Employer Concentration and Outside Options

Gregor Schubert, Anna Stansbury, and Bledi Taska *

February 5, 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 2.6% on average and by 7.3% for workers in the lowest quartile of outward occupational mobility. We also find meaningful effects of changes in the value of outside-occupation job options on wages. Our findings imply that policymakers should take employer concentration seriously, but that measures of employer concentration – typically calculated for single occupations – should be considered alongside occupational mobility and the availability of outside-occupation options.

*Corresponding author: G. Schubert. Address: Finance Group, UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095. Email: gregor.schubert@anderson.ucla.edu. Other author affiliations: 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, 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, and Queen Mary University London. 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 for whom does it matter the most?* We estimate the effects of employer concentration on average hourly wages across over 100,000 US SOC 6-digit occupation-by-metropolitan-area labor markets over 2011–2019, following 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, forthcoming; Gabaix and Koijen, 2020). Specifically, we instrument for employer concentration within a particular local occupation with the predicted change in employer concentration based on the national hiring patterns of large firms (exclud-

[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).

ing 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 or decrease 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),[3] 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 identify effects of changes in this outside-occupation option index on wages, instrumenting

[3]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.

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 2.6 log points lower wages.[4] This average masks substantial heterogeneity: within-occupation employer concentration matters substantially more 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 7.3 log points lower wages; for occupations in the highest quartile of occupational mobility, like counter attendants or bank tellers, our point estimate is close to zero and the confidence interval rules out any decrease in wages greater than 1.3 log points. As expected, we 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 half.

A back-of-the-envelope calculation, using our coefficient estimates, suggests that more than 10% 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 health care employment concentration and low occupational mobility.

We also find a positive and significant effect of an increase in the value of outside-occupation options, holding constant within-occupation employer concentration: for the median occupation, moving from the 25th to the 75th percentile value of outside-occupation options across metro areas is associated with 4.8 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. Overall, our paper demonstrates that the availability of job options outside workers' own firm – which is affected both by within-occupation employer concentration and the quality of outside-occupation job options – matters substantially for workers' own wages.

[4] Our instrumental variable estimates are about 20% larger than our OLS estimates, suggesting that some combination of omitted variables or measurement error bias the coefficient towards zero in simple regressions of wages on employer concentration.

Related literature: 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 (2018), and Rinz (2018),[5] as well as a growing theoretical literature demonstrating an effect of employer concentration on wages (Berger, Herkenhoff and Mongey, 2019; Jarosch, Nimczik and Sorkin, 2019; Azkarate-Askasua and Zerecero, 2020). We make two contributions to this literature: we use a new instrument to estimate plausibly causal effects of employer concentration on wages, and we show that simple wage-concentration regressions which do not do consider outside-occupation options are likely biased and obscure important heterogeneity.

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 Petrongolo, 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., 2019; 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]

[5] As well as Lipsius (2018), Hershbein, Macaluso and Yeh (2019), 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.

[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 (2020), 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 (2013), Matsudaira (2014), Naidu, Nyarko and Wang (2016),

2 Conceptual framework

There are a number of models of the labor market in which employer concentration matters for wages. First, employer concentration can generate upward-sloping labor supply curves to individual firms, leading to wage markdowns (e.g. Berger et al., 2019). Second, the presence of a small number of firms can facilitate collusion to suppress wages. Third, in a bargaining model of the labor market, employer concentration reduces the number of feasible outside options for workers, as the average worker in a given labor market has few distinct firms as alternative possible employers: this reduces workers' relative bargaining position and therefore the wage (Jarosch et al., 2019).[7] In this paper, the conceptual framework guiding our empirical analysis is the third: **employer concentration worsens workers' outside options**. While the main focus of our paper is empirical, we outline below a stylized framework where we formalize this intuition (developed further in Appendix A).

Wage bargaining. At the start of each period, each employed worker Nash-bargains with her employer i. The outcome is wage w_i, equal to the value of the worker's outside option if she leaves her job, oo_i, plus share β of the match surplus (where p_i is the product of the match):[8]

$$w_i = \beta p_i + (1 - \beta) oo_i \tag{1}$$

Job search. Once bargaining with incumbent workers has concluded, firms post vacancies to fill empty positions – vacated either by labor force exit, or by bargaining breakdown with an existing worker. Each posted vacancy offers a wage equal to the wage the firm has bargained with its incumbent workers. Job seekers – workers whose previous wage bargain broke down, who were unemployed in the last period, or who newly entered the labor force – are each paired randomly with a vacancy within their labor market. (For now, consider

Bassier, Dube and Naidu (2019), Goolsbee and Syverson (2019), and Dube, Jacobs, Naidu and Suri (2020).

[7]Several recent papers specifically demonstrate that an HHI is a relevant statistic to measure firms' labor market power. Berger et al. (2019) show in a general equilibrium oligopsony model that firms with higher market share have lower labor supply elasticities, and that the the wage-bill HHI is a relevant statistic for assessing the welfare effects of firms' labor market power. Azkarate-Askasua and Zerecero (2020) also show that firms' employment shares affect the elasticity of labor supply to the firm, in a model which also features worker bargaining 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. Jarosch et al. (2019) show that in a search model with bargaining, the presence of large employers worsens workers' outside option and so reduces wages, with the effect determined by a concentration index closely related to an HHI.

[8]We assume all workers at the same firm i have the same outside option in expectation.

a situation where all workers and all jobs are perfectly substitutable.) Random matching means that the chance of any given worker receiving a job offer from a particular firm j is equal to firm j's share of vacancies in the labor market, σ_j (in the spirit of Burdett and Mortensen (1980) and Jarosch et al. (2019)). If a worker does not receive any job offers, she moves to unemployment for the period and receives b.

Outside option value. The outside option for an employed worker is to leave her current job and become a job seeker. Her outside option value is therefore a weighted average of the wages paid by each local firm (weighted by the probability of being matched with each firm), and the unemployment benefit (weighted by the probability of receiving no job offers):[9]

$$oo_i = \sum_j \sigma_j \cdot w_j + \sigma_i \cdot b. \tag{2}$$

Average wage. To a second order approximation, we can therefore express the average wage in the labor market ($\bar{w} = \sum_i \sigma_i w_i$) as a function of the sum of the squares of employer shares ($HHI = \sum_i \sigma_i^2$):

$$\begin{aligned}\bar{w} &= \beta\bar{p} + (1-\beta)\bar{oo} \\ &= \beta\bar{p} + (1-\beta)\left((1-HHI)\cdot\bar{p} + HHI\cdot b\right) - \beta(1-\beta)\sum_i \sigma_i^2 \hat{p}_i \end{aligned} \tag{3}$$

where $\bar{p} = \sum_i \sigma_i p_i$ is average productivity across firms, and $\hat{p}_j = p_j - \bar{p}$ is the difference between firm j's productivity and the market average. As in the Nash bargain expression we started with, the average wage is a weighted average of the average worker's productivity $\bar{p}$, and the value of the average worker's outside option. The outside option value is now in turn a weighted average of productivity in the labor market $\bar{p}$, which reflects the value of transitioning to other jobs, and the value of unemployment b. The higher is employer concentration HHI, the less likely it is that the average worker will be able to find a job outside her firm, and so the outside option value of other jobs is down-weighted. This

[9] Note that this implies the wage at firm i will depend *negatively* on its share of the labor market σ_i. Why? In a labor market with atomistic firms, every job seeker would be matched with a feasible employer each period. But in a labor market with some large employers, the chance that a worker is re-matched with the firm she just quit is non-zero, and we assume this firm does not re-hire her. This means that the probability that a worker at firm i receives a job offer from *any other* firm if she leaves her job is proportional to the share of all jobs which are outside her firm, $(1-\sigma_i)$. This intuition can be translated to a less stylized setting with on-the-job search, where workers can only receive outside job offers from firms which are *not* their own firm – meaning that workers are large firms are less likely to receive outside job offers.

expression illustrates that (1) the wage declines as employer concentration increases, since higher employer concentration reduces the value of workers' outside option as workers are less likely to receive job offers from other firms, and (2) there is an interaction between employer concentration and worker bargaining power, since the outside option matters less if workers have more bargaining power over the match surplus.[10]

Outside-occupation job options. The framework above 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 direct analog of the market definition approach to calculate HHIs in antitrust in product markets).[11] This is not, however, how most labor markets work in practice. We therefore now adapt our earlier framework, defining the baseline labor market as a worker's local occupation and incorporating workers' option to switch occupation. The value of the outside option is still the weighted average of the wage in each other local firm. The weights for firms within workers' occupation o are now a product of (i) the probability that a worker from occupation o will be matched with a job within occupation o (ζ_o), and (ii) the probability that, conditional on staying in occupation o, the worker will be matched with firm j (which we once again proxy for using the vacancy share $\sigma_{j,o}$ of firm j in occupation o). For firms outside workers' occupation, similarly, the value of the outside option is a weighted average of wages, where the weights are a product of (i) the probability that a worker from occupation o will be matched with a job outside occupation o ($1 - \zeta_o$), (ii) the probability that conditional on leaving their occupation o, a worker will be matched with *some* job in occupation p ($Prob(o \to p)$), and (ii) the probability that conditional on being matched with a firm in occupation p, the worker will be matched specifically with firm j ($\sigma_{j,p}$).[12] The outside option for workers in

[10] The third term depends on the joint distribution of employment shares and productivity. If their correlation is sufficiently small the wage is simply a concentration- and bargaining power-weighted average of productivity p and unemployment benefit b.

[11] 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., 2020b,a). Jarosch et al. (2019) 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.

[12] Note: this assumes that employment decisions are taken at the firm-by-occupation level.

firm i and occupation o is therefore:

$$oo_{i,o} = \underbrace{\zeta_o \sum_{j \neq i}^{N_o} \sigma_{j,o} w_{j,o}}_{\text{own-occ options}} + \underbrace{(1-\zeta_o) \sum_{p \neq o}^{N^{occs}} Prob(o \to p) \sum_{l}^{N_p} \sigma_{l,p} w_{l,p}}_{\text{outside-occ options}} + \underbrace{\zeta_o \sigma_{i,o} b}_{\text{unemployment}}$$

where N_o denotes the set of firms in occupation o, and N^{occs} denotes the set of occupations. Next, we define the value of outside-occupation job options as the probability-weighted average of wages in other occupations, $oo^{occs} = \sum_{p \neq o}^{N^{occs}} Prob(o \to p) \sum_{l}^{N_p} \sigma_{l,p} \cdot w_{l,p}$.[13] We can then express the average wage in occupation o, to a second order approximation, as a function of employer concentration within workers' occupation, and the value of outside-occupation job options, as follows:

$$\bar{w}_o = \tilde{\psi}_o \left(\alpha \bar{p}_o + (1-\alpha) oo^{occs}\right) + (1-\tilde{\psi}_o) b - \tilde{p}_o \tag{4}$$

where $\tilde{\psi}_o = 1 - (1-\beta)\zeta_o HHI_o$, $\alpha = \frac{\beta}{1-\zeta_o(1-\beta)}$, and $\tilde{p}_o = \beta(1-\beta)\zeta_o \sum_i \sigma_i^2 \hat{p}_{i,o}$. This expression suggests that the average wage in occupation o is, roughly, a weighted average of average productivity in occupation o ($\bar{p}_o$), the value of jobs outside occupation o (oo_o^{occs}), and unemployment benefits (b), where the weights depend on employer concentration within workers' occupation, the likelihood a worker will remain in her own occupation, and worker bargaining power.[14] As before, higher within-occupation employer concentration reduces the wage by reducing the average availability of feasible job options outside workers' own current firm. There is also an interaction with $Prob(o \to o)$: the less likely it is that a worker can find a job in a different occupation, the more employer concentration in her own occupation matters.

Implications for empirical analysis. Typically, regressions of wages on employer concentration at the level of a local labor market take the binary market definition approach: they do not take into account workers' (differential) ability to switch occupation. The discussion above illustrates two problems that this could cause when estimating the effect of con-

[13] Note that in practice the wages in firms in other occupations p will depend on the wages set in occupation o, and vice versa – the "reflection problem". We segment workers' outside options at the boundary of their occupation for empirical tractability. In our empirical analysis we will construct a measure of outside-occupation options and avoid the endogeneity caused by the reflection problem with an instrumental variable strategy to estimate the effect of plausibly exogenous variation in wages in other local occupations p on wages in occupation o.

[14] Ignoring the final term $\tilde{p}_o$, which is once again small if the average productivity of individual firms is not strongly correlated with their employment shares

centration on wages. First, heterogeneity: one should expect the effect of within-occupation employer concentration on wages to be different for occupations with low outward occupational mobility (high ζ_o) as compared to occupations with high outward occupational mobility. Second, omitted variable bias: if the degree of employer concentration within a local occupation (HHI) is correlated with the quality of outside options *outside* the local occupation (oo^{occs}), then estimation of the effect of concentration on the wage may be biased without controlling for outside-occupation options. We take both of these into account in our empirical analysis.

To summarize, the analysis above implies four empirically testable predictions: (1) higher employer concentration reduces wages, (2) better outside-occupation options increase wages, (3) the wage-HHI relationship is stronger for occupations with limited outward mobility, and (4) the estimated wage-HHI relationship may be biased if within-occupation HHI is correlated with outside-occupation job options. These hypotheses are tested in the empirical sections that follow.

3 Empirical Approach

We jointly estimate the effect on wages of (1) within-occupation employer concentration and (2) the value of outside-occupation job options. Specifically, in our baseline specification we regress the log of the average hourly wage in a SOC 6-digit occupation, metro area, and year, on the log of the HHI of employer concentration ($HHI_{o,k,t}$), the log of an index of the value of outside-occupation options ($oo^{occs}_{o,k,t}$), and a set of occupation-by-year and metro area-by-year fixed effects ($\alpha_{o,t}$, $\alpha_{k,t}$):

$$\ln \bar{w}_{o,k,t} = \alpha + \alpha_{o,t} + \alpha_{k,t} + \gamma_1 \ln HHI_{o,k,t} + \gamma_2 \ln oo^{occs}_{o,k,t} + \xi_{o,k,t} \tag{5}$$

We run this regression allowing 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 resume data from Burning Glass Technologies), interacting $HHI_{o,k,t}$ and $oo^{occs}_{o,k,t}$ with an indicator variable for the applicable quartile of outward mobility of occupation o. Our approach takes into account the two empirical issues raised in the previous section: we allow for heterogeneity of the effect of employer concentration on wages according to the occupation's degree of outward mobility, and we correct for omitted variable bias in the coefficient on the HHI (arising from failure to consider workers' option to switch occupations) by including the outside-occupation option index as an independent variable.

We run these 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 445,975 occupation-metro area-year observations.[15] We use BLS OES data for average hourly wages by occupation, metro area and year for the dependent variable $\bar{w}_{o,k,t}$. We construct the HHI from Burning Glass Technologies' vacancy posting data (discussed further in section 3.1), 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 3.2). To account for possible endogeneity, we develop instruments for both the HHI and the oustide-occupation option index, discussed further in sections 3.3 and 3.4.

3.1 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. (2019).[16] 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:

$$HHI_{o,k,t} = \sum_{i=1}^{N} \left(\frac{v_{i,o,k,t}}{\sum_{i=1}^{N} v_{i,o,k,t}} \right)^2 \tag{6}$$

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).[17]

[15]This includes 367 metro areas and 707 occupations, with 106,792 occupation-metro area labor markets appearing in at least one year from 2011–2019. 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.

[16]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 than employment. In equilibrium, one would expect vacancy and employment HHIs to be highly correlated. Indeed, Marinescu et al. (2021) show in France that an HHI of employment flows (reflecting filled vacancies) is highly correlated with employment HHIs.

[17]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

Representativeness. 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 reasons for concern. 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.[18] 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.[19]

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.[20] For further discussion of the BGT vacancy data, see Appendix B.

3.2 Measuring outside-occupation options

To jointly estimate the effect of *both* within-occupation employer concentration *and* outside-occupation job options on wages, we also need a measure of outside-occupation options – the empirical counterpart of the outside-occupation option measure outlined in our concep-

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 B.

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

[19]If there is a big difference in the ratio of hires per job posting between large and 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 may therefore be less reliable for occupations for which there are many large employers who hire a lot of workers for undifferentiated job roles.

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

tual framework section. We use data on workers' mobility patterns between occupations to identify these outside-occupation options.[21]

BGT resume data. Since there is no existing US occupational mobility data with high enough granularity to study transitions between SOC 6-digit 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.[22] This resume data was collected by labor market analytics company Burning Glass Technologies ("BGT"), who sourced the resumes from a variety of BGT partners, including recruitment and staffing agencies, workforce agencies, and job boards.

Transition share. We use this data to construct our measure of occupational transitions $\pi_{o \to p}$, which approximates the probability of a worker moving from occupation o to occupation p conditional on leaving her occupation:[23]

$$\begin{aligned}\pi_{o \to p} &= \frac{\text{\# in occ } o \text{ in year } t \text{ observed in occ } p \text{ in year } t+1}{\text{\# in occ } o \text{ in year } t \text{ observed in any new occ in year } t+1} \\ &\approx Prob(\text{move from occ } o \text{ to occ } p | \text{leave occ } o)\end{aligned} \tag{7}$$

Leave share. We also construct the 'occupation leave share', approximating the share of people who leave their occupation when they leave their job:[24]

$$\begin{aligned}\text{leave share}_o &= \frac{\text{\# in occ } o \text{ in year } t \text{ \& no longer in occ } o \text{ in year } t+1}{\text{\# in occ } o \text{ in year } t \text{ \& in a new job in year } t+1} \\ &\approx Prob(\text{leave occ } o | \text{leave job})\end{aligned} \tag{8}$$

[21] 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 E we compare our approach to approaches based on task or skill similarity.

[22] The CPS has at least an order of magnitude fewer occupational transition observations than our data over the same time period. This matters: with 705,600 possible transition pairs between the SOC 6-digit occupations, data sets with even a few million observations are not big enough to capture many transition paths.

[23] 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.

[24] 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 $t+1$.

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.[25]

Representativeness. 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.

Five facts about occupational mobility. Is it sensible to use occupational transitions to infer the scope of workers' labor markets? Are job options outside the narrow occupation even an important part of most workers' labor markets? Using the BGT data, we document five stylized facts about occupational mobility which suggest the answer to both questions is yes.

1. Occupational mobility is high, suggesting that the SOC 6-digit occupation fails to capture many workers' true labor markets: the average probability of a worker leaving her 6-digit occupation when she leaves her job - the "occupation leave share" defined above - is 23%.
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).[26]
3. Aggregating up the SOC classification hierarchy - which groups ostensibly similar occu-

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

[26] Almost all of the occupations with low leave shares are highly specialized, including various medical, legal and educational occupations (see Appendix Table A4). In contrast, many high leave share occupations require more general skills, including restaurant hosts/hostesses, cashiers, tellers, counter attendants, and food preparation workers.

pations - still fails to capture most occupational transitions, suggesting that this cannot solve the market definition problem.[27]

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).[28]
5. Empirical occupation transitions reflect similarities between occupations in terms of their task requirements, wages, amenities, and leadership responsibilities, suggesting that occupational transitions do indeed reflect the underlying feasibility of an occupation as an outside option (see Figure 3).[29]

Measuring outside-occupation options. The facts above suggest that we can use occupational transitions to infer workers' outside option occupations. Informed by our conceptual framework, we therefore define a measure of the outside option value of jobs in other occupations - the **outside-occupation option index** oo^{occs} – as a weighted average of the wage in each alternative occupation, weighted by a measure of the likelihood that the worker will move to a job in each of those alternative occupations if she leaves her current occupation: $oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} Prob(o \rightarrow p)_{o,k,t} \cdot \mathrm{w}_{p,k,t}$, where subscripts refers to the metro area (k), current occupation (o), possible destination occupations (p), and year (t). To proxy

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

[28] See Appendix Figure A7 and Appendix Table A5. The asymmetry partly reflects the fact 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.

[29] 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 3. Similarly, Macaluso (2019) finds that mobility between U.S. SOC 2-digit occupations is highly correlated with task similarity. See Appendix F for more details on our analysis.

for $Prob(o \to p)$, we use the product of two variables: (1) the national average empirical occupation transition share $\pi_{o \to p}$, and (2) the relative employment share of occupation p in metro area k compared to the national average, $\frac{s_{p,k}}{s_p}$.[30] Our empirical outside-occupation option index is therefore:

$$oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} \pi_{o \to p} \cdot \frac{s_{p,k,t}}{s_{p,t}} \cdot \bar{w}_{p,k,t} \tag{9}$$

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 $\bar{w}_{p,k,t}$).[31]

3.3 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.[32]

We therefore instrument for local labor market concentration, creating an instrumental

[30] 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.

[31] See summary stats of our index in Table 2. Note: we use "metro areas" to refer to the CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas) for which data is available in the BLS OES. 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. To create a consistent panel of occupations over time we crosswalk SOC classifications over time: see Appendix D.

[32] Concerns like these are raised in many of the 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). 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. Hsieh and Rossi-Hansberg (2019) show that over recent decades large national firms have expanded into more local labor markets, reducing local employer concentration and possibly increasing productivity.

variable which leverages differential local occupation-level exposure to large national firms' hiring decisions, 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 $g_{o,k,t}$. For firms j which are sufficiently large,[33] 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 $\tilde{g}_{j,o,t}$). Our instrument for the HHI, $Z^{HHI}_{o,k,t}$, is therefore:

$$Z^{HHI}_{o,k,t} = log\left(\sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1\right)\right) \tag{10}$$

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.[34] 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 local labor market k. Through the lens of shift-share instruments (Borusyak et al., forthcoming), our instrument features plausibly exogenous 'shocks' (a function of firms' national hiring growth),

[33] In our baseline specification, we define large firms as firms which have vacancies in occupation o in at least five metropolitan areas in the year in question. In Appendix Table A16 we present results with different shock definitions and size restrictions.

[34] Note that by taking the log of the instrument, we implicitly 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.

and possibly endogenous exposure 'shares' (the last-period local occupational vacancy shares of each of those firms).

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 A7 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.

One concern with this instrumental variable is that differential local exposure to national firms' growth may differentially affect total labor demand, not just employer concentration. In our model, 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 or productivity is linear. Thus, 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 (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 should be 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.[35]

[35]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 A11 and A13). 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

A second concern is bias due to the fact that our exposure 'shares' do not sum to one. As such, following Borusyak et al. (forthcoming) we introduce an "exposure control": the sum of the squared local vacancy shares of the large national firms j which feature in the instrument.[36] We further discuss identification conditions in Appendix G.

Third, our instrument is unlikely to be strong for small changes in employer concentration in initially unconcentrated labor markets – if each firm has only a trivial share of local employment, even substantial hiring growth will not much change local employer concentration. 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 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 (2018); Qiu and Sojourner (2019); Marinescu et al. (2021); Gibbons et al. (2019)). This circumvents some endogeneity concerns, 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.[37] Our strategy 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 hospital mergers. 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

expand more nationally tend to be located in weaker labor markets where wage growth is lower.

[36]That is, in the baseline, the sum of the squared local vacancy shares of any firms j which have vacancy postings in at least four other metro areas in the occupation in question in the year in question, and for whom vacancy growth in at least one of these metro areas was non-zero from one year to the next. This controls for the fact that different local occupations may have different initial shares of employment accounted for by large national firms.

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

local labor demand.[38] 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.

3.4 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 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 ($\frac{s_{p,k,1999}}{s_{p,1999}}$).[39] Our instrument for the log of the oo^{occs} index is:

$$Z^{oo}_{o,k,t} = log\left(\sum_{p}^{N_{occs}}\left(\pi_{o\to p}\cdot\frac{s_{p,k,1999}}{s_{p,1999}}\cdot\bar{w}_{p,\not{k},t}\right)\right) \tag{11}$$

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.[40] For our instrument

[38] 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.

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

[40] 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'

to be valid, the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p must be correlated with the local wage of occupation p in location k (relevance condition), 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).[41] We discuss conditions for identification further in Appendix G, following the approach to shift-share IVs of Borusyak et al. (forthcoming).

4 Results

Our analysis suggested four testable predictions: (1) higher employer concentration reduces wages, (2) better outside-occupation options increase wages, (3) the wage-HHI relationship is stronger for occupations with limited outward mobility, and (4) the estimated wage-HHI relationship may be biased if within-occupation HHI is correlated with outside-occupation job options. We find empirical support for all four.

4.1 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 4, 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 -0.012 (Table 3, column *a*).[42]

When instrumenting for the HHI, the coefficient magnitude increases by around 25% relative to the OLS specification (Table 3, column *c*). 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.[43] 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,

industrial composition.

[41] 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 A11).

[42] Similarly Hershbein et al. (2019) 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.

[43] The first stage is shown in Table A9, column (a).

confirming that outside-occupation job options matter for wages. After introducing the outside-occupation option index the coefficient on the HHI falls by 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 estimates in Table 3, column d – instrumenting for both employer concentration and outside-occupation job options – suggest 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 2.6 log points lower hourly wage.[44] As noted, we caution against applying these coefficient estimates to labor markets with very low initial levels of employer concentration, since our instrument is weak in these cases.[45]

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).[46] 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 7.3 log points lower wages.[47] For the quartile of occupations with the highest outward mobility on the other hand, 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 would

[44] Calculated as $(ln(1882) - ln(137)) \cdot -0.010 = -0.026$. 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 SOC 6-digit occupation by MSA level) leads to a 0.1% decrease in wages on average.

[45] Appendix Figure A15 illustrates that the correlation between our (log) HHI instrument and the (log) HHI for occupation-metro area labor markets in 2019 somewhat breaks down for occupation-metro area cells with a very low value of the HHI or of our HHI instrument.

[46] In Appendix Table A15, we explore whether there is evidence for heterogeneity of the effect of concentration on wages for different occupation groups, but do not find patterns which are clearly statistically significantly different across these groups.

[47] Calculated as $(ln(1882) - ln(137)) \cdot -0.028 = -0.073$.

be associated with at most a 1.3 log point lower wage.[48]

Robustness checks. We explore a number of additional variations on our baseline analyses, illustrated in Figure 6 and Appendix Tables A11-A14. First, we show that our coefficient estimates for the effect of employer concentration are similar if we remove the controls for vacancy growth (Tables A11 and A13 column a), if we remove the exposure control (column c), or if we follow Gabaix and Koijen (2020) in adding an additional control for the equal-weighted vacancy growth of local firms ($g^e_{o,k,t} = \frac{1}{N}\sum_j^N g_{j,o,k,t}$) to reflect common local occupation-specific shocks (column b). Second, we control for an industry Bartik shock to proxy for local metro area occupation exposure to common national industry trends (column d), and similarly find little change in our estimates.[49] Third, we re-run our baseline regressions with fixed effects for occupation-metro area and year, rather than occupation-year and metro area-year (column e). Here, identifying variation comes from year-to-year changes in employer concentration within the same occupation-metro area labor market over the period 2011–2019. The coefficient estimate for the average effect has a small point estimate but a relatively large standard error; however, for the estimates by quartile of outward occupational mobility, there is still a large, negative, and statistically significant coefficient for the employer concentration effect on the lowest-mobility quartile. Fourth, we show that our coefficient estimates are similar, although a little smaller, if we regress without employment weights (Appendix Tables A12 and A14, column a). Fifth, we drop occ-metro area cells with very low HHIs (less than 50), since the logic of our HHI instrument holds less well for very low HHI cells (Appendix Tables A12 and A14, column b). Sixth, to address concerns about representativeness in our BGT vacancy data, we re-run our baseline regression estimates *excluding* any occupations which are substantially underrepresented in the vacancy data (column c),[50] and also weight each occupation or metropolitan area

[48] Calculated as: $(ln(1882) - ln(137)) \cdot (0.002 - 1.96 \cdot 0.003) = -0.013$. 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.

[49] See Appendix G for details of the construction of the local occupation level industry Bartik shocks.

[50] We exclude occupations with a 'represented-ness' less than 0.5 in the BGT vacancy data, corresponding

respectively by its represented-ness in the BGT data (columns d and e).

4.2 Results: outside-occupation options

We can also use our baseline regressions to ask: how big are the effects of outside-occupation options on wages? Our baseline 2SLS IV coefficient estimate (Table 3 column d) suggests that a 10 log point higher outside-occupation option index leads to 1.2 log points higher wages in workers' own occupation. This implies that moving from the 25th to the 75th percentile value of outside-occupation options across metro areas for the median occupation leads to 4.8 log points higher wages.[51] This is 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.[52] In addition, finding a large, significant, and positive effect of shocks to outside-occupation options on wages reinforces our conclusions that workers' true labor markets are broader than their narrow 6-digit SOC occupations, and that our "probabilistic" method of identifying relevant outside options can capture workers' true labor markets relatively well.[53]

Robustness checks. 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.[54] To control for this possibility we construct a shift-share "industry

to about one third of occupations.

[51]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 0.12. With the same exercise for employer concentration, we find that for the median occupation, moving from the 25th to the 75th percentile HHI across metro areas would result in a wage increase of 1.4 log points.

[52]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.

[53]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.

[54]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 because more accountants in SF already work in tech.

Bartik" shock that captures the predicted impact of industry level wage trends on local occupation wages and include it in our baseline regressions.[55] Coefficients on the outside-occupation option index remain robust to its inclusion (Appendix Tables A11 and A13, column *d*).

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 A17). 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 A18, 1999–2016), if we control for local occupational employment (Appendix Table A19), if we split our analysis into three time periods (Appendix Table A20), or if we remove employment weights (Appendix Table A21).

5 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),[56] as follows:

$$\text{wage effect}_{o,k,t} = (log(HHI)_{o,k,t} - log(150)) \cdot \gamma_1^q \tag{12}$$

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,

[55] 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 G.

[56] 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).

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 13.7 million workers – 11.7% of the workers in occupation-metro area labor markets covered by our data – have wages which are at least 2% lower as a result of above-median employer concentration.[57] In Table 5, we show the average estimated wage effect for different combinations of employer concentration and outward occupational mobility, illustrating 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 medium levels of employer concentration. 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 A18 and A19). Note that our estimates focus only on wages, but employer concentration may also affect non-wage benefits and workplace amenities.[58]

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 nearly two million registered nurses, licensed practical and vocational nurses, and nursing assistants, and more than 500,000 pharmacists and pharmacy technicians.[59] According to our estimates, large numbers of security guards and hairdressers,

[57] This may be an overestimate since the figure include some occupations which are underrepresented in the BGT vacancy posting data. Excluding all occupations with a 'represented-ness' of less than 0.5 in the BGT vacancy data, there remain 8.5 million workers whose wages are suppressed by at least 2% as a result of employer concentration. On the other hand, our data 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. One might expect these workers to face greater wage suppression from employer concentration than the average in our data.

[58] 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.

[59] This is in keeping with recent work that has found large effects of hospital mergers on wages of nursing

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).[60] 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 counter attendants, and substantial underestimation of the effects of employer concentration for low-mobility occupations like nurses and pharmacy technicians (as illustrated in Appendix Table A25).

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 can be done is antitrust.[61] 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.[62]

---

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).

[60]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 non-compete and no-poaching agreements may make this distinction moot in practice.

[61]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.

[62]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

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.[63] 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.[64] An alternative might be strengthening minimum wages or benefits standards

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 local SOC 6-digit occupation. Our finding that the wage suppressive effects of employer concentration are so much higher for low outward mobility occupations suggests that screening based only on local within-occupation HHI without considering outward occupational mobility will lead to some mergers being scrutinized which may have little effect on wages, while others which may have serious anti-competitive effects may go unnoticed.

[63] As emphasized by Hovenkamp (2018), Berger et al. (2019), 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.

[64] In our conceptual framework outlined in section 2, higher worker bargaining power β 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 suggestive empirical evidence that labor markets with higher unionization rates see smaller effects of employer concentration on wages. When we re-run our baseline regression of the effect of employer concentration on wages with an interaction with states' right-to-work status – a proxy for the ease of forming a union – we find larger effects of both employer concentration and outside-occupation options on wages in right-to-work states (Appendix Table A22). 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. (2018) find a stronger relationship between employer concentration and wages in U.S. manufacturing firms where unionization rates are lower.

in local labor markets characterized by high employer concentration.[65]

Implications: promoting mobility. Our results suggest that employer concentration within a local occupation matters 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.[66] 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. In addition, restrictions on worker mobility *within* an occupation (like non-compete clauses) could exacerbate the effects of employer concentration on wages.[67]

Incidence of employer concentration effects. Our estimates suggest that increases in employer concentration reduce local wages, but cannot tell us whether the 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). Similarly, 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.

6 Conclusion

Our findings point to a middle ground between two prominent views about the effects of employer concentration in the US labor market. On the one hand, employer concentration is *not* a niche issue confined to a few factory towns: we find large, negative, and significant

[65]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.

[66]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 costs of mobility, could be acting to increase the effects of employer concentration.

[67]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, and Starr, Prescott and Bishara (2021); Johnson, Lavetti and Lipsitz (2020) on the effect of non-competes.

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 more than 10% 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).[68] 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.

[68]Similarly, Rinz (2018), Berger et al. (2019), 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*, 2013, (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 Bergman, and Hyunseob Kim, "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?," 2018.

Berger, David W, Kyle F Herkenhoff, and Simon Mongey, "Labor Market Power," 2019.

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, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *Review of Economic Studies*, forthcoming.

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.

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," 2019.

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.

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

Hershbein, Brad, Claudia Macaluso, and Chen Yeh, "Concentration in U.S. local labor markets: evidence from vacancy and employment data," *Working Paper*, 2019.

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.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The industrial revolution in services," Technical Report, National Bureau of Economic Research 2019.

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?," 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 Inequality, and Earnings Mobility," *U.S. Census Bureau Carra Working Paper Series*, 2018, *2018-10.*

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.

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," 2020.

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 3.2 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, a vacancy HHI, and an outside-occupation option index. This comprised 109,582 occupation-by-metro area labor markets and 117,286,314 workers in 2019 (according to the BLS OES data). 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	-0.012***	-0.008***	-0.015***	-0.010***
	(0.002)	(0.002)	(0.003)	(0.003)
Log outside-occ. options		0.139***		0.120***
		(0.012)		(0.018)
Vacancy growth			-0.142*	-0.106*
			(0.075)	(0.060)
Predicted vacancy growth			0.047	-0.002
			(0.045)	(0.038)
Exposure control			0.024	0.010
			(0.015)	(0.014)
Observations	445,681	445,681	445,681	445,681
F-Stat			486	202

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	-0.021***	-0.016***	-0.030***	-0.028***
X Q1 occ mobility	(0.004)	(0.004)	(0.005)	(0.006)
Log HHI	-0.018***	-0.009***	-0.021***	-0.012***
X Q2 occ mobility	(0.002)	(0.002)	(0.003)	(0.003)
Log HHI	-0.007***	-0.004**	-0.007	-0.002
X Q3 occ mobility	(0.002)	(0.002)	(0.004)	(0.004)
Log HHI	-0.001	-0.001	-0.002	0.001
X Q4 occ mobility	(0.003)	(0.002)	(0.004)	(0.003)
Log outside-occ options		0.131***		0.097***
X Q1 occ mobility		(0.012)		(0.017)
Log outside-occ options		0.151***		0.135***
X Q2 occ mobility		(0.013)		(0.015)
Log outside-occ options		0.124***		0.119***
X Q3 occ mobility		(0.013)		(0.024)
Log outside-occ options		0.118***		0.107***
X Q4 occ mobility		(0.015)		(0.031)
Vacancy growth			-0.146*	-0.109*
			(0.078)	(0.062)
Predicted vacancy growth			0.041	-0.000
			(0.047)	(0.038)
Exposure control			0.027**	0.013
			(0.013)	(0.013)
Constant	3.072***	2.615***		
	(0.009)	(0.041)		
Observations	445,681	445,681	445,681	445,681
F-stat			116	47

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* Q*i* 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 *i*th 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	1.7%	4.8%	7.3%	9.8%
	Employment (m)	9.9	7.7	5.2	1	.98
Q2 mobility	Avg. wage effect	0	0.7%	2.0%	3.1%	4.1%
	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 3 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 wage impact of employer concentration on their wage is 9.8%. 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)

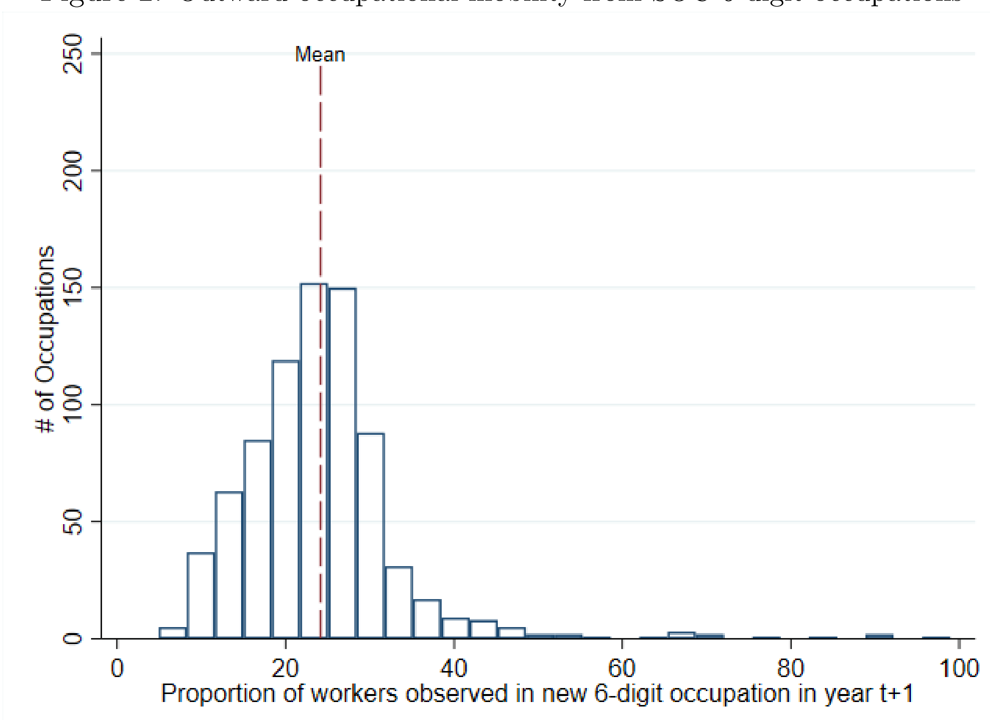
Occupation	National employment	Share of occupation w/ estimated effect 2% or greater	Number in occupation w/ estimated effect 2% or greater
Registered nurses	2,982,280	.35	1,043,200
Security guards	1,126,370	.84	950,680
Nursing assistants	1,419,920	.44	629,320
Hairdressers, hairstylists, and cosmetologists	385,960	.9	345,680
Pharmacy technicians	417,780	.74	309,440
Pharmacists	311,200	.67	209,880
Medical assistants	712,430	.26	184,430
Licensed practical and licensed vocational nurses	697,510	.25	172,760
Fitness trainers and aerobics instructors	325,500	.48	156,490
Emergency medical technicians and paramedics	521,200	.28	145,630
Radiologic technologists	207,360	.64	132,750
Heavy and tractor-trailer truck drivers	1,856,130	.066	123,160
Medical and clinical laboratory technologists	332,542	.35	115,679
Phlebotomists	128,290	.86	110,270
Aircraft mechanics and service technicians	133,310	.7	93,930
Lawyers	657,170	.14	91,810
Massage therapists	107,240	.86	91,770
Management analysts	709,750	.13	88,900
Nurse practitioners	200,600	.43	86,490
Manicurists and pedicurists	111,780	.73	81,700
Respiratory therapists	132,090	.53	70,500
Physician assistants	120,090	.55	65,850
Surgical technologists	109,000	.56	61,390
Physicians and surgeons, all other	390,680	.15	59,530
Secretaries and administrative assistants (except legal, medical, and executive)	2,038,340	.028	56,870

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 A24 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 (i) heavily public sector, and/or (ii) 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). From this list, the excluded occupations based on these criteria are a mix of primarily public or quasi-public sector occupations (Bus drivers (school or special client), Teachers and instructors (all other), Police and sheriff's patrol officers, Firefighters, Postal service mail carriers, First-line supervisors of police and detectives, Self-enrichment education teachers, Court, municipal, and license clerks, Highway maintenance workers, Librarians, Social and human service assistants), 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), occupations with the "all other" classification, for which vacancies may be hard to parse accurately (Information and record clerks (all other), First-line supervisors of protective service workers (all other), Assemblers and fabricators (all other)), and finally Operating engineers and other construction equipment operators, and Opticians, dispensing. We also exclude Childcare workers (which ranks number 24 on this list), since our vacancy data to some extent may pick up platform-based listings for childcare workers which may be better considered postings by separate employers rather than a single employer.

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

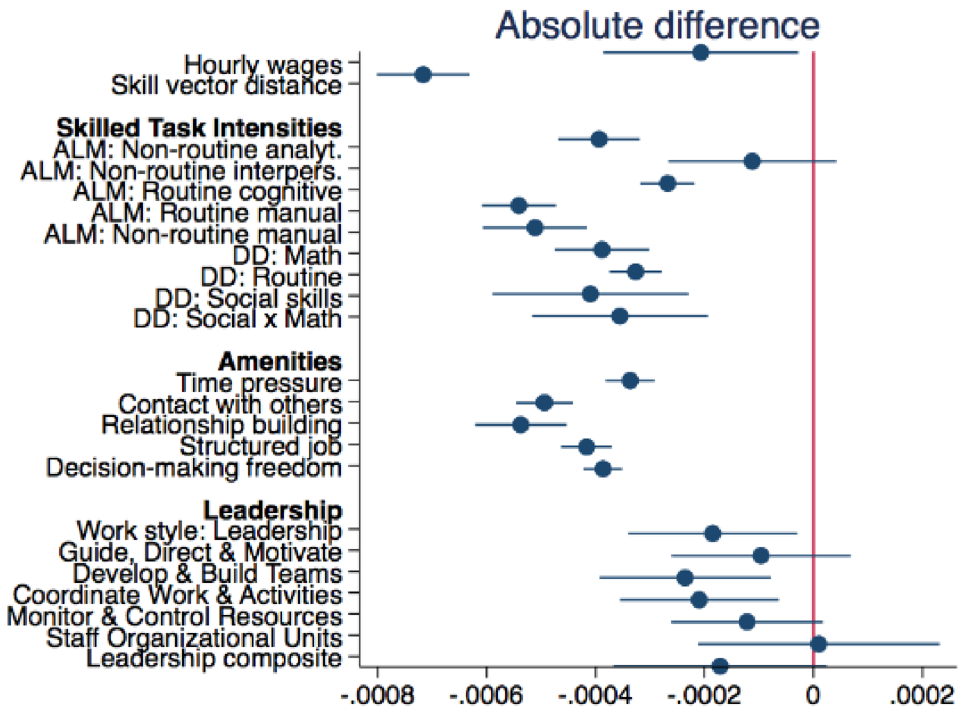
Note: HHI is measured using Burning Glass Technologies vacancy data, at the level of a SOC 6-digit occupation by metro area labor market. Our data covers occupation-metro area labor markets which include 117m of the 151m workers in the U.S. labor market in 2019. Left panel shows the distribution of HHIs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



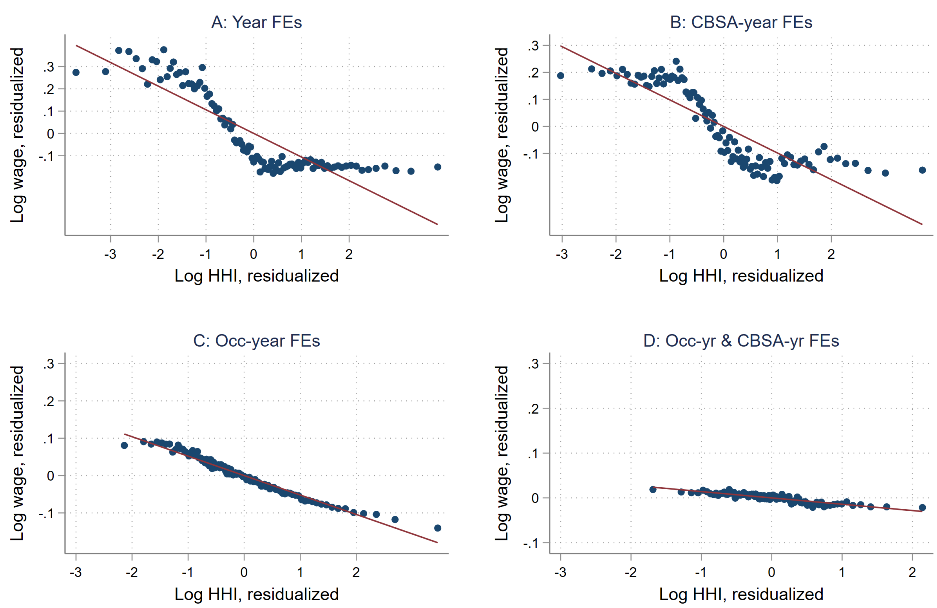
Distribution of the "occupation leave share" – the probability that a worker will leave their occupation conditional on leaving their job – by occupation. Occupation leave share is calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample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

Figure 3: Occupational transitions and occupational characteristic similarity



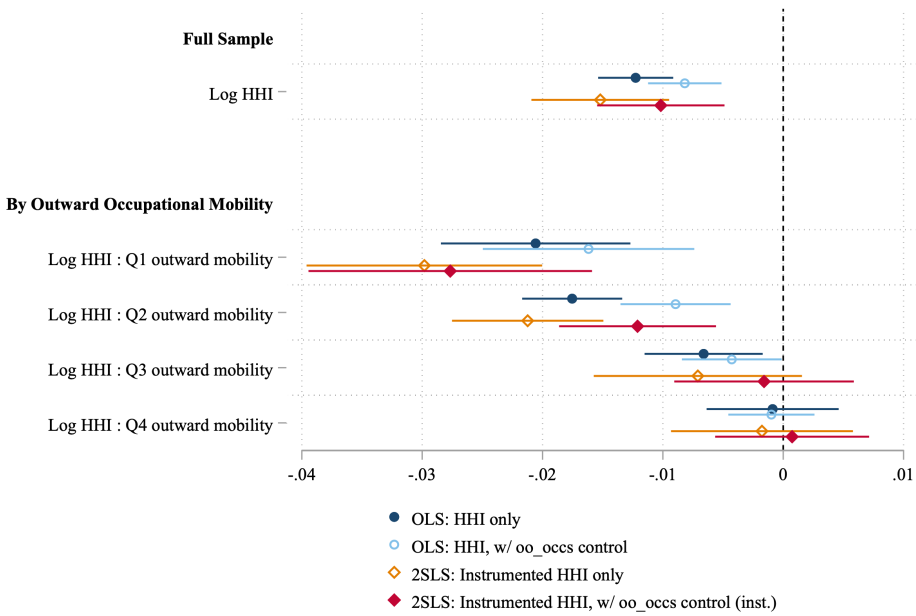
Note: This plot shows coefficients and 95% confidence intervals from regressions of occupation transition shares $\pi_{o\to p}$ on occupational characteristics: $\pi_{o\to p} = \alpha_o + \beta f(X_{occ\ o\to p}) + \gamma f(\Delta w_{o\to p}) + \epsilon_{op}$, where $f(\cdot)$ represents the absolute difference in characteristic X between occupation o and p, and α_o is occupation o fixed effect. Regressions also include absolute avg. hourly wage differences (except for amenities regressions). Standard errors are clustered by origin occupation. Regressions are described in more detail in Appendix F; Appendix Figure A11, shows an analogous exercise with the *relative* difference in characteristics.

Figure 4: Correlations between wage and HHI



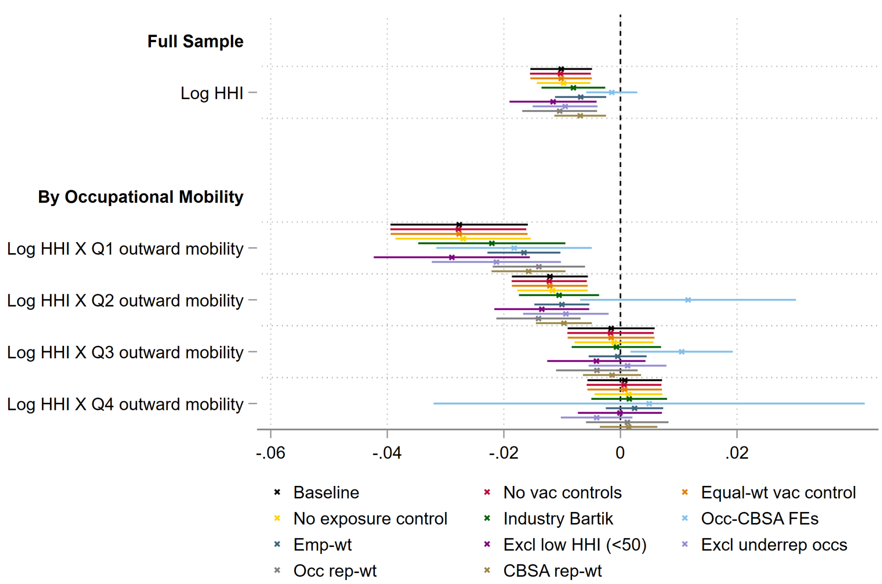
Note: Figure shows binned scatter plots of the relationship between log wages and log employer HHI for occupation-metro area cells over 2011–2019, weighted by occupation-metro area average employment and residualized on different combinations of fixed effects. Slopes for the line of best fit on each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

Figure 5: Coefficients on wage-HHI regressions



Note: Coefficients on log HHI and 95% confidence intervals from our baseline regressions of occupation-metro area wages 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 wages on instrumented employer HHI, across various robustness checks (as in Appendix Tables A11-A14).
Black shows the baseline estimates from Figure 5.

APPENDIX

A Appendix: Conceptual framework: more detail

This section expands on the conceptual framework presented briefly in section 2 in the main body of the paper.

Conceptual framework: effect of concentration on wages

Timing. Each period has two phases: the hiring phase and the production phase. During the hiring phase, workers exit or enter the labor market, employed workers bargain with their employer, and new workers are hired. During the production phase, employed workers produce at the firms they were hired at, receiving the wage determined in the hiring phase, and unemployed workers receive unemployment benefit b. At the start of the next period (in the next hiring phase), employed workers may leave their firms and/or renegotiate their wage with their current employer, and unemployed workers may search for a job. More details of each step in the process follow. Note that in this initial simple framework, we consider a clearly-defined labor market where all workers and all jobs are perfect substitutes. We relax this assumption later.

Firms. There are N firms in the labor market. Each firm j can employ up to n_j workers and has a constant returns to scale production technology with labor productivity p_j over a period. Firm size and productivity are both exogenously determined from the perspective of this model, and do not change over the period we are considering. In addition, no new firms can enter.[69]

Labor market exit and entry. At the start of each period, during the hiring phase, a fraction ξ of workers from each firm 'die' – that is, they leave their jobs *and* the labor market, for exogenous reasons (for example, family reasons, relocation, retirement, ill health, or death). These workers are replaced by an equal number of workers who are 'born' – i.e., they are new to the local labor market (perhaps they have moved, or newly entered or re-entered the labor force), who enter as job seekers.

Wage bargaining. Each worker who is currently employed at the start of the period Nash-bargains with her employer i over the wage.[70] The outcome is a wage w_i, equal to the

[69] We choose this extremely simple, static set-up to illustrate as cleanly as possible the impact of employer concentration on wages. With the possibility for firm entry, and for firm growth – with some entry and/or adjustment costs – the general intuition of a negative impact of employer concentration on outside options and therefore on wages would persist, but the exact impact would depend on the degree to which new firms can enter, the speed at which incumbent firms can grow, and the distribution of production technologies across new and incumbent firms, and small and large firms.

[70] The Nash bargaining outcome can be derived as the outcome of a bargaining problem where the firm

value of the worker's outside option if bargaining breaks down and she leaves her job, oo_i, plus a share β – reflecting worker bargaining power – of the match surplus:

$$w_i = \beta(p_i - oo_i) + oo_i = \beta p_i + (1-\beta)oo_i \tag{13}$$

If wage bargaining breaks down, the worker leaves her job and becomes a job seeker. We assume that all workers at a given firm have the same set of outside options in expectation, such that all workers at any given firm will in equilibrium receive the same wage.

Vacancies. After labor market exit and wage bargaining have happened, firms post vacancies to fill the positions which have been vacated by either worker exit or the breakdown of wage bargaining. There is no cost to post vacancies to fill existing positions, but firms cannot post vacancies for new positions (i.e. there is no firm growth).[71] A firm can only post one vacancy for each position they wish to fill. Vacancies are posted as take-it-or-leave-it wage offers, with the posted wage equal to the wage the firm is paying to its other workers (which was decided on in the wage bargaining process described above). This constraint – that similar workers are paid similar amounts – is often observed in practice in firms. It may be motivated by fairness concerns or explicit internal pay hierarchies or bargaining agreements.[72]

Job seekers. Job seekers are comprised of workers who have newly entered the labor market, workers who have left their previous job because their match broke down at the start of the period, and workers who were unemployed in the previous period. Job search can only take place at the start of the period, during the hiring phase. There is no on-the-job-search: only workers who have left their current employer may search for a new job. Each job seeker applies to all local employers $j \in N$, but search and matching frictions mean that each job seeker will receive exactly one job offer during each hiring phase. Specifically, a ball-urn matching procedure randomly matches pairs of job seekers and firms (in a procedure similar

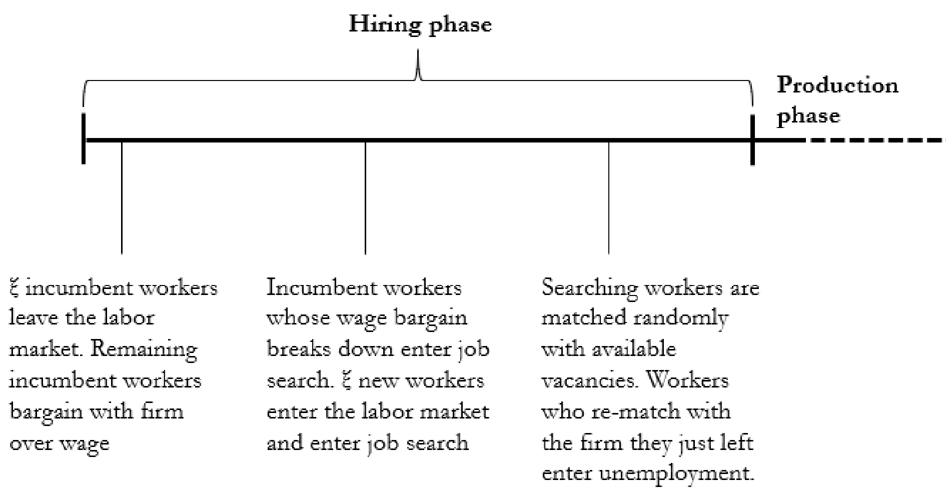
---

and worker both wish to maximize their joint surplus from the match, where the surplus generated is the difference between the product of the match p_i and the worker's outside option oo_i. The specific bargaining problem which generates the Nash outcome in our framework is one where the wage satisifes $w_i = argmax_w (w_i - oo_i)^{\beta}(p_i - w_i)^{(1-\beta)}$, as shown in Jaeger, Schoefer, Young and Zweimueller (2020) or Manning (2011). (This is a particularly simple formula which arises in part from the assumption that the firm's outside option is zero).

[71] This assumption is not necessary for the qualitative direction of our conclusions but keeps things simple.

[72] There is a large literature on the role of internal vs. external factors in determining the wages of new hires as compared to incumbent workers. For some examples: Bewley (1999), in interviews with firms in New England, assembles a range of evidence that conceptions of internal fairness and equity are extremely important in wage setting. Galuscak, Keeney, Nicolitsas, Smets, Strzelecki and Vodopivec (2012) use firm-level data in 15 EU countries and find strong evidence that firms do not wish to differentiate between the wages of newly hired workers and similarly qualified incumbents even if external labor market conditions change. Our assumption that new hires receive a 'take-it-or-leave-it' wage offer conforms to the survey evidence of Hall and Krueger (2012), who find in a survey of 1,300 US workers that two thirds of workers considered their offers to be 'take-it-or-leave-it' and did not bargain over the wage.

to the framework developed by Jarosch, Nimczik and Sorkin (2019)). For a given job seeker i, define the probability of receiving an offer from each firm j as α_{ij}. The probability of receiving no offers is therefore $1 - \sum_j \alpha_{ij}$. If job seekers do not accept the job offer they receive, or if they receive no job offers, they remain unemployed for the period, receiving unemployment benefit b. We assume b is strictly lower than the wages offered by any feasible employer, such that a job seeker who receives an offer always accepts it. In our model this condition will always be satisfied as long as the productivity of all jobs is greater than the unemployment benefit: this is because the wage offered to each worker by firm j is equivalent to the wage bargained by the existing workers at firm j, which itself is strictly greater than unemployment benefit b. [73] The timing of the wage bargain and job search process is illustrated in the figure below.



Outside option value for employed workers. The wage at each firm is determined by the bargain with employed workers, which in turn depends partly on the value of the outside option for these employed workers. What is this outside option value? The outside option for an employed worker bargaining with her employer is to leave her current job and become a job seeker. She does not know with certainty what her outcome will be as she will be matched with at most one feasible job if she leaves her current job. Her expected wage if she leaves her job is therefore a weighted average of the wages paid by each firm

[73] The analysis above assumes that workers care only about money. If we consider instead utility (which may include a value of leisure or work), the productivity of all jobs must be greater than the money-equivalent utility value of unemployment (including the unemployment benefit and any utility or disutility of unemployment relative to work) for all workers.

j, w_j, weighted by the probability of being matched with each firm j, α_j, as well as the unemployment benefit b multiplied by the probability of receiving no job offers $1 - \sum_{j\neq i}^{N} \alpha_j$:

$$oo_i = \sum_{j\neq i}^{N} \alpha_j \cdot w_j + \left(1 - \sum_{j\neq i}^{N} \alpha_j\right) \cdot b \tag{14}$$

We assume that each firm-worker match has weakly positive surplus, such that the bargained wage is always weakly greater than the outside option value. This means that, in equilibrium, no bargaining session will break down.[74]

Equilibrium wage. The bargained wage for workers at firm i (which is also paid to new hires at firm i) satisfies:

$$\begin{aligned} w_i &= \beta p_i + (1-\beta) oo_i \\ &= \beta p_i + (1-\beta)\left(\sum_{j} \alpha_{j\neq i} \cdot w_j + \left(1 - \sum_{j\neq i} \alpha_j\right) \cdot b\right), \end{aligned} \tag{15}$$

where the sums are over the set of firms N in the labor market. What do the probabilities of being matched with each feasible firm, α_j, correspond to? Since each job seeking worker is randomly matched with one vacancy, and there are equal numbers of vacancies and job searchers, the probability of the offer a job seeker receives being from a particular firm j is proportional to the share of vacancies posted by that firm j, as a share of all vacancies in the labor market, σ_j. This is in the same spirit as Burdett and Mortensen (1980) and Jarosch et al. (2019) who use the employment share to proxy for the likelihood of getting a job offer from a given firm.

In a labor market with infinitesimally small firms, this procedure would lead to every job seeker receiving a match from a feasible employer each period. However, in a labor market with some large employers, a worker who left firm j at the start of the period has some non-zero chance of being re-matched with firm j in the job search process (in fact, her chance of being re-matched with her former employer is σ_j). Knowing this, during the wage bargain process a large employer can threaten *not* to re-employ their worker if that worker leaves the firm but is re-matched with them in the job search process. We assume that this threat is credible and is exercised: if bargaining breaks down between a given worker and her employer j, and if the random job search process re-matches that worker to her former employer j, employer j will refuse to re-hire her (as in Jarosch et al., 2019) and she will instead have to move to unemployment for the period.[75] These assumptions imply the following wage

[74] Note that we assume complete information about the outside option for both the worker and the firm.

[75] In a single-period game, this strategy would not be time-consistent for the employer, as they are unable to fill the vacancy for the period and so lose the potential for positive surplus from re-hiring the worker. However, in a multiple period game with repeated interactions between firms and workers, and/or firm

equation for a worker at firm i:

$$w_i = \beta p_i + (1-\beta)\left(\sum_{j\neq i}\sigma_j \cdot w_j + \sigma_i \cdot b\right) \tag{16}$$

Note that the outcome of the wage bargain reached by workers at firm i depends on the outcome of the wage bargain reached by workers at all other local firms j, but the wage bargained by workers at firms j depends on the wage bargained by workers at firm i. To solve this "reflection problem", we iteratively substitute for w_j. The wage expression becomes

$$\begin{aligned} w_i = &\beta p_i + \beta(1-\beta)\sum_{j\neq i}\sigma_j p_j + \beta(1-\beta)^2\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k p_k + ... \\ &+ (1-\beta)\sigma_i b + (1-\beta)^2\sum_{j\neq i}\sigma_j^2 b + (1-\beta)^3\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k^2 b + ... \end{aligned} \tag{17}$$

expanded out to the third order, where the ellipses (...) signify further expansions to higher orders.

Since we are interested in the *average* wage in the labor market, we then take the average wage across all firms in the labor market, $\bar{w} = \sum_i \sigma_i w_i$. This gives us the expression:

$$\begin{aligned} \bar{w} = &\beta\left(\sum_i \sigma_i p_i + (1-\beta)\sum_i\sigma_i\sum_{j\neq i}\sigma_j p_j + (1-\beta)^2\sum_i\sigma_i\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k p_k + ...\right) \\ &+ (1-\beta)b\left(\sum_i\sigma_i^2 + (1-\beta)\sum_i\sigma_i\sum_{j\neq i}\sigma_j^2 + (1-\beta)^2\sum_i\sigma_i\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k^2 + ...\right) \end{aligned} \tag{18}$$

To simplify this expression, define $\bar{p} = \sum_i \sigma_i p_i$ as the average productivity across firms, and denote $\hat{p}_j = p_j - \bar{p}$ as the difference between firm j's productivity and the market average. In addition, define the rth order concentration index Ω_r as the concentration index with r "steps" as in the expression

$$\Omega_r = \underbrace{\sum_i \sigma_i \sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k ... \sum_{m\neq n}\sigma_m\sum_{p\neq m}\sigma_p^2}_{\text{with } r \text{ summation terms or "steps" in the expression}} \tag{19}$$

such that the first order concentration index Ω_1 is the sum of the squared employer shares (the HHI: $\Omega_1 = \sum_i \sigma_i^2$), the second order concentration index is $\Omega_2 = \sum_i \sigma_i \sum_{j\neq i}\sigma_j^2$, and so

reputations, it can be in the firm's interest to avoid re-hiring a worker who has quit the firm in order to maintain the firm's reputation in future bargaining rounds with this or other workers - even if this comes at the cost of an unfilled vacancy in the current period. A less stylized setting producing a similar outcome could be a setting with on-the-job search, where workers can receive job offers from firms other than their own, and when they do so, they can use these as outside options to bargain for a higher wage. In this setting, workers at the largest firms will receive the fewest outside offers.

on. Also define $\Omega_0 = 0$. We can then rewrite the average wage equation (18) as

$$\bar{w} = \beta\bar{p}\left(1 + \sum_{n=1}^{\infty}(1-\beta)^n\left(1 - \sum_{r=1}^{n}\Omega_r\right)\right) + (1-\beta)b\left(\sum_{n=1}^{\infty}(1-\beta)^{n-1}\Omega_n\right)$$
$$-\beta\left((1-\beta)\sum_i \sigma_i^2\hat{p}_i + (1-\beta)^2\sum_i\sigma_i\sum_{j\neq i}\sigma_j^2\hat{p}_j + (1-\beta)^3\sum_i\sigma_i\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k^2\hat{p}_k + ...\right) \tag{20}$$

That is, the average wage in a given labor market is a function of three terms: the average productivity in the labor market $\bar{p}$, multiplied by a function of worker bargaining power and employer concentration; the employment benefit b, multiplied by a function of worker bargaining power and employer concentration; and a third term which reflects the interaction between employer share σ_j and employer relative productivity $\hat{p}_j$.

The first and second terms of our wage expression illustrate that the wage declines as average employer concentration increases: as different aspects of average employer concentration increase, Ω_r increases, meaning that less weight in the wage equation falls on average productivity ($\bar{p}$) and more weight falls on the unemployment benefit (b). This in turn is because large employers can credibly threaten not to re-hire workers who quit, reducing these workers' bargaining power by making their outside option worse (i.e. making it more likely that they will enter unemployment if they quit their firm). So, higher employer concentration suppresses wages by worsening workers' outside option in the wage bargain.

Note also that the relationship between employer shares σ_j and firm productivity p_j factors into the wage in both the first and third terms. Since average productivity $\bar{p}$ is determined by the productivity of each employer, and the share of that employer in the labor market, the first term illustrates that average productivity will be higher and therefore the wage will be higher if the high productivity firms are also the large firms. However, the third term mitigates this effect somewhat: it reflects the fact that, if it is the largest firms which are the most productive, the passthrough of average productivity to average wages via the outside option channel will be lower than if it is the smallest firms which are the most productive, because the largest firms are in fewer workers' outside option set.

To obtain the average wage expression we cite in the main body of the paper, we take a second order approximation of equation (20) in the employer shares σ_j (i.e. removing all terms with σ_j^n where $n > 2$). This reduces the expression to become a function of the squares of the employer shares – and, therefore, the commonly-used Herfindahl Hirschmann Index or HHI:

$$\bar{w} = \beta\bar{p} + (1-\beta)\bar{o}o \tag{21}$$

$$= \beta\bar{p} + (1-\beta)\left((1-HHI)\cdot\bar{p} + HHI\cdot b\right) - \beta(1-\beta)\sum_i \sigma_i^2\hat{p}_i \qquad (22)$$

That is, this expression suggests that the wage is a weighted average of average productivity $\bar{p}$ and the value of unemployment b. Average productivity both affects the portion of the wage determined by the internal product of the worker's firm (the first term), and factors into the value of the worker's outside option (as it affects the average value of getting a job at another firm). Within the outside option term, higher employer concentration increases the weighting on the unemployment benefit b, and decreases the weighting on the productivity of labor $\bar{p}$, relative to a world with no employer concentration. If all firms have the same productivity $p_i = \bar{p}\forall i$, or if the correlation of market shares and firm productivity is small, the last term is close to zero and the wage is simply a concentration- and bargaining power-weighted average of productivity p and unemployment benefit b.[76]

Note, then, that one might mean two different things when one asks "what is the effect of employer concentration on wages?". First, one might be asking "what is the effect of this labor market having become more concentrated, relative to the past?". On the one hand, rising concentration exerts downward pressure on wages by worsening worker outside options. On the other hand, rising concentration may well exert upward pressure on wages by increasing average productivity if the firms with growing market shares are relatively more productive.

Alternatively, one might be asking "what is the effect of the employer concentration in this labor market, in terms of suppressing workers' wages below their productivity?". In that case, the answer to this question takes the degree of productivity in the labor market *as given* and looks only at the effect of employer concentration in reducing wages below that level of productivity by worsening outside options. In a sense, this is trying to *isolate* the effect of employer concentration on outside options from its potential effect on productivity. This *latter* question is the one we are focusing on in this paper.

[76]Note: in our framework, for simplicity the *only* way the worker can end up unemployed is if they do not get matched with a firm, and the only way they do not get matched with a firm is if their 'match' in the random matching process is their current employer, who refuses to re-hire them. This means that in an unconcentrated labor market ($HHI = 0$) in this model, there would be no unemployment and the average wage would equal the average product $\bar{w} = \bar{p}$. On the other hand in a labor market with only one firm ($HHI = 1$) the wage would be $\beta\bar{p} + (1-\beta)b$, the Nash bargaining formula when the only outside option for workers is unemployment. In our framework, therefore, the only reason there is a markdown of the wage from the marginal product is because there is employer concentration, which results in some probability of becoming unemployed if bargaining breaks down with thew worker's own firm. This simplification is for clarity of exposition only; one could extend this framework to incorporate steady state unemployment even in an unconcentrated labor market, but where higher employer concentration increases the probability of unemployment if a match breaks down.

Incorporating outside-occupation options

The conceptual framework explained above, however, assumes that all jobs and workers are perfectly substitutable in a clearly delineated labor market. As discussed in this paper, this is rarely the case in practice. Workers can switch between jobs in different occupations and locations, but differentially so for different options. In this paper, we focus on occupations within a given metro area. One can easily extend our framework to incorporate also the option to move metro area.[77]

Ideally, we would be able to delineate which firms in outside occupations are in a worker's feasible labor market for any given worker, and estimate the probability that a worker would receive a job in that outside occupation. In practice, we cannot. To work with commonly-available data definitions and publicly-available data, we must instead work with the occupational definitions from the SOC 6-digit classification scheme. We therefore extend our framework above by defining the primary labor market for workers as their local occupation o and incorporating an outside option term reflecting the value of moving to jobs in other occupations. Refer back to our equation (16) for the bargained wage in firm i, given by

$$w_i = \beta p_i + (1-\beta)\left(\sum_{j \neq i} \sigma_j \cdot w_j + \sigma_i \cdot b\right)$$

but now edit this equation to reflect explicitly that some feasible jobs are in workers' own occupation o whereas some are in other occupations p (where employer share $\sigma_{j,o}$ now refers to firm j's share of vacancies *within* occupation o), such that the wage for workers in firm i and occupation o is:

$$w_{i,o} = \beta p_i + (1-\beta)\underbrace{\zeta_o \sum_{j \neq i}^{N_o} \sigma_{j,o} w_{j,o}}_{\text{own-occ options}} + (1-\beta)\underbrace{(1-\zeta_o)\sum_{p \neq o}^{N^{occs}} Prob(o \to p) \sum_{l}^{N_p} \sigma_{l,p} w_{l,p}}_{\text{outside-occ options}} + (1-\beta)\underbrace{\zeta_o \sigma_{i,o} b}_{\text{unemployment}}$$

Here, N_o denotes the set of firms in occupation o, and N^{occs} denotes the set of occupations. This expression states that the bargained wage for workers in firm i in occupation o is a function of their productivity at firm i $p_{i,o}$, the outside option value of moving to other

[77] While the outside option to move location certainly matters, the data suggests that for most workers this will be less important than the outside option to get a job in another occupation in workers' current metro area. Occupational mobility is substantially higher than geographic mobility: only around 3% of U.S. workers move between metropolitan areas each year (according to IRS county-to-county migration data), and over 80% of job applications by workers are sent to jobs within their metropolitan area Marinescu and Rathelot (2018). In addition, geographic mobility has declined over time (Molloy, Smith and Wozniak, 2011), and the proliferation of state-level occupational licensing has made geographic mobility more difficult for many workers (Johnson and Kleiner, 2020). There are, however, a subset of more mobile workers – like highly educated professionals – for whom the failure to consider job options outside their own metro area may be more problematic.

firms j in their own occupation o, the outside option value of moving to other firms l in other occupations p, and the outside option value of moving to unemployment and receiving benefit b.[78]

The value of the outside option in equation A is still the weighted average of the wage in each other local firm. The weights for firms within workers' occupation o are now a product of (i) the probability that a worker from occupation o will be matched with a job within occupation o (ζ_o), and (ii) the probability that, conditional on staying in occupation o, the worker will be matched with firm j (which we once again proxy for using the vacancy share of firm j in occupation o). For firms outside workers' occupation, similarly, the value of the outside option is a weighted average of wages, where the weights are a product of (i) the probability that a worker from occupation o will be matched with a job outside occupation o $(1-\zeta_o)$, (ii) the probability that conditional on leaving their occupation o, a worker will be matched with *some* job in occupation p ($Prob(o \to p)$), and (ii) the probability that conditional on being matched with a firm in occupation p, the worker will be matched specifically with firm j ($\sigma_{j,p}$).[79]

To take this expression to the data, we note that if employers' vacancy shares are relatively similar to their current employment shares, $\sum_l^{N_p} \sigma_{l,p} \cdot w_{l,p}$ can be approximated simply by the average wage in local occupation p. This eliminates the need for us to consider the reflection problem that wages in occupation o affect wages in occupation p and vice versa - instead, we can use data on the average wage in each local occupation p to control directly for the effect of wages in occupation p on occupation o (as we do in our empirical implementation). Next, we define the value of outside-occupation job options as the probability-weighted wages in other occupations, $oo_o^{occs} = \sum_{p \neq o}^{N^{occs}} Prob(o \to p) \sum_l^{N_p} \sigma_{l,p} \cdot w_{l,p}$. As in the simple framework, we iteratively substitute for wages in other local firms j in the same occupation o, rearrange, take the average wage in the local occupation, and write the resulting wage expression in terms of our higher order concentration indices $\Omega_{r,o}$ (where subscript o denotes that this is the concentration index for employers in local occupation o). This gives us an expression for the average wage in occupation o:

$$\begin{aligned}\bar{w}_o = (\beta \bar{p}_o + (1-\beta) oo_o^{occs}) \left(1 + \sum_{n=1}^{\infty} (1-\beta)^n \zeta_o^n \left(1 - \sum_{r=1}^{n} \Omega_{r,o}\right)\right) \\ + b \left(\sum_{n=1}^{\infty} (1-\beta)^n \zeta_o^n \Omega_{n,o}\right)\end{aligned}$$

[78]Note also that this expression assumes implicitly that each firm only employs workers of one occupation (or, alternatively, that the worker's own firm i can only refuse to re-employ that worker if she is re-matched with a job in her initial occupation o, but not if she is re-matched with firm i with a job in a new occupation p.

[79]Note: this assumes that employment decisions are taken at the firm-by-occupation level.

$$- \beta \left((1-\beta)\zeta_o \sum_i \sigma_i^2 \hat{p}_{i,o} + (1-\beta)^2 \zeta_o^2 \sum_i \sigma_i \sum_{j \neq i} \sigma_j^2 \hat{p}_{j,o} + ... \right) \tag{23}$$

Once again taking a second order approximation in employer shares – that is, only considering concentration index $\Omega_{1,o} = \sum_i \sigma_{i,o}^2 = HHI_o$ – we can simplify the expression for the average wage in occupation o to become

$$\begin{aligned} \bar{w}_o = &(1-(1-\beta)\zeta_o HHI_o)\left(\alpha \bar{p}_o + (1-\alpha) oo^{occs}\right) + (1-\beta)\zeta_o HHI \cdot b \\ &- \beta(1-\beta)\zeta_o \sum_i \sigma_i^2 \hat{p}_{i,o} \end{aligned} \tag{24}$$

where $\alpha = \frac{\beta}{1-\zeta_o(1-\beta)}$.

Ignoring the final term, which is very small if the average productivity of individual firms is not strongly correlated with their vacancy shares, the average wage in occupation o is a weighted average of the average productivity in occupation o, $\bar{p}_o$, the value of jobs outside occupation o, oo_o^{occs}, and unemployment benefit b. The weights are a function of worker bargaining power β, the probability that workers are matched with another firm in their own occupation if they leave their job ζ_o, and employer concentration HHI_o.

As before, employer concentration within workers' own occupation increases the relative likelihood of workers ending up unemployed if they quit their job, increasing the weighting on b the unemployment benefit in the wage bargain and reducing the weighting on other jobs (productivity $\bar{p}_o$ and the value of moving to outside options outside workers' own occupation oo^{occs}). In addition, though, there is now an interaction with $\zeta_o = Prob(o \to o)$, the likelihood of the worker staying in her occupation if she leaves her job. The more likely she is to stay in her own occupation if she leaves her job – i.e., the less likely she is to be able to find a job in a different occupation – the more employer concentration in her own occupation matters for her wage. Finally, as before, note that there is an interaction with worker bargaining power β. The more bargaining power a worker has over the match surplus, the less the outside option matters in the wage bargain and therefore the less employer concentration matters for the wage.[80]

B 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 3.2). (We also use a different data set from BGT – the resume data set – to construct our measures of occupational mobility. We discuss the BGT resume

[80]Note that if there is no possibility of finding a job in another occupation (i.e. $\zeta_o = 1$, and $oo_o^{occs} = 0$), this expression becomes identical to the expression for the average wage in the first part of this section, equation (21).

data set in more detail in Appendix C.)

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).

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 (2018) 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".[81] 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), 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

[81] 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%). We show in Appendix Table A8 Panel C that, just as there is a strong correlation between local occupational wages and the HHI (even with occupation-year and metro area-year fixed effects), there is also a strong correlation between local occupational wages and the inverse number of firms in a local occupation.

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 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.

C 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.[82]
- Eliminate all jobs held before 2001.
- Eliminate all resumes with non-US addresses.
- Eliminate any jobs which are listed as having lasted less than 6 months, to ensure that we are only capturing actual jobs rather than short-term internships, workshops etc.

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

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

To identify occupational transitions from year to year, we match all sequential pairs of worker-job-occupation-year observations. For instance, if a worker had a job as a Purchasing Manager in the period 2003-2005, and a job as a Compliance Officer in 2005-2007, we would record sequential occupation patterns of the form shown in the table below.

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

Illustrative example of sequential job holding data.

Year:	2004	2005	2006
Occ. in year t	*Occ. in year t+1*		
Purchasing Mgr. (11-3061)	11-3061		
	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).[83] 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, and 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 under the same employer). The median

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

length job was 2 years, with the 25th percentile just under 1 year and the 75th percentile 4 years. The median span of years we observe on a resume (from date started first job to date ended last job) is 12 years. Table A2 shows more information on the distribution of job incidences and job durations on our resumes.

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 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.[84]

[84]The occupations with more than 2 million observations are: General and Operations Managers; Sales

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 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.

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.

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.[85]

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

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

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 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? 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 slightly 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 slightly 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.

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.

D 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.[86] 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.[87]

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.[88]

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 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

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

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

[88] 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.

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.

E Appendix: Alternative approaches to estimating occupational similarity

In Section 3.2 of this paper, we define workers' baseline labor market as a SOC 6-digit occupation within a metropolitan area.[89] 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 (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

[89]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.

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.

F Appendix: Determinants of occupational mobility

In section 3.2 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.[90]

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

[90] 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.

[91] 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.

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:

$$\pi_{o\to p} = \alpha_o + \beta^{abs}|X_{occ\,p} - X_{occ\,o}| + \gamma|\Delta w_{o\to p}| + \epsilon_{op}. \tag{25}$$

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,[92] 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 3.[93]

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 (25), but now using the *relative* (target minus origin) difference in occupational characteristics as the independent variable:

$$\pi_{o \to p} = \alpha_o + \beta^{rel}(X_{occ\ p} - X_{occ\ o}) + \gamma \Delta w_{o \to p} + \epsilon_{op}. \tag{26}$$

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 (26) for the different measures are shown in Figure A11. (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 with more routine task requirements;[94] and (4) that workers have on average been moving

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

[93] 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.

[94] 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).

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. Appendix Table A3 shows the adjusted R-squared statistics from regressions of $\pi_{o \to p}$ on our measures of skill distance, wage difference, amenity difference (temporal flexibility), leadership difference, and a composite skill measure. In all of these cases, while the correlation is strong and positive, 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.

G 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 (forthcoming), 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

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

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.[95] In our baseline specification, we use only large firms j in this instrument, 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 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 (5), the exclusion restriction for the instrument on the HHI concentration index is then equivalent to

$$Cov[Z^{HHI}_{o,k,t}, \xi_{o,k,t} | \Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{o}^{N^{occs}} \sigma^2_{j,o,k,t-1} \tilde{G}^{\perp}_{j,o,k,t} \xi_{o,k,t} \right] \to 0$$

where $\tilde{G}^{\perp}_{j,o,k,t}$ represents $\tilde{G}_{j,o,k,t}$ 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

[95] 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.

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,

$$\mathbb{E}[\sum_j \sigma^2_{j,o,k,t-1}\tilde{G}_{j,o,k,t}|\xi_{o,k,t},\Gamma_{kt},\Gamma_{ot},g_{o,k,t},\tilde{g}_{o,k,t},e_{o,k,t}] = \tau_1\Gamma_{kt}+\tau_2\Gamma_{ot}+\tau_3 g_{o,k,t}+\tau_4\tilde{g}_{o,k,t}+\tau_5 e_{o,k,t} \quad \forall o \in N^{occs}$$

for some constant parameters τ_1, τ_2, τ_3, τ_4, and τ_5. 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.[96]

Second, there needs to be a large number of independent firm-level shocks, that is,

$$\mathbb{E}[(\tilde{G}_{j,o,k,t} - E[\tilde{G}_{j,o,k,t}])(\tilde{G}_{j,o,k,t} - E[\tilde{G}_{j,o,k,t}])|\phi_{pt},\phi_{jt},\Gamma_{kt},\Gamma_{ot}] = 0$$

for all $p, j \in N^{occs}$ if $p \neq j$.

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 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

[96] 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).

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.[97] 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, for a given employer, there is a negative relationship between the initial vacancy share in an occupation-metro area labor market and the next year's vacancy growth rate, but this relationship is not sufficiently strong to invalidate our first stage (for each one percentage point increase in the initial vacancy share, there is roughly a 1.4 percentage point lower vacancy growth rate from one year to the next). Empirically, our first stage holds for occupation-metro area labor markets with HHIs above all but very low levels.

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 3, 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, $\frac{s_{p,k,1999}}{s_{p,1999}}$,[98] and the national occupation transition shares from the

[97]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; and Employer X has 65% of the market in another metro area, with the rest of the market comprised of atomistic firms. 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 (only 6% of which have HHIs of greater than 6,800 – and only 2% of workers in our data face HHIs of greater than 5,000). In practice our first stage regressions are positive and strongly significant even when segmenting our data to analyze only cells with high HHIs of 2,500 of more, meaning this concern is not hugely relevant in practice for our empirical analysis.

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

worker's occupation o to each of the other occupations, $\pi_{o\to p}$:

$$Z^{oo}_{o,k,t} = \sum_{p}^{N_{occs}} \left(\pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\not{k},t} \right) \tag{27}$$

To make the assumptions transparent under which this wage instrument identifies the coefficient on our outside-occupation option index in equation (5), we again follow the framework presented in Borusyak et al. (forthcoming).[99] Note that we can write the instrument as

$$Z^{oo}_{o,k,t} = \sum_{p=1}^{N^{occs}} s_{okp}\bar{w}_{p,\not{k},t}$$

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 (5), the exclusion restriction for the instrument for outside-occupation options is then equivalent to

$$Cov[Z^{oo}_{o,k,t}, \xi_{o,k,t}|\Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = \sum_{t=1}^{T}\sum_{p=1}^{N^{occs}} \bar{s}_{okp} w^{\perp}_{p,\not{k},t} \phi^{oo}_{pt} \to 0$$

where $\bar{s}_{okp} = \mathbb{E}[s_{okp}]$ is the average exposure to occupation p, and $\phi^{oo}_{pt} \equiv \mathbb{E}[s_{okp}\xi_{o,k,t}]/\mathbb{E}[s_{okp}]$ is an exposure-weighted expectation of the structural wage residuals. Moreover, $w^{\perp}_{p,\not{k},t}$ represents $\bar{w}_{p,\not{k},t}$ 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 the concentration index and control variables.

Borusyak et al. (forthcoming) 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. That is,

$$\mathbb{E}[\bar{w}_{p,\not{k},t}|\phi^{oo}_{pt}, \Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = \tau_1\Gamma_{kt} + \tau_2\Gamma_{ot} + \tau_3 g_{o,k,t} + \tau_4\tilde{g}_{o,k,t} + \tau_5 e_{o,k,t} \quad \forall p \in N^{occs}$$

for some constant parameters τ_1 through τ_5. Second, there needs to be a large number of independent occupational shocks, that is,

$$\mathbb{E}[(\bar{w}_{p,\not{k},t} - \mu)(\bar{w}_{j,\not{k},t} - \mu)|\phi_{pt}, \phi^{oo}_{jt}, \Gamma_{kt}, \Gamma_{ot}, g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}] = 0$$

for all $p, j \in N^{occs}$ if $p \neq j$, and also $\sum_{p=1}^{N^{occs}} \bar{s}^2_p \to 0$.

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

[99]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.

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.[100] The inclusion of these fixed effects increases our confidence that the assumptions for instrument validity hold.

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 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

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

[100] 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.

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 Panel A of Table A11.

H Appendix: Stata commands

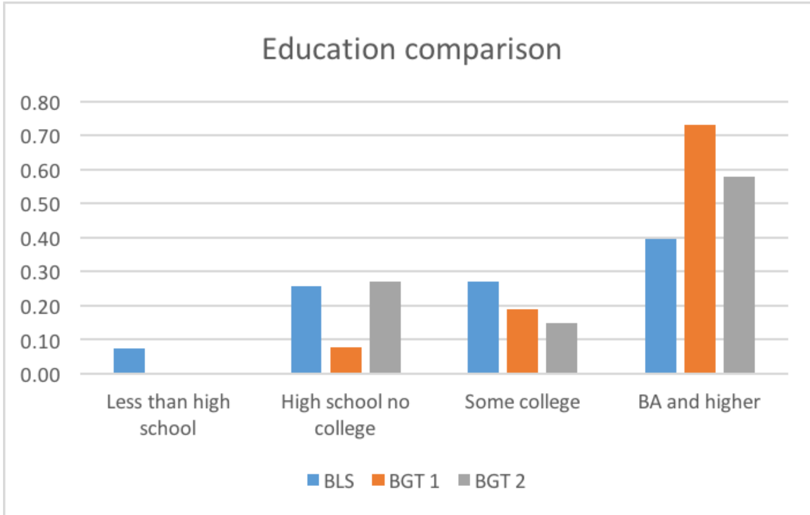
In our estimation, we used a number of user-written Stata commands: *reg2hdfe* (Guimaraes and Portugal, 2010), *reghdfe* (Correia, 2016), *ivreg2hdfe* (Bahar, 2014), *binscatter* (Stepner, 2013), *binscatter2* (Droste, 2019), and *coefplot* (Jann, 2013).

I Appendix: Figures

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

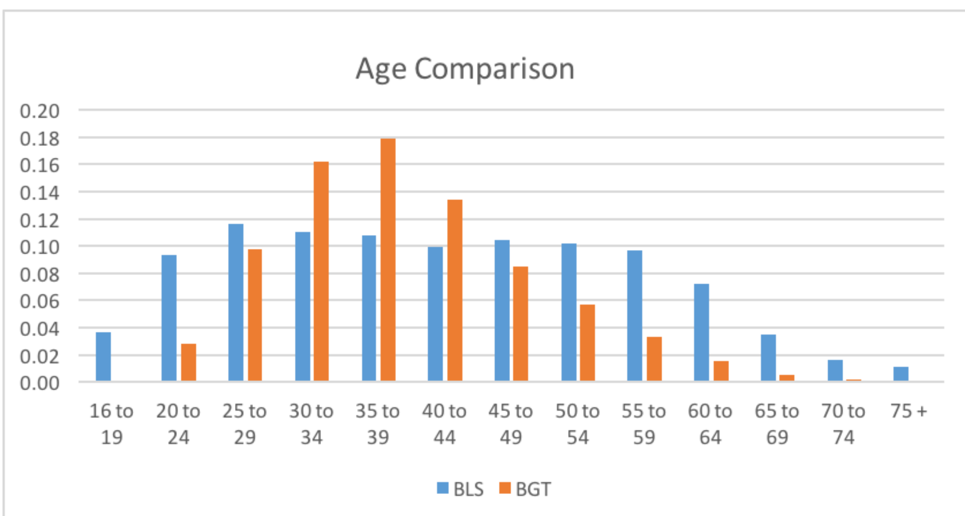
Note: Comparison of distribution of share of vacancies accounted for by each SOC 6-digit occupation in the BGT vacancy data in 2019, relative to that occupation's share of total employment in 2019 in the BLS occupational employment statistics. Occupations comprising greater than 1% share of either data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

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



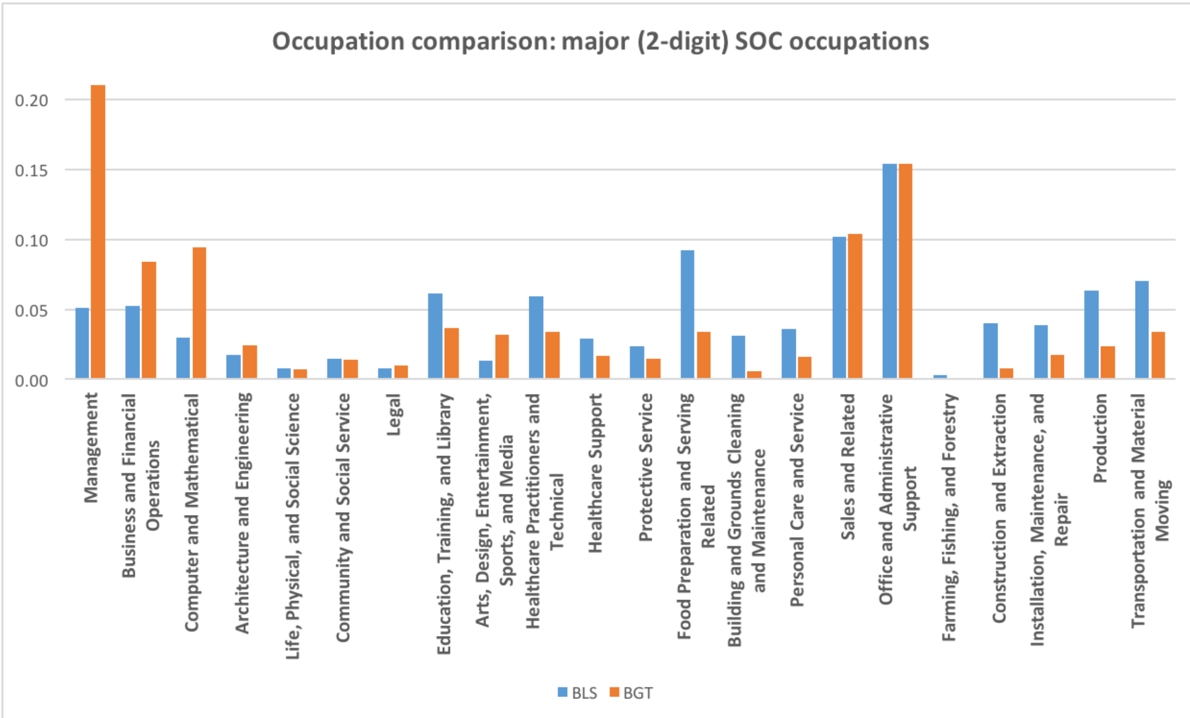
Note: Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT resume data. Two versions are shown: BGT 1 excludes all resumes missing educational information, while BGT 2 assumes all resumes missing educational information have high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

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



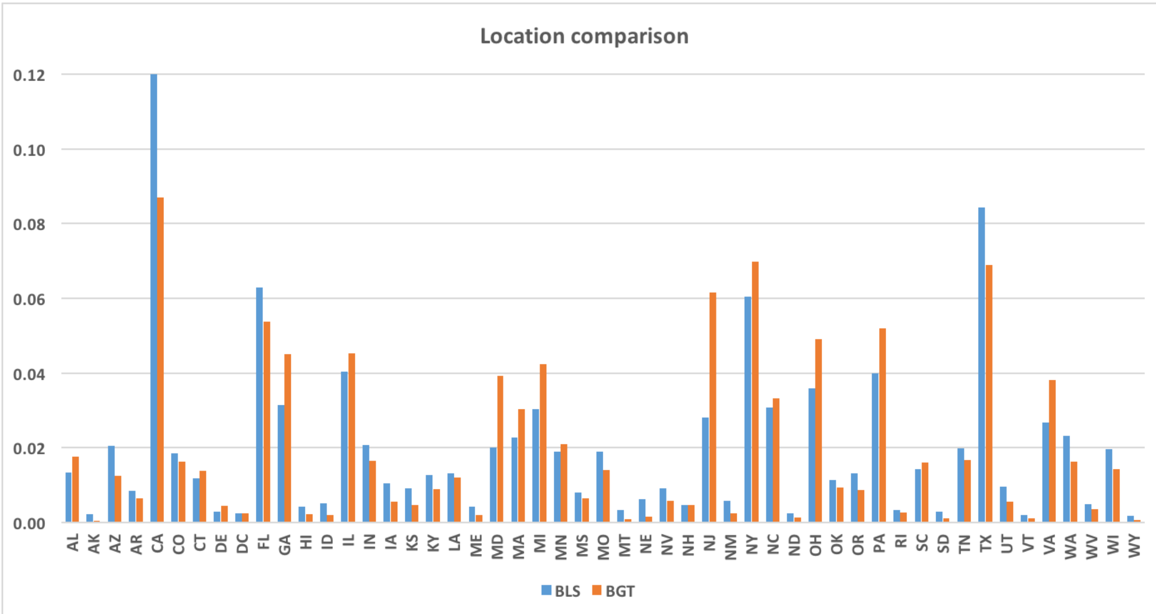
Note: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT resume data. The resume data is discussed in detail in Appendix C.

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



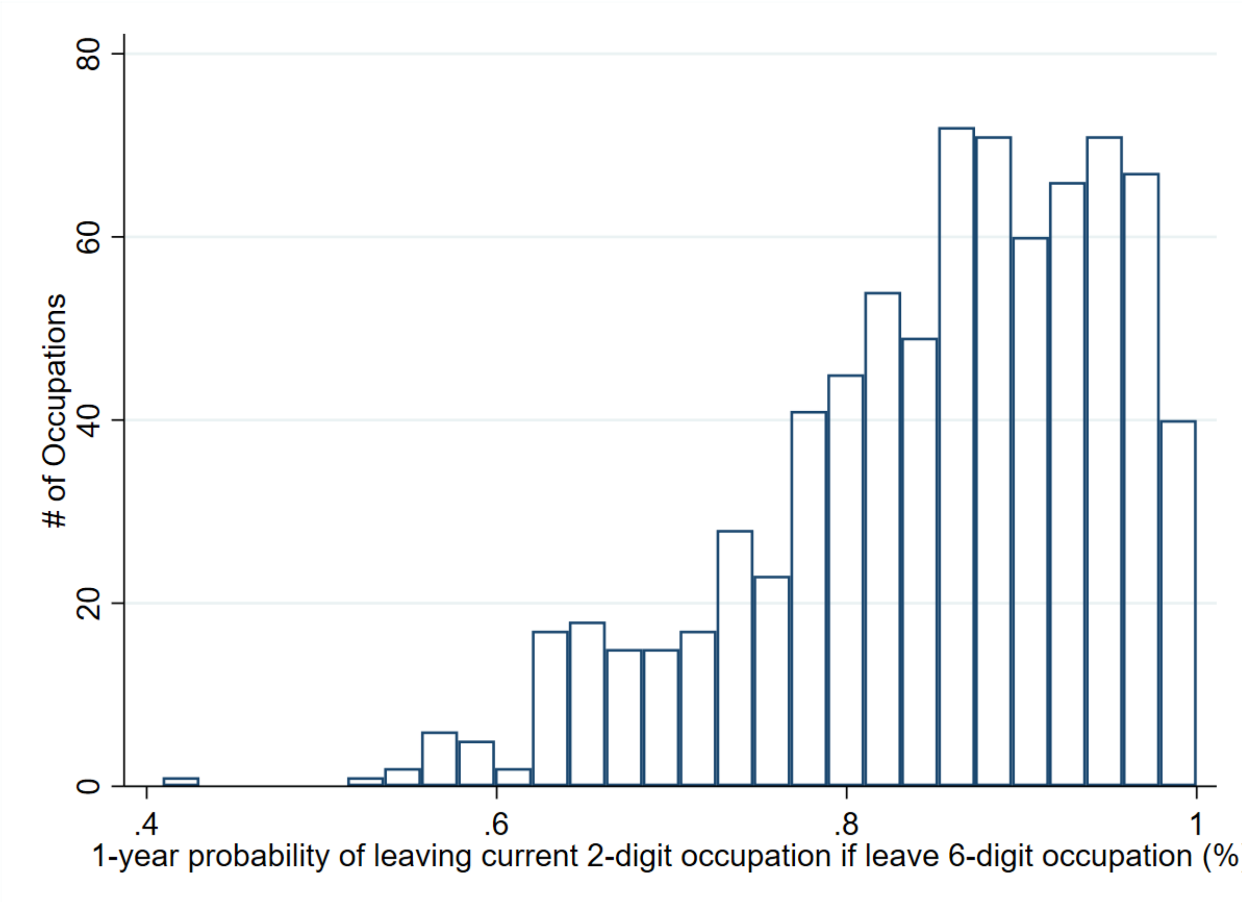
Note: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT resume data. The resume data is discussed in detail in Appendix C.

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



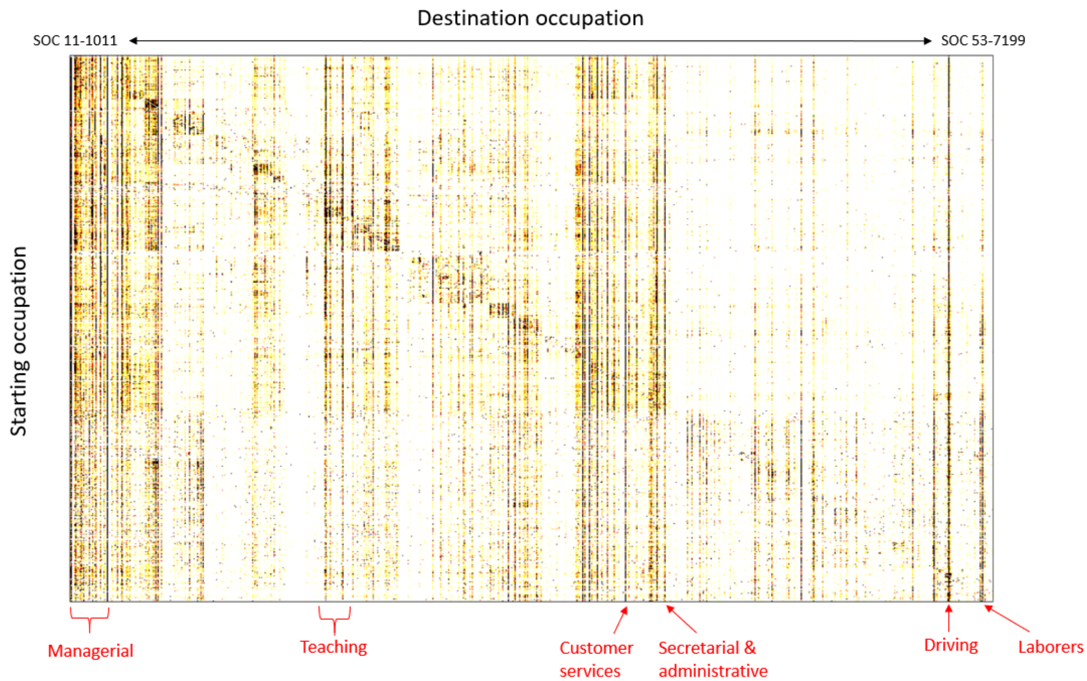
Note: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT resume data. Graph shows share of total in each state. The resume data is discussed in detail in Appendix C.

Figure A6: Occupational mobility: SOC 6-digit moves that are also 2-digit moves



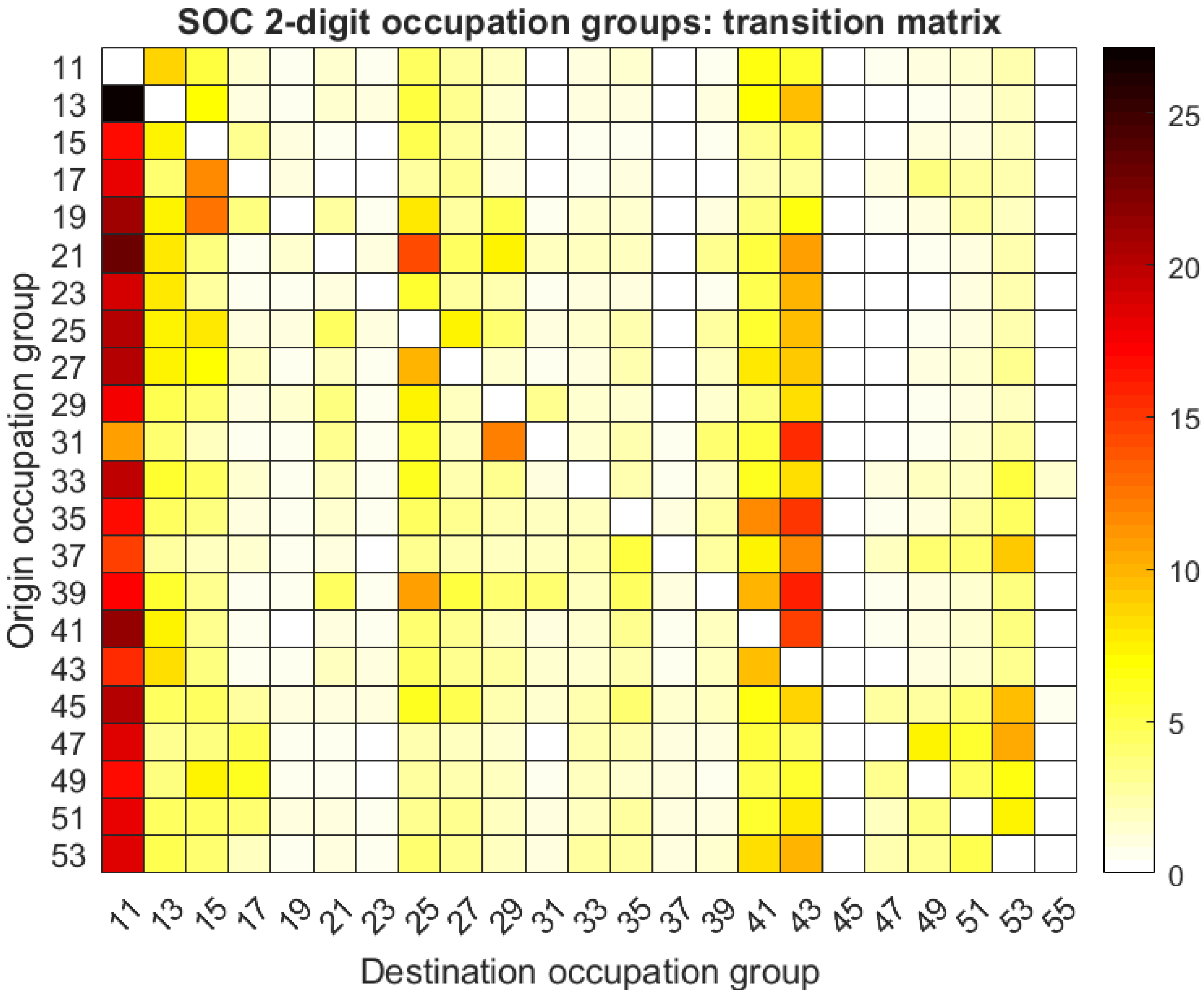
Note: Distribution of the proportion of workers moving 6-digit SOC occupation who *also* move 2-digit SOC occupation, by occupation, calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations. The resume data is discussed in detail in Appendix C.

Figure A7: 6-digit SOC occupational transition matrix



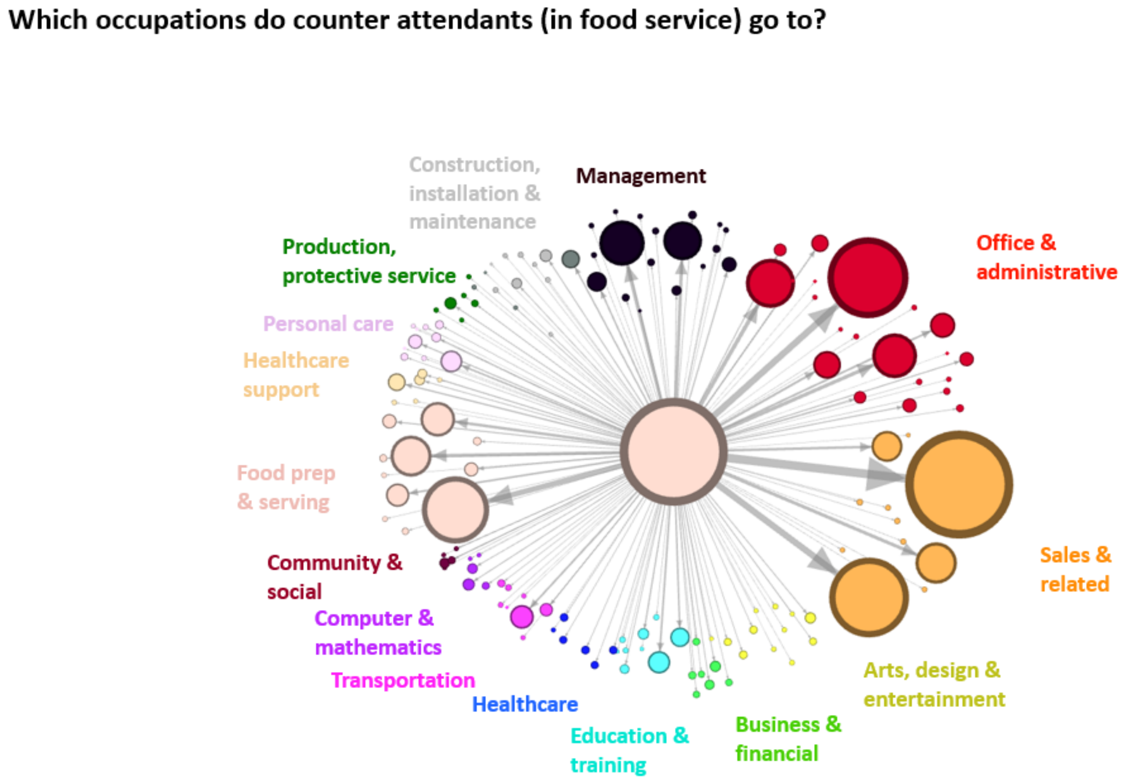
Note: Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial job. Occupations are sorted in SOC numerical order. Cells colored black have a transition probability of 1% or greater conditional on leaving the initial job. Transitions 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 C.

Figure A8: 2-digit SOC occupational transition matrix



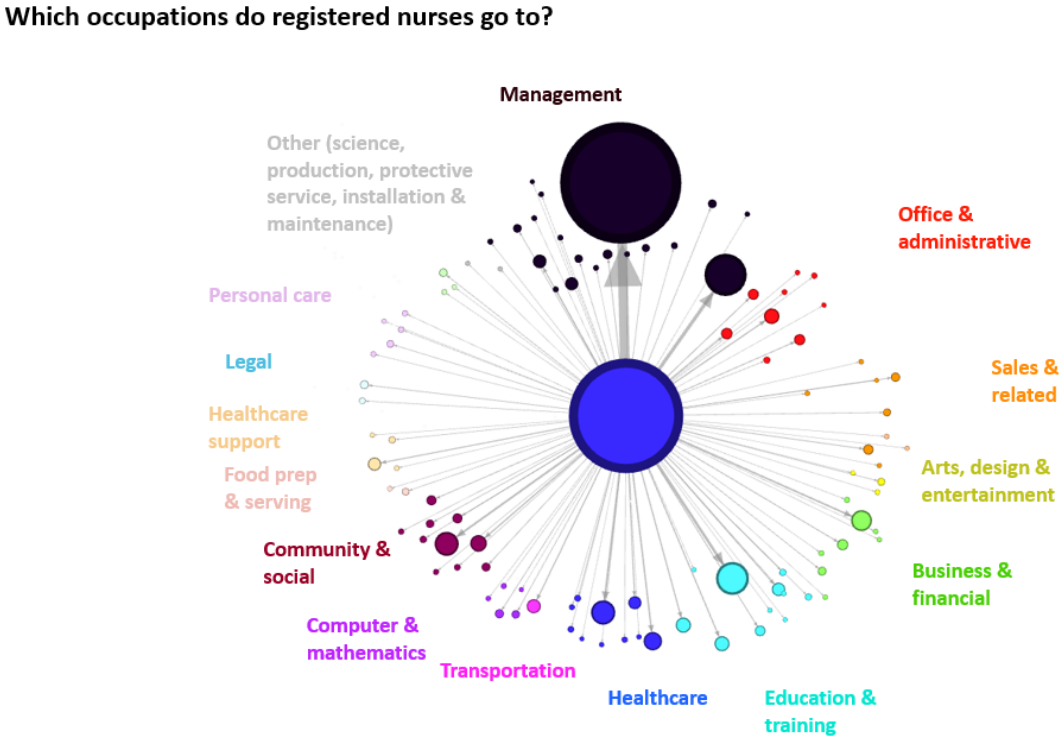
Note: Occupational transition matrix showing transition probability between 2-digit SOC occupation groups conditional on leaving the initial job. Cells colored black have a transition probability of 25% or greater conditional on leaving the initial job. Job transitions within an occupation group are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

Figure A9: Examples of probabilistic labor markets: counter attendants



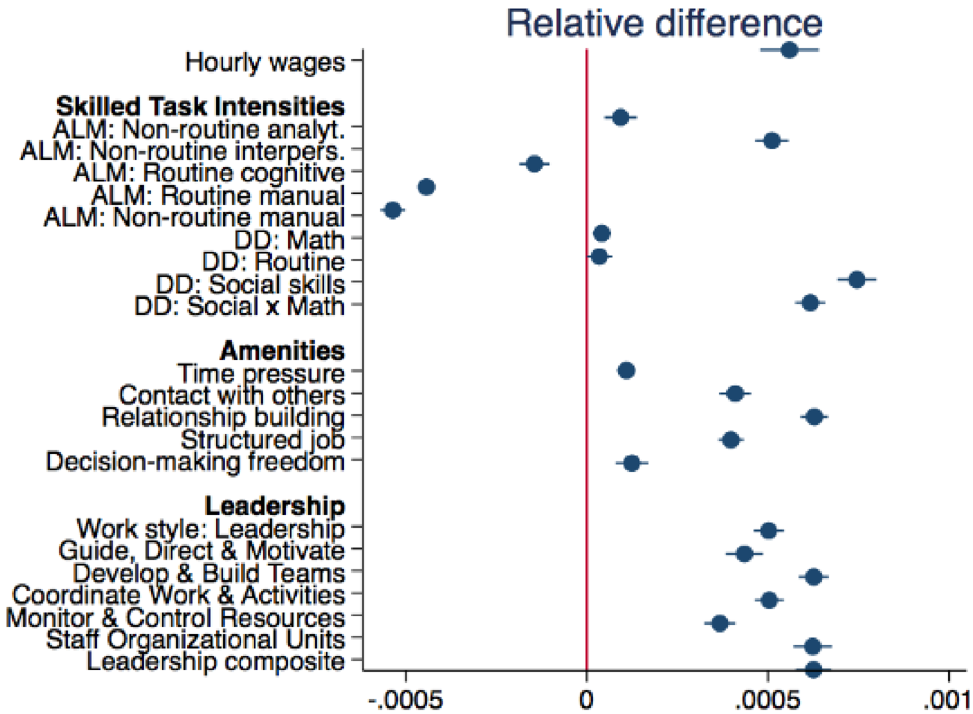
Note: Example visualization of occupational transitions for counter attendants in the food industry. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of counter attendants in the BGT data who switch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

Figure A10: Examples of probabilistic labor markets: registered nurses



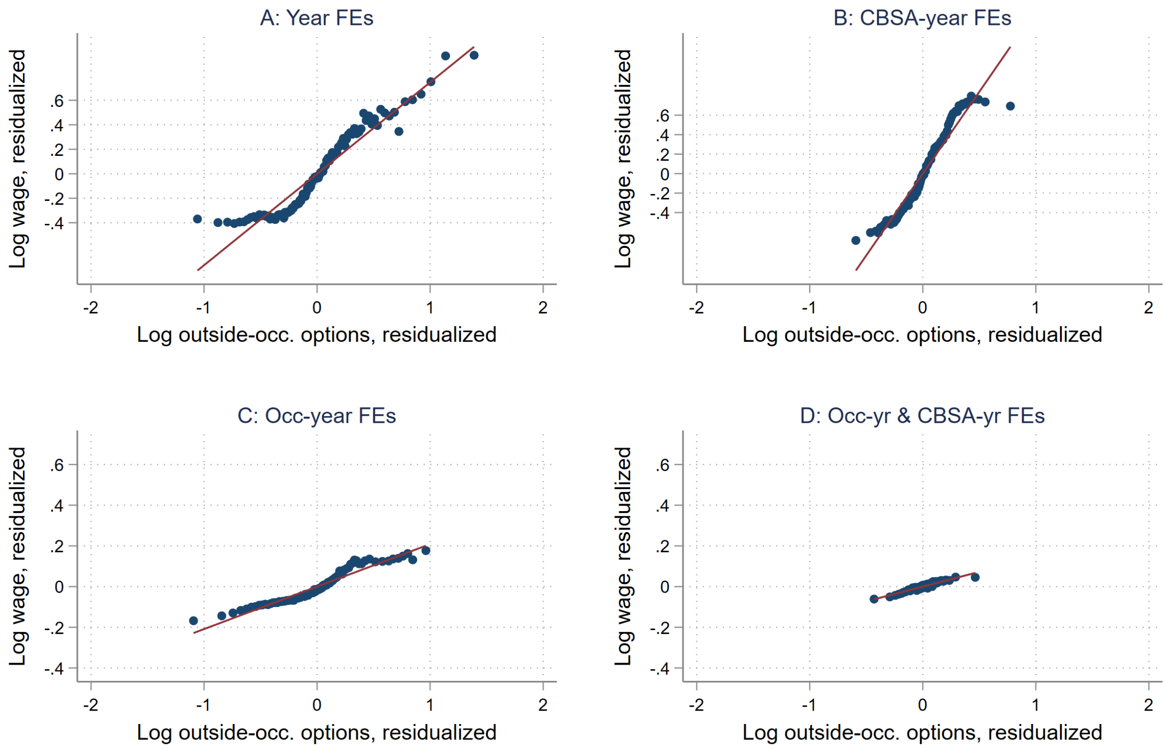
Note: Example visualization of occupational transitions for registered nurses. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of registered nurses in the BGT data who switch occupation, who are observed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

Figure A11: Determinants of occupational mobility



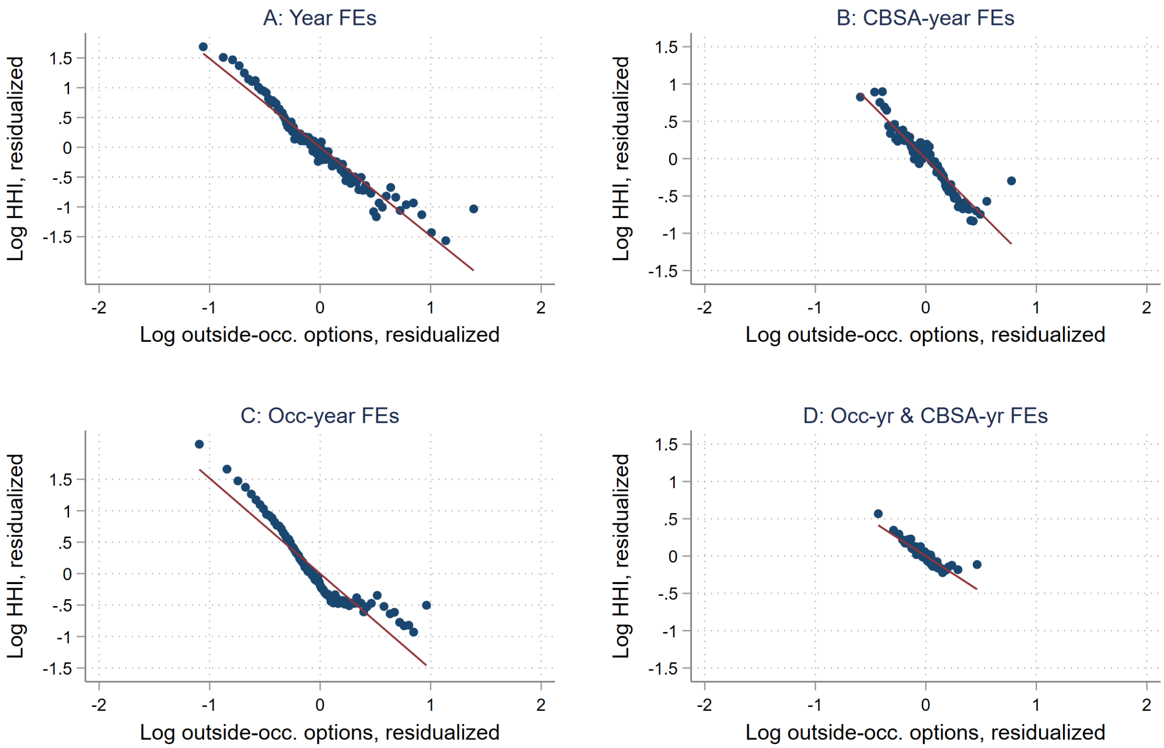
Note: This plot shows the coefficients and 95% confidence intervals from regressions of occupation transition shares $\pi_{o\to p}$, calculated from Burning Glass Technologies Resume data, on *relative* differences in occupational characteristics: $\pi_{o\to p} = \alpha_o + \beta f(X_{occ\ o\to p}) + \gamma f(\Delta w_{o\to p}) + \epsilon_{op}$ where the function $f(\cdot)$ represents the difference in characteristic between starting occupation o and destination p, and α_o is occupation o fixed effect. Regressions also include absolute avg. hourly wage differences (except for the amenities regressions). Standard errors are clustered at the origin occupation level. This is the analog of Figure 3, which shows coefficients on the regression of occupation transition shares on the *absolute* difference in characteristics between the pairs of occupations. These analyses are discussed in more detail in Appendix F.

Figure A12: Correlations between wage and outside-occupation option index



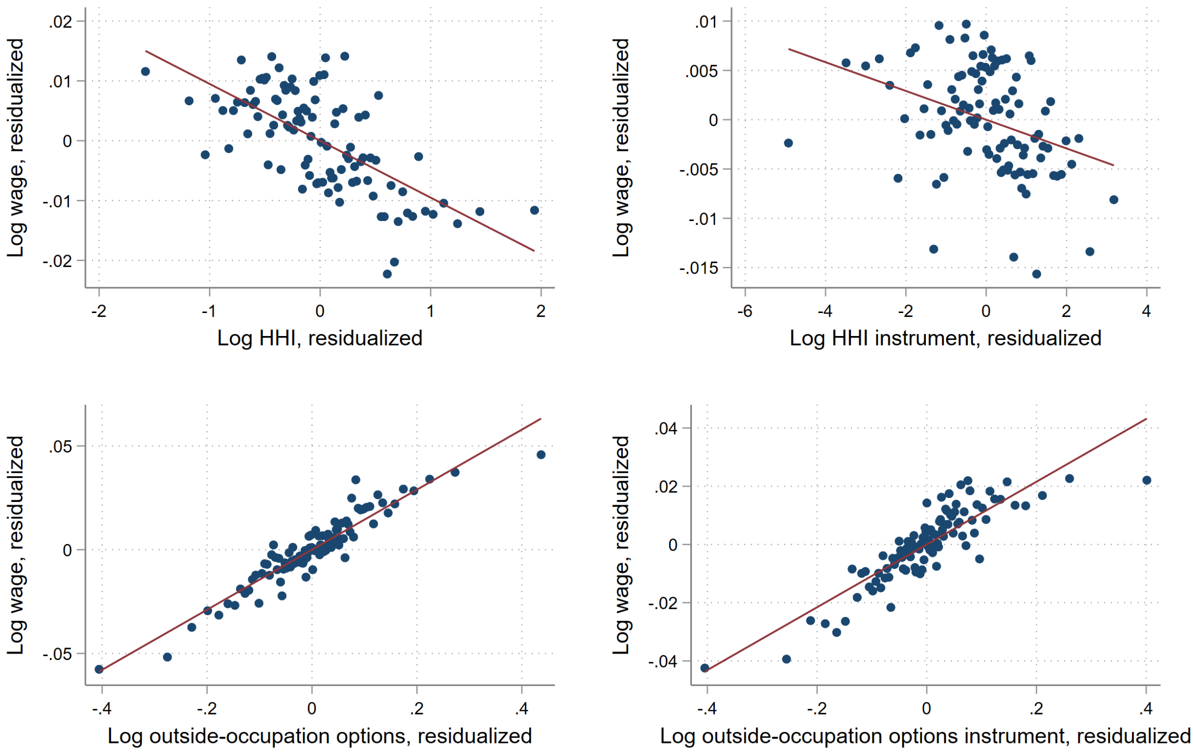
Note: Figure shows binned scatter plots of the employment-weighted relationship between average log wages and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on different combinations of fixed effects (as described by the panel titles). Regression coefficients 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 A13: Correlations between HHI and outside-occupation option index



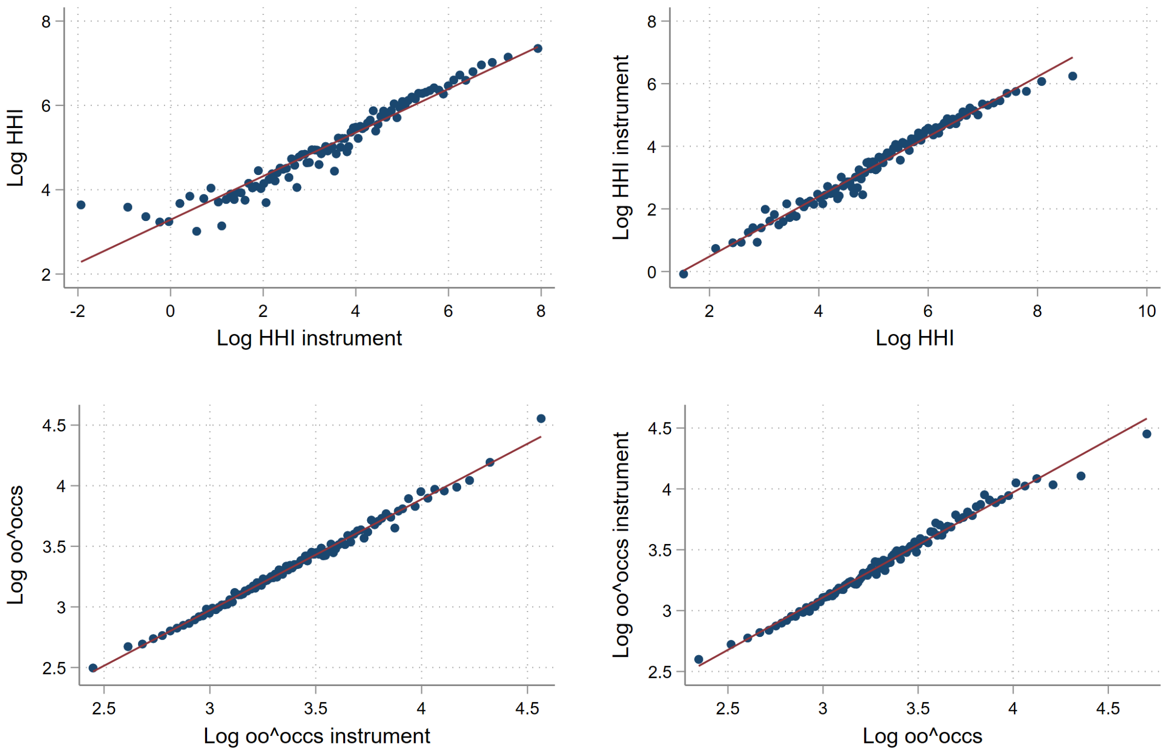
Note: Figure shows binned scatter plots of the employment-weighted relationship between average log HHI and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on different combinations of fixed effects (as described by the panel titles). Regression coefficients for the line of best fit on each graph are: A: -1.49, B: -1.48, C: -1.52; D: -0.96.

Figure A14: Visualization of baseline regression results



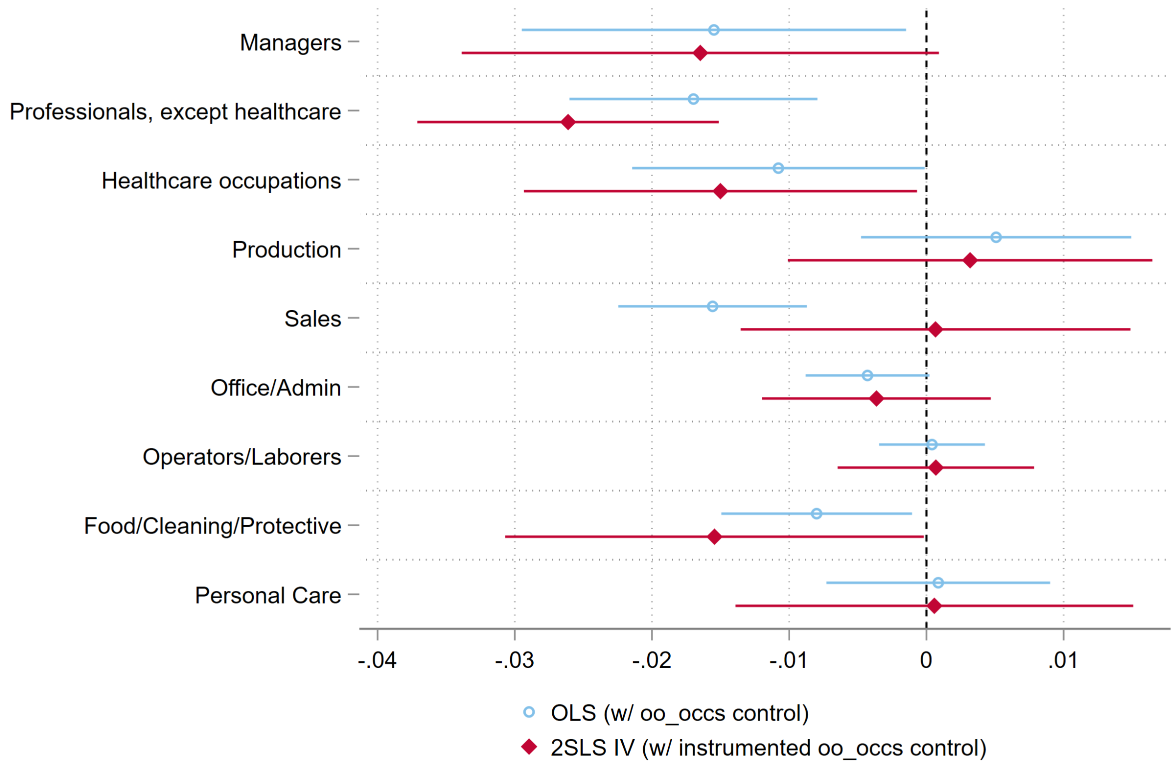
Note: Binned scatter plots of the log wage on log HHI and log outside-occupation options, and instrumented, including full set of fixed effects and controls, and employment-weighted, as in baseline regression specification (i.e. the left panels correspond to coefficient estimates in Table 3 column (b), and the right panels correspond to the reduced form equivalent of the 2SLS IV coefficient estimates in Table 3 column (d)).

Figure A15: Correlation between instruments and independent variables



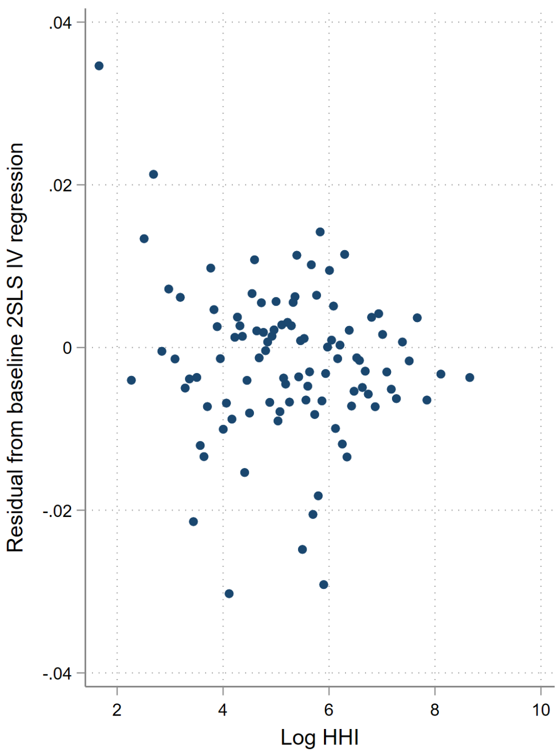
Note: Binned scatter plots of the employment-weighted relationships between the HHI instrument and raw variable (top panel) and outside-occupation option index instrument and raw variable (bottom panel) for occupation-metro area cells in 2019.

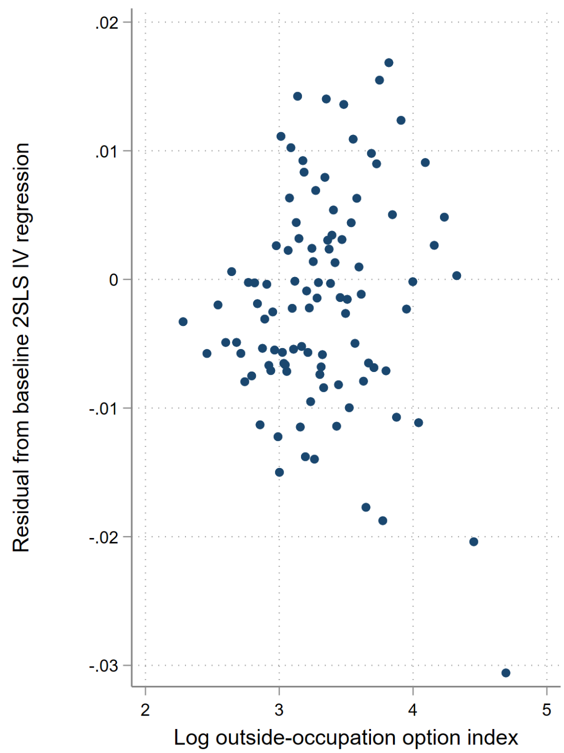
Figure A16: Coefficients on wage-HHI regressions: by occupation group



Note: Coefficients on HHI and 95% confidence intervals from regressions of occupation-metro area wages on instrumented local employer HHI, controlling for (instrumented) outside-occupation job options, with coefficient on HHI allowed to vary by occupation group. Regressions span 2011-2019, are employment-weighted, and include occupation-year and metro area-year fixed effects, as well as the other controls included in our baseline 2SLS IV regressions (described in Section 3). Standard errors are clustered at the metro area level. Occupation groups are listed in descending order of average wage. Occupation groups map from SOC 2-digit occupations as defined in Appendix Table A6. Coefficient estimates correspond to those in Appendix Table A15 column (a).

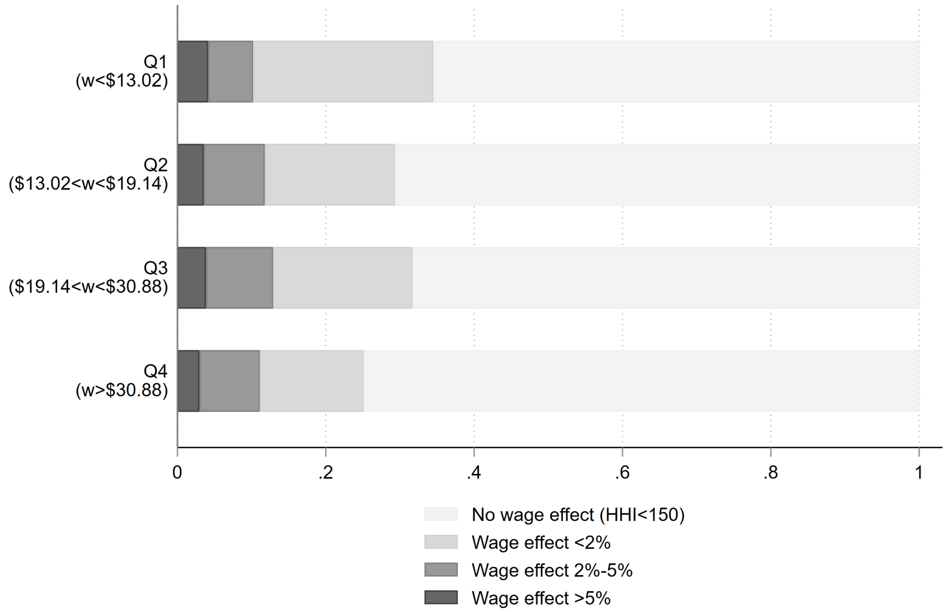
Figure A17: Baseline regressions: residual plots





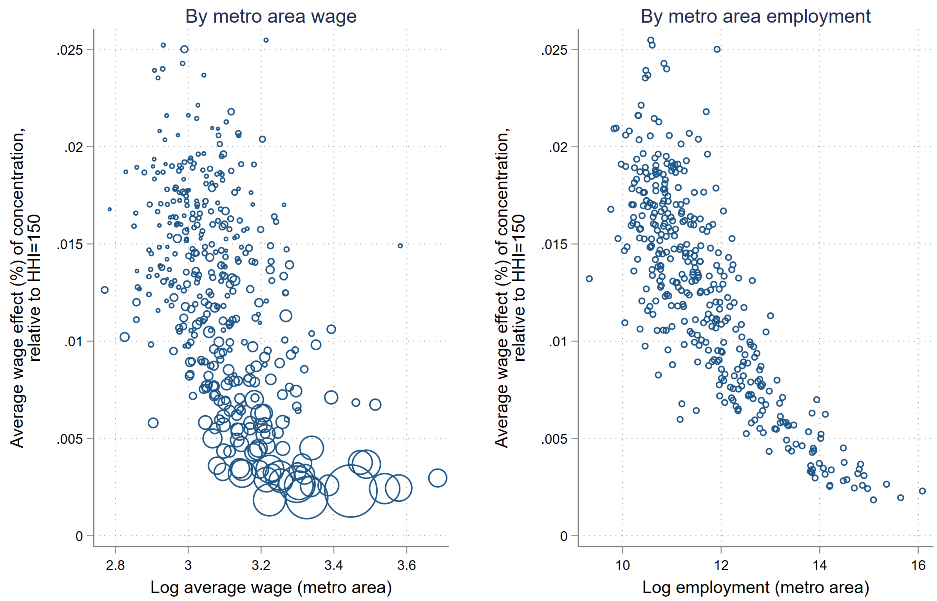
Binned scatter plots of residuals (employment-weighted), from baseline 2SLS IV employment-weighted regression of wage on log HHI and log outside-occupation option index, against the log HHI and log outside-occupation option index respectively. (The baseline regression results are reported in column (d) of Table 3).

Figure A18: Average estimated effect of employer concentration relative to HHI=150, by hourly wage



Note: This figure shows what share of workers may have experienced different degrees of wage suppression as a result of employer concentration, across the US hourly wage distribution in 2019, relative to a counterfactual where HHI was 150 (holding all else constant). We estimate the wage effect of employer concentration as described in section 5.

Figure A19: Average estimated 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 5: 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.

J Appendix: Tables

Table A1: Summary statistics in 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: Distribution of number of jobs on resume and duration of jobs in BGT resume data set.

Percentile	10th	25th	50th	75th	90th
# Jobs on resume	2	3	4	6	9
Job duration (months)	4	12	24	48	98

Note: This table shows some summary statistics from the BGT resume data: the distribution of the number of jobs in each resume (across all 16 million resumes in our data set), and the distribution of average job duration in months (across all the jobs reported in our data set).

Table A3: Adjusted R-squared from regressions of occupational transitions on occupational similarity (based on tasks, skills, amenities, or wages)

Dependent variable:	$\pi_{o \to p}$	
Included characteristic	No FE	Incl. origin occupation FE
Skill distance	0.011	0.025
Wages	0.003	0.021
Job amenities	0.021	0.039
Leadership	0.017	0.033
Skill composites	0.035	0.058

Note: Table shows adjusted R-squared from regressions of the form $\pi_{o \to p} = \kappa + \alpha_o + \beta \Delta X_{occ\ p-o} + \epsilon_{op}$. Here, $\pi_{o \to p}$ is the share of job changers in the origin occupation o that move into target occupation p, and α_o are origin occupation fixed effects (included only in the second column). All regressions contain a constant. The variable $\Delta X_{occ\ p-o}$ represents the group of included characteristic differences noted in the table, which are included in relative target-minus-origin form and as absolute distances, with the exception of skill distance. All regressions are weighted by the average 2002-2015 national employment in the origin occupation. Note that the underlying occupational transition matrix is sparse, with many cells that show zero transitions, which is why the linear regression fit yields a relatively small R-squared. These analyses are discussed in more detail in Appendix Section F.

Table A4: 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 3.2) - 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 A5: 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 3.2). Only occupations with at least 500 observations in the BGT data and 2017 OES employment data are shown.

Table A6: Assignment of SOC 2-digit occupation groups into 10 larger occupation categories

Occupation category	2-digit SOC occupation group	SOC code
Managers	Management occupations	11-0000
Managers	Business and financial operations occupations	13-0000
Professionals, except healthcare	Computer and mathematical operations	15-0000
Professionals, except healthcare	Architecture and engineering occupations	17-0000
Professionals, except healthcare	Life, physical, and social science occupations	19-0000
Professionals, except healthcare	Community and social service occupations	21-0000
Professionals, except healthcare	Legal occupations	23-0000
Professionals, except healthcare	Education, training, and library occupations	25-0000
Professionals, except healthcare	Arts, design, entertainment, sports, and media occupations	27-0000
Healthcare occupations	Healthcare practitioners and technical support occupations	29-0000
Healthcare occupations	Healthcare support occupations	31-0000
Food, Cleaning, and Protective Service	Protective service occupations	33-0000
Food, Cleaning, and Protective Service	Food preparation and serving related occupations	35-0000
Food, Cleaning, and Protective Service	Building and grounds cleaning and maintenance occupations	37-0000
Personal Care	Personal care and service occupations	39-0000
Sales	Sales and related occupations	41-0000
Office/Administrative	Office and administrative support occupations	43-0000
Operators / Laborers	Farming, fishing, and forestry occupations	45-0000
Operators / Laborers	Transportation and material moving occupations	53-0000
Production	Construction and extraction occupations	47-0000
Production	Installation, maintenance, and repair occupations	49-0000
Production	Production occupations	51-0000

Note: Table shows our allocation of 2-digit SOC occupational groups into 9 larger occupation categories for the purpose of analyzing whether the relationship between wages and HHI differs across occupation groups. These occupation categories are drawn primarily from Acemoglu and Autor (2011).

Table A7: Large employer-occupation pairs with highest national growth in job postings

Year	Employer	Occupation	Total job postings	YoY growth
2013	Macy's	Retail salespersons	36,050	2.9
2014	C.R. England	Heavy and tractor-trailer truck drivers	87,000	5.1
2014	Family Dollar	First-line supervisors of retail sales workers	19,749	2.1
2014	Starbucks Coffee	First-line supervisors of food preparation and serving workers	52,316	4.0
2014	Starbucks Coffee	Counter attendants, cafeteria, food concession, and coffee shop	55,445	8.6
2015	Accenture	Computer occupations, all other	24,425	3.8
2015	Dollar Tree	First-line supervisors of retail sales workers	16,044	2.1
2017	Anthem	Medical and health services managers	20,828	2.1
2017	Anthem	Managers, all other	20,086	2.2
2017	Anthem	Management analysts	32,524	2.7
2017	Anthem	Computer occupations, all other	42,730	3.3
2017	CACI International	Software developers, applications	28,270	3.6
2017	CACI International	Computer occupations, all other	26,431	3.3
2017	CRST The Transportation Solution	Heavy and tractor-trailer truck drivers	50,417	2.0
2017	USA Truck	Heavy and tractor-trailer truck drivers	25,985	2.3
2018	Anthem	Registered nurses	68,571	3.7
2018	AutoZone	Retail salespersons	23,606	3.0
2019	Dollar Tree	First-line supervisors of retail sales workers	21,960	2.7
2019	Regis Corporation	Hairdressers, hairstylists, and cosmetologists	18,738	2.1
2019	Starbucks Coffee	First-line supervisors of food preparation and serving workers	31,184	4.0
2019	Starbucks Coffee	Counter attendants, cafeteria, food concession, and coffee shop	37,246	3.2

Note: Table shows all employer-occupation-year total annual job posting observations in Burning Glass Technologies data that (1) Experienced an increase of at least 10,000 postings since the previous year, (2) Had at least 5,000 postings in the previous year, and (3) Represent at least a 2% increase in postings YoY.

Table A8: Wage-Concentration correlations: Different measures of employer concentration

Dependent variable: log wage				
	(a)	(b)	(c)	(d)
Panel A: Log HHI				
Log HHI	-0.120***	-0.052***	-0.111***	-0.012***
	(0.007)	(0.007)	(0.006)	(0.002)
Panel B: HHI				
HHI*	-1.032***	-0.408***	-0.735***	-0.033***
	(0.114)	(0.067)	(0.086)	(0.009)
Panel C: Inverse number of employers				
1 / Num. employers	-0.744***	-0.488***	-0.180***	-0.019**
	(0.098)	(0.080)	(0.031)	(0.008)
Observations	445,975	445,681	445,975	445,681
Fixed effects	Year	Occ-yr	Metro area-yr	Occ-yr, Metro area-yr

Note: This table shows regressions of the log hourly wage on different measures of employer concentration, with different combinations of fixed effects. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019, inclusive, employment-weighted. HHI*=HHI/10,000 – rescaled so that the coefficients are legible. Inverse number of employers is the reciprocal of the number of unique employers in our vacancy data in an occupation-metro area cell in a given year. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A9: 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	0.133***				
	(0.006)				
Log outside-occ. options instrument	-0.848***				
	(0.067)				
Log HHI instrument		0.136***			
X Q1 outward mobility		(0.008)			
Log outside-occ options instrument		-0.519***			
X Q1 outward mobility		(0.135)			
Log HHI instrument			0.142***		
X Q2 outward mobility			(0.007)		
Log outside-occ options instrument			-0.951***		
X Q2 outward mobility			(0.092)		
Log HHI instrument				0.117***	
X Q3 outward mobility				(0.008)	
Log outside-occ options instrument				-0.846***	
X Q3 outward mobility				(0.058)	
Log HHI instrument					0.124***
X Q4 outward mobility					(0.007)
Log outside-occ options instrument					-0.969***
X Q4 outward mobility					(0.099)
Vacancy growth	-3.679**	-6.274***	-1.903	-7.405***	-5.599***
	(1.732)	(1.437)	(1.495)	(0.903)	(1.193)
Predicted vacancy growth	1.972***	2.763***	0.655*	3.513***	2.627***
	(0.338)	(0.808)	(0.344)	(0.837)	(0.742)
Exposure control	2.835***	3.460***	2.350***	3.317***	2.455***
	(0.104)	(0.164)	(0.116)	(0.152)	(0.092)
Observations	445,681	116,883	112,442	121,693	94,662

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 3.2). 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 A10: 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	0.828***				
	(0.027)				
Log HHI instrument	-0.001***				
	(0.000)				
Log outside-occ options instrument		0.760***			
X Q1 outward mobility		(0.033)			
Log HHI instrument		-0.001**			
X Q1 outward mobility		(0.001)			
Log outside-occ options instrument			0.866***		
X Q2 outward mobility			(0.029)		
Log HHI instrument			-0.000		
X Q2 outward mobility			(0.001)		
Log outside-occ options instrument				0.824***	
X Q3 outward mobility				(0.029)	
Log HHI instrument				-0.000	
X Q3 outward mobility				(0.000)	
Log outside-occ options instrument					0.834***
X Q4 outward mobility					(0.026)
Log HHI instrument					-0.001***
X Q4 outward mobility					(0.000)
Vacancy growth	-0.078	-0.127**	-0.038	-0.249***	0.016
	(0.048)	(0.055)	(0.038)	(0.079)	(0.057)
Predicted vacancy growth	0.094***	0.139**	0.071**	0.135	0.115***
	(0.033)	(0.056)	(0.029)	(0.093)	(0.042)
Exposure control	0.006	0.019**	0.014	0.023**	-0.032***
	(0.007)	(0.010)	(0.011)	(0.010)	(0.009)
Observations	445,681	116,883	112,442	121,693	94,662

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 3.2). 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 A11: Regression of wage on HHI and outside-occupation options: robustness

Dependent variable:	Log wage				
	(a) No vac controls	(b) Equal-wt vac	(c) No exposure control	(d) Industry Bartik	(e) Occ-Metro area FEs
Log HHI, instrumented	-0.010***	-0.010***	-0.010***	-0.008***	-0.002
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Log outside-occ. options, instrumented	0.120***	0.120***	0.120***	0.124***	0.044
	(0.018)	(0.018)	(0.018)	(0.019)	(0.027)
Exposure control	0.009	0.010		0.005	-0.000
	(0.014)	(0.014)		(0.015)	(0.006)
Vacancy growth		-0.105*	-0.102*	-0.216***	-0.100**
		(0.059)	(0.058)	(0.069)	(0.047)
Predicted vacancy growth		-0.001	-0.003	-0.002	-0.129***
		(0.038)	(0.038)	(0.051)	(0.027)
Equal-weighted vacancy growth		-0.000			
		(0.000)			
Industry Bartik				-0.015***	
				(0.004)	
Observations	445,681	445,681	445,681	403,512	429,282
F-Stat	200	203	293	194	3
Fixed effects	Occ-year Metro-year	Occ-year Metro-year	Occ-year Metro-year	Occ-year Metro-year	Occ-Metro Year

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) excludes our controls for local actual and predicted vacancy growth. Column (b) includes an additional control for equal-weighted vacancy growth of local firms in the relevant occupation. Column (c) excludes our HHI exposure control. Column (d) includes an industry Bartik shock to control for correlated industry shocks across occupation-metro area cells. All these specifications feature occupation-year and metro area-year fixed effects. Column (e) runs our baseline regression specification, but with occupation-metro area and 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. F-Stat is Kleibergen-Paap Wald F statistic. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A12: Regression of wage on HHI and outside-occupation options: robustness (2)

Dependent variable:	Log wage				
	(a) No emp. weight	(b) Drop low HHI	(c) Drop low rep. occs	(d) Occ rep. weight	(e) Metro area rep. weight
Log HHI, instrumented	-0.007***	-0.012***	-0.009***	-0.010***	-0.007***
	(0.002)	(0.004)	(0.003)	(0.003)	(0.002)
Log outside-occ. options, instrumented	0.088***	0.118***	0.131***	0.096***	0.089***
	(0.008)	(0.015)	(0.012)	(0.009)	(0.008)
Vacancy growth	-0.008	-0.097*	-0.199***	-0.002	-0.008
	(0.014)	(0.058)	(0.076)	(0.015)	(0.014)
Predicted vacancy growth	0.004	-0.023	0.076*	0.066	0.013
	(0.021)	(0.038)	(0.042)	(0.066)	(0.022)
Exposure control	0.004	0.015	0.007	0.022**	0.006
	(0.006)	(0.014)	(0.019)	(0.010)	(0.006)
Observations	445,681	438,789	329,842	445,681	445,681
F-stat	491	426	208	539	510

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. Column (a) does not weight by employment (instead giving each occ-metro area-year cell equal weight). Column (b) drops all occupation-metro area cells with an average HHI over 2011–2019 of less than 50. Column (c) drops all occupations with average represented-ness in the BGT vacancy data of 0.5 or less. Columns (d) and (e) weight the regressions by average represented-ness of the occupations, and metro areas, in the BGT vacancy data (respectively). 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. About one third of occupations in our data have occupation represented-ness of 0.5 or less in the BGT data. 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. All regressions are employment-weighted, unless otherwise specified (cols (a), (d), and (e)). F-Stat is Kleibergen-Paap Wald F statistic. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A13: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness checks

Dependent variable:	Log wage				
	(a) No vac controls	(b) Equal-wt vac	(c) No exposure control	(d) Industry Bartik	(e) Occ-Metro area FEs
Log HHI *X* Q1 outward mobility	-0.028*** (0.006)	-0.028*** (0.006)	-0.027*** (0.006)	-0.022*** (0.006)	-0.018*** (0.007)
Log HHI *X* Q2 occ mobility	-0.012*** (0.003)	-0.012*** (0.003)	-0.012*** (0.003)	-0.011*** (0.003)	0.012 (0.009)
Log HHI *X* Q3 occ mobility	-0.002 (0.004)	-0.002 (0.004)	-0.001 (0.003)	-0.001 (0.004)	0.010** (0.004)
Log HHI *X* Q4 occ mobility	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.002 (0.003)	0.005 (0.019)
Log outside-occ options *X* Q1 occ mobility	0.097*** (0.017)	0.097*** (0.017)	0.097*** (0.017)	0.104*** (0.016)	-0.238 (0.310)
Log outside-occ options *X* Q2 occ mobility	0.135*** (0.015)	0.135*** (0.015)	0.136*** (0.016)	0.139*** (0.016)	0.226 (0.261)
Log outside-occ options *X* Q3 occ mobility	0.119*** (0.024)	0.119*** (0.024)	0.120*** (0.024)	0.119*** (0.026)	0.102* (0.058)
Log outside-occ options *X* Q4 occ mobility	0.107*** (0.031)	0.107*** (0.031)	0.108*** (0.031)	0.106*** (0.033)	0.135 (0.211)
Exposure control	0.013 (0.013)	0.013 (0.013)		0.006 (0.014)	-0.003 (0.016)
Vacancy growth		-0.107* (0.061)	-0.103* (0.060)	-0.223*** (0.069)	-0.056 (0.086)
Predicted vacancy growth		0.001 (0.037)	-0.002 (0.037)	0.001 (0.047)	-0.147*** (0.050)
Equal-weighted vacancy growth		-0.000 (0.000)			
Industry Bartik				-0.016*** (0.004)	
Observations	445,681	445,681	445,681	403,512	429,282
F-stat	47	48	68	45	.09
Fixed effects	Occ-year Metro-year	Occ-year Metro-year	Occ-year Metro-year	Occ-year Metro-year	Occ-Metro Year

Note: This table repeats the robustness checks in Table A11, 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 3.2). 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. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A14: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness checks (2)

Dependent variable:	Log wage				
	(a) Emp weight	(b) Log emp weight	(c) Drop low rep. occs	(d) Occ rep. weight	(e) Metro area rep. weight
Log HHI *X* Q1 occ mobility	-0.017*** (0.003)	-0.029*** (0.007)	-0.021*** (0.006)	-0.014*** (0.004)	-0.016*** (0.003)
Log HHI *X* Q2 occ mobility	-0.010*** (0.002)	-0.014*** (0.004)	-0.009** (0.004)	-0.014*** (0.004)	-0.010*** (0.002)
Log HHI *X* Q3 occ mobility	-0.000 (0.003)	-0.004 (0.004)	0.001 (0.003)	-0.004 (0.004)	-0.001 (0.003)
Log HHI *X* Q4 occ mobility	0.002 (0.003)	-0.000 (0.004)	-0.004 (0.003)	0.001 (0.004)	0.001 (0.003)
Log outside-occ options *X* Q1 occ mobility	0.064*** (0.009)	0.095*** (0.015)	0.106*** (0.013)	0.067*** (0.010)	0.069*** (0.009)
Log outside-occ options *X* Q2 occ mobility	0.085*** (0.008)	0.136*** (0.013)	0.146*** (0.014)	0.102*** (0.009)	0.088*** (0.008)
Log outside-occ options *X* Q3 occ mobility	0.102*** (0.008)	0.119*** (0.019)	0.130*** (0.012)	0.109*** (0.010)	0.100*** (0.008)
Log outside-occ options *X* Q4 occ mobility	0.103*** (0.008)	0.105*** (0.026)	0.143*** (0.013)	0.107*** (0.011)	0.099*** (0.008)
Vacancy growth	-0.008 (0.015)	-0.100* (0.060)	-0.219*** (0.080)	-0.000 (0.014)	-0.008 (0.015)
Predicted vacancy growth	0.002 (0.021)	-0.022 (0.038)	0.083* (0.044)	0.061 (0.065)	0.011 (0.022)
Exposure control	0.004 (0.006)	0.016 (0.012)	0.017 (0.018)	0.019* (0.010)	0.006 (0.006)
Observations	445,681	438,789	329,842	445,681	445,681
F-stat	116	100	55	128	121

Notes: This table repeats the robustness checks in Table A12, 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 3.2). Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. Units of observation are 6 digit occupation 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. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A15: Regression of wage on HHI and outside options, by occupation group

	OLS	2SLS IV
Log HHI, instrumented X Managers	-0.015**	-0.016*
	(0.007)	(0.009)
Log outside-occ. options, instrumented X Managers	0.136***	0.139***
	(0.022)	(0.019)
Log HHI, instrumented X Professionals, excl managers	-0.017***	-0.026***
	(0.005)	(0.006)
Log outside-occ. options, instrumented X Professionals, excl managers	0.120***	0.088***
	(0.013)	(0.022)
Log HHI, instrumented X Healthcare	-0.011**	-0.015**
	(0.005)	(0.007)
Log outside-occ. options, instrumented X Healthcare	0.052***	0.007
	(0.012)	(0.020)
Log HHI, instrumented X Production	0.005	0.003
	(0.005)	(0.007)
Log outside-occ. options, instrumented X Production	0.139***	0.112***
	(0.019)	(0.023)
Log HHI, instrumented X Sales	-0.016***	0.001
	(0.003)	(0.007)
Log outside-occ. options, instrumented X Sales	0.080***	0.095***
	(0.011)	(0.026)
Log HHI, instrumented X Office/Admin	-0.004*	-0.004
	(0.002)	(0.004)
Log outside-occ. options, instrumented X Office/Admin	0.096***	0.082**
	(0.022)	(0.038)
Log HHI, instrumented X Operators/Laborers	0.000	0.001
	(0.002)	(0.004)
Log outside-occ. options, instrumented X Operators/Laborers	0.086***	0.075***
	(0.013)	(0.026)
Log HHI, instrumented X Food/Cleaning/Protective Service	-0.008**	-0.015**
	(0.004)	(0.008)
Log outside-occ. options, instrumented X Food/Cleaning/Protective Service	0.113***	0.068***
	(0.014)	(0.022)
Log HHI, instrumented X Personal Care	0.001	0.001
	(0.004)	(0.007)
Log outside-occ. options, instrumented X Personal Care	0.110***	0.087***
	(0.011)	(0.021)
Vacancy growth		-0.098*
		(0.057)
Predicted vacancy growth		-0.007
		(0.038)
Exposure control		0.007
		(0.012)
Observations	445,681	445,681

Notes: Column (a) repeats our baseline OLS regression and column (b) repeats our baseline 2SLS IV regression (Table 3 columns (b) and (d) respectively), but allow the coefficients on the HHI and outside-occupation option index to vary for different aggregated occupation groups. Occupation groups map from SOC 2-digit occupations as defined in Appendix Table A6. Other 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. Regressions have occupation-year and metro area-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A16: Regression of wage on HHI and outside-occupation options: alternative HHI istruments

Dependent variable:	Log wage					
	HHI instrument construction:					
						Baseline
Shocks	Pos. only	All	Pos. only	All	Pos. only	All
Firm size	≥ 3	All	≥ 5	≥ 3	All	≥ 5
	(a)	(b)	(c)	(d)	(e)	(f)
Log HHI	-0.015***	-0.012***	-0.015***	-0.009**	-0.015***	-0.010***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Log outside-occ. options	0.109***	0.112***	0.109***	0.113***	0.114***	0.120***
	(0.022)	(0.019)	(0.023)	(0.021)	(0.022)	(0.018)
Vacancy growth	-0.103	-0.058	-0.105	-0.060	-0.303***	-0.106*
	(0.069)	(0.035)	(0.069)	(0.037)	(0.073)	(0.060)
Predicted vacancy growth	-0.031	-0.053*	-0.004	-0.009***	0.034	-0.002
	(0.031)	(0.028)	(0.006)	(0.004)	(0.047)	(0.038)
Exposure control	0.015	0.016	0.014	0.002	0.015	0.010
	(0.018)	(0.015)	(0.021)	(0.017)	(0.018)	(0.014)
Observations	535,532	498,283	579,668	564,279	492,757	445,681

Notes: This table reproduces our baseline 2SLS IV regression results, but with alternative formulations of the HHI instrument. Column (f) is our baseline specification. The other columns construct the HHI instrument differently along 2 dimensions: whether only positive national firm hiring shocks ("pos. shocks") are used, or all hiring shocks are used, and whether there is a restriction on firm size (where the baseline specification restricts to only use as instruments hiring shocks from firms with vacancies in 3 or more metro areas in a given occupation in a given year). Other 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. Regressions have occupation-year and metro area-year fixed effects and are weighted by average employment in each occ-metro area over the period. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A17: 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)
Panel C: First stage regressions				
oo^{occs} instrument	1.093***	0.861***	0.868***	0.823***
	(0.050)	(0.025)	(0.021)	(0.121)
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 A18: 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 A19: 2SLS IV Regressions of wages on outside-occupation options, incorporating employment share

Dependent var.:	Employment share	Log wage	
oo^{occs} , instrumented	-0.236***	0.030***	0.027***
	(0.030)	(0.012)	(0.012)
Empl. share			-0.012***
			(0.002)
Fixed effects	Occ-Metro area, Occ-Year	Occ-Metro area, Occ-Year	Occ-Metro area, Occ-Year
Observations	*2,255,447*	*2,255,447*	*2,255,447*

* p<0.10, ** p<0.05, *** p<0.01

Notes: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses. 2SLS IV regression results shown with instrumented outside-occupation options. Units of observation are 6 digit SOC by metro area by year, employment-weighted (by occ-metro area average employment 1999–2019), 1999–2019.

Table A20: Regressions of wage on outside-occupation option index, sample split into three periods, with different combinations of fixed effects

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: 1999–2006:				
OLS: oo^{occs}	0.415***	0.070***	0.084***	0.028***
	(0.014)	(0.004)	(0.005)	(0.005)
IV: oo^{occs}, instrumented	0.426***	0.055***	0.062***	0.018***
	(0.015)	(0.005)	(0.006)	(0.004)
Observations	778,532	778,474	778,474	762,260
Panel B: 2007–2012:				
OLS: oo^{occs}	0.440***	0.086***	0.093***	0.030***
	(0.016)	(0.005)	(0.006)	(0.004)
IV: oo^{occs}, instrumented	0.503***	0.079***	0.083***	0.013***
	(0.019)	(0.006)	(0.007)	(0.005)
Observations	684,254	684,206	684,206	671,149
Panel C: 2013–2019:				
OLS: oo^{occs}	0.478***	0.093***	0.103***	0.021***
	(0.015)	(0.005)	(0.006)	(0.003)
IV: oo^{occs}, instrumented	0.545***	0.084***	0.089***	0.007
	(0.019)	(0.007)	(0.007)	(0.005)
Observations	805,220	805,159	805,159	791,536
Fixed effects	Year	Occ-Year Metro area	Metro area-Year Occ-Year	Occ-Year Occ-Metro area

Notes: This table shows 2SLS IV regressions of wages on outside-occupation options, splitting our sample over three periods. Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}\ p < .01$. Units of observation are 6-digit SOC by metro area by year, over 1999–2019 inclusive (split into three roughly equal length periods), weighted by average SOC-metro area employment over 1999–2019. 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 (outside-occupation option index) in one regression specification, with included fixed effects held constant within each column.

Table A21: Regressions of wage on outside-occ. option index, *non-employment-weighted*, with different combinations of fixed effects

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: OLS				
oo^{occs}	0.445***	0.084***	0.094***	0.047***
	(0.014)	(0.004)	(0.005)	(0.005)
Panel B: 2SLS				
oo^{occs}, instrumented	0.491***	0.072***	0.077***	0.031***
	(0.017)	(0.005)	(0.006)	(0.006)
First stage:				
Coeff. on $oo^{occs,inst}_{o,k,t}$	0.962***	0.850***	0.804***	0.969***
	(0.023)	(0.018)	(0.012)	(0.046)
Fixed effects	Year	Occ-Year	Metro area-Year	Occ-Year
		Metro area	Occ-Year	Occ-Metro area
Observations	2,268,006	2,267,839	2,267,839	2,255,447

Note: This replicates Table A17, without employment weights. Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses: $*p < .1$,$**p < .05$,$***$ $p < .01$. Units of observation are 6 digit SOC by metro area by year, over 1999–2019.

Table A22: Regression of wage on HHI and oo^{occs}: Right-to-work and non right-to-work states

Dependent variable:	Log wage			
	(a) OLS	(b) OLS	(c) 2SLS IV	(d) 2SLS IV
Log HHI X Non right-to-work	-0.011*** (0.001)	-0.007*** (0.001)	-0.014*** (0.003)	-0.010*** (0.002)
Log HHI X right-to-work	-0.016*** (0.003)	-0.011*** (0.002)	-0.018*** (0.004)	-0.012*** (0.003)
Log outside-occ options X Non right-to-work		0.132*** (0.017)		0.109*** (0.018)
Log outside-occ options X right-to-work		0.162*** (0.017)		0.149*** (0.030)
Vacancy growth			-0.147* (0.077)	-0.110* (0.061)
Predicted vacancy growth			0.045 (0.045)	-0.003 (0.038)
Exposure control			0.028* (0.015)	0.015 (0.014)
Observations	445,681	445,681	445,681	445,681

Notes: This table shows our baseline regression specifications as in Table 3 but allowing the coefficients on the HHI and outside-occupation options to differ in right-to-work states and non right-to-work states. States are classified as right-to-work or non-right-to-work according to data from the National Conference of State Legislatures. States which passed right-to-work laws in 2015 or 2016 (Wisconsin and West Virginia) are coded as non-right-to-work in our sample, under the assumption that it would take some time for the passage of right-to-work statutes to affect labor market behavior. Other info: Heteroskedasticity-robust standard errors clustered at the metro area level shown in parentheses: $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Units of observation are 6 digit SOC occupation by metro area by year, for all observations with available data over 2011–2019 inclusive. All specifications include occupation-by-year and metro area-by-year fixed effects, and are weighted by average occ-metro area employment over the period.

Table A23: 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	1.7%	4.7%	7.3%	9.6%
	Employment(m)	7.3	5.7	3.7	.59	.44
Q2 mobility	Avg. wage effect	0	0.6%	2.0%	3.1%	4.1%
	Employment(m)	12	4	1.7	.35	.32
Q3 mobility	Avg. wage effect	0	0	0	0	0
	Employment(m)	15	5.5	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 A24: 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
Medical assistants	.83
Licensed practical and licensed vocational nurses	1.3
Fitness trainers and aerobics instructors	.63
Emergency medical technicians and paramedics	.59
Radiologic technologists	.76
Heavy and tractor-trailer truck drivers	2.1
Medical and clinical laboratory technologists	1.1
Phlebotomists	1.6
Aircraft mechanics and service technicians	.72
Lawyers	.87
Massage therapists	1
Management analysts	1.9
Nurse practitioners	4.1
Manicurists and pedicurists	.51
Respiratory therapists	1.2
Physician assistants	2.6
Surgical technologists	2
Physicians and surgeons, all other	1.8
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 A25: 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	**Considerin mobility** *(our baseline: ap HHI effect by qu of occupational m* No. of workers 2% wage effe
Nursing assistants	157,560	629,320
Medical assistants	141,470	184,430
Registered nurses	130,210	1,043,200
Tellers	123,460	0
Hairdressers, hairstylists, and cosmetologists	93,820	345,680
Pharmacy technicians	82,690	309,440
Security guards	81,620	950,680
Massage therapists	58,100	91,770
Stock clerks and order fillers	57,670	0
Butchers and meat cutters	57,120	0
Postal service clerks	52,720	52,720
Pharmacists	50,520	209,880
Radiologic technologists	49,330	132,750
Medical secretaries	48,540	0
Tax preparers	44,580	44,710
Tire repairers and changers	43,900	0
Emergency medical technicians and paramedics	41,910	145,630
Veterinary technologists and technicians	37,960	55,370
Industrial machinery mechanics	36,780	0
Phlebotomists	35,740	110,270
Fitness trainers and aerobics instructors	35,690	156,490
Licensed practical and licensed vocational nurses	35,470	172,760
Secretaries and administrative assistants, except legal, medical, and executive	34,710	56,870
Electrical power-line installers and repairers	33,450	0
Securities, commodities, and financial services sales agents	33,280	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 $(log(HHI) - log(150)) \cdot -0.010$, where *-0.010* 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).

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.

Bewley, Truman, *Why Wages Don't Fall During a Recession*, Harvard University Press. Cambridge, Massachusets, and London, England, 1999.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *Review of Economic Studies*, forthcoming.

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

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.

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

Galuscak, Kamil, Mary Keeney, Daphne Nicolitsas, Frank Smets, Pawel Strzelecki, and Matija Vodopivec, "The determination of wages of newly hired employees: Survey evidence on internal versus external factors," *Labour Economics*, 2012, *19* (5), 802–812.

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).

Hall, Robert E and Alan B Krueger, "Evidence on the incidence of wage posting, wage bargaining, and on-the-job search," *American Economic Journal: Macroeconomics*, 2012, *4* (4), 56–67.

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.

Jaeger, Simon, Benjamin Schoefer, Sammy Young, and Josef Zweimueller, "Wages and the Value of Nonemployment," *Quarterly Journal of Economics*, 2020.

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.

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

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," Technical Report, National Bureau of Economic Research 2018.

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

Manning, Alan, "Imperfect competition in the labor market," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 973–1041.

Marinescu, Ioana and Roland Rathelot, "Mismatch Unemployment and the Geography of Job Search," *American Economic Journal: Macroeconomics*, 2018.

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

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

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.

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.