Employer Concentration and Outside Options

Gregor Schubert, Anna Stansbury, and Bledi Taska *

August 19, 2022

Abstract

We study the effect of within-occupation employer concentration and outside-occupation job options on wages in the US, identifying outside-occupation options using new occupational mobility data from 16 million resumes. Using shift-share instruments to identify plausibly exogenous local variation, we find that moving from the median to 95th percentile of employer concentration reduces wages by 3.9% on average and by 11.3% for the occupations in the lowest quartile of outward mobility – those for whom the local occupation is a good approximation to their true labor market. We also find meaningful effects of changes in the value of outside-occupation job options on wages: an exogenous 1 percentage point higher wage in outside option occupations leads to a 0.1 percent higher wage in workers' own occupation. Our findings imply that employer concentration affects a non-trivial minority of workers, that policymakers should take the effects of employer concentration seriously for these workers, and that when identifying labor markets where employer concentration may be a concern, measures of employer concentration within an occupation should be considered alongside occupational mobility and the availability of outside-occupation options.

*G. Schubert: UCLA Anderson School of Management, gregor.schubert@anderson.ucla.edu; A. Stansbury: MIT Sloan School of Management, amms@mit.edu; B. Taska: Burning Glass Technologies, btaska@burning-glass.com. Previous versions of this paper have been circulated under the titles "Monopsony and Outside Options", and "Getting Labor Markets Right". 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 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).

1 Introduction

In recent years, concerns about employer concentration have increased. Employer concentration has been posited as a possible explanation for inequality, low pay, and stagnant pay growth. Antitrust authorities have been called on to consider employer concentration in merger and acquisition reviews. Concerns have been raised that employer concentration facilitates restrictions on competition like no-poaching agreements. And, since employer concentration can be a source of monopsony power,[1] concerns around high employer concentration have bolstered calls to raise minimum wages and strengthen collective bargaining.[2]

To assess whether – or in which cases – policy should respond to employer concentration, we need to understand the nature and effects of employer concentration in the US. In this paper, we seek to answer the question: *To what extent does employer concentration matter for US workers' wages, and how does this depend on workers' other outside options?* We estimate the effect of employer concentration on average hourly wages across over 100,000 US SOC 6-digit occupation-by-metropolitan-area labor markets over 2011–2019.[3] We follow Azar, Marinescu, Steinbaum and Taska (2020a) in measuring employer concentration with a Herfindahl-Hirschman Index (HHI) constructed from Burning Glass Technologies' online job postings database. Our empirical strategy addresses two common empirical issues: endogeneity and market definition.

The first empirical issue is endogeneity. While recent research has documented a negative relationship between local employer concentration and wages, the extent to which this is causal – and the magnitude of any such causal effect – is unclear: employer concentration may be correlated with other local economic conditions which also affect wages, complicating the estimation of any underlying wage-concentration relationship.

To respond to this issue, we propose a new identification approach for the effects of employer concentration on wages, drawing on shift-share and granular IV methodology (Borusyak, Hull and Jaravel, 2022; Gabaix and Koijen, 2020). Specifically, we instrument for employer concentration within a particular local occupation with the predicted change

[1] Other possible sources of monopsony power include search frictions, switch costs, and worker and job heterogeneity (Robinson, 1933; Manning, 2003).

[2] Authors making the arguments in this paragraph include, 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).

[3] Where 'SOC' refers to Standard Occupational Classification.

in employer concentration based on firms' national hiring patterns (excluding their hiring decisions in the local area in question). This enables us to construct shocks to local employer concentration that are plausibly orthogonal to local occupation-specific productivity, with the key identifying assumption being that each large firm's decision to increase its hiring nationwide is exogenous with respect to the local economic conditions in the occupation in question.

The second empirical issue is market definition. Assessing the effect of local employer concentration on wages, and pinpointing the workers who are most affected by it, requires a good definition of the relevant local labor market for workers. Using new, highly-granular occupational mobility data constructed from 16 million US workers' resumes (obtained by Burning Glass Technologies),[4] we show that occupational mobility is high and highly heterogeneous across occupations. This suggests that regressing wages on within-occupation employer concentration – as much recent research does – without considering the availability of these outside occupation job options (1) may obscure heterogeneity, as some occupations are a better approximation of workers' true labor market than others, and (2) may lead to biased estimates, as workers who are in high-concentration labor markets (*within* their local occupation) also tend to have poor local job options *outside* their occupation.

To respond to this issue, we introduce two new factors into our baseline regressions of wages on within-occupation employer concentration. First, we allow the estimated coefficient on within-occupation employer concentration to vary by occupations' outward mobility. This enables us to estimate different effects of employer concentration on wages for low-mobility vs. high-mobility occupations (for whom the SOC 6-digit occupation is less likely to be a good approximation to their true labor market). Second, we develop a measure of the value of workers' outside job options in other occupations – an "outside-occupation option index" – and estimate its effect on wages in our baseline regression alongside the effect of within-occupation employer concentration. Our outside-occupation option index is the weighted average of local wages in all occupations except the worker's own, with each weight the product of: (i) occupational mobility flows to each outside occupation and (ii) the local relative employment share in each outside occupation. We use a shift-share IV approach to

[4]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.

identify effects of changes in this outside-occupation option index on wages, instrumenting for local occupational wages with the leave-one-out national mean wage in outside option occupations.

How much does employer concentration matter for wages? Our baseline results suggest that moving from the median to the 95th percentile HHI (as faced by workers) results in 3.9 log points lower wages. Our instrumental variable estimates are about 30% larger than our OLS estimates, suggesting that omitted variables and/or measurement error bias the coefficient towards zero in simple regressions of wages on employer concentration. As expected, we also find that regressions of wages on employer HHI suffer from omitted variable bias if the availability of outside-occupation options is not included in the analysis, with an upward bias in the coefficient size of about one third.

The average effect of employer concentration masks important heterogeneity: within-occupation employer concentration matters *only* for workers who are less able to find comparably good jobs in other occupations. For occupations in the bottom quartile of occupational mobility, like registered nurses and security guards, moving from the median to 95th percentile HHI is associated with on average 11.3 log points lower wages. For occupations in the second quartile, we find that moving from the median to 95th percentile HHI is associated with on average 3.9 log points lower wages. On the other hand, for occupations in the third quartile or in the highest quartile of occupational mobility (like counter attendants or bank tellers), our point estimates are close to zero and not statistically significantly different from zero.

A back-of-the-envelope calculation, using our coefficient estimates, suggests that almost one in six of the 117 million workers covered by our data in 2019 experience wage suppression of 2% or more as a result of employer concentration. Many of the most-affected workers are healthcare workers, reflecting both high healthcare employment concentration and low occupational mobility.

How much do outside-occupation options matter for wages? We find a large, positive, and significant effect of an increase in the value of outside-occupation options: an exogenous 1 percentage point increase in the wage in outside option occupations leads to a roughly 0.1 percent higher wage in workers' own occupation, and for the median occupation, moving from the 25th to the 75th percentile value of our outside-occupation options index across metro areas is associated with 4.4 log points higher wages. These magnitudes are meaningful relative to the degree of geographic wage dispersion across metro areas: for the median occupation, moving from the 25th to the 75th percentile metro area by average wage was associated with a 20 log points higher wage in 2019. These results not only demonstrate that outside-occupation options are a meaningful influence on workers' labor market outcomes, but also that occupational mobility flows can be used to identify these options.

Related literature: Our work relates to four areas of research in labor economics. First, we build on a growing body of work demonstrating an empirical relationship between wages and employer concentration, which began in recent years with Azar et al. (2020a), Azar, Marinescu and Steinbaum (2020b), Benmelech, Bergman and Kim (2022), and Rinz (2022),[5] as well as a growing theoretical literature demonstrating a negative effect of employer concentration on wages (Berger, Herkenhoff and Mongey, 2022; Jarosch, Nimczik and Sorkin, 2019; Azkarate-Askasua and Zerecero, 2020). We make three contributions to this literature: (i) we use a new instrument to estimate plausibly causal negative effects of employer concentration on wages (using different identifying variation than estimates based on M&A activity from Arnold (2020) and Prager and Schmitt (2021)), (ii) we show that simple wage-concentration regressions which do not do consider outside-occupation options are biased, and (iii) we show that negative effects of within-occupation employer concentration on wages are driven entirely by occupations with low outward mobility – i.e. those for which the local occupation is a meaningful approximation to their true local labor market.

Second, in estimating the effect of outside-occupation options on wages, we add to a literature on outside options in the labor market, including Beaudry, Green and Sand (2012), who show local spillovers from changes in industrial employment in the US; Caldwell and Danieli (2018), who find wage effects of workers' outside options in Germany, estimated from the diversity of jobs held by similar workers; Macaluso (2019), who shows that the skill mix of local employment affects laid-off workers' outcomes in the US; and Alfaro-Urena, Manelici and Vasquez (2020), who estimate the outside option value of jobs at multinational corporations in Costa Rica.

Third, in using occupational transitions to identify outside options we build on papers which use worker flows to identify the scope of workers' labor markets (Manning and Petron-

[5]As well as Lipsius (2018), Hershbein, Macaluso and Yeh (2020), Gibbons, Greenman, Norlander and Sørensen (2019), and Qiu and Sojourner (2019) in the US, Abel, Tenreyro and Thwaites (2018) in the UK, Marinescu, Ouss and Pape (2021) in France, Martins (2018) in Portugal, and Dodini, Lovenheim, Salvanes and Willén (2020) in Norway.

golo, 2017; Nimczik, 2018), and to study skill similarity across occupations and industries (Shaw, 1987; Neffke, Otto and Weyh, 2017; Arnold, 2020).

Finally, we contribute to a broader literature on imperfect competition in labor markets, including both the literature on labor market monopsony and the elasticity of the labor supply curve to the firm (e.g. Boal and Ransom, 1997; Manning, 2003; Azar, Berry and Marinescu, 2019b; Berger et al., 2022; Azkarate-Askasua and Zerecero, 2020), and the large search-and-matching literature which features outside options in the worker-firm wage bargain (e.g. Burdett and Mortensen, 1980; Cahuc, Postel-Vinay and Robin, 2006).[6]

2 Empirical Approach

Why might employer concentration matter for wages? The core intuition is that employer concentration can reduce the availability of feasible outside job options for workers. This reduces the strength of workers' relative bargaining position with employers, and can therefore reduce their wage. For example, Jarosch et al. (2019) show that employer concentration reduces wages in a random search model with large ("granular") employers. In their model, workers bargaining with large firms have worse outside options: there are fewer other feasible job opportunities outside the firm they are currently bargaining with, because firms do not compete with their own vacancies. In models with wage posting rather than bargaining, employer concentration can generate upward-sloping labor supply curves to individual firms, leading to wage markdowns (Berger et al., 2022; Azkarate-Askasua and Zerecero, 2020).[7] Moreover, explicit collusion to suppress wages can be facilitated by the presence of a small number of firms.

Several recent papers specifically demonstrate that a Herfindahl-Hirschmann Index ("HHI") across employers is a relevant statistic to measure labor market power arising from employer size. In Jarosch et al. (2019) the effect of concentration on wages is determined by an index which is a function of firm employment shares. To a second-order approximation, this index

[6] Additional work on monopsony includes the empirical estimates of the elasticity of labor supply to the firm in Webber (2015) and Sokolova and Sorensen (2021), and empirical analyses of specific industries, firms, or worker classes in Hirsch and Schumacher (2005), Staiger, Spetz and Phibbs (2010), Ransom and Sims (2010), Ashenfelter, Farber and Ransom (2010), Matsudaira (2014), Naidu, Nyarko and Wang (2016), Bassier, Dube and Naidu (2019), Goolsbee and Syverson (2019), and Dube, Jacobs, Naidu and Suri (2020).

[7] These models have different implications for employment: in Jarosch et al. (2019), there is no effect of employer concentration on employment, while in Berger et al. (2022) and other monopsony or oligopsony models, employment is suppressed.

reduces to an HHI.[8] In Berger et al. (2022), an HHI of the wage bill across employers is a relevant statistic for assessing the welfare effects of firms' labor market power. Arnold (2020) and Naidu and Posner (2021) show that an employer HHI is related to the size of the wage markdown under Cournot competition.

In this paper, we estimate the effect of employer concentration on the wage, where employer concentration is measured by an HHI of employer vacancy shares within a given SOC 6-digit occupation, metropolitan area, and year. A natural starting point is an OLS regression, where we regress the log of the average hourly wage in a SOC 6-digit occupation, metro area, and year ($\bar{w}_{o,k,t}$), on the log of the HHI of employer concentration ($HHI_{o,k,t}$). This, however, does not deal with two empirical issues: market definition and endogeneity.

Market definition. Regressing a wage on an HHI within a specific labor market - whether a local occupation, industry, or cluster of firms - assumes a clearly-delineated labor market, with all workers and jobs within the labor market equally substitutable, and all those outside the labor market irrelevant. This is typically the approach taken in theoretical and empirical work on employer concentration and wages (and the analog of the market definition approach to calculate HHIs in antitrust in product markets).[9] In practice, however, the outside option value of a job is not binary but rather a spectrum, with different jobs differently valuable as outside options. The local SOC 6-digit occupation may be too narrow a measure of the labor market for many workers who are able to switch occupation easily. This creates two issues for empirical estimation. First, heterogeneity: An increase in employer concentration in local occupations where it is easy to switch may not have much effect on the wage, while an increase in employer concentration in local occupations which are good measures of workers' true labor market may have a large effect on the wage. To address this, we allow the coefficient on the within-occupation HHI to vary according to the occupation's degree of outward mobility. Second, there may be bias: to the extent that the quality of outside options outside a worker's occupation are correlated with the quality of their options within the occupation (as proxied by the HHI), regressions of the wage on the

[8]As they note, "in a random search setting the sum of squared market shares captures the ex-ante probability of a worker twice encountering the same firm".

[9]Labor markets have typically been defined as a single occupation or industry within a given local area (commuting zone, metropolitan area, or county), and debate has focused on how narrow an occupational or industrial definition to draw (e.g. Azar et al., 2020a,b). Jarosch et al. (2019), Arnold (2020), and Dodini et al. (2020) define local labor markets more flexibly as clusters of firms inferred using worker flows or common skill requirements (respectively), but still use a binary concept of the labor market.

within-occupation HHI may be biased. To address this, we develop a new index of the value of workers' job options outside their occupation ($oo^{occs}_{o,k,t}$), jointly estimating its effect on the wage alongside the effect of the within-occupation HHI. We explain this outside-occupation option index in section 2.2.

Endogeneity. Employer concentration in a local occupation is in part an outcome of productivity, demand, and supply conditions. This results in an endogeneity concern: employer concentration may be correlated with the wage without having a causal effect on it. The outside-occupation option index suffers from a similar endogeneity issue. We therefore develop instruments for both the HHI and the outside-occupation option index, discussed further in sections 2.4 and 2.5.[10]

Baseline specification. Our baseline empirical specification is as follows:

$$\ln \bar{w}_{o,k,t} = \alpha + \alpha_{o,t} + \alpha_{k,t} + \gamma_1 \ln HHI_{o,k,t} + \gamma_2 \ln oo^{occs}_{o,k,t} + \xi_{o,k,t} \quad (1)$$

where subscripts refers to the metro area (k), occupation (o), and year (t), and $\alpha_{o,t}$ and $\alpha_{k,t}$ are a set of occupation-by-year and metro area-by-year fixed effects. Our wage measure $\bar{w}_{o,k,t}$ is the average hourly wage for all workers in a given occupation, metro area, and year, from BLS OES data. We construct the HHI from Burning Glass Technologies' vacancy posting data (discussed further in section 2.1), and we construct the outside-occupation option index using wage and employment data from BLS OES and using occupational transition shares from Burning Glass Technologies' resume data (discussed further in section 2.2). We allow the coefficients γ_1 and γ_2 on the HHI and outside-occupation option index to vary according to the occupation's degree of outward mobility (which we estimate using the "'leave share"

[10]Critiques of the empirical literature which finds a negative correlation between local employer concentration and wages – including Berry, Gaynor and Scott Morton (2019) and Rose (2019), following older critiques of the structure-conduct-performance paradigm (e.g. Schmalensee, 1989) – focus on one or more of (1) conceptual clarity, (2) market definition, and (3) endogeneity. The conceptual concern is that a regression of market concentration on an outcome like wages is not well-defined, because there is no single theoretical channel by which concentration would affect wages - it depends on the circumstances. Jarosch et al. (2019) and others illustrate that there is a clear conceptual channel by which concentration will always ceteris paribus exert downward pressure on wages, as it reduces the value of workers' outside option set. The market definition concern is that there is no appropriate definition of a market on which a meaningful concentration index can be calculated; we attempt to address this concern as discussed above. The endogeneity concern is that employer concentration is determined by, as well as affecting, local economic conditions. Rose (2019) argues that empirical strategies attempting to identify a causal effect of employer concentration on wages must isolate the effect of employer concentration from changes in labor demand; our identification strategy attempts to do this.

calculated from resume data from Burning Glass Technologies and described in section 2.2). Specifically, we interact $HHI_{o,k,t}$ and $oo^{occs}_{o,k,t}$ with an indicator variable for the applicable quartile of outward mobility of occupation o. We run our regressions across the largest possible subset of U.S. occupation-metro area-year cells for which we can obtain all our key variables: Our full data set for our baseline regressions over 2011–2019 comprises 579,986 occupation-metro area-year observations.[11]

2.1 Measuring employer concentration

To measure employer concentration, we use Burning Glass Technologies' ("BGT") database of online vacancy postings, following Azar et al. (2020a) and Hershbein et al. (2020).[12] We calculate the Herfindahl-Hirschman Index (HHI) of each employer's share of vacancy postings within individual SOC 6-digit occupations and metropolitan areas, in each year 2011–2019:

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

where $v_{i,o,k,t}$ denotes the number of vacancy postings from employer i in occupation o and metropolitan area k in year t. The BGT vacancy posting data covers the near-universe of online job postings, drawn from over 40,000 distinct sources including company websites and online job boards, with no more than 5% of vacancies from any one source (Hazell and Taska, 2019).[13]

[11]This includes 367 metro areas and 714 occupations, with 108,594 occupation-metro area labor markets appearing in at least one year from 2011–2019. Our baseline regressions have 579,668 observations since there are 318 singleton observations when controlling for occupation-year and metro-year fixed effects. We have data on the wage, HHI, and outside-occupation option index (*but not* the instruments) for a larger set of occupation-metro area-year labor markets. We calculate summary statistics and counterfactuals on this larger set. Nonetheless, this larger data set is missing many occupations and metro areas. This is because, of the possible 786,335 occupation-metro area cells, wage data in the BLS OES only exists for approximately 115,000 each year. The missing occupations and metro areas are primarily the smaller ones. Note: To create a consistent panel of occupations over time we crosswalk SOC classifications over time: see Appendix C. Note also: we use "metro areas" to refer to CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas).

[12]Why use vacancies rather than employment data? First, we are not able to obtain firm-level employment data within local occupations. Second, vacancies may be a better reflection of workers' feasible outside options. In equilibrium, vacancy and employment HHIs should be highly correlated. Marinescu et al. (2021) show in France that an HHI of employment flows (reflecting filled vacancies) is highly correlated with an HHI of employment.

[13]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

Since the Burning Glass Technologies vacancy data covers the near-universe of online job postings, it is relatively representative of the vacancies which are advertised online. There are, however, two possible concerns. First, not all vacancies are posted online. Azar et al. (2020a) estimate that in 2016, the BGT vacancy database captured around 85% of all job vacancies both online and offline (as measured from the Help Wanted Online database), but this is likely substantially lower for certain occupations where a large share of jobs are advertised offline or informally.[14] 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 a higher number of people per vacancy posting.[15]

To understand the degree to which each of these might be an issue, 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 (as per BLS OES). By this metric, occupations which are particularly underrepresented include low-wage food service jobs, cleaners, home health aides, laborers, and cashiers. In our estimates of the effect of employer concentration on wages, we carry out a number of sensitivity checks to account for underrepresentation of certain occupations.[16] For further discussion of the BGT vacancy data, see Appendix A.

2.2 Using occupational mobility to identify outside options

To what extent are jobs outside workers' occupations part of their labor market? We use occupational mobility data to answer this question.[17] Since there is no existing US occupational mobility data with high enough granularity to study transitions between SOC 6-digit

---

by which BGT obtains, parses, and deduplicates this data can be found in Carnevale, Jayasundera and Repnikov (2014). To identify jobs at the same employer, we largely group jobs by employer name. We discuss the data and our process for identifying employers in detail in Appendix A.

[14]If the missing vacancies disproportionately come from small firms or households, which seems likely, we will overestimate employer concentration for underrepresented occupations.

[15]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 there are many large employers who hire a lot of workers for undifferentiated job roles.

[16]We also control for occupation-by-year fixed effects, which should assuage concerns about the relative representativeness of the data for different occupations.

[17]We see occupational mobility patterns as a transparent, non-parametric way to capture the value of a different occupation as an outside option, since they capture a combination of both feasibility and desirability. In Appendix D we compare our approach to approaches based on task or skill similarity.

occupations, we construct a new data set of occupational transitions using 16 million unique US resumes, which enable us to observe longitudinal snapshots of workers' job histories over 2002–2018.[18] This resume data was collected by Burning Glass Technologies, who sourced the resumes from a variety of BGT partners, including recruitment and staffing agencies, workforce agencies, and job boards.

We use this data to construct an occupation **'transition share'** $\pi_{o \to p}$, which approximates the probability of a worker moving from occupation o to occupation p conditional on leaving her occupation:[19]

$$\begin{aligned} \pi_{o \to p} &= \frac{\#\text{ in occ } o \text{ in year } t \text{ observed in occ } p \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \text{ observed in any new occ in year } t+1} \\ &\approx Prob(\text{move from occ } o \text{ to occ } p | \text{leave occ } o) \end{aligned} \tag{3}$$

Intuitively, $\pi_{o \to p}$ captures the probability that someone working in an occupation other than their current occupation o in the next year will be working in occupation p in particular, such that p can be considered a likely outside option in the case of a change in occupation. How likely is such a change in occupation? We also construct an **'occupation leave share'**, approximating the share of people who leave their occupation when they leave their job:[20]

$$\begin{aligned} \text{leave share}_o &= \frac{\#\text{ in occ } o \text{ in year } t \text{ \& no longer in occ } o \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \text{ \& in a new job in year } t+1} \\ &\approx Prob(\text{leave occ } o | \text{leave job}) \end{aligned} \tag{4}$$

We calculate these as averages across the whole US and across all years in our data, to capture the underlying degree of occupational similarity rather than transitory fluctuations from year to year.[21]

[18] The CPS has at least an order of magnitude fewer occupational transition observations over the same time period. 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.

[19] Specifically, $\pi_{o \to p}$ is the share of people observed in occupation o at some point in year t who are also observed in occupation p at some point in year $t+1$, as a fraction of all those in occupation o in t who are observed in any new occupation at some point in $t+1$. We exclude jobs lasting 6 months or less. Our measure includes people with jobs in two different occupations at the same time – implicitly assuming that this indicates viability as an outside option.

[20] Specifically, this measure captures the share of people observed in occupation o in year t who are *no longer* observed in occupation o at any point in year $t+1$, as a share of those observed in occupation o in year t who are observed in some new job in year $t+1$.

[21] We estimate transition shares $\pi_{o \to p}$ and leave shares for a large proportion of the possible pairs of SOC 6-digit occupations. We exclude the occupations for which we have fewer than 500 observations in the BGT

The BGT resume data set is largely representative of the U.S. labor force in its distribution by gender and location. However, it over-represents younger workers and white-collar occupations. Since we use this data set to estimate occupational transitions paths from one occupation to another, the over-representation by occupation is not a substantial concern as long as we still have sufficient data for most occupations to have some degree of representativeness *within* each occupation. The over-representation of younger workers, however, might be a concern if younger workers tend to be more mobile or to have different occupational mobility patterns than older workers. We therefore adjust for the over-representation by age by re-weighting our observed occupational transitions to match the distribution of employment by age within each U.S. occupation, provided by the BLS for 2012-2017. We discuss the BGT resume data in more detail in Appendix C.

Using the BGT data, we document five stylized facts about occupational mobility which suggest that jobs outside workers' occupation are an important part of the labor market for many workers, and that occupational transitions can be used to identify these options.

1. Occupational mobility is high, suggesting that the SOC 6-digit occupation fails to capture many workers' true labor markets: the median probability of a worker leaving her 6-digit occupation when she leaves her job - the "occupation leave share" defined above - is 24% (Table 1).
2. Mobility is heterogeneous across occupations, suggesting that the SOC 6-digit occupation is a better approximation of the labor market for some occupations than others: a quarter of occupations have a leave share lower than 19%, and a quarter higher than 28% (Table 1, Figure 2).[22]
3. Aggregating up the SOC classification hierarchy - which groups ostensibly similar occupations - still fails to capture most occupational transitions, suggesting that this cannot solve the market definition problem.[23]

data (roughly the bottom 10% of occupations), resulting in 786 origin SOC 6-digit occupations in our data.

[22]Almost all of the occupations with low leave shares are highly specialized, including various medical, legal and educational occupations (see Appendix Table A2). In contrast, many high leave share occupations require more general skills, including restaurant hosts/hostesses, cashiers, tellers, counter attendants, and food preparation workers.

[23]For the median occupation, 87% of moves to a different 6-digit occupation are also to a different 2-digit occupation, but with substantial variation (see Table 1). For example, only 39% of systems software developers leave their 2-digit occupation group when they move across 6-digit occupations, compared to 95% of flight attendants. Note that management roles are often considered a separate 2-digit occupational group from non-management roles in the same field. Excluding transitions to and from management, at the

4. The occupational transition matrix is sparse, suggesting that workers' relevant labor markets are mostly comprised of only a few occupations, and is highly asymmetric, suggesting that the relevance of occupations as outside options is not symmetric across occupation pairs (unlike in many task- and skill-based measures of occupational similarity).[24]
5. Empirical occupation transitions reflect similarities between occupations in terms of their task requirements, wages, amenities, and leadership responsibilities, suggesting that occupational transitions reflect a combination of many different aspects of the underlying feasibility and desirability of an occupation as an outside option (see Figure A12).[25]

2.3 Measuring outside-occupation options

To measure the value of workers' job options outside their occupation, we propose a new **outside-occupation option index** oo^{occs}. This is a weighted average of the wage in each alternative occupation in a worker's metro area, $\mathrm{w}_{p,k,t}$, weighted by a measure of the relevance of each other occupation in the worker's outside option set, where this relevance is the likelihood that the worker moves to a job in occupation p if she leaves her job in occupation o, $Prob(o \to p)_{o,k,t}$:

$$oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} Prob(o \to p)_{o,k,t} \cdot \mathrm{w}_{p,k,t}.$$

This measure – a probability-weighted average of wages in other local occupations – is an intuitive measure of the expected value of local job options. It corresponds to the idea that

median 67% of SOC 6-digit occupational transitions cross SOC 2-digit boundaries.

[24]See Appendix Figure A6 and Appendix Table A3. The asymmetry partly reflects the fact that workers in an occupation with specialized skills may be able to move to occupations which require generalist skills (e.g. retail salespersons) but the reverse flow is less feasible.

[25]To show this, we regress our measure of occupational transitions on a number of different occupational characteristics derived from the O*Net database: the vector difference in the importance scores for all "Skill" task content items (see Macaluso (2019)); task composites capturing the distinction between cognitive vs. manual, routine vs. non-routine task contents, and social skills, based on Autor, Levy and Murnane (2003) and Deming (2017); characteristics that proxy for flexibility on the job (Goldin, 2014), such as time pressure and the need for establishing and maintaining interpersonal relationships; and characteristics measuring leadership responsibilities. In every pairwise regression of occupational mobility on the absolute difference in characteristics (controlling for the difference in wages), the coefficients are significantly negative or statistically insignificant, as shown in Figure A12. Similarly, Macaluso (2019) finds that mobility between U.S. SOC 2-digit occupations is highly correlated with task similarity. See Appendix E for more details on our analysis.

the outcome of a worker's wage bargain with her firm depends in part on her job options outside her occupation, with the value of each of those outside options a function of the likelihood of it being exercised if she leaves her current job.

The facts we document from the BGT resume data above suggest that we can use occupational transitions to infer workers' outside option occupations. To proxy for $Prob(o \rightarrow p)$ – the likelihood that a worker in occupation o will move to a new job in occupation p if she leaves her job – in our outside-occupation option index, we therefore use the product of two variables: (1) the national average empirical occupation transition share from the BGT resume data, $\pi_{o\rightarrow p}$, and (2) the relative employment share of occupation p in metro area k compared to the national average, $\frac{s_{p,k}}{s_p}$. The national occupation transition share proxies for the likelihood that, nationwide, the average worker's best job option outside her occupation would be in each other occupation p. The local relative employment share adjusts this for the local availability of jobs in each occupation p. Our empirical outside-occupation option index is therefore:

$$oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} \pi_{o\rightarrow p} \cdot \frac{s_{p,k,t}}{s_{p,t}} \cdot \bar{w}_{p,k,t} \tag{5}$$

We construct this outside-occupation option index for each year 1999-2019 for as many SOC 6-digit occupations and metro areas as our data allows, using the BLS Occupational Employment Statistics (OES) to obtain relative employment shares $\frac{s_{p,k,t}}{s_{p,t}}$ and average wages $\bar{w}_{p,k,t}$. We show summary statistics of this index in Table 2.

2.4 Identification: employer concentration

When estimating the effect of local occupational employer concentration on wages, endogeneity issues may bias the estimated coefficients on the HHI. The direction of the bias is ambiguous: an increase in employer concentration could reflect the expansion of a highly productive large firm, which would result in higher employer concentration (expected to reduce wages) but also higher average productivity (expected to increase wages). Or, an increase in employer concentration could reflect a lack of local dynamism, with few new firms, which may lead to higher employer concentration alongside falling productivity.

We therefore instrument for local labor market concentration, creating an instrumental variable which leverages differential local occupation-level exposure to large national firms' hiring growth, in a strategy which builds on both the "granular" instrumental variable approach (GIV) of Gabaix and Koijen (2020) (which uses plausibly exogenous idiosyncratic firm-level variation to instrument for changes in market-level aggregates), and on the shift-share 'Bartik' approach. Our strategy is based on the facts that (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.

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, $g_{j,o,k,t}$ (leaving aside firm entry): $\Delta HHI_{o,k,t} = \sum_j \sigma^2_{j,o,k,t} - \sum_j \sigma^2_{j,o,k,t-1} = \sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+g_{j,o,k,t})^2}{(1+g_{o,k,t})^2} - 1\right)$. The increase in local occupational employer concentration is a function both of initial concentration and of the growth rates of firm-level vacancies $g_{j,o,k,t}$ relative to overall vacancy growth in the labor market $g_{o,k,t}$. We instrument for the vacancy growth for each firm j in occupation o and metro area k with the national vacancy growth of that firm j in occupation o, leaving out the metro area in question k, (which we denote $\tilde{g}_{j,o,t}$).[26] Our instrument for the HHI, $Z^{HHI}_{o,k,t}$, is therefore:

$$Z^{HHI}_{o,k,t} = log\left(\sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1\right)\right) \tag{6}$$

where $\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1}\tilde{g}_{j,o,t}$ 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. The key assumptions for this instrument to be valid are that the firm's national leave-one-out vacancy growth is (i) correlated with its local vacancy growth, but (ii) uncorrelated with the determinants of occupation-specific productivity growth in any given metro area k,

[26] In our baseline specification, we use only positive shocks (i.e. where firm j's national leave-one-out vacancy growth in occupation o was positive from year $t-1$ to year t). In robustness analyses, we present results using all shocks and using only shocks to large firms (with vacancies in occupation o in at least 5 metro areas in year t). Note that by taking the log of the instrument, we exclude observations where the predicted change in HHI based on national firm-level growth is negative. Note also that we are instrumenting for the local level of the HHI with an instrument derived from an expression for the change in the HHI. In a robustness check, we formulate an alternative instrument based on the predicted level rather than change in HHI, $Z^{HHI,alt}_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \log\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2}\right)$. This alternative formulation may also reduce bias introduced from the non-linear log transformation in our baseline instrument (Borusyak and Hull, 2020).

conditional on our occupation-year and metro area-year fixed effects. Through the lens of shift-share instruments (Borusyak et al., 2022), our instrument features plausibly exogenous 'shocks' (a function of firms' national hiring growth), and possibly endogenous exposure 'shares' (the last-period local occupational vacancy shares of each of those firms).[27]

For intuition about the instrument, consider a hypothetical example: assume that in Bloomington, IL, State Farm has a large employment share of insurance sales agents, while in Amarillo, TX, employment is more concentrated in other large insurance companies. In years where State Farm grows substantially faster than other insurance companies nationwide, 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. When looking at examples of large national shocks driving the instrument variation in our data (see Appendix Table A4 for some large examples), it also becomes clear that the underlying large movements in hiring at the national level can to some extent be mapped to broader trends in the labor market, such as the rise in demand for truck drivers, the expansion of coffee chains, and the rise of dollar stores. 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.[28]

One concern with this instrumental variable is that greater local exposure to fast-growing national firms may increase total local labor demand, as well as increasing employer concentration. This would be expected to bias any estimated negative coefficient upward (toward zero). We therefore control for (1) the growth rate of local vacancies in the occupation-

[27]Further, for these changes in local employer concentration driven by large national firms to affect local wages, large national firms must set wages in response to local conditions rather than setting fixed wages nationally. Hazell, Patterson, Sarsons and Taska (2021) show that about 65% of multi-establishment firms set wages locally as opposed to nationally, and that national wage setting is more common for higher wage workers and workers where firms report that they are hiring on a national labor market (workers for whom local employer concentration is likely to be less relevant for wages).

[28]For our first stage to be positive and strong, the occupation vacancy growth rate of firm j in metro areas outside metro area k must be positively correlated with the occupation vacancy growth rate of firm j in metro area k. This is the case in our data. A regression of the latter on the former gives a coefficient of 0.02 (standard error 0.0001); when restricting to firms with vacancies in 5 or more metro areas in the same occupation, the coefficient is 0.35 (standard error 0.0006). Giroud and Mueller (2019) provide evidence consistent with the logic that a shock to a national firm, orthogonal to local economic conditions in a specific metro area, can affect local hiring decisions by that firm. Specifically, they show that even in non-tradable industries, establishment-level employment in a specific local area is sensitive to consumer demand shocks in distant regions in which the establishment's parent firm operates.

metro area labor market ($g_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} g_{j,o,k,t}$), 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: $\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} \tilde{g}_{j,o,t}$ 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.[29]

A second concern is bias due to the fact that our exposure 'shares' do not sum to one as not all firms in a local area operate in multiple labor markets, which means that we can only compute our instrument for a subset of local firms. To address this issue, following Borusyak et al. (2022) we introduce an "exposure control": the sum of the squared local vacancy shares of the large national firms j which feature in the instrument.[30] We further discuss identification conditions in Appendix F.

Third, our instrument is unlikely to be strong for small changes in employer concentration in initially unconcentrated labor markets – if each firm has initially only a trivial share of local employment, even substantial hiring growth will not much change local employer concentration. In our discussion of the implications of our estimates, we therefore apply our estimates of the effect of employer concentration on wages only to local labor markets with above-median employer concentration.

Notwithstanding these caveats, we see our approach as a novel contribution with regard to the problem of estimating the effect of employer concentration on wages. Some recent empirical work instruments for changes in employer concentration in a given local occupation

[29]The effect of a large firm's growth on local labor market concentration is quadratic, whereas the effect of a large firm's growth on local labor demand is linear. 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). While the assumption that these linear terms capture demand effects is relatively strong, note that their inclusion does not affect our baseline coefficient estimates (Appendix Tables A7 and A9). One plausible threat to our identification is that the growth of a large national firm, locally, pushes up wages by more in areas with inelastic labor supply to the local occupation than in areas with elastic labor supply to the local occupation, and that the elasticity of labor supply to the local occupation is correlated with the initial local employment share of the large national firm in that occupation. An alternative plausible threat to our identification is that the firms which expand more nationally tend to be located in weaker labor markets where wage growth is lower. However, the fact that our analysis is robust to the inclusion of metro area-by-year fixed effects in the baseline analyses and occupation-by-metro area fixed effects in further robustness checks indicates that this is likely not the variation driving our results.

[30]That is, in the baseline, the sum of the squared local vacancy shares of any firms j which have vacancy postings in at least one other metro area in the occupation in question in the year in question, and for whom vacancy growth in at least one of these other metro areas was non-zero from one year to the next.

with changes in (the inverse of) the number of employers in the same occupation in other local areas (e.g. Azar et al. (2020a,b); Rinz (2022); Qiu and Sojourner (2019); Marinescu et al. (2021); Gibbons et al. (2019)). This circumvents some endogeneity issues, but a 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.[31] 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, including Arnold (2020) for all industries, and Prager and Schmitt (2021) for hospitals. 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. Our approach allows us to examine the effects of various sources of variation across broad swathes of the US labor market, and to control at least somewhat for effects on local labor demand.[32] 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.

2.5 Identification: outside-occupation options

Endogeneity issues may also bias the coefficients on our outside-occupation option index: a positive local demand shock for an occupation similar to a worker's own may come at the same time as a positive local demand shock for her own occupation (driven, for example, by a common product market shock or a regulatory change). In addition, there is a reverse causality problem: if occupation p and occupation o are good outside options for each other, then a wage increase in o will increase wages in p and vice versa. To identify causal effects, we need exogenous shocks to the wages in workers' outside-occupation options which do not affect, and are not affected by, the local wages in their own occupation.

We use a 'Bartik' shift-share approach, instrumenting for local wages in each outside option

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

[32] Dodini et al. (2020) adopt an additional different strategy, demonstrating that workers laid-off in mass layoffs see larger wage losses in more concentrated labor markets in Norway.

occupation p in metro area k with the leave-one-out national mean wage for occupation p excluding its wage in metro area k ($\bar{w}_{p,\not{k},t}$). We also instrument for the local relative employment share in each occupation using the initial employment share in that occupation in 1999, the first year for which we have data ($\frac{s_{p,k,1999}}{s_{p,1999}}$).[33] Our instrument for the log of the oo^{occs} index is:

$$Z^{oo}_{o,k,t} = log\left(\sum_{p}^{N_{occs}}\left(\pi_{o\to p}\cdot\frac{s_{p,k,1999}}{s_{p,1999}}\cdot\bar{w}_{p,\not{k},t}\right)\right) \tag{7}$$

The identifying variation within a given occupation across different metro areas comes from differences in each metro area's initial occupational employment composition. Identifying variation over time within the same occupation-metro area cell comes from national (leave-one-out) changes over time in wages of local outside-option occupations.[34] For our instrument to be valid, the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p must be positively correlated with the local wage of occupation p in location k, but must not affect the local wage in initial occupation o through a direct channel other than increasing the quality of local outside-occupation options (conditional on controlling for occupation-year and metro area-year fixed effects).[35] We discuss conditions for identification further in Appendix F, following the approach to shift-share IVs of Borusyak et al. (2022).[36]

3 Results

Our results suggest that higher employer concentration causally reduces wages on average, and in particular that this effect is driven entirely by occupations with low outward occupational mobility – occupations for which the local occupation may be a relatively good

[33]Or the first year the occupation-metro area is in the data, if it is not present in 1999.

[34]That is, in a year when there is a national wage shock to one of occupation o's outside option occupations p, metro areas which had a higher proportion of their jobs in occupation p in 1999 should see bigger increases in the wage of occupation o (because they were more exposed to the shock to their outside options). This instrumental variable strategy is closely related to that of Beaudry et al. (2012), who use national industry wage premia to substitute for metro area-level industry wages when estimating spillover effects of cities' industrial composition.

[35]The inclusion of these fixed effects means that differences in metro area-level trends or national productivity of different occupations do not represent an issue for our identification strategy. An additional concern may be that groups of local occupations that share similar labor markets experience similar location-specific industry shocks. We show that our results are robust to controlling for common exposure to industry shocks (see Appendix Table A7).

[36]Appendix Table A16 also contains answers to common questions about our empirical approach and identification strategy.

approximation to the actual labor market that workers face. Our results also suggest that better outside-occupation options increase wages. These results are robust to a range of alternate specifications and control variables. We discuss these findings further below.

3.1 Results: Employer concentration

In our data, there is a robust negative correlation between log vacancy HHIs and log wages at the occupation-metro area level (Figure 3, as found by others including Azar et al. (2020a)). In a regression with occupation-year and metro area-year fixed effects, the OLS relationship is strongly statistically significant, with a coefficient of -0.014 (Table 3, column *a*).[37]

When instrumenting for the HHI, the coefficient magnitude increases by around one third relative to the OLS specification (Table 3, column *c*).[38] This suggests that some combination of omitted variable bias or measurement error biases the coefficient toward zero in simple OLS regressions of wages on HHI. We also introduce a control for our outside-occupation option index (Table 3, columns *b* and *d* for OLS and 2SLS respectively). The coefficient on the instrumented outside-occupation option index is positive and highly statistically significant, confirming that outside-occupation job options matter for wages (which we discuss further in the next section). After introducing the outside-occupation option index the coefficient on the HHI falls by between a quarter and a third in both the OLS and IV regressions, consistent with omitted variable bias. This is because the vacancy HHI is negatively correlated with workers' outside-occupation options: workers with worse options *within* their occupation also have worse options *outside* their occupation (as illustrated in Appendix Figure A13).

How big is the average effect of employer concentration on wages? Our baseline coefficient estimate of -0.015 in Table 3, column *d* – instrumenting for both employer concentration and outside-occupation job options – suggests that going from the HHI faced by the median worker to the HHI faced by the worker at the 95th percentile in 2019 (from an HHI of 137 to 1,882) would be associated with a 3.9 log points lower hourly wage.[39] As noted, we caution

[37]Similarly Hershbein et al. (2020) find a coefficient of -0.014, regressing wages on vacancy concentration in local SOC 6-digit occupations over 2010–2017. Other papers' estimates are not directly comparable because of different labor market definitions or wage measures.

[38]The first stage is shown in Table A5, column (a).

[39]Calculated as $(ln(1882) - ln(137)) \cdot -0.015 = -0.039$. This is at the low end of the range presented in Marinescu and Hovenkamp (2019). Reviewing existing evidence, they suggest that a 10% increase in employer concentration (at the SOC 6-digit occupation by commuting zone level) leads to a 0.3% to 1.3% decrease in wages. Our point estimate in Table 3, column *d* suggests a 10% increase in concentration (at the

against applying these coefficient estimates to labor markets with very low initial levels of employer concentration, since our instrument is weak in these cases.

Heterogeneity by occupational mobility. Re-running our baseline regression, but allowing the coefficients on the HHI and outside-occupation option index to vary for occupations with different degrees of outward occupational mobility, we find that the average effect of within-occupation employer concentration on wages conceals substantial heterogeneity (Table 4, Figure 5). For the quartile of occupations with the lowest outward mobility, as proxied by our occupation "leave share", our coefficient estimate suggests that going from the median to the 95th percentile HHI faced by workers would be associated with 11.3 log points lower wages.[40] Since occupations with low "leave shares" are more likely to be good approximations to workers' 'true' labor market, this suggests that high employer concentration when calculated over workers' true effective labor market can have a very large impact on wages. On the other hand, for the quartile of occupations with the highest outward mobility, the point estimate is very close to and not statistically significantly different from zero, and the confidence interval suggests that an equivalent increase in the HHI in the current occupation would be associated with at most a 3.1 log point lower wage.[41]

Robustness. We explore a number of additional variations on our baseline analyses, illustrated in Figure 6 (with coefficient estimates and standard errors shown in Appendix Tables A7-A10).

First, we run our regressions with different specifications, showing that our coefficient estimates for the effect of employer concentration are similar both on average and by quartile of occupational mobility if we do not weight by employment, if we remove the controls for vacancy growth and exposure, or if we follow Gabaix and Koijen (2020) in adding an additional control for the equal-weighted vacancy growth of local firms ($g^e_{o,k,t} = \frac{1}{N}\sum_j^N g_{j,o,k,t}$)

SOC 6-digit occupation by MSA level) leads to a 0.15% decrease in wages on average.

[40] Calculated as $(ln(1882) - ln(137)) \cdot -0.043 = -0.113$.

[41] Calculated as: $(ln(1882) - ln(137)) \cdot (0.006 - 1.96 \cdot 0.003) = -0.031$. The pattern of our results are consistent with Prager and Schmitt (2021), who find that hospital mergers which induce large increases in concentration reduce nursing and pharmacy workers' wages substantially, somewhat suppress wages of non-medical hospital professionals, and have no detectable effect on wages for the remainder of hospital workers (in maintenance and repairs, operations, housekeeping, catering, and medical records). They interpret these differentials as reflecting the degree to which workers have industry-specific skills. Our estimates would similarly suggest that nursing and pharmacy workers would experience substantially higher wage effects of employer concentration than maintenance, housekeeping, and catering workers, since the former tend to have lower occupational mobility.

to reflect common local occupation-specific shocks. These results are shown in Appendix Tables A7 and A9 , columns (a), (b), and (c) respectively.

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 (less than 50). Moreover, to address concerns about representativeness in our BGT vacancy data, we re-run our baseline regression estimates *excluding* any occupations or metro areas which are substantially underrepresented in the vacancy data.[42] One might also be concerned that our instrument works less well for occupations producing non-tradable goods or services, since local firm-level hiring decisions may be mostly driven by local occupation-specific economic conditions rather than being driven by (orthogonal) decisions made by the firm for its nationwide hiring. We therefore also run a specification with only occupations which produce tradable goods or services – those for which it is not necessary to produce and consume the relevant good or service in the same place.[43] In all these cases, our coefficient estimates are similar to the baseline both for the full sample and by occupational mobility quartile. These result are shown in Appendix Tables A7 and A9 , columns (d), (e), and (f) respectively.

Third, we run our regressions with alternate versions of our HHI instrument. One might be concerned that the exclusion restriction in our instrument is weak if our results are driven by firms which are only present in a few metro areas. For example, if a firm only hires workers in occupation o in two metro areas, $k1$ and $k2$, the firm's hiring decisions in metro area $k1$ may not be orthogonal to the occupation-specific economic conditions in metro area $k2$. To address this, we create a version of our instrument which only uses the national leave-one-out occupational vacancy growth of *large* firms (firms with vacancies in at least five metro areas in an occupation o in year t) to instrument for local firm-level vacancy growth $g_{j,o,k,t}$ in occupation o. Another concern might be that by excluding negative shocks (i.e. where a firm's national occupation-specific hiring contracts from one year to the next), we exclude useful information. We therefore create a version of our instrument which uses all hiring shocks, positive and negative. A further concern might

[42] We exclude any occupations or metro areas in the bottom third of 'represented-ness' in the BGT vacancy data.

[43] We define tradable occupations as all occupations in manufacturing or production, extraction, or farming, 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).

be that the log transformation in our instrument introduces bias, as discussed by Borusyak and Hull (2020). To address this, we create an alternative transformation of our instrument $Z^{HHI,alt}_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \log\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2}\right)$, a transformation which Borusyak and Hull (2020) propose in the case of logged shift-share IVs to reduce bias. The magnitude and significance of coefficients, as well as their pattern across occupational mobility quartiles, remains similar to our baseline with each of these three altered HHI instruments. These results are shown in Appendix Tables A8 and A10, columns (a), (b), and (c) respectively.

In further robustness analyses, we control for an industry Bartik shock to proxy for local metro area occupation exposure to common national industry trends (Appendix Tables A8 and A10 column (d)), and re-run our regressions by occupation mobility quartile using a measure of outward occupational mobility we construct from the CPS rather than from the BGT resume data (Appendix Table A10 column (f)).[44]

Finally, we run our baseline regressions with fixed effects for occupation-metro area and year, rather than occupation-year and metro area-year (Appendix Tables A8 and A10 column (e)). 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. The coefficient estimate for the average effect, at -0.07 is smaller than our baseline estimate of -0.15, and is statistically significant at the 10% level (with a p-value of 0.07). For the estimates by quartile of outward occupational mobility, the coefficient estimate for the lowest-mobility quartile is almost as large as our baseline estimates at -0.035, and is statistically significant at the 5% level (while estimates for other quartiles are noisier and not statistically significantly different from zero or from our baseline coefficient estimates).[45]

Employment. The primary focus of our paper is the effect of employer concentration on wages. We also investigate whether employer concentration affects employment. Different models have different predictions, with monopsony models suggesting higher employer con-

[44]See Appendix F for details of the construction of the local occupation level industry Bartik shocks, and Appendix B for more discussion of the differences between CPS and BGT occupational mobility estimates.

[45]The BLS Occupational Employment Statistics produces estimates on local occupation-specific wages and employment in a given year t using data from six semi-annual surveys conducted in year t, year $t-1$, and year $t-2$. Thus, identifying only off year-to-year variation within an occupation-metro area labor market would be expected to attenuate our coefficients somewhat, since our dependent variable is best thought of as a moving average over time.

centration will reduce both employment and wages as large employers reduce their hiring in response to their wage-setting power, while models arising from bargaining frameworks suggest that higher employer concentration may simply alter the split of surplus between workers and firms without altering employment levels (Berger et al., 2022; Jarosch et al., 2019). We repeat our baseline regressions and our robustness checks with the log of occupation-metro area employment as the dependent variable, with results illustrated in Appendix Figures A17 and A18. In our baseline specifications, we find a negative and significant employment effect, which would tend to support monopsonistic models of employer concentration. However, in our robustness check with occupation-metro area fixed effects, the coefficient estimates become very noisy and not statistically significantly different from zero, suggesting that these results may not be robust to alternative specifications. We therefore tentatively conclude that employer concentration reduces employment alongside wages – but with less certainty than our conclusions on wages.

3.2 Results: outside-occupation options

We can also use our baseline regressions to ask: how big are the effects of changes in the value of outside-occupation options on wages? Finding a positive and significant coefficient on our outside-occupation option index is a joint test of two aspects: first, that job options outside workers' occupation matter for the wages inside their occupation, and second, that occupational mobility flows (from our BGT resume data) can be used to infer what those relevant outside occupations are. Our baseline 2SLS IV coefficient estimate (Table 3 column d) is indeed large, positive, and significant, and suggests that a 10 log point higher outside-occupation option index leads to 1.1 log points higher wages in workers' own occupation. Since the outside-occupation option index approximates a weighted average of wages in other local occupations, this implies roughly that 10 percentage point higher wages in workers' outside option occupations is associated with 1.1 percent higher wages in their own occupation in the same year. Applying the coefficient to understand spatial wage differences, we see that moving from the 25th to the 75th percentile value of outside-occupation options across metro areas for the median occupation leads to 4.4 log points higher wages.[46] This is

[46]We calculate this by estimating the interquartile range of the log outside-occupation option index for each occupation across metro areas in 2019, taking the median across occupations (0.40), and applying our coefficient estimate of 0.11. With the same exercise for employer concentration, we find that for the median

quite large in the context of the geographic variation of wages: for the median occupation, the interquartile range of average wages across metro areas in 2019 was 20 log points.[47]

Finding a large, significant, and positive effect of shocks to outside-occupation options on wages also demonstrates both that workers' true labor markets are broader than their narrow 6-digit SOC occupations, and that our "probabilistic" method of identifying relevant outside options using observed occupation switches can capture workers' true labor markets relatively well.[48]

Robustness. There may be concerns that our coefficients are biased by exposure to correlated industry shocks which affect both a workers' own occupation and her outside option occupations.[49] To control for this possibility we construct a shift-share "industry Bartik" shock that captures the predicted impact of industry level wage trends on local occupation wages and include it in our baseline regressions.[50] Coefficients on the outside-occupation option index remain robust to its inclusion (Appendix Tables A8 and A10, column d). We also find large, statistically significant, and positive effects of outside-occupation options on wages without employment weighting (Appendix Tables A7 and A9, column a), with occupation-metro area and year fixed effects (Appendix Tables A8 and A10, column e), or with an alternate specification of the outside-occupation option instrument which addresses the possibility for bias introduced by the log transformation, following the recommenda-

occupation, moving from the 25th to the 75th percentile HHI across metro areas would result in a wage increase of 2.2 log points.

[47]For a specific example where outside-occupation options might be relevant, consider Baltimore, MD, and Houston, TX. They are a similar size with a similar average hourly wage, but statisticians in Baltimore earned 12 log points more than statisticians in Houston in 2019. Applying our baseline coefficient estimate suggests that around 6.5 log points of this difference – around half – may be attributable to differential availability of outside-occupation job options.

[48]While we do not consider the effects of outside-metro area options on wages in this paper, our methodology could easily be extended to do so. The wage effect of local employer concentration and outside-occupation options is limited by workers' option to move.

[49]For example, if (1) the finance industry and the tech industry are disproportionately likely to employ both accountants and data scientists, (2) San Francisco has a large share of tech employment while New York has a large share of finance employment, and (3) being a data scientist is a good outside option occupation for an accountant, then in years where tech is booming nationwide, this will impact SF more than NY. Accountants in SF will see wages rising by more than accountants in NY but this may be driven simply by the fact that more accountants in SF already work in tech.

[50]The shock is constructed such that the exposure of occupation o in metro area k to each industry ι is defined as the employment share of industry ι in occupation o nationwide, multiplied by the employment share of industry ι in metro area k. relative to the national average. See Appendix F.

tion of Borusyak and Hull (2020) (Appendix Tables A8 and A10, column *c*).[51] In addition, while our HHI data only covers 2011–2019, we can calculate our outside-occupation option index from 1999 onwards. Over this longer period, we find large, positive, and significant effects of outside-occupation options on wages, even with both occupation-by-metro area and occupation-by-year fixed effects (Appendix Table A11). We also find large effects if we calculate the outside-occupation option index using occupational mobility at the SOC 2-digit or 3-digit level instead of 6-digit level (Appendix Table A12, 1999–2016).

4 Discussion and Implications

What might our results suggest about the aggregate effects of 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 HHI labor market in 2019, relative to a scenario where their HHI is reduced to 150 (roughly the median in our data in 2019),[52] as follows:

$$\text{wage effect}_{o,k,t} = (log(HHI)_{o,k,t} - log(150)) \cdot \gamma_1^q \tag{8}$$

where γ_1^q denotes the estimated coefficient on the log(HHI) in our baseline regression specification in Table 4 column *d*, for the appropriate quartile q of outward occupational mobility. 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.

Roughly 56 million of the 117 million workers in our data set were in occupation-metro area labor markets with an HHI greater than 150 in 2019. Of these, our counterfactual wage exercise suggests that roughly 16.5 million workers – 14% of the workers in occupation-

[51]Specifically this instrument takes the log of the wage inside the sum: $Z^{oo,alt}_{o,k,t} = \sum_p^{N_{occs}} \pi_{o \to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \log \bar{w}_{p,\cancel{k},t}$.

[52]An HHI of 150 could represent, for example, a labor market with roughly 67 equal-sized employers, or with two large employers each with 7.5% of workers and an atomistic 'fringe' of firms employing the rest. This level of concentration is not typically thought to be a concern in product markets. Note: There may be monopsony power even in unconcentrated labor markets arising from employer heterogeneity or search frictions (Naidu and Posner, 2021).

metro area labor markets covered by our data – have wages which are at least 2% lower as a result of above-median employer concentration.[53] In Table 5, we show the average estimated wage effect for different combinations of employer concentration and outward occupational mobility. The Table illustrates that the most affected workers include not only those in local labor markets with high employer concentration, but also those with low outward occupational mobility in local labor markets with moderate levels of employer concentration - levels which may not typically be thought high enough to be of concern. Our estimates suggest that employer concentration slightly widens inequality: the share of workers affected by employer concentration is smaller than average in the highest quartile of the wage distribution and in high-wage cities (Appendix Figures A15 and A16).[54]

Which occupations are most affected by employer concentration? In Table 6, we list the twenty-five occupations with the largest number of workers who see an estimated wage effect of 2% or greater in their local occupational labor market (excluding occupations which are substantially under-represented in the BGT vacancy data). A large share of these are healthcare occupations, including 2.3 million registered nurses, licensed practical and vocational nurses, and nursing assistants, and more than 500,000 pharmacists and pharmacy technicians.[55] According to our estimates, large numbers of security guards and hairdressers,

[53]This may be an overestimate since the figure includes some occupations which are underrepresented in the BGT vacancy posting data, meaning that we may overestimate their HHIs. Excluding all occupations with a 'represented-ness' of less than 0.5 in the BGT vacancy data, our data covers 69,698 occupation-metro area labor markets with a total of 68.9 million workers. Of these, we estimate there are 10.3 million workers (15%) whose wages are suppressed by at least 2% as a result of employer concentration. On the other hand, our full data set only covers 117 million of the 151 million nonfarm employees in 2019, and those not represented in our data are disproportionately in non-metropolitan areas or small occupations, where one would expect employer concentration to be higher than average. These workers would therefore be expected to experience greater wage suppression from employer concentration than the average in our data.

[54]Note that our estimates focus only on wages, but employer concentration may also affect non-wage benefits and workplace amenities. Qiu and Sojourner (2019) and Marinescu, Qiu and Sojourner (2020) find negative relationships between employer concentration and the receipt of employment-based health insurance, and labor rights violations respectively. Note also that while our estimates suggest 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 (and the balance likely depends on the nature of product market competition. Kahn and Tracy (2019) 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.

[55]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 et al., 2010).

hairstylists, and cosmetologists are also affected by employer concentration, as large shares of their local labor markets are comprised of employment by a few large companies or chains (although, note that both occupations are somewhat underrepresented in our vacancy data).[56] Importantly, the list of most-affected occupations is very different if occupational mobility is taken into account: simply applying the *average* estimated effect of employer concentration, without accounting for the fact that this effect is very different for high outward mobility and low outward mobility occupations, leads to substantial overestimation of the effect of employer concentration for high-mobility occupations like bank tellers and retail salespersons, and substantial underestimation of the effects of employer concentration for low-mobility occupations like nurses and pharmacy technicians (as illustrated in Appendix Table A15).

Our back-of-the-envelope exercise suggests that while employer concentration suppresses wages for several million workers, the majority of American workers 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 both concentrated labor markets within their occupation and limited opportunities for occupational mobility.

Implications: antitrust. One area where this analysis can be applied is antitrust.[57] Marinescu and Hovenkamp (2019) and Naidu et al. (2018) argue that antitrust authorities should use measures of employer concentration as a preliminary screen for anticompetitive effects of mergers in labor markets (as they already do in product markets). Our analysis suggests that this screen should involve two variables: the HHI in a local 6-digit SOC occupation, and the degree of outward mobility from that occupation.[58]

[56]Note also that we consider employer concentration at the level of a salon chain, many of which are franchised: one might argue that it is better to consider employer concentration at the level of individual franchised salons, though in some cases covenants not to compete within a franchise may make this distinction moot in practice.

[57]Several scholars have called for antitrust authorities to pay attention to employer concentration (Marinescu 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: the topic has featured in FTC and DoJ hearings, the FTC is expanding its retrospective merger review to scrutinize labor market power, and the FTC raised concerns about wage suppression for nurses in a September 2020 public comment on a proposed hospital merger in Hendrick TX.

[58]The screen should also evaluate whether employer concentration in outside-occupation options will be affected by the merger – a concern in occupations whose outside options are predominantly in the same industry, like healthcare. Our proposal differs slightly from Marinescu and Hovenkamp (2019), who argue that antitrust authorities should screen for anti-competitive effects of mergers based only on the HHI in a

However, 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.[59] 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.

Implications: policy to raise wages. In many cases, rather than seeking to reduce employer concentration it may be more appropriate to recognize the fact that employer concentration may give large firms scope to pay a wage which is marked down relative to productivity – and to design labor market policies to counteract this. One such way to do this might be equipping workers with countervailing power by bolstering support for collective bargaining.[60] An alternative might be strengthening minimum wages or benefits standards in local labor markets characterized by high employer concentration.[61]

Implications: promoting mobility. Our results suggest that employer concentration

local SOC 6-digit occupation. Screening based only on local within-occupation HHI without considering outward occupational mobility would likely lead to some mergers being scrutinized which may have little effect on wages, while others which may have anti-competitive effects may go unnoticed.

[59]As 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.

[60]In Jarosch et al. (2019) for example, higher worker bargaining power reduces the weight placed on the outside option in the wage bargain and therefore reduces the importance of employer concentration in wage determination. There is some empirical evidence that labor markets with higher unionization rates see smaller effects of employer concentration on wages. Prager and Schmitt (2021) find larger effects of hospital mergers on nursing wages when nursing unionization rates are lower and in right-to-work states, and Benmelech et al. (2022) find a stronger relationship between employer concentration and wages in U.S. manufacturing firms where unionization rates are lower.

[61]Indeed, higher minimum wages, would be expected to have less of a negative effect on employment in labor markets where employers have monopsony power. Azar, Huet-Vaughn, Marinescu, Taska and Von Wachter (2019a) find that US labor markets with higher employer concentration see smaller employment effects of minimum wage increases.

within a local occupation matters substantially less if workers can find similarly good jobs outside their occupation. By the same logic, this would also be true if workers can easily move geographically. This suggests that policies which make it easier to switch occupation and/or to work in different geographic areas may – by increasing workers' outside options – reduce the degree to which employer concentration can suppress wages.[62] These could include reducing any disproportionate barriers to acquiring training, licensing, or certification in occupations, increasing reciprocal recognition of state-specific licenses and certifications, and increasing affordable housing supply in high-cost cities. [63]

Implications: labor market definition. 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 occupation. 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. 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.[64] While studies of product markets have long recognized the need to define product markets well, the same care is needed when discussing labor markets.

[62]Indeed, the decline in occupational and geographic mobility in the U.S. (Molloy, Smith and Wozniak, 2011; Xu, 2018, documented by), which may partly reflect an increase in the costs of mobility, could be acting to increase the effects of employer concentration.

[63]See Johnson and Kleiner (2020) on the effect of state licensing standards on mobility, Ganong and Shoag (2017) on the effect of housing costs on mobility. In addition, restrictions on worker mobility *within* an occupation (like non-compete clauses as studied by Starr, Prescott and Bishara (2021); Johnson, Lavetti and Lipsitz (2020)) could exacerbate the effects of employer concentration on wages.

[64]For these purposes, we (the authors) make our matrix of occupational mobility, linking SOC 6-digit occupations, freely available for download from our websites, to be used as an input into other studies of occupational labor markets.

5 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 almost 15% of the U.S. private sector workforce experiences 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).[65] 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 concepts of the labor market 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.

[65] 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

Abel, Will, Silvana Tenreyro, and Gregory Thwaites, "Monopsony in the UK," 2018.

Alfaro-Urena, Alonso, Isabela Manelici, and Jose P. Vasquez, "The Effects of Multinationals on Workers: Evidence from Costa Rica," 2020.

Arnold, David, "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes," 2020. Working paper.

Ashenfelter, Orley, Henry Farber, and Michael Ransom, "Modern Models of Monopsony in Labor Markets: A Brief Survey," *IZA Discussion Paper*, 2010, (4915).

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

Azar, José A, Ioana Marinescu, Marshall I Steinbaum, and Bledi Taska, "Concentration in US Labor Markets: Evidence From Online Vacancy Data," *Labour Economics*, 2020, *66*, 101886.

Azar, José, Emiliano Huet-Vaughn, Ioana Marinescu, Bledi Taska, and Till Von Wachter, "Minimum wage employment effects and labor market concentration," Technical Report, National Bureau of Economic Research 2019.

_ , Ioana Marinescu, and Marshall Steinbaum, "Labor market concentration," *Journal of Human Resources*, 2020, pp. 1218–9914R1.

Azar, José, Steven Berry, and Ioana Marinescu, "Estimating Labor Market Power," 2019.

Azkarate-Askasua, Miren and Miguel Zerecero, "The Aggregate Effects of Labor Market Concentration," *Unpublished Working Paper*, 2020.

Bahn, Kate, "Understanding the importance of monopsony power in the U.S. labor market," Technical Report, Washington Center for Equitable Growth 2018.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu, "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies," 2019. Working paper.

Beaudry, Paul, David A Green, and Benjamin Sand, "Does industrial composition matter for wages? A test of search and bargaining theory," *Econometrica*, 2012, *80* (3), 1063–1104.

Benmelech, Efraim, Nittai K Bergman, and Hyunseob Kim, "Strong Employers and Weak Employees How Does Employer Concentration Affect Wages?," *Journal of Human Resources*, 2022, *57* (S), S200–S250.

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

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

Boal, William M and Michael R Ransom, "Monopsony in the labor market," *Journal of Economic Literature*, 1997, *35* (1), 86–112.

Borusyak, Kirill and Peter Hull, "Non-random exposure to exogenous shocks: Theory and applications," Technical Report, National Bureau of Economic Research 2020.

_ , _ , and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Burdett, Kenneth and Dale T. Mortensen, "Search, Layoffs, and Labor Market Equilibrium," *Journal of Political Economy*, 1980, *88* (4), 652–672.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage bargaining with on-the-job search: A structural econometric model," *Econometrica*, 2006, *74* (2), 323–364.

Caldwell, Sydnee and Oren Danieli, "Outside Options in the Labor Market," 2018. Working paper.

Carnevale, Anthony P., Tamara Jayasundera, and Dmitri Repnikov, "Understanding Online Job Ads Data: A Technical Report," 2014.

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

Deming, David J, "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

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

Dodini, Samuel, Michael F. Lovenheim, Kjell G. Salvanes, and Alexander Willén, "Monopsony, Skills, and Labor Market Concentration," *Unpublished Working Paper*, 2020.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri, "Monopsony in Online Labor Markets," *American Economic Review: Insights*, 2020, *2* (1), 33–46.

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

Gabaix, Xavier and Ralph Koijen, "Granular Instrumental Variables," 2020. Working paper.

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90.

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

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

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Goolsbee, Austan and Chad Syverson, "Monopsony Power in Higher Education:

A Tale of Two Tracks," 2019.

Hazell, Jonathon and Bledi Taska, "Downward Rigidity in the Wage for New Hires," 2019.

_ **, Christina Patterson, H Sarsons, and B Taska**, "National wage setting," Technical Report, Working Paper 2021.

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

Hershbein, Brad, Claudia Macaluso, and Chen Yeh, "Monopsony in the US labor market," *Federal Reserve Bank of Richmond Working Paper*, 2020.

Hirsch, Barry T. and Edward J. Schumacher, "Classic or new monopsony? Searching for evidence in nursing labor markets," *Journal of Health Economics*, 2005, *24* (5), 969–989.

Hovenkamp, Herbert, "Whatever Did Happen to the Antitrust Movement," *Notre Dame L. Rev.*, 2018, *94*, 583.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin, "Granular Search, Market Structure, and Wages," 2019. Working paper.

Johnson, Janna E and Morris M Kleiner, "Is Occupational Licensing a Barrier to Interstate Migration?," *American Economic Journal: Economic Policy*, 2020.

Johnson, Matthew S, Kurt Lavetti, and Michael Lipsitz, "The labor market effects of legal restrictions on worker mobility," *Available at SSRN 3455381*, 2020.

Kahn, Matthew E and Joseph Tracy, "Monopsony in spatial equilibrium," Technical Report, National Bureau of Economic Research 2019.

Krueger, Alan B. and Eric A. Posner, "A Proposal for Protecting Low-Income Workers from Monopsony and Collusion," 2018.

Lipsius, Ben, "Labor Market Concentration does not Explain the Falling Labor Share," 2018. Working paper.

Macaluso, Claudia, "Skill remoteness and post-layoff labor market outcomes," 2019.

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

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

Marinescu, Ioana and Herbert Hovenkamp, "Anticompetitive Mergers in Labor Markets," *Indiana Law Journal*, 2019.

Marinescu, Ioana Elena and Eric A. Posner, "Why Has Antitrust Law Failed Workers?," *Cornell Law Review*, 2020.

_ **, Ivan Ouss, and Louis-Daniel Pape**, "Wages, Hires, and Labor Market Concentration," *Journal of Economic Behavior and Organization*, 2021.

_ **, Yue Qiu, and Aaron Sojourner**, "Wage Inequality and Labor Rights Violations," 2020.

Martins, Pedro S., "Making their own weather? Estimating employer labour-market power and its wage effects," 2018.

Matsudaira, Jordan D., "Monopsony in the Low-Wage Labor Market? Evidence from Minimum Nurse Staffing Regulations," *The Review of Economics and Statistics*, 2014, *96* (1), 92–102.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak, "Internal migration in the United States," *Journal of Economic Perspectives*, 2011, *25* (3), 173–196.

Naidu, Suresh and Eric A. Posner, "Labor Monopsony and the Limits of the Law," *Journal of Human Resources*, 2021.

_ , _ , and E. Glen Weyl, "Antitrust Remedies for Labor Market Power," *Harvard Law Review*, 2018.

_ , Yaw Nyarko, and Shing-Yi Wang, "Monopsony power in migrant labor markets: evidence from the United Arab Emirates," *Journal of Political Economy*, 2016, *124* (6), 1735–1792.

Neffke, Frank, Anne Otto, and Antje Weyh, "Inter-industry labor flows," *Journal of Economic Behavior & Organization*, 2017, *142* (C), 275–292.

Nimczik, Jan Sebastian, "Job Mobility Networks and Endogenous Labor Markets," 2018.

Phillips, Noah Joshua, "Prepared Statement before the U.S. House of Representatives on Antitrust and Economic Opportunity: Competition in Labor Markets," 2019.

Prager, Elena and Matthew Schmitt, "Employer Consolidation and Wages: Evidence from Hospitals," *American Economic Review*, 2021.

Qiu, Yue and Aaron Sojourner, "Labor-Market Concentration and Labor Compensation," 2019.

Ransom, Michael R. and David P. Sims, "Estimating the Firm's Labor Supply Curve in a "New Monopsony" Framework: Schoolteachers in Missouri," *Journal of Labor Economics*, 2010, *82* (2), 331–355.

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

Robinson, Joan, *The Economics of Imperfect Competition*, Springer, 1933.

Rose, Nancy, "Thinking Through Anticompetitive Effects of Mergers on Workers," Technical Report, American Bar Association 2019 Antitrust Spring Meeting 2019.

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

Shambaugh, Jay, Ryan Nunn, Audrey Breitwieser, and Patrick Liu, "The State of Competition and Dynamism: Facts about Concentration, Start-Ups, and Related Policies," 2018.

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

Sokolova, Anna and Todd Sorensen, "Monopsony in labor markets: A meta-analysis," *ILR Review*, 2021, *74* (1), 27–55.

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

Stansbury, Anna 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.

Starr, Evan, James J Prescott, and Norman Bishara, "Noncompetes in the US labor force," *Journal of Law and Economics*, 2021.

Steinbaum, Marshall and Maurice E Stucke, "The Effective Competition Standard: A New Standard for Antitrust," *University of Chicago Law Review*, 2020.

Webber, Douglas, "Firm market power and the earnings distribution," *Labour Economics*, 2015.

Xu, Ming, "Understanding the Decline in Occupational Mobility," *Working Paper*, 2018.

Figures and Tables

Table 1: Summary statistics: BGT occupational mobility data

Percentile (occ.)	1	5	10	25	50	75	90	95	99
Panel A: Number of obs. in the BGT occ. mobility data in '000s, by occ. (2002-2015)									
Observations	0.6	1.1	1.6	4.9	20.8	112.3	466.8	853.9	3,471.9
Panel B: Share leaving job and occupation, by occ. (2002-2015)									
Share in diff. job	0.30	0.35	0.37	0.40	0.45	0.52	0.61	0.66	0.74
Share leaving 6d. occ.	0.047	0.062	0.074	0.90	0.10	0.12	0.14	0.18	0.29
Leave share	0.09	0.11	0.14	0.19	0.24	0.28	0.33	0.38	0.69
Panel C: Share of occupational transitions which cross SOC 2d boundary (2002-2015)									
All occ. transitions	0.55	0.65	0.70	0.79	0.87	0.93	0.97	0.98	1.00
Excl. management	0.40	0.48	0.51	0.59	0.67	0.75	0.80	0.83	0.87

Notes: We exclude occupations with <500 observations in the BGT resume data. In Panel A, an observation is a person-year unit that is also observed in the data the following year. Panel B shows the share of workers observed in a new job or new occupation from one year to the next, and the "leave share", defined in section 2.2 as the share leaving their occupation conditional on leaving their job. Panel C shows the share – by origin occupation – of all SOC 6-digit occupational transitions which also span SOC 2-digit boundaries. The percentiles refer to percentiles across occupations, such that (for example) the median occupation in our data has 20,800 observations (Panel A).

Table 2: Summary statistics: main data set

Percentile (occ.-metro area)	1	5	10	25	50	75	90	95	99
Panel A: Employer concentration HHI (2019)									
HHI	26	80	140	349	905	2,200	5,000	7,813	10,000
HHI, emp-wt	6	14	21	53	137	408	1,049	1,882	5,047
Panel B: Outside-occupation option index oo^{occs} (2019)									
oo^{occs}	9.4	12.0	13.6	17.0	21.7	28.1	36.0	42.6	62.0
$\frac{oo^{occs}}{wage}$	0.24	0.38	0.48	0.69	0.98	1.30	1.63	1.85	2.39
$\frac{oo^{occs}}{wage}$, emp-wt	0.39	0.59	0.73	1.00	1.35	1.70	2.02	2.25	2.92
Panel C: Occupation-metro area wages and employment (2019)									
Employment	30	40	50	90	220	660	1,970	3,930	14,860
Mean hourly wage	9.73	11.65	13.23	16.77	22.9	32.48	45.51	55.14	88.07
Wage, emp-wt	9.54	10.86	12.23	14.76	19.87	32.17	47.92	59.42	84
Panel D: National hourly wage distribution (2019) from BLS OES									
Hourly wage	–	–	10.35	13.02	19.14	30.88	48.57	–	–

Notes: Panels A, B, and C show summary statistics for our main data set in 2019, calculated over all occupation-metro area-year cells for which we have wage data (from BLS OES), a vacancy HHI (from BGT vacancy data), and an outside-occupation option index (constructed from BLS OES and BGT resume data). This comprised 109,582 occupation-by-metro area labor markets, which collectively contained 117,286,314 workers in 2019 according to BLS OES. Rows labeled "emp-wt" show the percentiles weighted by occupation-metro area employment. Panel D shows the national 10th, 25th, 50th, 75th, and 90th percentile of the hourly wage distribution according to the full BLS OES data set, for comparison.

Table 3: Regression of wage on HHI and oo^{occs}, full sample

Dependent variable:	Log wage			
	(a) OLS	(b) OLS	(c) 2SLS IV	(d) 2SLS IV
Log HHI	-0.014***	-0.010***	-0.020***	-0.015***
	(0.002)	(0.001)	(0.004)	(0.004)
Log outside-occ. options		0.136***		0.109***
		(0.012)		(0.023)
Vacancy growth			-0.126	-0.105
			(0.082)	(0.069)
Predicted vacancy growth			-0.002	-0.004
			(0.008)	(0.006)
Exposure control			0.029	0.014
			(0.021)	(0.021)
Observations	579,668	579,668	579,668	579,668
F-Stat			252	107

Notes: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC occupation by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions feature occupation-by-year and metro area-by-year fixed effects. Regressions are employment-weighted by average employment in the occ-metro area over the 2011–2019 period. Columns (a) and (b) show OLS regressions and columns (c) and (d) show 2SLS IV regressions (where both the log HHI and log outside-occ. option index are instrumented). The reported F-stat for the 2SLS IV regressions is the Kleibergen-Paap Wald F statistic. The vacancy growth and predicted vacancy growth variables are *rescaled* by dividing by 10 (so that 1% vacancy growth in a local area corresponds to a value of 0.001, rather than 0.01), such that the coefficient estimates can be seen in the table for most specifications. See text for detailed explanation of instruments and controls.

Table 4: Regression of wage on HHI and oo^{occs}, by quartile of occupation leave share

Dependent variable:	Log wage			
	(a) OLS	(b) OLS	(c) 2SLS IV	(d) 2SLS IV
Log HHI	-0.025***	-0.021***	-0.042***	-0.043***
X Q1 occ mobility	(0.004)	(0.005)	(0.006)	(0.008)
Log HHI	-0.019***	-0.010***	-0.027***	-0.015***
X Q2 occ mobility	(0.002)	(0.002)	(0.005)	(0.004)
Log HHI	-0.008***	-0.006**	-0.010*	-0.003
X Q3 occ mobility	(0.003)	(0.002)	(0.006)	(0.007)
Log HHI	-0.001	-0.001	-0.003	-0.001
X Q4 occ mobility	(0.002)	(0.002)	(0.004)	(0.006)
Log outside-occ options		0.128***		0.076***
X Q1 occ mobility		(0.014)		(0.024)
Log outside-occ options		0.150***		0.130***
X Q2 occ mobility		(0.015)		(0.020)
Log outside-occ options		0.115***		0.111***
X Q3 occ mobility		(0.015)		(0.028)
Log outside-occ options		0.121***		0.098**
X Q4 occ mobility		(0.019)		(0.043)
Vacancy growth			-0.129	-0.106
			(0.084)	(0.069)
Predicted vacancy growth			-0.003	-0.004
			(0.007)	(0.006)
Exposure control			0.040**	0.026
			(0.019)	(0.019)
Observations	579,668	579,668	579,668	579,668
F-stat			67	30

Notes: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC occupation by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions feature occupation-by-year and metro area-by-year fixed effects. Regressions are employment-weighted by average employment in the occ-metro area over the 2011–2019 period. Columns (a) and (b) show OLS regressions and columns (c) and (d) show 2SLS IV regressions. The reported F-stat for the 2SLS IV regressions is the Kleibergen-Paap Wald F statistic. The vacancy growth and predicted vacancy growth variables are *rescaled* by dividing by 10 (so that 1% vacancy growth in a local area corresponds to a value of 0.001, rather than 0.01), such that the coefficient estimates can be seen in the table for most specifications. Independent variables labelled "*X* Qi outward mobility" show the coefficient on an interaction term between the HHI or outside-occupation option index (respectively) with an indicator variable which takes the value 1 if the occupation in question is in the ith quartile of outward occupational mobility (where "Q1" represents the least outwardly mobile occupations, and so on). See text for detailed explanation of variables.

Table 5: Counterfactual wage effects of setting HHI to 150 (& number affected)

		0< HHI <150	150< HHI <500	500< HHI <1,500	1,500< HHI <2,500	2,500< HHI <10,000
Lowest mobility	Avg. wage effect	0	2.6%	7.5%	11.5%	15.4%
	Employment (m)	9.9	7.7	5.2	1	.98
Q2 mobility	Avg. wage effect	0	0.8%	2.5%	3.9%	5.2%
	Employment (m)	19	6.9	3.4	.84	.87
Q3 mobility	Avg. wage effect	0	0	0	0	0
	Employment (m)	21	8	3.3	.65	.68
Q4 mobility	Avg. wage effect	0	0	0	0	0
	Employment (m)	12	8.1	5	1.3	1.5

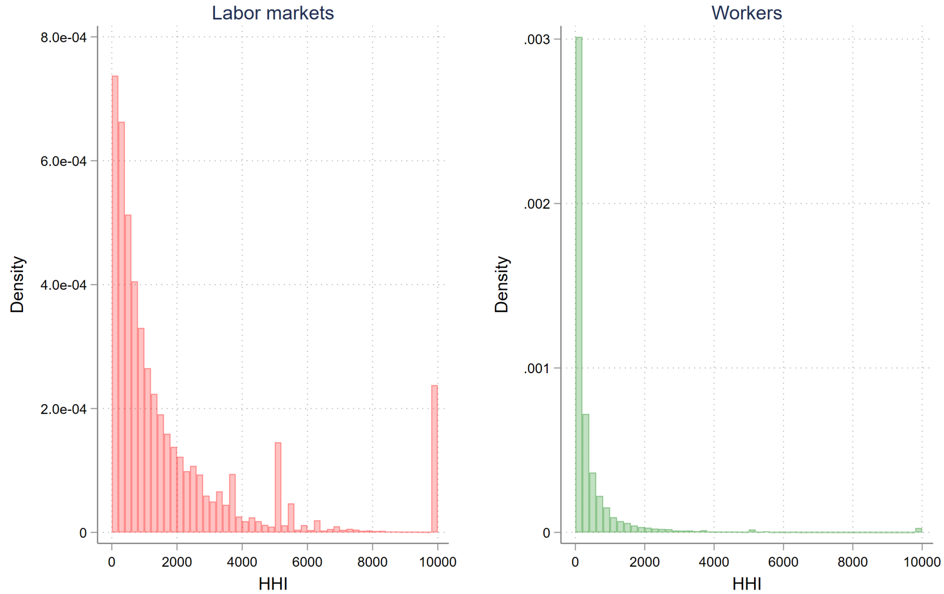
Notes: This table shows the estimated wage impact, and number of people affected, of lowering the HHI to 150 (roughly the median as experienced by workers in 2019) in all occupation-metro area cells where it was greater than 150 in 2019. The estimated wage impact is calculated as the difference between the *actual* log HHI and the log of 150, multiplied by the estimated coefficient in our wage-HHI regressions (with the coefficient used corresponding to the appropriate quartile of occupational outward mobility, as estimated in Table 4 column (d), and using only significant coefficients). The impact number in each cell in the table is the average impact across all workers in that cell: so, for example, for the 0.98 million workers in our data who are in occupations in the lowest quartile of outward mobility (Q1), and who are in occupation-metro area labor markets with an HHI greater than 2500, the *average* estimated impact of employer concentration on their wage is 15.4%. Note (1) this exercise implicitly holds productivity constant, and (2) our data set covers around 117 million workers in total, from the BLS OES occupation-by-metropolitan area employment and wage data.

Table 6: Twenty-five occupations with most people affected by employer concentration (based on a predicted occupation-metro area wage effect of 2% or greater)

Occupation	National employment	Share of occupation w/ estimated effect 2% or greater	Number in occupation w/ estimated effect 2% or greater
Registered nurses	2,982,280	.46	1,373,090
Security guards	1,126,370	.92	1,034,590
Nursing assistants	1,419,920	.53	751,870
Hairdressers, hairstylists, and cosmetologists	385,960	.92	353,710
Pharmacy technicians	417,780	.83	346,480
Pharmacists	311,200	.8	250,130
Licensed practical and licensed vocational nurses	697,510	.31	219,010
Medical assistants	712,430	.31	217,870
Fitness trainers and aerobics instructors	325,500	.58	187,920
Heavy and tractor-trailer truck drivers	1,856,130	.09	167,280
Emergency medical technicians and paramedics	521,200	.31	162,720
Radiologic technologists	207,360	.74	154,050
Medical and clinical laboratory technologists	332,541.996	.38	125,062
Lawyers	657,170	.17	113,650
Phlebotomists	128,290	.86	110,270
Nurse practitioners	200,600	.52	103,770
Aircraft mechanics and service technicians	133,310	.76	100,690
Management analysts	709,750	.14	96,920
Software developers, applications	927,925.088	.099	92,181
Massage therapists	107,240	.86	91,770
Maids and housekeeping cleaners	926,960	.099	91,580
Light truck or delivery services drivers	923,050	.096	89,050
Respiratory therapists	132,090	.66	86,940
General and operations managers	2,400,280	.035	83,750
Secretaries and administrative assistants (except legal, medical, and executive)	2,038,340	.04	81,060

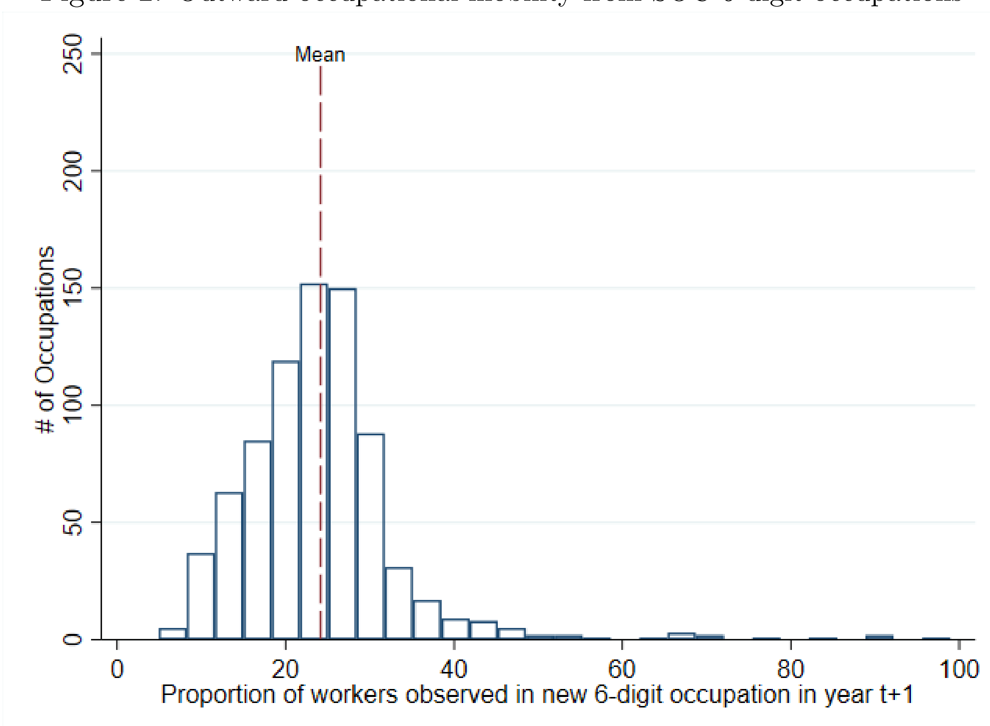
Notes: This table lists the twenty-five occupations with the largest number of workers who experience an estimated wage impact of 2% or more as a result of employer concentration in 2019 (see Table 5 for description of how this effect is calculated). The columns list, respectively, each occupation's total employment in the BLS OES data in 2019, the share of those workers who are in metro areas with an estimated wage impact of 2% or more, and the number of workers in metro areas with an estimated wage impact of 2% or more (the product of the first two columns). In Appendix Table A14 we list the degree of represented-ness of each of these occupations in the BGT vacancy data. We exclude occupations from this list that are very under-represented in the BGT vacancy data relative to overall employment (with a cutoff with represented-ness<0.5, or around the 33rd percentile). The excluded occupations are a mix of primarily public or quasi-public sector occupations (Bus drivers, Teachers and instructors (all other), Police and sheriff's patrol officers, Firefighters, Postal service mail carriers, First-line supervisors of police and detectives, Court, municipal, and license clerks, Librarians), lower-wage occupations for which we have low representativeness in the data (Personal care aides, Farmworkers and laborers (crop, nursery, and greenhouse), Dental hygienists, Waiters and waitresses, Bartenders, Home health aides, First-line supervisors of personal service workers, Dental assistants, Lifeguards and ski patrol, Ushers, lobby attendants, and ticket takers, Janitors and cleaners, except maids and housekeeping cleaners, Manicurists and pedicurists, Landscaping and groundskeeping workers), occupations with the "all other" classification, for which vacancies may be hard to parse accurately (Information and record clerks (all other)), and finally Operating engineers and other construction equipment operators.

Figure 1: Histogram of employer HHI across occ-metro area labor markets and across workers, 2019



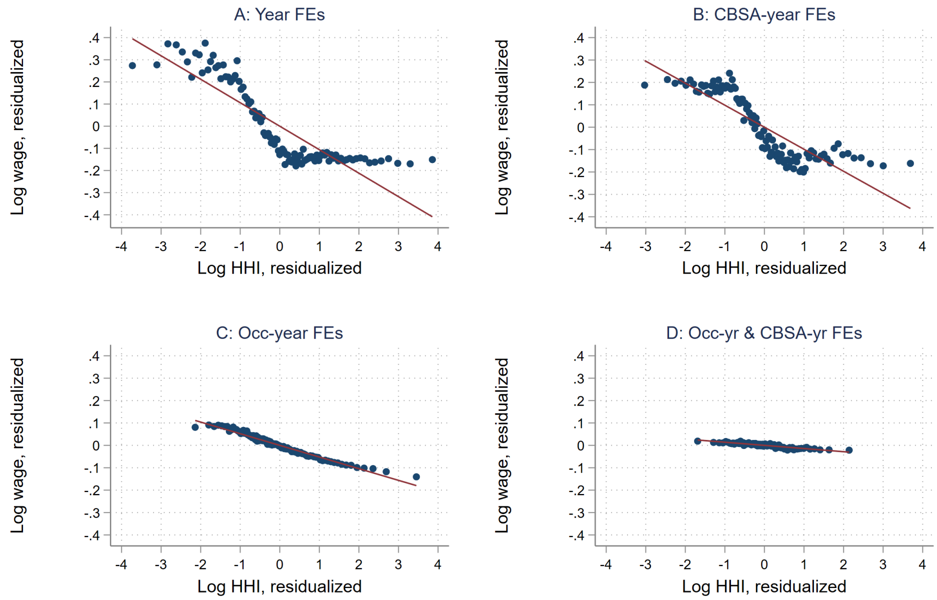
Note: HHI is measured using Burning Glass Technologies vacancy data, at the level of a SOC 6-digit occupation by metro area labor market. Our data covers occupation-metro area labor markets which include 117m of the 151m workers in the U.S. labor market in 2019. Left panel shows the distribution of HHIs across occ-metro area labor markets in 2019. Right panel shows the distribution of HHIs across workers in 2019 (i.e. the distribution of HHIs across occ-metro area labor markets, weighted by employment in each of these labor markets).

Figure 2: Outward occupational mobility from SOC 6-digit occupations



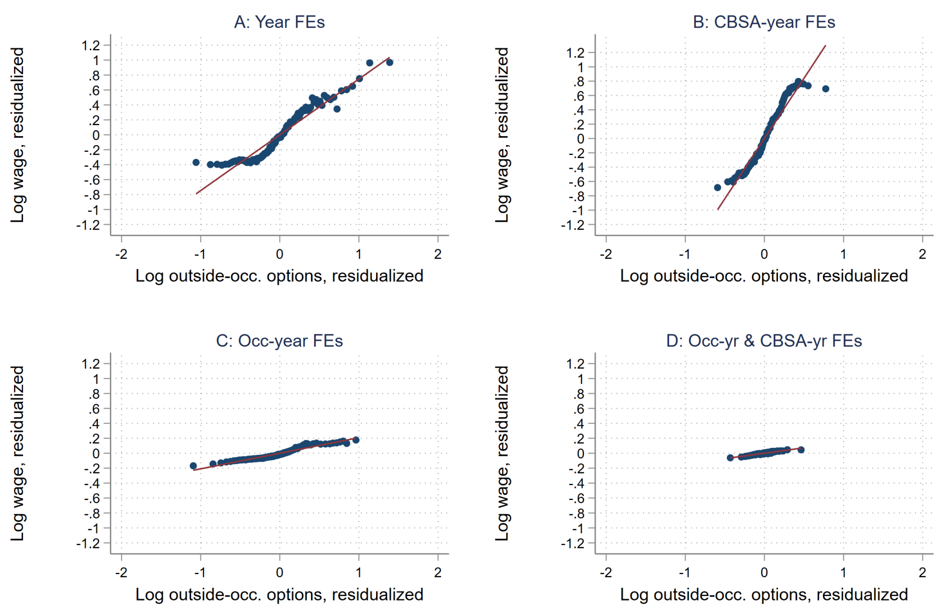
Distribution of the "occupation leave share" – the probability that a worker will leave their occupation conditional on leaving their job – by occupation. Occupation leave share is calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample mean.

Figure 3: Correlations between wage and HHI



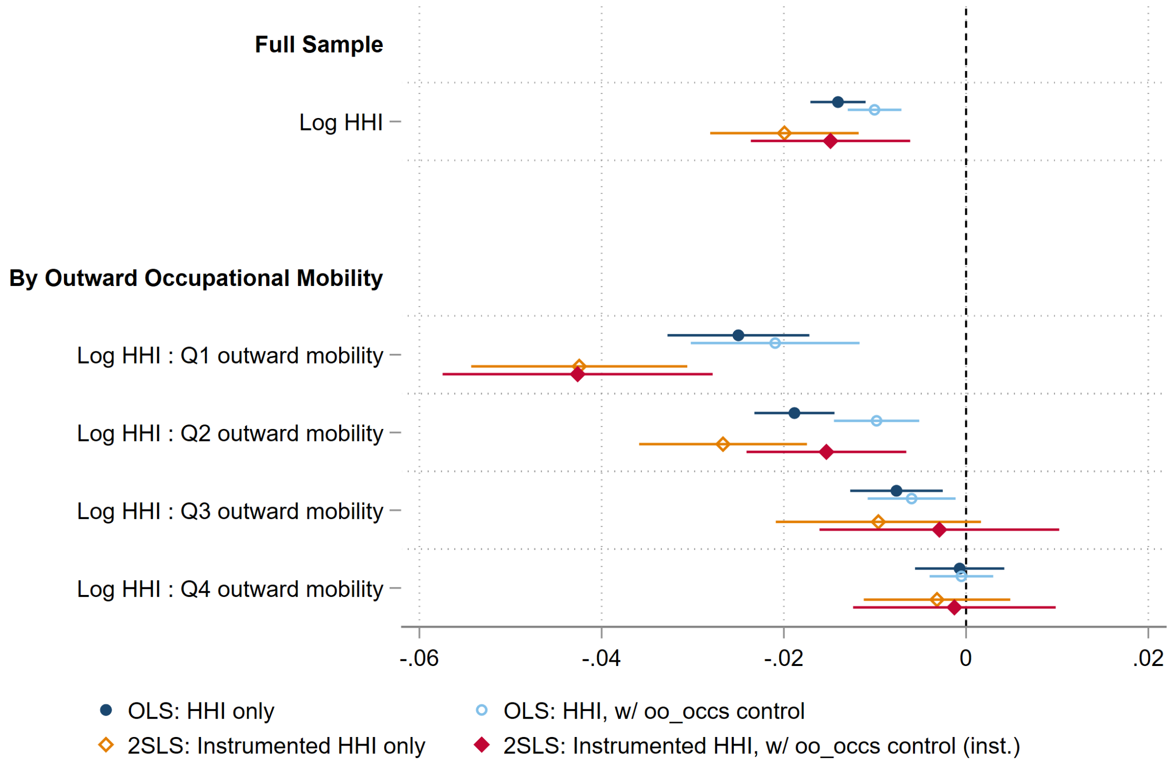
Note: Figure shows binned scatter plots of the relationship between log wages and log employer HHI for occupation-metro area cells over 2011–2019, weighted by occupation-metro area average employment and residualized on different combinations of fixed effects. Slopes for the line of best fit on each graph are: A: -0.11, B: -0.10, C: -0.05; D: -0.01.

Figure 4: Correlations between wage and outside-occupation option index



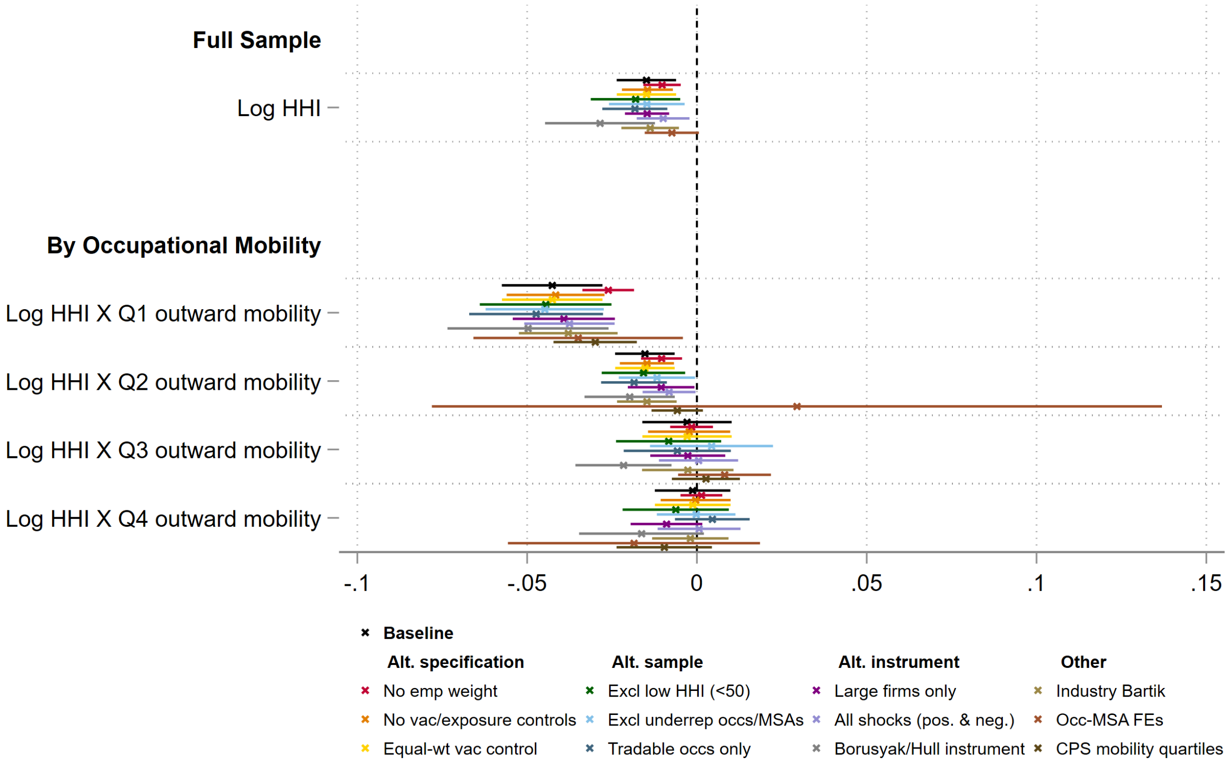
Note: Figure shows binned scatter plots of the employment-weighted relationship between average log wages and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on different combinations of fixed effects (as described by the panel titles). Regression coefficients for the line of best fit on each graph are: A: 0.75, B: 1.68, C: 0.21; D: 0.15. The non-linear shape of the figures without occupation fixed effects (panels A and B) is explained by healthcare occupations which tend to have both low outward mobility and high pay.

Figure 5: Coefficients on wage-HHI regressions



Note: Coefficients on log HHI and 95% confidence intervals from our baseline regressions of occupation-metro area log hourly wage on employer HHI. Navy and light blue represent OLS regression coefficient of wages on HHI, without (navy) and without (light blue) control for outside-occupation job options. Orange and red represent 2SLS IV regression coefficient of wages on instrumented HHI without (orange) and with (red) control for instrumented outside-occupation job options. Top panel presents coefficients for the full sample (as in Table 3); bottom panels present the coefficients estimated separately by quartile of outward occupational mobility (as in Table 4). Regressions use annual data for occupation-by-metro area labor markets over 2011-2019, and include occupation-year and metro area-year fixed effects as well as controls described in the text. Employment-weighted by average occ-metro area employment over 2011-2019. Standard errors clustered at metro area level.

Figure 6: Coefficients on wage-HHI regressions: robustness checks



Note: Coefficients on log HHI and 95% confidence intervals from our baseline 2SLS IV regressions of occupation-metro area log hourly wage on instrumented employer HHI, across various robustness checks (as in Appendix Tables A7-A10). Black shows the baseline estimates from Figure 5.

APPENDIX

A Appendix: Burning Glass Technologies Vacancy Posting 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 2.2). Burning Glass Technologies 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, Marinescu, Steinbaum and Taska (2020) 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 B.)

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).

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.

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. 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. 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 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. (2020)). 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 (Azar et al. (2020) 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).

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 A1). 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 A1): 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.

Vacancy postings, job vacancies, and employment

A natural question is how our data on vacancy postings relates to total job vacancies and to total employment. In theory, when calculating an HHI of employer concentration, one would either like to use data on the share of job vacancies or the share of employment accounted for by each employer. Instead, we have the share of job *postings* accounted for by each employer at the level of each SOC 6-digit occupation, metropolitan area, and year.

BGT estimates that its vacancy data covers the near-universe of online job postings. 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. This may be a particular concern when measuring employer concentration, 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".[66] 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. If an occupation has a particularly low ratio of job postings to job openings, one would expect it to be underrepresented in our data relative to its employment in the general workforce: As discussed in the "representativeness" section below, our data appears to be underrepresentative particularly for certain large low-wage occupations like laborers, cashiers, and food serving and preparation workers, for whom this might be a particularly common phenomenon. Ideally, we would be able to calculate employer concentration at the level of true job openings/vacancies, or employment, rather than vacancy postings, but we are not aware of a data set that enables us to observe firm-level local occupational employment or vacancies in the US.

Representativeness

To what extent is the online job *posting* data representative of all job *openings*? 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. 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. (2020),

[66] 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).

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. (2020) for more details. Note in particular that Azar et al. (2020) use the BGT vacancy data for the same purposes as we do: to calculate employer HHI concentration indices at the level of local SOC 6-digit occupations.

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: one is the degree to which the BGT vacancy *posting* data is representative of the totality of vacancy postings in the US, one is the degree to which vacancy *postings* are representative of true vacancies (job openings), and one 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 OES, in Appendix Figure A1.

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

B 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.

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.[67]
- 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

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

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:

Illustrative example of a resume.

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

We consider the switch from being a purchasing manager to 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:

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		
	13-1040		
Compliance Off. (13-1040)		13-1040	13-1040

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 as follows. For each pair of (different) 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).[68] 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.

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 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

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

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 A2). 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 A3 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.[69]

Figure A4 compares the prevalence of occupations at the 2-digit SOC level in our BGT 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

[69]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.

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 A5), 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 OES 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.

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 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.[70]

[70]In addition, the length of many work histories in the data allows for inferring a broader range of latent

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 executives 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

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).

publications listed by a sample of urology residency applicants on their resumes could not be verified.

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.

Comparability with CPS occupational mobility

The average occupation "leave share" in our BGT resume data is 23%. This is roughly the probability that a worker leaves their SOC 6-digit occupation when they leave their job. This is constructed from the average share of workers leaving their occupation (11%) and the average share of workers leaving their job (46%) in any given year.

To what extent is our measure similar to measures of occupational mobility constructed from the CPS? It is first important to note that our measure is 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.

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 average share of workers leaving their occupation in any given year (11%) reflects the total number of workers who are observed in occupation o in year t who are *not* observed in occupation p at any point in year $t + 1$. This makes it a more conservative measure of occupational mobility than the annual occupational mobility concept commonly constructed from the CPS.

With these caveats in mind: our measure of occupational mobility – the share of workers leaving their occupation being 11% from one year to the next – is somewhat lower than the occupational mobility estimate from Kambourov and Manovskii (2008), who find occupational mobility of 0.20 at the Census 3-digit level in the CPS for the late 1990s. Our measure is, however, in a similar range to Xu (2018) who finds occupational mobility of 0.08 in 2014. The fact that our measure is relatively low compared to Kambourov and Manovskii (2008) is interesting, since sample selection bias might be expected to *overstate* occupational mobility in our data set if the people applying for jobs (whose resumes we observe) are more mobile than average.

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. An employment-weighted binned scatter plot of the 722 occupations for which we are able to calculate both a BGT and a CPS outward mobility measure is shown in Appendix Figure A9. As can be seen, there is a very strong positive correlation between the measures. As would be expected (given the discussion above), CPS outward mobility is higher than BGT outward mobility for any given occupation.

Our "occupation leave share" represents not *unconditional* annual occupational mobility but rather the degree of outward occupational mobility *conditional* on leaving the worker's initial job. We find that 46% of workers in our data are observed in some new job from one year to the next. This is consistent with the average length of a job in our data being 2 years. Note that according to the definition of a job we have chosen to work with, leaving your job does not necessarily entail leaving your firm: moving occupation or job title at the same firm would entail leaving your job. The CPS reports that median employee tenure at their firm in 2018 was 4.2 years, so an average job duration of 2 years in our data is consistent with workers working on average 2 consecutive jobs at the same employer.

C Appendix: OES Occupational Code Crosswalk

In our analysis of the effect of outside-occupation options on wages, we run some regressions over a longer period of 1999–2019. To construct our data set of wages and employment at the occupation-metro area level over this period, we need to create a crosswalk for OES occupational codes from SOC 2000 to SOC 2010.

We start from the crosswalk provided by the BLS for matching occupation codes. The crosswalk is based on an exact match if a SOC 2000 code corresponds to exactly one SOC 2010 code.

When SOC 2000 codes map into multiple SOC 2010 codes, or vice versa, we create a probabilistic mapping. This mapping is based on relative employment shares between the target occupation codes as of 2009 and 2012, obtained at a national level from the BLS.

When one SOC 2000 code splits into multiple SOC 2010 codes, its employees are split based on the relative employment shares in the resulting SOC 2010 codes as of 2012.

When there are multiple SOC 2000 codes mapping into multiple SOC 2010 codes, the number of employees in 2009 and 2012 are counted for the whole cluster of ambiguous assignments. Then, unique assignments within the cluster are made based on the ratio of total 2012 to 2009 employees in the cluster. The remaining employees are apportioned based on their relative share in the remainder. For 2010 and 2011 numbers, the OES combines data collected under both the old and new classification system, and grouped them under either SOC 2010 codes or hybrid identifiers.[71] Where this combination did not result in ambiguity with regard to the meaning of the SOC 2010 code used, this difference in collection methods was ignored and the content of the OES 2010 code transferred one-to-one into the applicable SOC 2010.[72]

Where the OES 2010 code is more aggregated than the SOC 2010 code, it was split based on 2012 employment shares in the target codes.[73]

Similarly, the BLS created hybrid codes for 2017 and 2018, and separately for 2019 OES data during the transition to 2018 SOC codes. We use a BLS mapping between these code

[71] Detailed breakdown of the affected codes available at: `https://www.bls.gov/oes2010_and_2011_oes_classification.xls`

[72] This was the case for the following OES 2010 codes: 11-9013, 15-1799, 51-9151

[73] This was the case for the following OES 2010 codes: 13-1078, 15-1150, 15-1179, 21-1798, 25-2041, 25-3999, 29-1111, 29-1128, 29-2037, 29-2799, 31-1012, 31-9799, 39-4831, 41-9799, 43-9799, 47-4799, 49-9799, 51-9399.

structures and the SOC 2010 codes to crosswalk the OES data for those years to SOC 2010 codes. We use employment data for 2016 to compute relative employment shares under the SOC 2010 codes before the switch to these hybrid codes, and employment data for 2018 and 2019 to capture relative employment under the two hybrid code structures, and then use the same methodology as above to split codes probabilistically, where this is required.

Using these occupational crosswalks, we can stack the OES occupational employment and wage data by metro area provided by the BLS, creating an unbalanced panel of 2.3 million occupation-by-metro area-by-year data points of employment and mean hourly and annual wages for the years 1999-2019.

D Appendix: Alternative approaches to estimating occupational similarity

In Section 2.2 of this paper, we define workers' baseline labor market as a SOC 6-digit occupation within a metropolitan area.[74] We then use occupational transitions to identify workers' outside options. There are two other possible methods of estimating occupational similarity to infer which jobs are good options for workers' outside their occupation: skill-and task-based similarity measures, and demographic- and qualification-based similarity measures. Why do we use occupational mobility?

To answer this question, we ask: What makes jobs in a given occupation a good outside option? Good outside option jobs should be both *feasible* in the sense that the worker can relatively easily become as productive as an average worker in that job, and should be at least somewhat *desirable* to work in (relative to the worker's current job). We show that occupational mobility measures capture the underlying feasibility of a job transition, in the sense that they represent moves that people actually made. This means that they can capture many dimensions of feasibility of a transition – including task, skill, and amenity similarity, but also including other constraints that prevent moves in practice but may not be observed in task or skill data (e.g. regulation, occupational licensing barriers, etc.). Since occupational transitions also reflect moves people have (mostly) chosen to make, they also incorporate the desirability of moves between different occupations.

Skill- and task-based occupational similarity measures define two occupations as more similar, the more similar the skills and tasks are that they require. For example, Macaluso

[74]We choose local SOC 6-digit occupations as our baseline labor market, rather than industries, since research on human capital specificity suggests that occupations are a more accurate approximation of the set of jobs open to workers (Kambourov and Manovskii, 2009; Sullivan, 2010). We choose a metropolitan area as an approximation of the jobs that are available to workers without having to move. A Commuting Zone would be a better geographic measure than a metropolitan area, but unfortunately the BLS data does not include wages by SOC 6-digit occupation at the Commuting Zone level.

(2019) measures occupational skill similarity using the vector difference of occupational skill content, and Gathmann and Schönberg (2010) use the angular separation of occupations' task content vectors. A skill- or task-based measure of the similarity between two occupations does indeed capture many dimensions of the feasibility of an occupational transition. However, it has a number of weaknesses relative to a transition-based measure.

First, a skill- or task-based similarity measure cannot capture non-skill-related aspects which affect the feasibility of moving from one occupation to another occupation, such as occupational licensing or certification barriers between two occupations which may have similar skill requirements. Second, a skill- or task-based similarity measure cannot capture the desirability of moving from one occupation to another: it may be that two occupations are very similar in terms of the skills and tasks that they require, but the amenities may differ (for example, long or unpredictable hours being required may make an occupation less desirable for parents of young children) – so that the kind of people that work in one occupation may not want to work in the other.

Third, skill- or task-based similarity measures are (usually) symmetric between occupation pairs, whereas transitions data can capture the asymmetry of the value of different occupations as outside options for each other: occupation p may be a relevant outside option for occupation o but not the other way around, perhaps because of generalist/specialist skill differentials, differences in job hierarchy or status, or specific requirements for experience, training or certification. Fourth, skill- or task-based similarity measures require both the ability to *measure* the underlying skill and task requirements for each occupation with some accuracy *and* substantial assumptions as to how skill and task data should be combined to create a similarity measure. Skill- and task-based similarity measures can be highly sensitive to these assumptions. In contrast, a transition-based measure has the advantage of being non-parametric. This allows us to capture the equilibrium job choice policy function without having to impose a particular model of how workers and firms choose to offer and accept jobs, or about equilibrium play (Bajari, Benkard and Levin, 2007).

Demographic- and qualification-based occupational similarity measures define two occupations as more similar, the more similar are their workers based on their observable demographic and educational characteristics. (This is a simplified version of the approach used by Caldwell and Danieli (2018), who probabilistically identify workers' outside options using the distribution of other similar workers across jobs and locations). This type of measure can capture occupational similarity in terms of the skills required, based on workers' inherent characteristics and education/training, and in terms of preferences determined by these factors. It also has the advantage of requiring substantially fewer assumptions than a skill- and task-based measure, since it uses workers' actual labor market choices to reveal their outside options. Since it does not consider career paths, however, a demographic- and qualification-based occupational similarity measure cannot capture the role of occupation-specific experience and learning, or obstacles to occupational transitions, in determining future employment options. In that sense, a demographic- and qualification-based measure of occupational similarity can be thought of as a static approach to defining a 'revealed' labor market, whereas a transition-based measure can be thought of as a dynamic approach. In addition, as with skill- and task-based approaches, this approach in practice requires assumptions on which observables are relevant for job choices and parametric assumptions on the functional form of the choice function.

Our transitions-based measure does have a major potential drawback relative to a skill- or task-based measure: off-equilibrium outside options are not observed if bargaining is efficient. It may be the case that another occupation is very feasible but slightly less desirable, which makes it a relevant outside option for a worker but one that is rarely exercised in equilibrium. However, if the number of workers and firms is large enough to observe rare transitions, worker preferences are continuous, and idiosyncratic shocks have enough variance to induce many workers to change occupations, these off-equilibrium options will on average still be revealed by the transition data - and we believe these conditions hold for job transitions.

More specifically, there are three conditions under which the above concern about off-equilibrium options in the 'revealed labor market' approach based on observed occupational transitions is not significant. First, there is a continuous distribution of worker heterogeneity with regard to preferences over different firms, and so any given worker's closest outside options (off-equilibrium option) are revealed by the actual equilibrium paths of similar workers (similar to the way that choice probabilities map to expected value functions in discrete choice models with i.i.d. preference shocks (McFadden, 1974)). Second, there has to be a sufficient number of similar workers and firms to observe these transitions. Third, that the only *relevant* off-equilibrium outside options for workers in the wage bargaining process are those which are quite similar to their existing job or skill set in expected match quality (i.e. that cashier jobs are not relevant outside options for engineers), such that the variance of worker preferences beyond the expected match quality is large enough to manifest in different job matches for all relevant outside options. If these conditions are satisfied, the expected relevant off-equilibrium options for workers in a given occupation can be inferred by the equilibrium choices of other workers in the same occupation.

E Appendix: Determinants of occupational mobility

In section 2.2 we showed that empirical occupational transitions reflect underlying similarity in occupations' task and skill requirements and in their amenities. We explain this analysis in more detail here.

Occupation characteristics: measures

Task requirements. To measure occupational similarity in terms of tasks required, we use two different approaches from prior literature.

First, we use the vector difference between the importance scores for "Skill" task content items provided by the O*Net database of occupational characteristics, as proposed by Macaluso (2019). In our measure, as in Macaluso (2019), dissimilarity is measured as the average difference in importance scores (scaled to lie between zero and ten) across the full set of 35 tasks. For a similar notion of task distance, see Gathmann and Schönberg (2010).

Our measure of average task distance $\bar{D}_{op}$ between occupations o and p is defined as:

$$\bar{D}_{op} = \frac{1}{35}\sum_{k=1}^{35} |S_{k,occ\,p} - S_{k,occ\,o}|,$$

where $S_{k,occ\,p}$ is the standardized skill k measure for occupation p.

Second, we use composite task measures from recent literature relating occupational task content to important economic outcomes. We consider six task composites (denoted "ALM") first introduced in Autor, Levy and Murnane (2003) and updated to the most recent O*Net version in Acemoglu and Autor (2011). These composites mainly capture the distinction between cognitive vs. manual and routine vs. non-routine task contents. We also consider a categorization by Deming (2017) (denoted "DD"), which recasts the occupational task composites and also introduces a composite capturing social skill-related task intensity.[75]

Job amenities. We measure similarity in the "temporal flexibility" of different occupations using the 5 O*Net occupation characteristics that Goldin (2014) identifies as proxies for the ability to have flexibility on the job: time pressure, contact with others, establishing and maintaining interpersonal relationships, structured vs. unstructured work, and the freedom to make decisions.[76] These amenities are particularly important because, as Goldin (2014) notes, "certain occupations impose heavy penalties on employees who want fewer hours and more flexible employment" (p. 1106), which in turn may contribute to gender gaps in earnings. Note that higher scores in each of these domains imply more rigid time demands as a

[75]We update the task composites from Deming (2017) by using the latest source for task contents on O*Net, and computing the composites at the level of SOC 2010 occupational codes.

[76]The five characteristics correspond the following O*Net survey items: IV.C.3.d.1 - How often does this job require the worker to meet strict deadlines?; IV.C.1.a.4 - How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?; IV.A.4.a.4 - Developing constructive and cooperative working relationships with others; IV.C.3.b.8 - To what extent is this job structured for the worker, rather than allowing the worker to determine tasks, priorities, and goals?; IV.C.3.a.4 - Indicate the amount of freedom the worker has to make decisions without supervision.

result of business needs and make it less likely that workers are able to step away from their job whenever they need to.

Leadership responsibility. Another reason for observing occupational transitions may be career advancement (which is often reflected in a change of occupation). To study whether this appears in our data, we identify occupational characteristics measuring leadership responsibilities from the O*Net database, and create a new "leadership" composite measure defined at the level of each SOC 6-digit occupation. The measure incorporates the six characteristics most associated with leadership positions in the O*Net data, alongside the O*Net work style category for leadership. Since this is a new composite measure of an important occupational characteristic, we outline it in more detail here.

We used the following algorithm to determine which characteristics measure leadership responsibilities: On the O*Net website, we looked at the work activity characteristics that describe "Interacting with Others". For each of them, we considered the list of top 20 occupations with the highest level of that characteristic and counted how many of them are managerial positions, as evidenced by the words "supervisor", "manager", "director", or equivalents, in the occupation title. We selected all the characteristics for which the share of managerial positions among the top 20 occupations was greater than half, as these characteristics seem to be associated with "leadership" in some sense; we also added the O*Net work style category for leadership. The final list of characteristics contains the following O*Net items: I.C.2.b. - Leadership work style: job requires a willingness to lead, take charge, and offer opinions and direction; IV.A.4.a.2. - Communicating with Supervisors, Peers, or Subordinates; IV.A.4.b.1. - Coordinating the Work and Activities of Others; IV.A.4.b.2. - Developing and Building Teams; IV.A.4.b.4. - Guiding, Directing, and Motivating Subordinates; IV.A.4.c.3. - Monitoring and Controlling Resources; IV.A.4.c.2. - Staffing Organizational Units (We were reassured to note that for 6 of these 7 characteristics, "Chief Executives" are among the Top 20 occupations in terms of importance of this measure.). We use the mean score across these 7 characteristics as our "leadership" composite. All variables are converted into standardized Z-scores before including them in regressions, so coefficients represent the effect of a one standard deviation difference in the characteristic on the outcome variable.

Occupational similarity and mobility

To evaluate whether workers are more likely to move to occupations that have similar characteristics to their current occupation, we estimate the following regression:

$$\pi_{o\to p} = \alpha_o + \beta^{abs}|X_{occ\ p} - X_{occ\ o}| + \gamma|\Delta w_{o\to p}| + \epsilon_{op}. \tag{9}$$

where $\pi_{o \to p}$ is the share of job changers in the origin occupation o that move into target occupation p, $|X_{occ\,p} - X_{occ\,o}|$ is the absolute difference between the target and the origin occupation in each of the occupational characteristics X_o defined above, and α_o are origin occupation fixed effects to control for differences in outward mobility across occupations. We control for absolute wage differences between the occupations in all regressions except for those estimating the effect of wages or amenity differences on occupational mobility,[77] but note that the results are qualitatively similar without the wage controls.

We would expect the coefficient on the absolute difference in characteristics to be negative: the greater the difference between two occupations, the less likely we should be to observe the worker moving from one into the other. Our results bear this out: in every regression of pairwise occupational mobility on the absolute difference in characteristics, the coefficients are significantly negative or statistically insignificant, as shown in Figure A12.[78]

The previous results impose symmetry on the likelihood of occupational transitions – but between many pairs of occupations, the probability of moving in one direction is likely to be different than the probability of moving in the other direction. To study whether differences in characteristics also predict the direction of occupational flows, we estimate a similar regression equation to that shown in equation (9), but now using the *relative* (target minus origin) difference in occupational characteristics as the independent variable:

$$\pi_{o \to p} = \alpha_o + \beta^{rel}(X_{occ\,p} - X_{occ\,o}) + \gamma \Delta w_{o \to p} + \epsilon_{op}. \tag{10}$$

Again, we include origin occupation fixed effects and now control for relative wage differences between the occupations in all regressions except for the amenity differences and the wage regression. The β^{rel} coefficients obtained from estimating equation (10) for the different measures are also shown in Figure A12. (Note that this analysis involves directed relationships between occupations, so if the same share of moves in each direction is observed for an given occupation pair, the estimated effect of differences between them would be zero.)

A number of our predictions are borne out in the data: we find (1) that workers are more likely to move towards jobs with higher wages; (2) that workers transition on average *towards* jobs that require more leadership responsibility - as would be expected from moves up the career ladder; (3) that occupational transitions have on average been *towards* occupations that have higher analytical content and require more social skills, and out of occupations

[77]Amenities are most likely to be priced into wages (Goldin, 2014) and controlling for the latter would therefore be inappropriate.

[78]Our findings build on Macaluso (2019), who showed that greater skill distance between SOC 2-digit occupations is associated with lower occupational flows between these occupations: we demonstrate this relationship at the SOC 6-digit level with a larger variety of task and skill measures, and show that differences between occupations in temporal flexibility and leadership responsibilities also appear to determine workers' likelihood of moving between them.

with more routine task requirements;[79] and (4) that workers have on average been moving into occupations that require more contact and working relationships with others (and so have less time flexibility).

While occupational transitions therefore do reflect similarity in tasks, temporal flexibility, and leadership requirements, we note that there is substantial variation in occupational transitions which is not captured by these other occupational similarity measures. Indeed, in employment-weighted regressions of occupational transition likelihood $\pi_{o \to p}$ on various measures of differences in job characteristics, with origin occupation fixed effects, the adjusted R-squared is quite low: 0.025 for skill distance, 0.021 for wage difference, 0.039 for amenity difference (temporal flexibility), 0.033 for leadership difference, and 0.058 for a composite skill measure. That is, in all these cases the correlation between occupational transition probabilities and occupational characteristic similarities is strong and positive, but the explanatory power is low.

The failure of skill similarity measures to explain many occupational transitions can be illustrated by a few cases from our data. First, consider some occupation pairs that are very similar on a skill distance metric (in the lowest distance decile), but where our data shows almost no (less than 0.01%) chance of moving from one to the other when switching jobs, in either direction: Surveyors vs. Medical & clinical laboratory technologists; Carpenters vs. Dental assistants; Travel agents vs. Police, fire & ambulance dispatchers. In all of these occupational pairs it is intuitively clear why they may look similar in terms of an abstract description of the tasks involved, but in practice this skill distance does not make them relevant outside options for one another because of differences in other job characteristics or requirements. Second, consider another pair of occupations which are very similar on the skill distance metric (again, in the lowest distance decile): Pediatricians vs. Management analysts. When pediatricians change jobs, 8.7% of them become management analysts, but less than 0.01% of management analysts switching jobs become pediatricians. The skill distance metric misses the fact that one of these occupations requires extensive training and licensing which means that, in practice, the occupational move is only possible in one direction.

[79] These patterns could be in line both with career progression for individual workers, and/or with the aggregate decline of routine occupations over the same time period documented by, for example, Acemoglu and Autor (2011), and the increasing demand for social skills documented by Deming (2017).

F Appendix: IV analysis

Identification assumptions for concentration instrument

This section provides more formal details on the assumptions required for the IV identification of the effects of labor market concentration on wages. Our instrument can be interpreted as a type of granular IV following Gabaix and Koijen (2020), where market-level trends are instrumented for using idiosyncratic firm-level shocks (for details on the granular IV identification approach see Gabaix and Koijen (2020)). Or, it can be seen through the lens of the Bartik or shift-share IV approach, following Borusyak, Hull and Jaravel (2022), with exogenous 'shocks' in the form of differential national hiring patterns for large firms, and initial squared employer shares of each firm in a given local labor market determining the exposure to those shocks.

We can rewrite the concentration instrument as

$$\begin{aligned} Z^{HHI}_{o,k,t} &= \log\left(\sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1\right)\right) \\ &= \log\left(\sum_j \sigma^2_{j,o,k,t-1}\tilde{G}_{j,o,k,t}\right) \end{aligned}$$

where $\tilde{G}_{j,o,k,t} = \frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1$ is the predicted firm-level excess local vacancy growth relative to the average predicted local occupation vacancy growth - the time-varying shock - and $\sigma^2_{j,o,k,t-1}$ is the exposure of the local concentration index to that shock.[80] In our baseline specification, we use only positive vacancy growth shocks $\tilde{g}_{j,o,t} > 0$, as it is less clear that the logic of our instrument holds for negative shocks. In robustness checks, we include all shocks (positive and negative), and we also create a version of the instrument only using shocks to large firms, where large firms are defined as firms which have vacancies in that occupation o and year t in at least five different metropolitan areas k (and $\tilde{g}_{j,o,t}$ is set to zero for all other firms).

As noted in the main text, we add three controls to our baseline specification. To control for any effects on local labor demand of differential exposure to large national firms' hiring, we control for (1) the growth rate of local vacancies in the occupation-metro area labor market ($g_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} g_{j,o,k,t}$), and (2) the predicted growth rate of local vacancies based on large firms' national growth ($\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1}\tilde{g}_{j,o,t}$). To control for differential initial exposure to non-local firms, we introduce our "exposure control": the sum of squared vacancy shares in year $t-1$ of all firms in occupation o and metropolitan area t which were

[80] For simplicity of exposition, we assume here that employer concentration and outside-occupation options are not correlated – but the logic of this argument does not depend on this assumption.

nationally large enough to meet our definition of large firms (vacancies in occupation o and year t in at least five metropolitan areas k) and which had non-zero national vacancy growth between years $t-1$ and t: $e_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \cdot \mathbb{1}[\tilde{g}_{j,o,t} \neq 0]$.

In our fixed effects IV estimation of equation (1), the exclusion restriction for the instrument on the HHI concentration index is then equivalent to

$$Cov[Z^{HHI}_{o,k,t}, \xi_{o,k,t} | \Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{o}^{N^{occs}} \sum_{k}^{N^{cities}} (Z^{HHI}_{o,k,t})^{\perp} \xi_{o,k,t}\right] \rightarrow 0$$

where $(Z^{HHI}_{o,k,t})^{\perp}$ represents $Z^{HHI}_{o,k,t}$ after it has been residualized with regard to metro area-k-by-year-t fixed effects Γ_{kt} and occupation-o-by-year-t fixed effects Γ_{ot}, as well as our three control variables $g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}$, and $\xi_{o,k,t}$ represents the residual in the wage regression.

This orthogonality condition holds under two assumptions. First, we require that the national firm-level growth shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks $\xi_{o,k,t}$, the fixed effects Γ_{kt} and Γ_{ot}, actual and predicted average local vacancy growth $g_{o,k,t}$ and $\tilde{g}_{o,k,t}$, and initial exposure to non-local firms $e_{o,k,t}$. That is, once we account for the control variables, expected local squared exposure to excess national firm-level growth needs to be random in expectation.[81] Second, there needs to be a large enough number of independent firm-level shocks to drive sufficient variation in the instrument to identify the effect of interest.

The first assumption requires that the local size-squared-weighted exposure to national firm-level employment shocks does not affect the local wage in occupation o through a direct channel other than increasing the local labor market concentration $HHI_{o,k,t}$, conditional on the control variables. Note that this allows for different local occupations to have different average expected average growth rates based on national firm growth. It only requires that whether this growth is driven by the *national* growth of locally *large* firms vs. small firms varies across local occupations in a way that is uncorrelated with local wage residuals.

To be concrete, note the hypothetical example from the main text, which considered insurance sales agents in Bloomington, Illinois and in Amarillo, Texas. In each metro area, there are several insurance companies who employ insurance sales agents. Assume that in Bloomington, State Farm has a large share of local insurance sales agent employment, while in Amarillo employment is more concentrated in other large insurance companies. In years

[81] In a robustness check, we also include a control for average vacancy growth across firms within a local occupation, with each firm weighted equally, $\frac{1}{N}\sum_j g_{j,o,k,t}$. This is suggested by Gabaix and Koijen (2020) 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 already discussed).

where State Farm grows substantially faster than other major insurance companies nationwide, under most combinations of the distribution of that growth across metro areas and the initial distribution of employer shares in each metro area, employer concentration of insurance sales agents will grow by more in Bloomington IL than in Amarillo TX. Moreover, our granular IV identification approach controls for local growth rates of overall insurance sales agent employment in both metro areas. Thus, it allows 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 smaller employers is not correlated with local idiosyncratic wage shocks for insurance sales agents.

How does the first-stage assumption work? The first stage of our regression holds if, when large firm j grows nationally, local occupation-metro area labor markets with a higher share of vacancies accounted for by firm j in year $t-1$ see a larger increase in employer concentration. A sufficient condition for this to be the case under *most* initial employer share distributions is if firm j's new vacancies are allocated evenly 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.[82] 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. In our data there is a negative relationship between an employer's initial vacancy share in an occupation-metro area labor market and its next year's vacancy growth rate, but this relationship is not sufficiently strong to invalidate our first stage. Empirically, our first stage holds for occupation-metro area labor markets with HHIs above all but very low levels.

Borusyak and Hull (2020) demonstrate that nonlinear transformations in shift-share IVs risk introducing omitted variable bias. Specifically, they illustrate that a log transformation of a shift-share IV – a concave transformation – can introduce omitted variable bias since the expected instrument becomes systematically higher for units of observation where the shift-share sum has lower variance. Following Borusyak and Hull (2020), in a robustness

[82]Note that this is not the case – i.e. 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 for 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 (for example only 10% of occupation-metro area labor markets, containing only 1% of workers, had an HHI greater than 5,000 in our data in 2019).

check we therefore construct an alternative instrument as follows:

$$\begin{aligned} Z^{HHI,alt}_{o,k,t} &= \sum_j \sigma^2_{j,o,k,t-1} \log\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2}\right) \\ &\approx \sum_j \sigma^2_{j,o,k,t-1}\left(2(\tilde{g}_{j,o,t}-\tilde{g}_{o,k,t})\right) \end{aligned}$$

where the instrument is constructed by expressing the log HHI in period t as a function of growth rates and employer shares in $t-1$, $HHI_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+g_{j,o,t})^2}{(1+g_{o,k,t})^2}\right)$, defining the relative squared growth rate of each employer j as $\hat{g}_{j,o,t} = \frac{(1+g_{j,o,t})^2}{(1+g_{o,k,t})^2}$ and taking a log-linear approximation around $\hat{g}_{j,o,t} = 1$ (i.e. a log-linear approximation around the point at which the vacancy growth rate of each employer j is identical to the local labor market vacancy growth rate, so the HHI does not change from one year to the next).

Identification assumptions for outside-occupation option index instrument

This section provides more formal details on the assumptions required for identification of the outside-occupation options effect on wages using the instrumental variables strategy based on national leave-one out mean wages.

As described in Section 2, our instrument for the oo^{occs} index, Z^{oo}, is the weighted average of national leave-one out mean wages in occupation p, $\bar{w}_{p,\cancel{k},t}$, where the weights are the product of the year 1999 relative employment share in each of those occupations in the worker's own metro area, $\frac{s_{p,k,1999}}{s_{p,1999}}$,[83] and the national occupation transition shares from the worker's occupation o to each of the other occupations, $\pi_{o\to p}$:

$$Z^{oo}_{o,k,t} = \log\left(\sum_p^{N_{occs}} \pi_{o\to p}\cdot\frac{s_{p,k,1999}}{s_{p,1999}}\cdot\bar{w}_{p,\cancel{k},t}\right) \tag{11}$$

To make the assumptions transparent under which this wage instrument identifies the coefficient on our outside-occupation option index in equation (1), we again follow the framework presented in Borusyak et al. (2022).[84] Note that we can write the instrument as

$$Z^{oo}_{o,k,t} = \log\left(\sum_{p=1}^{N^{occs}} s_{okp}\bar{w}_{p,\cancel{k},t}\right)$$

where $s_{okp} = \pi_{o\to p}\cdot\frac{s_{p,k,1999}}{s_{p,1999}}$ is a measure of predicted local exposure to the shock. In our fixed effects IV estimation of equation (1), the exclusion restriction for the instrument for

[83]Or the first year in the data, if there is no data for the occupation-metro area cell in 1999.

[84]For simplicity, assume that the outside-occupation option index and the concentration index are not correlated - but the intuition for the identification does not depend on that.

outside-occupation options is then equivalent to

$$Cov[Z^{oo}_{o,k,t}, \xi_{o,k,t} | \Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = \sum_{t=1}^{T} \sum_{p=1}^{N^{occs}} \sum_{k}^{N^{cities}} (Z^{oo}_{o,k,t})^{\perp} \xi_{o,k,t} \rightarrow 0$$

where $(Z^{oo}_{o,k,t})^{\perp}$ represents $Z^{oo}_{o,k,t}$ after it has been residualized with regard to metro area-k-by-year-t fixed effects Γ_{kt} and occupation-o-by-year-t fixed effects Γ_{ot}, as well as our three control variables $g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}$.

Borusyak et al. (2022) show that this orthogonality condition holds under two assumptions. First, we require that the national occupation-level shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks ϕ_{pt}, the fixed effects Γ_{kt} and Γ_{ot}, and the control variables. Second, there needs to be a large number of independent occupational shocks. The first assumption requires that the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p is correlated with the local wage of occupation p in location k (relevance condition), but does not affect the local wage in initial occupation o through a direct channel other than increasing the quality of local outside options $Z^{oo}_{o,k,t}$. However, this lack of a direct effect only needs to hold *conditional* on controlling for fixed effects that include the national wage trend in occupation o itself and wage trends that are common to all occupations in metro area k.[85] The inclusion of these fixed effects increases our confidence that the assumptions for instrument validity hold.

Borusyak and Hull (2020) demonstrate that nonlinear transformations in shift-share IVs risk introducing omitted variable bias. As with our HHI instrument, following Borusyak and Hull (2020), in a robustness check we therefore construct an alternative instrument as follows:

$$Z^{oo,alt}_{o,k,t} = \sum_{p}^{N_{occs}} \pi_{o \rightarrow p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \log\left(\bar{w}_{p,\not{k},t}\right).$$

Industry Bartik shock

One possible concern with the identification assumptions required for our outside-occupation index – which may not entirely be picked up by our occupation-year, or metro area-year fixed effects – is that industry-level wage trends may differentially impact local occupations based on their metro area's direct exposure to those industries, rather than only based on indirect exposure through outside occupation job options. As discussed in the text, an example

[85] As an example, note that national-level correlation in the wages of a pair of occupations (e.g. Compliance Officers and Financial Analysts), perhaps due to common industry shocks, does *not* invalidate this identification strategy, because we are holding national wage trends constant for each occupation and are identifying outside option effects from the differences between metro areas *within* occupations.

of this could be the following. Imagine that the finance industry and the tech industry employ both accountants and data scientists to a disproportionate degree relative to other occupations, and that San Francisco has a large share of employment in tech while New York has a large share of employment in finance. Imagine further that being a data scientist is a good outside option occupation for an accountant. In years where the tech industry is booming nationwide, this will impact San Francisco more than New York. Accountants in San Francisco will see wages rising by more than accountants in New York – partly driven by the increase in the outside option value of becoming a data scientist, but partly simply because more accountants in SF already work in the tech industry, as compared to accountants in NY, and so they will see their wages rise by more. To control for this possible omitted variable bias, we incorporate an industry "Bartik" shock in a robustness check for our baseline regressions.

Shock construction. We construct this shock as the predicted impact of national industry wage trends for each occupation-metro area-year cell, with the limitation that we do not observe the exact industry exposure of each occupation at the local level. The industry Bartik shock for occupation o in city area k in year t is defined as

$$\sum_{\iota}^{industries} \underbrace{\frac{emp_{\iota,o,t-1}}{emp_{o,t-1}}}_{\substack{\text{Avg. occ. } o \text{ exposure} \\ \text{to ind. } \iota}} \cdot \underbrace{\frac{\frac{emp_{\iota,k,t-1}}{emp_{k,t-1}}}{\frac{emp_{\iota,t-1}}{emp_{t-1}}}}_{\substack{\text{Rel. exposure of city } k \\ \text{to ind. } \iota}} \cdot \underbrace{\left(\frac{\bar{w}_{\iota,t,-k} - \bar{w}_{\iota,t-1,-k}}{\bar{w}_{\iota,t-1,-k}}\right)}_{\substack{\text{LOO national growth in} \\ \text{ind. } \iota \text{ wages}}}$$

where ι denotes each NAICS 4-digit industry. The shock to each local occupation cell coming from industry trends is computed as the weighted sum of the exposure to wage shocks in each industry. The contribution of each industry ι to this sum is approximated as the product of (1) the national average exposure of the occupation to that industry ι, (2) the share of employment in the metro area k which is industry ι, relative to the national share of all employment in that industry, and (3) the leave-one-out growth in average wages in industry ι (omitting values from metro k itself). The exposure measures are lagged by one year to avoid the possibility of endogenous responses of employment to the industry-level shock in question. The use of industry Bartik shocks as instruments usually relies on the assumption that national industry-level wage shocks based on data from *other cities* are uncorrelated with local occupation-level wage trends, except to the extent that the former causes the latter. In our case, we only rely on it to represent a good proxy as a control variable for national industry trends affecting different cities and occupations in particular years, without any claim of a causal relationship. See Chodorow-Reich and Wieland (2020) for an example of a similar use of the Bartik industry shock as a control variable.

Data. We use data on national employment by NAICS 4-digit industry and SOC 6-

digit occupation from the Bureau of Labor Statistics Occupational Employment Statistics to construct the employment shares in each industry by occupation, and we use Quarterly Census of Employment and Wages data from the BLS to construct industry employment shares by metropolitan statistical area, national industry employment shares, and leave-one-out national industry wage growth.

We report our baseline regression results, controlling for this industry Bartik shock, in column d of Tables A8 and A10.

G Appendix: Stata commands

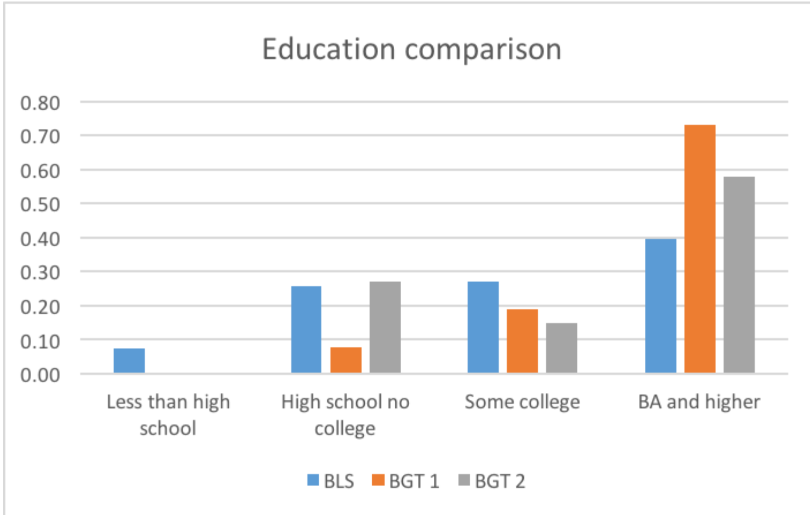
In our estimation, we used a number of user-written Stata commands: *reg2hdfe* (Guimaraes and Portugal, 2010), *reghdfe* (Correia, 2016), *ivreg2hdfe* (Bahar, 2014), *binscatter* (Stepner, 2013), *binscatter2* (Droste, 2019), and *coefplot* (Jann, 2013).

H Appendix: Figures

Figure A1: BGT Vacancy Data: representedness of occupations, relative to BLS OES, 2019

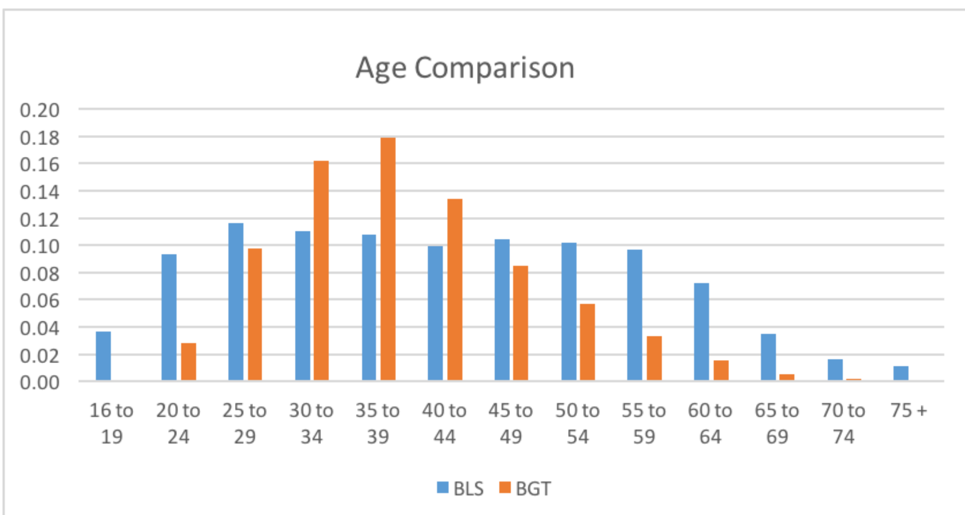
Note: Comparison of distribution of share of vacancies accounted for by each SOC 6-digit occupation in the BGT vacancy data in 2019, relative to that occupation's share of total employment in 2019 in the BLS occupational employment statistics. Occupations comprising greater than 1% share of either data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

Figure A2: BGT Resume Data: education relative to 2018 labor force



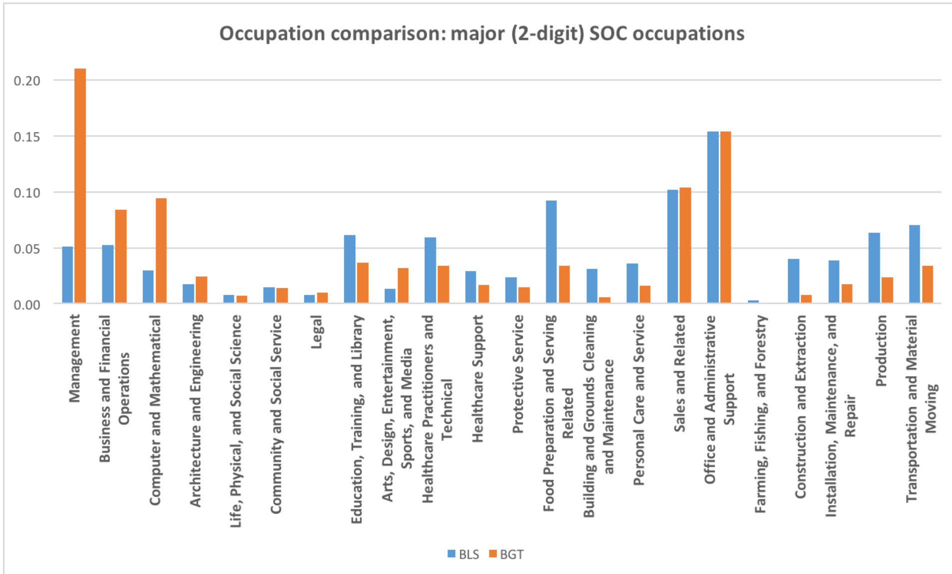
Note: Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT resume data. Two versions are shown: BGT 1 excludes all resumes missing educational information, while BGT 2 assumes all resumes missing educational information have high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

Figure A3: BGT Resume Data: age distribution relative to 2018 labor force



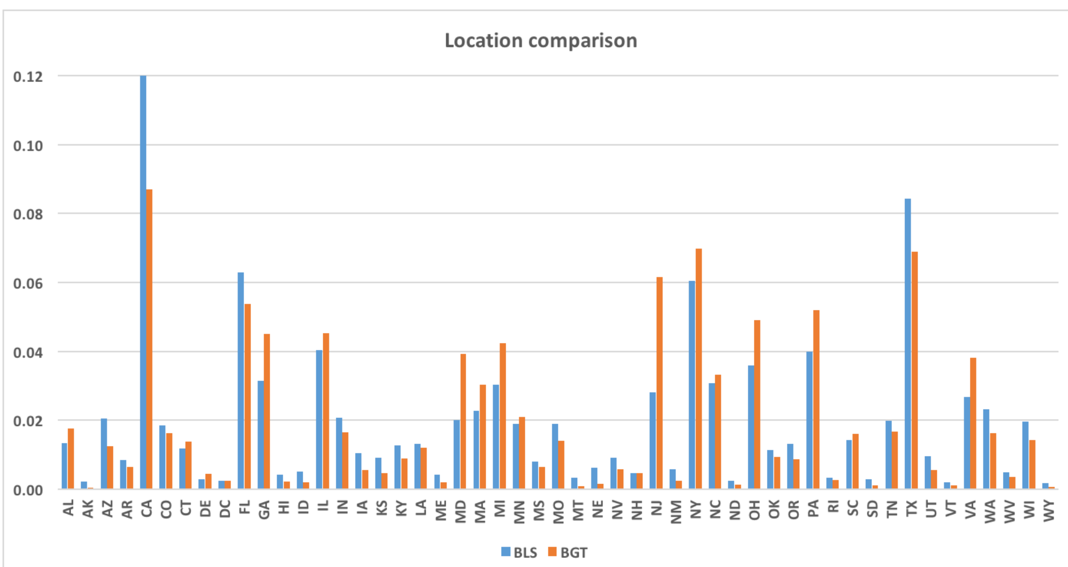
Note: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT resume data. The resume data is discussed in detail in Appendix B.

Figure A4: BGT Resume Data: occupations relative to 2017 labor force



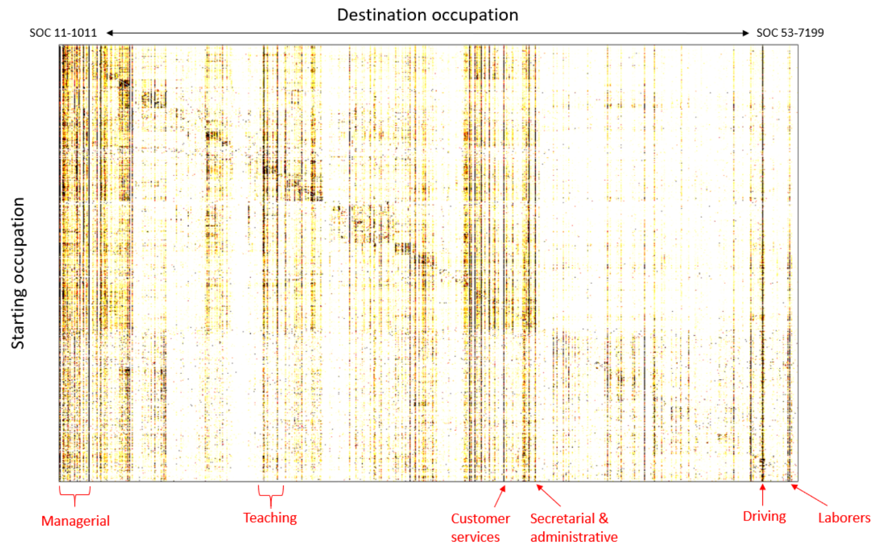
Note: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT resume data. The resume data is discussed in detail in Appendix B.

Figure A5: BGT Resume Data: locations relative to 2017 labor force



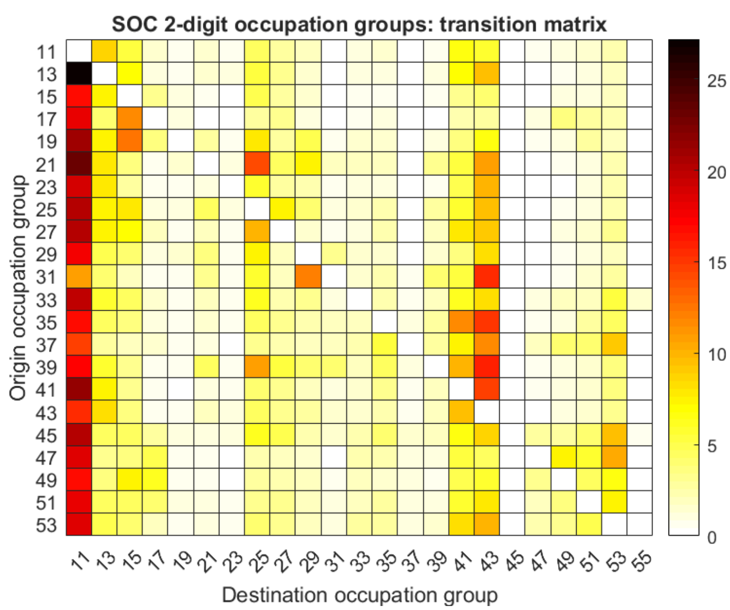
Note: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT resume data. Graph shows share of total in each state. The resume data is discussed in detail in Appendix B.

Figure A6: 6-digit SOC occupational transition matrix



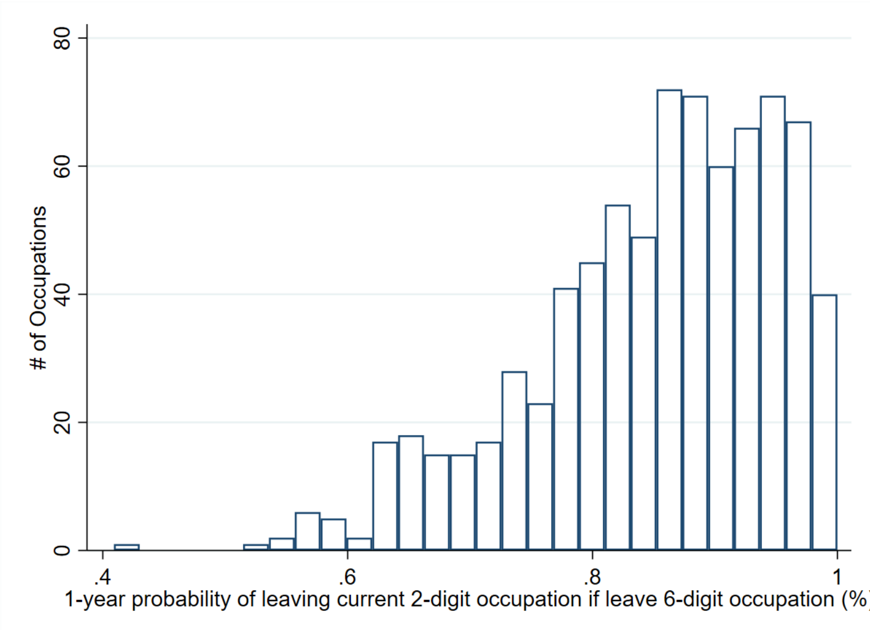
Note: Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial job. Occupations are sorted in SOC numerical order. Cells colored black have a transition probability of 1% or greater conditional on leaving the initial job. Transitions to own occupation are excluded. Data computed from BGT resume data set for 2002-2015. The annotation points out certain common destination occupations, which show up as darker vertical lines on the heatmap. The presence of a darker line along the diagonal suggests that workers commonly transition to occupations which are close to their own according to the numerical order of SOC codes. The resume data is discussed in detail in Appendix B.

Figure A7: 2-digit SOC occupational transition matrix



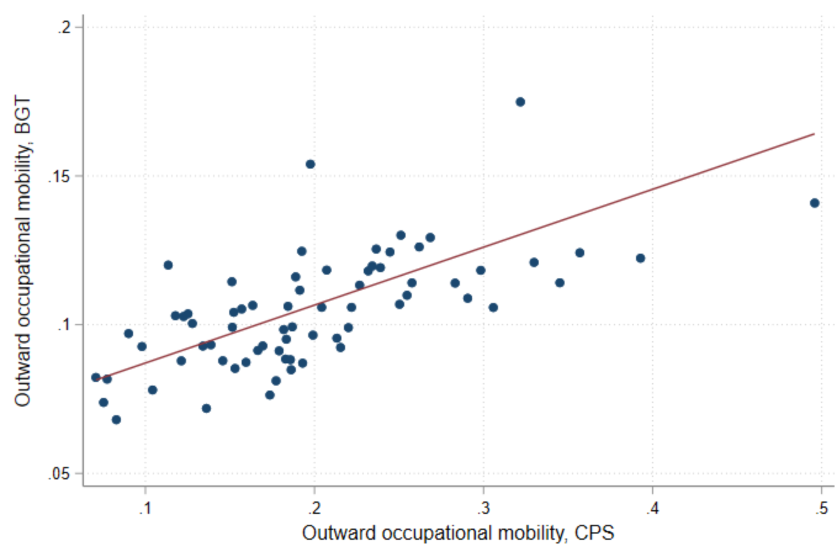
Note: Occupational transition matrix showing transition probability between 2-digit SOC occupation groups conditional on leaving the initial job. Cells colored black have a transition probability of 25% or greater conditional on leaving the initial job. Job transitions within an occupation group are excluded. Data computed from BGT resume data set for 2002-2015. The resume data is discussed in detail in Appendix B.

Figure A8: Occupational mobility: SOC 6-digit moves that are also 2-digit moves



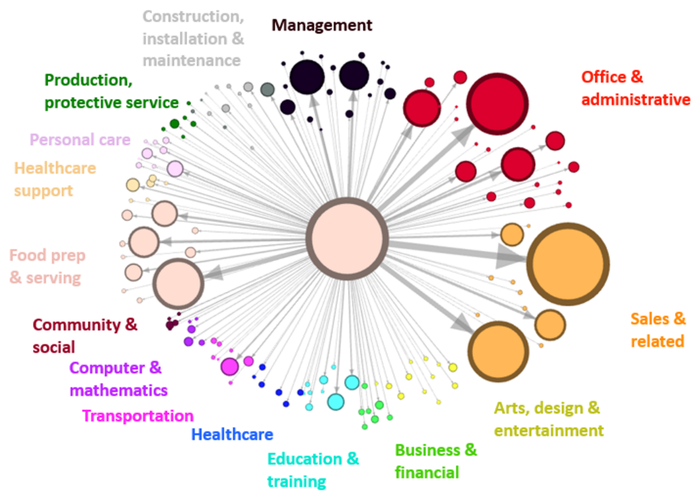
Note: Distribution of the proportion of workers moving 6-digit SOC occupation who *also* move 2-digit SOC occupation, by occupation, calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations. The resume data is discussed in detail in Appendix B.

Figure A9: Occupational mobility: comparing CPS and BGT measures



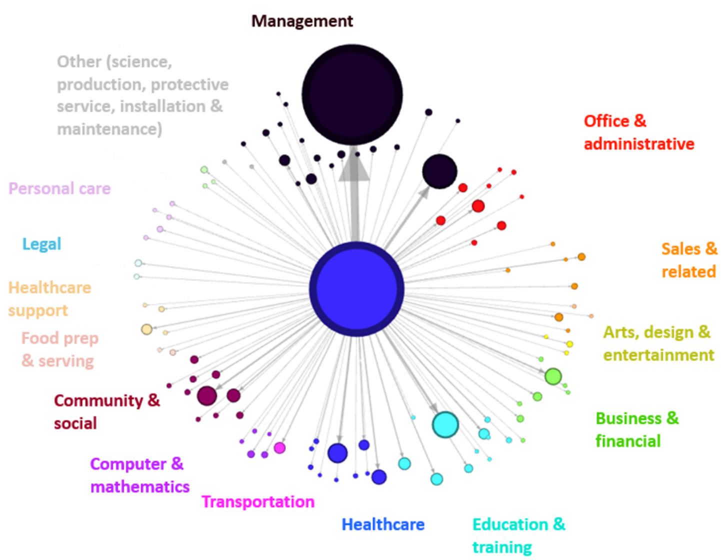
Note: Binned scatter plot of outward occupational mobility by occupation as calculated in our BGT resume data vs. in the CPS. Details of calculation in Appendix B.

Figure A10: Examples of probabilistic labor markets: counter attendants



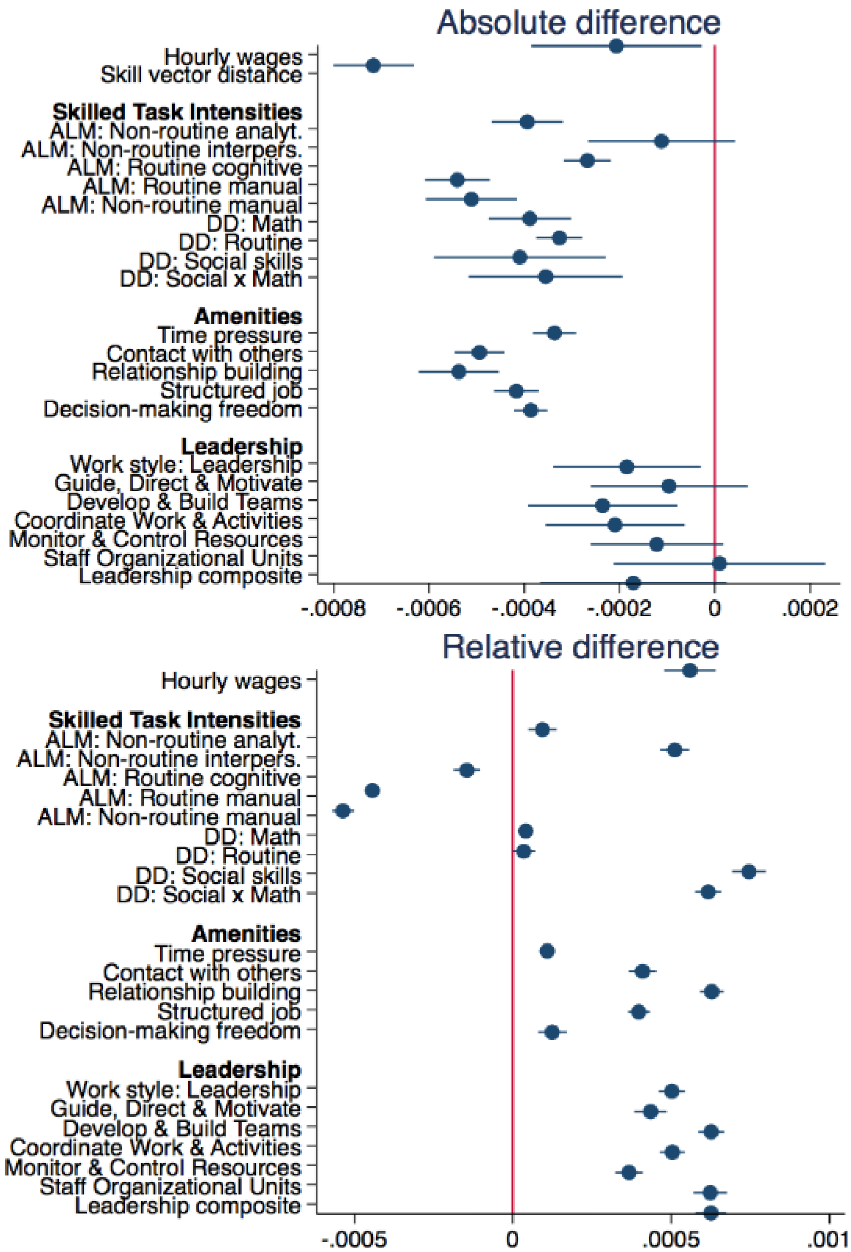
Note: Example visualization of occupational transitions for counter attendants in the food industry. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of counter attendants in the BGT data who switch occupation, who are observed in each destination occupation in the following year. The resume data is discussed in detail in Appendix B.

Figure A11: Examples of probabilistic labor markets: registered nurses



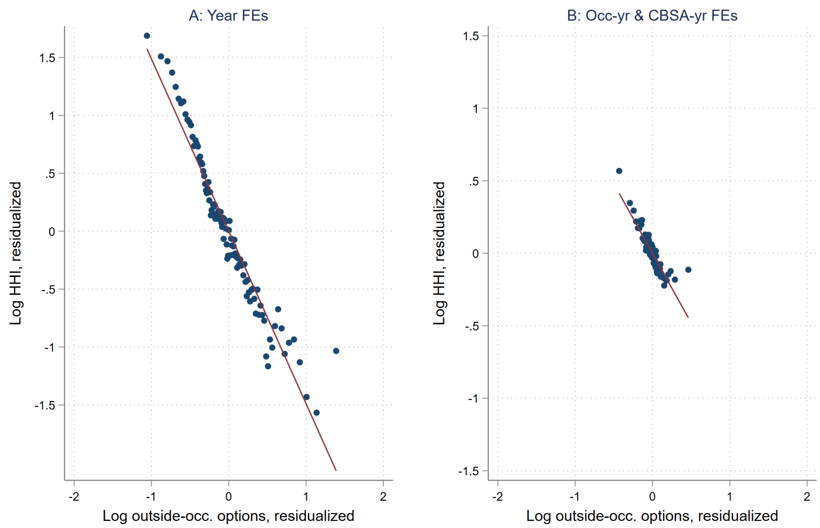
Note: Example visualization of occupational transitions for registered nurses. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of registered nurses in the BGT data who switch occupation, who are observed in each destination occupation in the following year. The resume data is discussed in detail in Appendix B.

Figure A12: Occupational transitions and occupational characteristic similarity



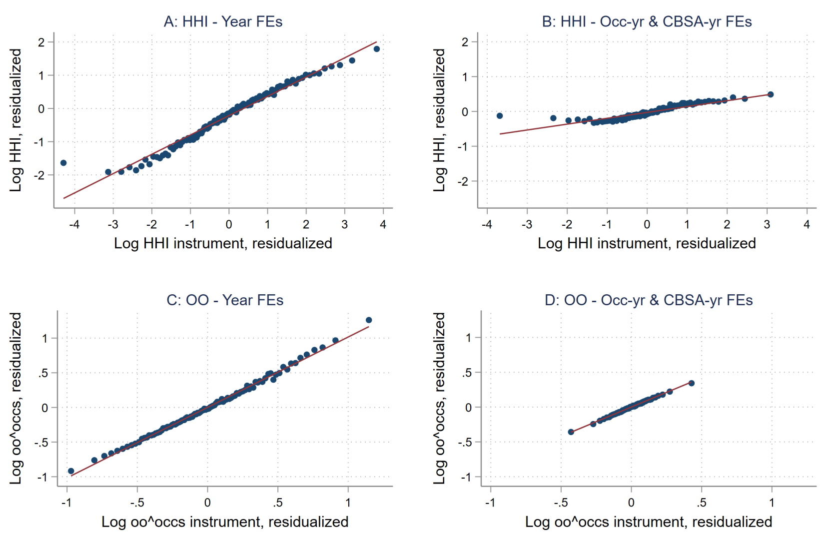
Note: This plot shows coefficients and 95% confidence intervals from regressions of occupation transition shares $\pi_{o \to p}$ on occupational characteristics: $\pi_{o \to p} = \alpha_o + \beta f(X_{occ\ o \to p}) + \gamma f(\Delta w_{o \to p}) + \epsilon_{op}$, where α_o is an occupation o fixed effect. In panel "Absolute difference", $f(\cdot)$ represents the absolute difference in characteristic X between occupation o and p. In panel "Relative difference", $f(\cdot)$ represents the raw difference in characteristic X between occupation o and p. Regressions also include avg. hourly wage differences (except for amenities regressions). Standard errors are clustered by origin occupation. Regressions are described in more detail in Appendix E

Figure A13: Correlations between HHI and outside-occupation option index



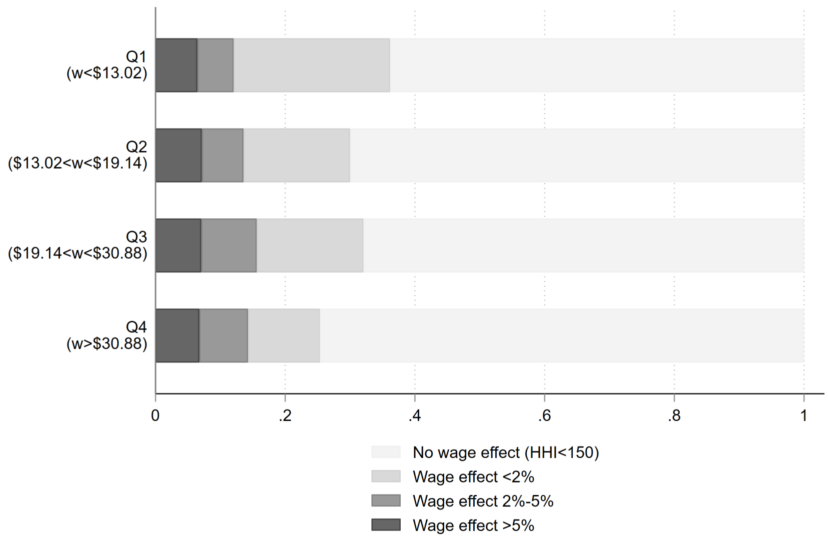
Note: Figure shows binned scatter plots of the employment-weighted relationship between log HHI and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on year FEs in panel A and occupation-year and metro area-year FEs in panel B.

Figure A14: Correlation between instruments and independent variables



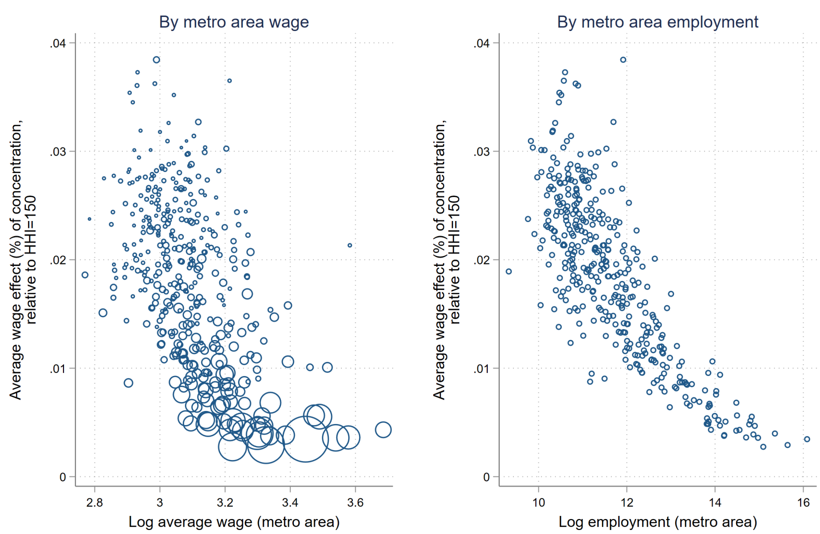
Note: Binned scatter plots of the employment-weighted relationships between the HHI instrument and raw HHI variable (top panel) and outside-occupation option index instrument and raw outside-occupation option variable (bottom panel) for occupation-metro area cells over 2011-2019, with year fixed effects (Panels A and C), and occupation-year and metro area-year fixed effects (Panels B and D).

Figure A15: Average estimated wage effect of employer concentration relative to HHI=150, by hourly wage



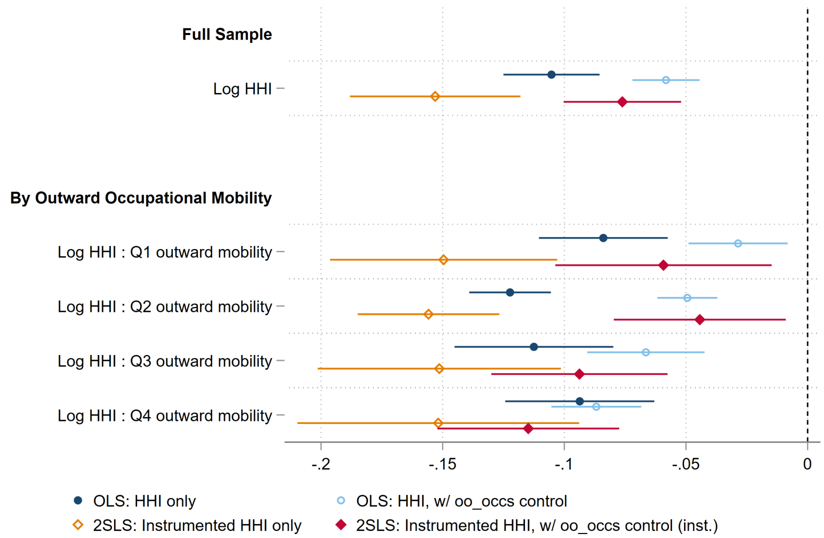
Note: This figure shows the share of workers in each quartile of the 2019 hourly wage distribution who we estimate experienced different degrees of wage suppression as a result of employer concentration, relative to a counterfactual where HHI was 150 (holding all else constant). We estimate the wage effect of employer concentration as described in section 4.

Figure A16: Average estimated wage effect of employer concentration relative to HHI=150, by metro area



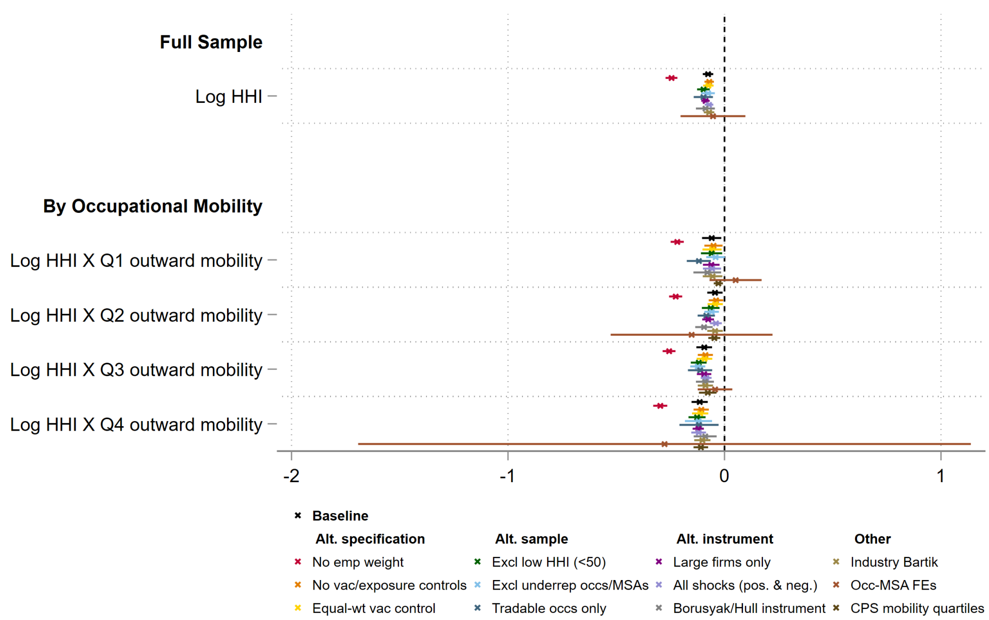
Note: This figure shows the average estimated wage effect of concentration in each metro area relative to a counterfactual HHI of 150 (holding all else constant), plotted against the average hourly wage in that metro area in 2019 (left) and the employment in that metro area in 2019 (right) according to BLS OES data. Bubble size in the left hand graph represents metro area employment in 2019. We estimate the wage effect of employer concentration as described in section 4: we use our coefficient estimates for the effect of the HHI on the wage, by quartile of outward occupational mobility, to calculate a counterfactual wage for each occupation-metro area labor market *if* the HHI had been 150.

Figure A17: Coefficients on employment-HHI regressions



Note: Coefficients on log HHI and 95% confidence intervals from our baseline regressions of occupation-metro area log employment on employer HHI, employment-weighted, 2011-2019, standard errors clustered at metro area level. Specifications are same as those in baseline log wage - HHI regressions in Figure 5.

Figure A18: Employment-HHI regressions: robustness



Note: Coefficients on log HHI and 95% confidence intervals from our baseline 2SLS IV regressions of occupation-metro area log employment on instrumented employer HHI, across the same set of robustness checks as for the wage regressions (shown in Figure 6.

I Appendix: Tables

Table A1: 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 OES (by occ.)	0.12	0.19	0.38	0.83	1.92	4.63	7.51
Occ. share relative to BLS OES (emp.-weight)	0.17	0.21	0.33	0.60	1.18	2.13	3.10
Metro area share relative to BLS OES (by metro area)	0.56	0.65	0.75	0.88	1.05	1.22	1.35
Metro area relative to BLS OES (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 OES' 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 OES 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.

Table A2: Twenty large occupations with lowest leave shares and highest leave shares

Initial occupation	Leave share	Employment (2017)	Obs. (BGT)	Modal new occupation
Dental hygienists	.062	211,600	17,458	Dental assistants
Nurse practitioners	.088	166,280	57,830	Registered nurses
Pharmacists	.09	309,330	121,887	Medical and health services managers
Firefighters	.098	319,860	60,039	Emergency medical technicians and paramedics
Self-enrichment education teachers	.1	238,710	169,369	Teachers and instructors, all other
Physical therapists	.11	225,420	44,314	Medical and health services managers
Postsecondary teachers, all other	.11	189,270	825,879	Managers, all other
Graphic designers	.12	217,170	439,953	Art directors
Emergency medical technicians and paramedics	.12	251,860	111,180	Managers, all other
Fitness trainers and aerobics instructors	.13	280,080	281,903	Managers, all other
Licensed practical and licensed vocational nurses	.13	702,700	254,787	Registered nurses
Lawyers	.13	628,370	667,960	General and operations managers
Registered nurses	.13	2,906,840	1,427,102	Medical and health services managers
Health specialties teachers, postsecondary	.13	194,610	41,963	Medical and health services managers
Physicians and surgeons, all other	.14	355,460	59,630	Medical and health services managers
Heavy and tractor-trailer truck drivers	.14	1,748,140	2,174,486	Managers, all other
Radiologic technologists	.14	201,200	80,347	Magnetic resonance imaging technologists
Hairdressers, hairstylists, and cosmetologists	.14	351,910	107,167	Managers, all other
Coaches and scouts	.14	235,400	533,082	Managers, all other
Chief executives	.15	210,160	1,425,400	General and operations managers
...				
Installation, maintenance, and repair workers, all other	.29	153,850	60,742	Maintenance and repair workers, general
Parts salespersons	.29	252,770	34,038	First-line supervisors of retail sales workers
Billing and posting clerks	.29	476,010	274,963	Bookkeeping, accounting, and auditing clerks
Data entry keyers	.29	180,100	288,523	Customer service representatives
Cashiers	.29	3,564,920	1,753,947	Customer service representatives
Insurance claims and policy processing clerks	.3	277,130	235,763	Claims adjusters, examiners, and investigators
Stock clerks and order fillers	.3	2,046,040	597,137	Laborers and freight, stock, and material movers, hand
Packers and packagers, hand	.3	700,560	101,025	Laborers and freight, stock, and material movers, hand
Cooks, institution and cafeteria	.3	404,120	5,174	Cooks, restaurant
Helpers–production workers	.31	402,140	112,759	Production workers, all other
Sales rep., wholesale & mfg., tech. & scient. products	.31	327,190	198,337	Sales rep., wholesale & mfg., exc. techn. & scient. products
Hosts and hostesses, restaurant, lounge, and coffee shop	.31	414,540	159,098	Waiters and waitresses
Shipping, receiving, and traffic clerks	.31	671,780	318,080	Laborers and freight, stock, and material movers, hand
Loan interviewers and clerks	.32	227,430	234,933	Loan officers
Counter attendants, cafeteria, food concession, and coffee shop	.32	476,940	118,131	Retail salespersons
Bill and account collectors	.32	271,700	310,951	Customer service representatives
Tellers	.32	491,150	468,829	Customer service representatives
Machine setters, operators, and tenders†	.32	154,860	6,805	Production workers, all other
Telemarketers	.36	189,670	47,409	Customer service representatives
Food servers, nonrestaurant	.45	264,630	13,199	Waiters and waitresses

Note: This table shows the twenty large occupations with the lowest and the highest occupation leave shares - defined as share of workers observed in one occupation in one year but not in the following year, divided by the share that leave their job over that period (see Section 2.2) - in the BGT data over 2002-2015, as well as total national employment in that occupation in 2017 from the OES, the number of occupation-year observations in the BGT data ('obs.') and the most popular occupation that workers who leave the initial occupation move to ('modal new occupation'). Large occupations are defined as those with national employment over 150,000 in 2017 (roughly the 75th percentile of occupations when ranked by nationwide employment). † Full occupation title is "Molding, coremaking, and casting machine setters, operators, and tenders, metal and plastic."

Table A3: Forty thickest occupational transition paths for large occupations

Initial occupation	New occupation	Transition share	Employment (2017)	Obs. (BGT data)
Licensed practical and licensed vocational nurses	Registered nurses	.3	702,700	254,787
Nurse practitioners	Registered nurses	.23	166,280	57,830
Construction managers	Managers, all other	.19	263,480	917,349
Sales rep., wholesale & mfg., tech. & scient. products	Sales rep., wholesale & mfg., exc. tech. & scient. products	.19	327,190	198,337
Physicians and surgeons, all other	Medical and health services managers	.19	355,460	59,630
Software developers, systems software	Software developers, applications	.19	394,590	53,322
Legal secretaries	Paralegals and legal assistants	.18	185,870	132,543
Accountants and auditors	Financial managers	.18	1,241,000	1,459,175
Registered nurses	Medical and health services managers	.16	2,906,840	1,427,102
Cost estimators	Managers, all other	.16	210,900	124,646
Human resources specialists	Human resources managers	.16	553,950	2,035,604
Physical therapists	Medical and health services managers	.16	225,420	44,314
Architectural and engineering managers	Managers, all other	.15	179,990	749,670
Computer programmers	Software developers, applications	.15	247,690	533,764
Software developers, applications	Computer occupations, all other	.15	849,230	2,110,229
Computer network architects	Computer occupations, all other	.15	157,830	407,591
Cooks, short order	Cooks, restaurant	.15	174,230	39,906
Cooks, institution and cafeteria	Cooks, restaurant	.14	404,120	5,174
First-line supervisors of construction trades and extraction workers	Construction managers	.14	556,300	186,747
Computer systems analysts	Computer occupations, all other	.14	581,960	1,152,614
Sales rep., wholesale & mfg., exc. tech. & scient. products	Sales managers	.13	1,391,400	4,377,654
Light truck or delivery services drivers	Heavy and tractor-trailer truck drivers	.13	877,670	226,349
Computer occupations, all other	Managers, all other	.13	315,830	3,515,188
Health specialties teachers, postsecondary	Medical and health services managers	.13	194,610	41,963
Meat, poultry, and fish cutters and trimmers	Heavy and tractor-trailer truck drivers	.13	153,280	2,383
Sales rep., wholesale & mfg., tech. & scient. products	Sales managers	.13	327,190	198,337
Operating engineers and other construction equipment operators	Heavy and tractor-trailer truck drivers	.13	365,300	55,317
Sales managers	Sales rep., wholesale & mfg., exc. tech. & scient. products	.13	371,410	3,471,904
Health specialties teachers, postsecondary	Registered nurses	.13	194,610	41,963
Industrial engineers	Engineers, all other	.13	265,520	171,358
Network and computer systems administrators	Computer occupations, all other	.13	375,040	1,103,700
Industrial production managers	Managers, all other	.12	171,520	750,609
Computer network support specialists	Computer user support specialists	.12	186,230	237,766
Software developers, systems software	Computer occupations, all other	.12	394,590	53,322
Financial analysts	Financial managers	.12	294,110	664,903
Legal secretaries	Secretaries and admin. assistants, except legal, medical, & exec.	.12	185,870	132,543
Mechanical engineers	Architectural and engineering managers	.12	291,290	408,178
Food batchmakers	Industrial production managers	.12	151,950	12,729
Licensed practical and licensed vocational nurses	Medical and health services managers	.11	702,700	254,787
Food batchmakers	Heavy and tractor-trailer truck drivers	.11	151,950	12,729

Note: This table shows the ‘thickest’ occupational transition paths from large occupations (defined as those with national employment greater than 150,000 in 2017). The transition share from occupation o to occupation p is defined as the share of all occupation leavers from the initial occupation o who move into that particular new occupation p (as in Section 2.2). Only occupations with at least 500 observations in the BGT data and 2017 OES employment data are shown.

Table A4: Large employer-occupation pairs: examples from BGT vacancy data

Employer	Occupation	Total job postings, 2017	YoY growth, 2017-18 (%)
Lowe's Companies, Inc.	Retail salespersons	73,899	-18
HCA Healthcare	Registered nurses	37,057	-9
Rent-A-Center	Sales representatives, wholesale and manufacturing*	6,095	-9
Dollar General	First-line supervisors of retail sales workers	29,228	-1
Deloitte	Computer occupations, all other	18,688	20
Roehl Transport	Heavy and tractor-trailer truck drivers	22,809	-10
Amazon	Software developers, applications	9,094	0
Anthem Blue Cross Blue Shield	Managers, all other	20,086	113
Lowe's Companies, Inc.	Customer service representatives	11,261	-34
Anthem Blue Cross Blue Shield	Medical and health services managers	20,828	46
Chipotle Mexican Grill	Combined food preparation and serving workers**	25,674	-57
Accenture	Sales managers	6,187	-49
Pizza Hut	Food service managers	10,433	121
Lowe's Companies, Inc.	Stock clerks and order fillers	17,155	8
Anthem Blue Cross Blue Shield	General and operations managers	9,321	14
Edward Jones	Secretaries and administrative assistants***	2,846	10
Anthem Blue Cross Blue Shield	Management analysts	32,524	27
McDonald's	Maintenance and repair workers, general	4,006	25
United Parcel Service	Laborers and freight, stock, and material movers, hand	9,880	30
Deloitte	Human resources specialists	2,333	-11

Note: Table shows examples of large employer-occupation pairs in the BGT vacancy data, total postings in 2017, and year-on-year postings growth 2017-18. The table is constructed by taking the twenty largest occupations in the BGT vacancy data in terms of total postings 2011-2019, then listing the largest employer within each occupation (in terms of total postings 2011-2019). *: Full occupation title is *Sales representatives, wholesale and manufacturing, except technical and scientific products.* **: Full occupation title is *Combined food preparation and serving workers, including fast food.* ***: Full occupation title is *Secretaries and administrative assistants, except legal, medical, and executive.*

Table A5: First stage regressions: HHI instrument

Dependent variable: log vacancy HHI (segmented by quartile of occ mobility in cols (b)-(e))					
	Full sample	By quartile of occ mobility			
		Q1	Q2	Q3	Q4
	(a)	(b)	(c)	(d)	(e)
Log HHI instrument	0.143***	0.116***	0.151***	0.150***	0.138***
	(0.010)	(0.007)	(0.011)	(0.015)	(0.013)
Log outside-occ. options instrument	-0.856***	-0.538***	-0.890***	-0.831***	-1.039***
	(0.076)	(0.157)	(0.111)	(0.074)	(0.102)
Observations	579,668	148,046	148,332	163,472	119,779
Controls					
Vacancy growth	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y
Exposure	Y	Y	Y	Y	Y
Fixed effects					
Occ-year	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y

Note: In column (a) we run a first-stage regression for our HHI instrument. In columns (b) through (e) we run separate first-stage regressions for our HHI instrument, segmenting our data into four quartiles by outward occupational mobility (the occupation "leave share" as defined in Section 2.2). Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions have occupation-year and metro area-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A6: First-stage regressions: outside-occ. options instrument

Dependent variable: log outside-occ. options (segmented by quartile of occ mobility in cols (b)-(e))					
	Full sample	By quartile of occ mobility			
		Q1	Q2	Q3	Q4
	(a)	(b)	(c)	(d)	(e)
Log outside-occ. options instrument	0.838***	0.770***	0.866***	0.844***	0.855***
	(0.029)	(0.037)	(0.031)	(0.030)	(0.028)
Log HHI instrument	-0.001	-0.001	-0.001	-0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	579,668	148,046	148,332	163,472	119,779
Controls					
Vacancy growth	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y
Exposure	Y	Y	Y	Y	Y
Fixed effects					
Occ-year	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y

Note: In column (a) we run a first-stage regression for our outside-occupation option index instrument. In columns (b) through (e) we run separate first-stage regressions for our outside-occupation option index instrument, segmenting our data into four quartiles by outward occupational mobility (the occupation "leave share" as defined in Section 2.2). Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions have occupation-year and metro area-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A7: Regression of wage on HHI and outside-occupation options: robustness

Dependent variable:	Log wage					
	(a) No emp. weight	(b) No controls	(c) Equal-wt. control	(d) Drop low HHI	(e) Drop low rep.	(f) Tradable occs
Log HHI, instrumented	-0.010***	-0.015***	-0.015***	-0.018***	-0.015***	-0.018***
	(0.003)	(0.004)	(0.004)	(0.007)	(0.006)	(0.005)
Log outside-occ. options, instrumented	0.090***	0.109***	0.109***	0.107***	0.122***	0.144***
	(0.008)	(0.022)	(0.023)	(0.020)	(0.022)	(0.017)
Observations	579,668	579,668	579,668	568,652	305,522	285,269
Controls						
Vacancy growth	Y		Y	Y	Y	Y
Predicted vac. growth	Y		Y	Y	Y	Y
Exposure	Y		Y	Y	Y	Y
Equal-wt. vac. growth			Y			
Fixed effects						
Occ-year	Y	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y	Y

Note: These represent robustness checks for our baseline regression specification (reported in column (d) of Table 3). Each column is a different robustness check. All columns are 2SLS IV regressions estimated using our HHI and outside-occupation option instruments. Column (a) runs the regression without employment weights. Column (b) excludes our controls for local actual vacancy growth, predicted vacancy growth, and exposure to large national firms. Column (c) includes an additional control for equal-weighted vacancy growth of local firms in the relevant occupation. Column (d) drops any occupation-metro area cells with HHI less than 50. Column (e) drops any occupations or metro areas with a represented-ness in the bottom third (across occupations or metro areas respectively) in our BGT vacancy data. Represented-ness by occupation (/metro area) in the BGT vacancy data is calculated as the share of all vacancies accounted for by a given occupation (/metro area) in the BGT vacancy data in a given year, divided by the share of employment accounted for by a given occupation (metro area) in the BLS OES in that same year, averaged over 2011–2019. Column (f) includes only occupations which provide a tradable product, defined as those in production, extraction, construction, agriculture, forestry, or fishing, as well as occupations which are classified as having some remote work possibility by Dingel and Neiman (2020). All specifications feature occupation-year and metro area-year fixed effects. Other regression info: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions are employment-weighted, unless otherwise specified. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A8: Regression of wage on HHI and outside-occupation options: robustness (2)

Dependent variable:	Log wage				
	(a) Alt. HHI inst. 1	(b) Alt. HHI inst. 2	(c) Alt. HHI inst. 3	(d) Industry Bartik	(e) Occ-MSA FEs
Log HHI, instrumented	-0.015*** (0.003)	-0.010** (0.004)	-0.029*** (0.008)	-0.014*** (0.004)	-0.007* (0.004)
Log outside-occ. options, instrumented	0.114*** (0.022)	0.112*** (0.022)	0.116*** (0.026)	0.113*** (0.023)	0.080* (0.046)
Observations	490,986	494,378	579,668	524,167	564,382
Controls					
Vacancy growth	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y
Exposure	Y	Y	Y	Y	Y
Industry Bartik				Y	
Fixed effects					
Occ-year	Y	Y	Y	Y	
Metro-year	Y	Y	Y	Y	
Occ-metro					Y
Year					Y

Notes: These represent robustness checks for our baseline regression specification (reported in column (d) of Table 3). Each column is a different robustness check. All columns are 2SLS IV regressions estimated using our HHI and outside-occupation option instruments. Columns (a), (b), and (c) use alternative formulations of our HHI instrument, with column (a) instrumenting only with national hiring growth for large firms (with vacancies in at least 5 metro areas in that occupation), column (b) using both national hiring growth and declines, and column (c) using the formulation proposed by Borusyak and Hull (2020) for logged shift-share IV instruments. Column (c) also uses the formulation proposed by Borusyak and Hull (2020) for the outside-occupation option instrument. Column (d) introduces a control for an industry Bartik shock to control for correlated industry shocks across occupation-metro area cells. Column (e) has occupation-metro area and year fixed effects instead of occupation-year and metro-year fixed effects. Other regression info: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019 inclusive. All specifications feature occupation-year and metro area-year fixed effects, unless otherwise specified. All regressions are employment-weighted. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A9: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness

Dependent variable:	Log wage					
	(a) No emp. weight	(b) No controls	(c) Equal-wt. control	(d) Drop low HHI	(e) Drop low rep.	(f) Tradable occs
Log HHI	-0.026***	-0.042***	-0.043***	-0.045***	-0.045***	-0.047***
X Q1 outward mobility	(0.004)	(0.007)	(0.008)	(0.010)	(0.009)	(0.010)
Log HHI	-0.010***	-0.015***	-0.015***	-0.016**	-0.012**	-0.019***
X Q2 occ mobility	(0.003)	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)
Log HHI	-0.002	-0.002	-0.003	-0.008	0.004	-0.006
X Q3 occ mobility	(0.003)	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)
Log HHI	0.001	-0.000	-0.001	-0.006	-0.000	0.005
X Q4 occ mobility	(0.003)	(0.005)	(0.006)	(0.008)	(0.006)	(0.006)
Log outside-occ options	0.050***	0.077***	0.076***	0.076***	0.086***	0.113***
X Q1 occ mobility	(0.010)	(0.024)	(0.024)	(0.022)	(0.025)	(0.027)
Log outside-occ options	0.096***	0.131***	0.130***	0.136***	0.154***	0.155***
X Q2 occ mobility	(0.009)	(0.020)	(0.020)	(0.016)	(0.023)	(0.016)
Log outside-occ options	0.105***	0.111***	0.111***	0.104***	0.127***	0.149***
X Q3 occ mobility	(0.009)	(0.028)	(0.028)	(0.024)	(0.023)	(0.016)
Log outside-occ options	0.112***	0.099**	0.098**	0.091**	0.162***	0.181***
X Q4 occ mobility	(0.009)	(0.043)	(0.043)	(0.041)	(0.023)	(0.028)
Observations	579,668	579,668	579,668	568,652	305,522	285,269
Controls						
Vacancy growth	Y		Y	Y	Y	Y
Predicted vac. growth	Y		Y	Y	Y	Y
Exposure	Y		Y	Y	Y	Y
Equal-wt. vac. growth			Y			
Fixed effects						
Occ-year	Y	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y	Y

Note: This table repeats the robustness checks in Table A7, but allowing the coefficient on the HHI and outside-occupation option index to vary by quartile of outward occupational mobility (the occupation "leave share" defined as in Section 2.2). Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A10: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness (2)

Dependent variable:	Log wage					
	(a) Alt. HHI inst. 1	(b) Alt. HHI inst. 2	(c) Alt. HHI inst. 3	(d) Industry Bartik	(e) Occ-MSA FEs	(f) CPS mobility
Log HHI	-0.039***	-0.038***	-0.050***	-0.038***	-0.035**	-0.030***
X Q1 occ mobility	(0.008)	(0.007)	(0.012)	(0.007)	(0.016)	(0.006)
Log HHI	-0.011**	-0.008**	-0.020***	-0.015***	0.029	-0.006
X Q2 occ mobility	(0.005)	(0.004)	(0.007)	(0.004)	(0.055)	(0.004)
Log HHI	-0.003	0.000	-0.022***	-0.003	0.008	0.003
X Q3 occ mobility	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.005)
Log HHI	-0.009*	0.001	-0.016*	-0.002	-0.019	-0.010
X Q4 occ mobility	(0.005)	(0.006)	(0.009)	(0.006)	(0.019)	(0.007)
Log outside-occ options	0.083***	0.075***	0.100***	0.084***	-0.229	0.081***
X Q1 occ mobility	(0.022)	(0.024)	(0.024)	(0.022)	(0.441)	(0.017)
Log outside-occ options	0.142***	0.134***	0.146***	0.134***	0.408	0.110***
X Q2 occ mobility	(0.017)	(0.020)	(0.023)	(0.020)	(0.828)	(0.030)
Log outside-occ options	0.113***	0.116***	0.106***	0.111***	0.151***	0.126***
X Q3 occ mobility	(0.027)	(0.026)	(0.029)	(0.030)	(0.056)	(0.032)
Log outside-occ options	0.088*	0.102**	0.102**	0.095**	-0.027	0.035
X Q4 occ mobility	(0.047)	(0.042)	(0.040)	(0.045)	(0.347)	(0.043)
Observations	490,986	494,378	579,668	524,167	564,382	579,668
Controls						
Vacancy growth	Y	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y	Y
Exposure	Y	Y	Y	Y	Y	Y
Industry Bartik				Y		
Fixed effects						
Occ-year	Y	Y	Y	Y		Y
Metro-year	Y	Y	Y	Y		Y
Occ-metro					Y	
Year					Y	

Notes: This table repeats the robustness checks in Table A8, but allowing the coefficient on the HHI and outside-occupation option index to vary by quartile of outward occupational mobility (the occupation "leave share" defined as in Section 2.2). It also includes an additional robustness check in column (f), which allows the coefficients to vary by the occupation's quartile of outward mobility as calculated in the CPS rather than in the BGT resume data. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A11: Regression of wage on outside-occupation options: 1999–2019

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: OLS regressions				
oo^{occs}	0.443***	0.088***	0.130***	0.028***
	(0.014)	(0.014)	(0.009)	(0.010)
Panel B: 2SLS IV regressions				
oo^{occs} , instrumented	0.748***	0.099***	0.106***	0.030***
	(0.076)	(0.013)	(0.015)	(0.012)
Observations	2,275,358	2,275,164	2,275,164	2,262,164
Fixed effects	Year	Occ-Year, Metro area	Occ-Year, Metro area-Year	Occ-Year, Occ-Metro area

Notes: This table repeats our baseline regressions with the outside-occupation option index only, over a longer period (1999-2019). Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, weighted by average employment in the occ-metro area over 1999-2019, for all observations with available data. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A12: Regressions of wage on outside-occupation option index: aggregated occupation codes, with different combinations of fixed effects

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: Minor SOC Group (3-digit) regressions:				
OLS: oo^{occs}	0.383***	0.091***	0.106***	0.071***
	(0.013)	(0.007)	(0.011)	(0.008)
IV: oo^{occs}, instrumented	0.400***	0.113***	0.105***	0.125***
	(0.018)	(0.012)	(0.015)	(0.014)
Observations	486,487	486,481	486,481	485,808
Panel B: Major SOC Group (2-digit) regressions:				
OLS: oo^{occs}	0.194***	0.081***	0.002	0.080***
	(0.015)	(0.009)	(0.023)	(0.008)
IV: oo^{occs}, instrumented	0.136***	0.079***	0.060**	0.327***
	(0.023)	(0.028)	(0.029)	(0.113)
Observations	137,650	137,650	137,650	137,609
Fixed effects	Year	Occ-Year Metro area	Metro area-Year Occ-Year	Occ-Year Occ-Metro area

Notes: This table reports 2SLS IV regressions of the wage on outside-occupation option index with outside options defined at the level of 3-digit or 2-digit occupations (rather than SOC 6-digit). Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses: $*p < .1$,$**p < .05$,$***\ p < .01$. Units of observation are 2-digit or 3-digit SOC by metro area by year, for all observations with available data over 1999–2016 inclusive. As noted in the paper, 'metro areas' refers to CBSAs (metropolitan and micropolitan statistical areas) or NECTAs (New England city and town areas). Each cell reports the coefficient for the variable of interest in one specification, with included fixed effects held constant within each column.

Table A13: Counterfactual wage effects of setting HHI to 150,
excluding occupations with a represented-ness<0.5 in our BGT vacancy data

		0< HHI <200	200< HHI <500	500< HHI <1,500	1,500< HHI <2,500	2,500< HHI <10,000
Lowest mobility	Avg. wage effect	0	2.6%	7.4%	11.4%	15.1%
	Employment(m)	7.3	5.7	3.7	.59	.43
Q2 mobility	Avg. wage effect	0	0.8%	2.5%	3.9%	5.2%
	Employment(m)	12.4	3.9	1.7	.34	.32
Q3 mobility	Avg. wage effect	0	0	0	0	0
	Employment(m)	14.6	5.6	1.9	.31	.27
Q4 mobility	Avg. wage effect	0	0	0	0	0
	Employment(m)	5.2	2.5	1.7	.31	.24

Notes: This table repeats the analysis in Table 5, but *dropping* any occupations with an average represented-ness in our BGT vacancy data of less than 0.5 (roughly the bottom third of occupations). This is because our concentration data on these occupations might be a substantial overestimate of the true degree of employer concentration, if our data is disproportionately sampled from large employers for these occupations.

Table A14: Represented-ness in BGT data for twenty-five occupations with most people affected by employer concentration (based on a predicted occupation-metro area wage effect of 2% or greater)

Occupation	Represented-ness in occupation in BGT vacancy data
Registered nurses	2.1
Security guards	.6
Nursing assistants	.68
Hairdressers, hairstylists, and cosmetologists	.63
Pharmacy technicians	.9
Pharmacists	1.2
Licensed practical and licensed vocational nurses	1.3
Medical assistants	.83
Fitness trainers and aerobics instructors	.63
Heavy and tractor-trailer truck drivers	2.1
Emergency medical technicians and paramedics	.59
Radiologic technologists	.76
Medical and clinical laboratory technologists	1.1
Lawyers	.87
Phlebotomists	1.6
Nurse practitioners	4.1
Aircraft mechanics and service technicians	.72
Management analysts	1.9
Software developers, applications	5.6
Massage therapists	1
Maids and housekeeping cleaners	.64
Light truck or delivery services drivers	.56
Respiratory therapists	1.2
General and operations managers	.54
Secretaries and administrative assistants (except legal, medical, and executive)	.78

Notes: This table lists the degree of represented-ness of each of these twenty-five occupations in the BGT vacancy data. Represented-ness is defined as the occupation's share of vacancy postings in the BGT database relative to the occupation's share of total employment (as per BLS OES). The twenty-five occupations in this table correspond to the occupations with the highest number of people affected by employer concentration, as listed in Table 6. The more underrepresented an occupation is in the BGT vacancy data, the more likely we are overestimating the degree of employer concentration in these occupations and therefore overestimating the effect of concentration. On the other hand, in better-represented occupations we might be more confident that we are accurately prioritizing these occupations.

Table A15: What would happen if we don't consider occupational mobility? Twenty-five occupations with most people estimated to experience a $\geq$ 2% wage effect of employer concentration, *without* taking into account occupational mobility

Occupation	**Not considering mobility** *(applying average estimated HHI effect)* No. of workers with 2% wage effect	**Considering mobility** *(our baseline: applying HHI effect by quartile of occupational mobility)* No. of workers with 2% wage effect
Security guards	599,700	1,034,590
Registered nurses	452,040	1,373,090
Nursing assistants	369,410	751,870
Tellers	297,880	0
Hairdressers, hairstylists, and cosmetologists	264,850	353,710
Stock clerks and order fillers	240,300	0
Securities, commodities, and financial services sales agents	222,330	0
Medical assistants	217,870	217,870
Pharmacy technicians	213,800	346,480
Retail salespersons	170,630	0
Pharmacists	139,330	250,130
Cooks, restaurant	108,310	0
Butchers and meat cutters	106,390	0
Licensed practical and licensed vocational nurses	99,060	219,010
First-line supervisors of food preparation and serving workers	96,100	0
Radiologic technologists	91,990	154,050
Maids and housekeeping cleaners	91,580	91,580
Emergency medical technicians and paramedics	90,580	162,720
Light truck or delivery services drivers	89,050	89,050
Medical secretaries	86,050	0
Phlebotomists	85,400	110,270
Massage therapists	85,330	91,770
General and operations managers	83,750	83,750
Secretaries and administrative assistants*	81,060	81,060
Tire repairers and changers	77,340	0

Notes: This table illustrates the difference it makes to take into account occupational mobility when estimating the effects of employer concentration on workers. The first column lists the twenty-five occupations which would be considered to have the most people affected by employer concentration, *if* we estimated the wage effect of employer concentration simply as $(log(HHI) - log(150)) \cdot -0.015$, where *-0.015* is the average wage effect of employer concentration in our data. The second column lists the number of workers in that occupation with an estimated wage effect of 2% or greater under that method. The third column lists the number of workers in that occupation with an estimated wage effect of 2% or greater *when taking occupational mobility into account*, i.e. using our methodology as outlined in the paper and in Table 6. As in Table 6, this table excludes occupations that are very heavily public sector (pre-school teachers), and also excludes occupations that are very under-represented in the BGT vacancy data relative to overall employment (with a cutoff with represented-ness<0.5, or around the 33rd percentile). *: Full title is *Secretaries and administrative assistants, except legal, medical, and executive.*

Table A16: Selected Challenges to the Empirical Approach and How They Are Addressed

This table summarizes how our empirical approach and robustness checks address some of the most common concerns around the identification of the causal effects of employer concentration and outside options on wages. For full details and a comprehensive list of robustness checks, please consult the relevant sections of the paper.

Concern	How it is addressed
Empirical approach	
1. Didn't critiques of the structure-conduct-performance paradigm show that regressions with concentration measures like HHI are not valid?	Older critiques of the structure-conduct-performance paradigm (e.g. Schmalensee, 1989) and more recent critiques of the empirical literature which finds a negative correlation between local employer concentration and wages – including Berry, Gaynor and Scott Morton (2019) and Rose (2019), – focus one one or more of (1) conceptual clarity, (2) market definition, and (3) endogeneity. The conceptual concern is that a simple regression of market concentration on an outcome like wages is not well-defined, because there is no single theoretical channel by which concentration would affect wages - it depends on the circumstances. Following Jarosch, Nimczik and Sorkin (2019) and others, we take the view that in our case, 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. The market definition concern is that there is no appropriate definition of a market on which a meaningful concentration index can be calculated; we attempt to address this concern as discussed above. The endogeneity concern is that employer concentration is determined by, as well as affecting, local economic conditions. Rose (2019) argues that empirical strategies attempting to identify a causal effect of employer concentration on wages must isolate the effect of employer concentration from changes in labor demand; our identification strategy attempts to do exactly that: the reason why we develop our IV strategy is to respond to the endogeneity concern that was raised by these critiques.

Table A16: (continued)

Concern	How it is addressed
2. How does your measure of outside options compare to that of Danieli & Caldwell (2018)?	Caldwell and Danieli (2018)'s index of the value of workers' outside options captures the degree to which workers of a given type are distributed across different job types. Intuitively, a worker has more outside options if other workers of the same type are distributed across a wider range of types of job, where job type is captured empirically by a combination of establishment characteristics, industry and occupation task requirements, and geographic location. (Worker type is captured with gender, age, secondary education qualification, citizenship status, and occupation at labor market entry.) Their index reflects an equilibrium distribution of workers across job types and space. This assumes that workers' outside option set is best captured by the outside option set of others who look similar to them in terms of underlying characteristics (rather than current occupation), and that the outside option set of these other workers is in turn best captured by the types of firms and task requirements in the jobs worked in by these other workers. In contrast, our outside option index captures the degree to which workers in the same occupation move to different occupations. Our primary assumption is therefore that workers' outside option set (outside their occupation) is best captured by the outside option set of others in the same occupation, which in turn is best captured by the revealed occupational switches of others in the same occupation.
3. To account for mobility, could we just define a labor market to be a cluster of more than one occupation?	While this kind of clustering may more closely approximate a worker's labor market than administrative classifications, such as SOC codes, it still represents a form of "binary" market definition that does not allow for differential weighting of a worker's options but rather limits the weights to 0% (out of the market) and 100% (in the market).
4. Instead of looking at both occupational mobility and concentration, could you construct a better HHI that combines both of these dimensions into one index that better measures concentration in probabilistic labor markets?	While we think the approach considering both mobility and concentration that we outline in this paper is more transparent and model-agnostic, some contexts (e.g. legal thresholds for antitrust purposes) may require a single index of concentration. Our heterogeneity results suggest that monopsony power has the highest negative effect on wages in low mobility occupations. Therefore, an index that represents the product or interaction of an occupational mobility measure (e.g. our *"leave share"*) and a conventional concentration measure like the HHI may be more suitable for those purposes. Our paper's contribution is to establish that these factors should be considered jointly However, the exact functional form and relevant thresholds will vary based on the context and usage, which is why we encourage other researchers and practitioners to adapt this general approach to their specific needs and further study the properties of such a combined index.

Table A16: (continued)

Concern	How it is addressed
Identification strategy	
5. What if an employer's expansion in *other* cities predicts an employment *contraction* in the focal city?	While this is certainly possible in individual cases, on average this is not true in our data: a regression of $g_{j,o,k,t}$ on $g_{j,o,\cancel{k},t}$ gives a coefficient of 0.02 (standard error 0.0001); when restricting to firms with vacancies in 5 or more metro areas in the same occupation, the coefficient is 0.35 (standard error 0.0006) (where $g_{j,o,k,t}$ is the growth in vacancy postings of firm j in occupation o and metro area k between years $t-1$ and t, and where $g_{j,o,\cancel{k},t}$ is the growth in vacancy postings of firm j in occupation o in all metro areas nationwide *except* k between years $t-1$ and t).
6. Does the fact that some cities have booming local economies while others do not - and that this affects wages - represent an issue for your identification approach?	The inclusion of metro-by-year fixed effects means that differences in metro area-level trends or national productivity of different occupations do not represent an issue for our identification strategy. Additionally, we show that our results are robust to controlling for common exposure to national industry shocks at the metro level (see Appendix Table A7).
7. What if the expansion of a national firm in a local labor market changes wages through a simple labor demand channel rather than through its effect on concentration?	We address this concern in two ways. (1) Mathematically, the effect of a large firm's growth on local labor market concentration is quadratic, whereas the effect of a large firm's growth on local labor demand is likely linear (i.e. a 10% increase in local vacancies by national firms increases labor demand twice as much as a 5% increase). Therefore, we should be able to control for the linear effect of higher labor demand by directly including the expected change in local vacancy postings in the regressions. Controlling for the exposure of local labor demand to national employment trends directly to prevent it from confounding a nonlinear IV is similar to the "double Bartik" approach in Chodorow-Reich and Wieland (2020). While the assumption that these linear terms capture the relevant demand effects is relatively strong, note that their inclusion does not affect our baseline coefficient estimates (Appendix Tables A7 and A9). (2) The most likely direction of correlation between an increase in concentration due to an increase in hiring by national firms and overall labor demand is positive. Note that, if the labor supply curve slopes upward, this means that our instrument should be positively correlated with wages, which would bias the negative estimates of the effect of concentration on wages towards zero. That is, if our attempts to control for labor demand are insufficient, this most likely means that our estimates are *too conservative*.
8. Does the fact that some occupations have recently seen an increase in labor demand and wages - and that this might correlate with increasing concentration in particular industries, such as the tech sector - represent an issue?	The inclusion of occupation-by-year fixed effects means that differences in occupation-level trends at the national level do not represent an issue for our identification strategy. Our estimates are identified off local variation at the occupation level due to local shocks that differ from those experienced by the average labor market for that occupation.

Table A16: (continued)

Concern	How it is addressed
9. Does your IV approach require exposure of local labor markets to large firms to be randomly assigned?	No. In the conventional framework of thinking of shift-share shocks as consisting of an "exposure" (here, the initial local employment concentration of different firms) and a "shifter" (here: national time-varying shocks to hiring by particular large firms), our causal effects are identified off the exogenous variation in the shifter, conditional on the included controls and fixed effects. We discuss conditions for identification further in Appendix F, following the approach to shift-share IVs of Borusyak et al. (2022).
10. Could your outside options results be driven by the fact that connected local occupations are affected by common *industry* shocks?	We explicitly try to control for this confounding factor by constructing a shock that reflects the common predicted impact of national industry wage trends for each occupation-metro area-year cell, similar to an approach to controlling for common trends in Chodorow-Reich and Wieland (2020). When we control for these common industry shocks, the coeffcient on the outside options index barely changes, suggesting that common industry variation is likely not a major driver of our results. See Appendix F and Tables A8 and A10 for details.

Appendix References

Acemoglu, Daron and David Autor, "Skills, tasks and technologies: Implications for employment and earnings," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1043–1171.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

Azar, José A, Ioana Marinescu, Marshall I Steinbaum, and Bledi Taska, "Concentration in US Labor Markets: Evidence From Online Vacancy Data," *Labour Economics*, 2020, *66*, 101886.

Bahar, Dany, "IVREG2HDFE: Stata module to estimate an Instrumental Variable Linear Regression Model with two High Dimensional Fixed Effects," *Statistical Software Components S457841, Boston College Department of Economics*, 2014.

Bajari, Patrick, C Lanier Benkard, and Jonathan Levin, "Estimating dynamic models of imperfect competition," *Econometrica*, 2007, *75* (5), 1331–1370.

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

Borusyak, Kirill and Peter Hull, "Non-random exposure to exogenous shocks: Theory and applications," Technical Report, National Bureau of Economic Research 2020.

—, —, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Caldwell, Sydnee and Oren Danieli, "Outside Options in the Labor Market," 2018. Working paper.

Carnevale, Anthony P., Tamara Jayasundera, and Dmitri Repnikov, "Understanding Online Job Ads Data: A Technical Report," 2014.

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

Correia, Sergio, "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator," Technical Report 2016. Working Paper.

Deming, David J, "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

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

Gabaix, Xavier and Ralph Koijen, "Granular Instrumental Variables," 2020. Working paper.

Gathmann, Christina and Uta Schönberg, "How general is human capital? A task-based approach," *Journal of Labor Economics*, 2010, *28* (1), 1–49.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Guimaraes, Paulo and Pedro Portugal, "A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects," *Stata Journal*, 2010, *10* (4).

Hazell, Jonathon and Bledi Taska, "Downward Rigidity in the Wage for New Hires," 2019.

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

Jann, Ben, "COEFPLOT: Stata module to plot regression coefficients and other results," *Statistical Software Components S457686, Boston College Department of Economics*, 2013.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin, "Granular Search, Market Structure, and Wages," 2019. Working paper.

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

— and —, "Occupational specificity of human capital," *International Economic Review*, 2009, *50* (1), 63–115.

— and —, "A Cautionary Note on Using (March) Current Population Survey and Panel Study of Income Dynamics Data to Study Worker Mobility," *Macroeconomic Dynamics*, 2013, *17*, 172–194.

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

Macaluso, Claudia, "Skill remoteness and post-layoff labor market outcomes," 2019.

McFadden, D, "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, 1974, pp. 105–142.

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

Rose, Nancy, "Thinking Through Anticompetitive Effects of Mergers on Workers," Technical Report, American Bar Association 2019 Antitrust Spring Meeting 2019.

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

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

Stepner, Michael, "BINSCATTER: Stata module to generate binned scatterplots," *Statistical Software Components S457709, Boston College Department of Economics*, 2013.

Sullivan, Paul, "Empirical Evidence on Occupation and Industry Specific Human Capital," *Labour Economics*, 2010, *17* (3), 567–580.

Weil, David, *The Fissured Workplace* 2014.

Xu, Ming, "Understanding the Decline in Occupational Mobility," *Working Paper*, 2018.