe certain jobs or education histories, and we may be more likely to observe "good" jobs and education histories than "bad" ones. The implication of this concern for our measure of job opportunities depends on the exact nature of this distortion. If workers generally inflate the level of occupation that they worked at, this would not necessarily distort our estimates of job transitions systematically, unless transition probabilities across occupations vary systematically with the social status / level of otherwise similar jobs. At the same time, if workers choose to highlight the consistency of their experiences by describing their jobs as more similar than they truly were, we may underestimate the ability of workers to transition across occupations. Conversely, if workers exaggerate the breadth of their experience, the occupational range of transitions would be overestimated. In any case, this issue is only likely to be significant if these types of distortions exist for many observed workers, do not cancel out, and differ systematically between workers in different occupations.

We are only aware of a very limited number of studies directly trying to estimate the incidence of misrepresentations on resumes. For instance, Sloane (1991) surveys HR execu-

[50] In addition, the length of many work histories in the data allows for inferring a broader range of latent occupational similarities by seeing the same individual work across different occupations, even when the jobs are decades apart (although we do not take advantage of this feature of the data in this paper).

tives in banking and finds that 51 responding executives were jointly aware of a total of 17 instances of meaningfully falsified job titles, which seems small given the presumably large number of resumes that these executives would have processed during their careers. All but one of the respondents estimated the incidence of falsification of *any* part of the resume to be below 20%, with most opting for lower estimates. Note that this study was done before online search made verification of basic resume information much faster and more affordable. More recently, Nosnik, Friedmann, Nagler, and Dinlenc (2010) found that 7% of the publications listed by a sample of urology residency applicants on their resumes could not be verified. Similarly, Kreisman, Smith, and Arifin (2023) document that some people strategically lie on their resume about their completed education. Cohen, Gurun, and Ozel (2023) document "manager" title inflation as a result of firms trying to avoid paying for overtime. Note, however, that even in these limited cases the joint occurrence of occupations on the same resume still reflects that either workers or firms consider the two occupations to be linked, which is what we are trying to measure.

While such low rates of misrepresentation seem unlikely to introduce systematic bias into our data, it is also important to keep in mind that we are trying to estimate the *plausibility* in a bargaining setting of other jobs constituting relevant outside options. If the skills of a job that they haven't actually held are plausibly consistent with *other* jobs on their resume in the eyes of jobseekers - and ultimately of employers - then this still constitutes evidence that these jobs are perceived as pertaining to the same labor market.

**3/ Parsing error:** Given the size of the dataset, BGT relies on an algorithmic parser to extract data on job titles, firms, occupations, education and time periods in different jobs and in education. Since there are not always standard procedures for listing job titles, education, dates etc. on resumes, some parsing error is likely to exist in the data. (For example, the database states that 25,000 resumes list the end date of the most recent job as 1900. We exclude these from the data, but there may be other parsing errors we are unable to detect).

**4/ Possible duplicates:** The resume data is collected from online job applications. If a worker over the course of her career has submitted multiple online job applications, it is possible that her resume appears twice in the raw database. BGT deduplicates the resume data based on matching name and address on the resume, but it is possible that there are people who have changed address between job applications. In these cases, we may observe the career history of the same person more than once in the data. Preliminary checks suggest that this is unlikely to be a major issue.

## E.5 Comparability with CPS occupational mobility

Our occupational mobility data is constructed from 16 million resumes obtained by BGT. As discussed above these data have the advantage of being very large (an order of magnitude larger than the CPS), but have the potential disadvantages of not being fully representative of the broader workforce, and of not capturing exact year-to-year occupational transitions (because people do not typically leave exact job dates on their resumes). To understand the degree to which our occupational mobility data may be biased, we can compare it to occupational mobility calculated from the CPS.

It is first important to note that our measures of occupational mobility are not strictly comparable to the concept of annual occupational mobility estimated from the CPS by Kambourov and Manovskii (2008) and Xu (2018) for two reasons.

First, the occupation categorization is different: we use SOC 6-digit occupations (of which there are a total of 840 in US data) and the CPS uses Census occupation codes, which are broader. This would push our estimate of outward occupational mobility upward relative to the CPS.

Second, because of the nature of our resume data, we cannot measure annual occupational mobility (share of workers whose main job was in occupation $o$ on date $d$ in year $t$ whose main job was no longer in occupation $o$ on date $d$ in year $t+1$). Instead, our measure of the transition share $\pi_{o\to o}$ is the share of consecutive occupation-year observations in year $t$ in occupation $o$ which are also in occupation $o$ in year $t+1$. Outward occupational mobility is under this definition $1-\pi_{o\to o}$. This would also push our estimate of outward occupational mobility upward relative to the CPS, since it includes any transitions that happen over almost a two-year period (from Jan in year $t$ to Dec in year $t+1$).

With these caveats in mind: our measure of occupational mobility ($1\text{-}\pi_{o\to o}$) is 33% on average in our data. As expected, this is somewhat higher than Kambourov and Manovskii (2008)'s estimate of annual occupational mobility of 20% at the Census 3-digit level in the CPS.

We can also compare how occupational mobility *by occupation* differs between our BGT measure and the CPS. To calculate outward occupational mobility in the CPS, we first map the CPS occupation codes to 6-digit SOC occupation codes using the crosswalk provided by the BLS. We then calculate outward occupational mobility in the CPS ASEC using the self-reported "occupation last year" variable. Specifically, we calculate outward mobility for each occupation $o$ as the number of people who report that their occupation last year was $o$ and their occupation this year was not $o$ (but they were employed), divided by the number of people who report that their occupation last year was $o$ and who are still employed this year.

There is a very strong positive correlation between the measures: the employment-weighted correlation coefficient between BGT and CPS outward mobility across the 722 SOC 6-digit occupations for which we can calculate both measures is 55%. As would be expected (given the discussion above), BGT outward mobility is higher than CPS outward mobility for any given occupation.