Employer Concentration and Outside Options

Gregor Schubert, Anna Stansbury, and Bledi Taska *

January 25, 2024

Abstract

We find that increases in employer concentration causally reduce wages, using a new instrument for employer concentration based on changes in large firms' national hiring patterns. We also show that measuring employer concentration within a single local occupation excludes important parts of workers' true labor markets. Moving from the median to the 95th percentile of employer concentration as experienced by workers causally reduces wages by 10.7 log points in low-outward-mobility occupations like registered nurses or security guards, and by 3 log points in high-outward-mobility occupations like bank tellers or counter attendants. We propose a new approach for defining mobility-adjusted labor markets, measuring employer concentration on clusters of local occupations identified through asymmetric mobility patterns (using new, highly granular data on occupational mobility from 16 million resumes). Overall, we estimate that around one in six U.S. workers face wage suppression of 2% or more as a result of employer concentration.

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

1 Introduction

In recent years, concerns about employer concentration have increased. Employer concentration has been posited as a possible explanation for inequality, low pay, and stagnant pay growth. Antitrust authorities have been called on to consider employer concentration in merger and acquisition reviews. Concerns have been raised that employer concentration facilitates restrictions on competition like no-poaching agreements. And, since employer concentration can be a source of monopsony power,[1] concerns around high employer concentration have bolstered calls to raise minimum wages and strengthen collective bargaining.[2] To assess whether – or in which cases – policy should respond to employer concentration, we need to understand how many workers are causally affected by employer concentration, who they are, and by how much.

In this paper, we estimate the causal effect of employer concentration on average hourly wages in the U.S. We contribute two important new insights to the debate around monopsony power.

First, we find a large negative causal effect of employer concentration on wages, using a new instrument for employer concentration based on changes in large firms' national hiring patterns. This estimate represents a major advance on existing work, since changes in employer concentration are usually endogenous to local economic conditions, and since our estimate covers over 100,000 US occupation-by-metropolitan-area labor markets representing more than 75% of the US workforce.

Second, we show that the effects of employer concentration on wages, if measured using conventional definitions of labor markets, vary substantially with worker occupational mobility. This means that most estimates of wage effects of concentration will be biased and will obscure substantial heterogeneity if labor markets are not defined carefully. We provide a new approach for adjusting labor market definitions based on worker mobility, using novel and highly granular data on occupation-to-occupation flows, and show how this approach can be used for empirical practice in academic, antitrust, and policy studies of monopsony power.

Overall, our estimates suggest around one in six U.S. workers experience wage suppression of 2% or more as a result of employer concentration, and that many of the most affected workers are in healthcare occupations. We outline our approach in more detail below.

Several recent studies have found a negative association between wages and employer

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

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

concentration (e.g. Azar, Marinescu, Steinbaum and Taska, 2020a; Azar, Marinescu and Steinbaum, 2020b; Benmelech, Bergman and Kim, 2022; Rinz, 2022). Yet, even with careful inclusion of controls, it is quite plausible that this relationship is simply the product of an omitted variable: for example, low productivity labor markets may attract few firms and have low wages. A core contribution of our paper is a new identification approach for the effects of employer concentration on wages. We draw on shift-share and granular IV methodology (Borusyak, Hull and Jaravel, 2022; Gabaix and Koijen, 2020), developing an instrument for employer concentration within a particular local occupation, which can be applied to a broad set of occupationally and geographically diverse labor markets.

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

Estimating the causal effect of employer concentration on wages also requires a good definition of the relevant local labor market. Why? We show in a simple conceptual framework that if workers have job options outside their current occupation, measuring employer concentration within a single occupation (or industry)—the norm in much of the existing literature—excludes important parts of workers' true labor markets. This leads to inaccurate estimates of the true effect of employer concentration. To illustrate the scope of this problem, we use new, highly-granular occupational mobility data constructed from 16 million US workers' resumes,[3] to show that occupational mobility is high and highly heterogeneous across occupations, making consideration of worker mobility a first order issue.

In our estimates of the effect of employer concentration on wages, we incorporate these

[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.

insights on worker mobility in two ways. First, we allow the estimated coefficient on within-occupation employer concentration to vary by occupations' outward mobility. Second, we develop a new measure of the value of workers' outside job options in other occupations: an "outside-occupation option index". Our index is the weighted average of wages in all other local occupations, with weights proxying for the relevance of each occupation as an outside job option (specifically, the product of occupational mobility flows and local relative employment shares). We use a shift-share IV approach to identify plausibly exogenous shocks to outside-occupation options, and estimate their effect on wages in our baseline regression alongside the effect of within-occupation employer concentration.

How much does employer concentration matter for wages? Our baseline IV results find that moving from the median to the 95th percentile HHI (as faced by workers) results in 6.5 log points lower wages. However, focusing on the average effect of employer concentration *within* occupations is misleading: within-occupation employer concentration matters much *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 reduces wages by 10.7 log points on average. In comparison, for occupations in the highest quartile of occupational mobility, like counter attendants or bank tellers, moving from the median to the 95th percentile HHI reduces wages by only around 3 log points—that is, there is a more-than-threefold difference in the estimated wage effect between low and high mobility workers. The results for the least outwardly mobile occupations are likely the most appropriate to infer the *true* effect of employer concentration in a well-defined labor market, since these occupations are most likely to reflect workers' true set of labor market options. Thus the effect of employer concentration, when measured only over well-defined labor markets, is substantially larger than a conventional estimation strategy would suggest. Conventional analyses using single-occupation (or single-industry) labor markets that ignore mobility lead us to both under-estimate the true effects of employer concentration for the average labor market *and* to overestimate it for more mobile occupations.

Supporting this conclusion, we find that our measure of outside-occupation options significantly impacts wages: an exogenous 1 percentage point increase in the wage in outside option occupations leads to a 0.1 percent higher wage in workers' own occupation in the same year. This jointly validates two postulates: first, that outside-occupation options are a meaningful part of workers' labor markets; and second, that occupational mobility flows can be used to identify these outside-occupation job options.

How should labor market researchers think about employer concentration in a context of high and heterogeneous mobility? We provide a simple and tractable solution to the issue of misspecification: a “mobility-adjusted labor market” (MALMA). This is a market definition which includes all other occupations that are likely targets for worker moves based on a chosen relevance threshold. These MALMAs differ from other cluster-based approaches to labor market definition because they are asymmetric: occupation p may be a relevant option for workers in occupation o, but not vice versa. This is an important distinction, informed by the asymmetry of occupation flows we observe in our new mobility data. Our estimates show that measuring employer concentration at the level of these MALMAs delivers much larger estimates of the effect of employer concentration on wages than using single occupations, with a move from the median to the 95th percentile of employer concentration as faced by workers leading to 14 log points lower wages. It also delivers estimates that are comparable across quartiles of outward occupational mobility (virtually eliminating the misspecification issue we identified earlier).

These mobility-based labor market adjustments represent a pragmatic improvement for future analyses of labor market power and antitrust policy. Moreover, these mobility-based labor market adjustments have broader applicability in studies of labor market outcomes wherever market definition is important, such as studying the transmission of economic shocks or the effects of labor market regulations.

Finally, we use our estimates to ask how much employer concentration matters in the U.S. labor market, and for whom. Our approach is uniquely suited to do so, given (i) our credible causal identification, using variation from a large share of the U.S.' labor markets, and (ii) our mobility-adjusted approach to labor market definition. A back-of-the-envelope calculation suggests that almost one sixth of the 117 million workers covered by our data in 2019 experienced wage suppression of 2% or more as a result of being in a labor market with higher-than-median employer concentration.[4] Many of the most-affected workers are healthcare workers, reflecting both high healthcare employment concentration and low occupational mobility.

Related literature. Our work relates to several active areas of research in labor economics. First, we build on a growing body of work demonstrating an empirical relationship between wages and employer concentration, which began in recent years with Azar et al. (2020a), Azar et al. (2020b), Benmelech et al. (2022), and Rinz (2022). We make two key contributions to this literature. First, we develop a new instrument to estimate plausibly causal negative effects of employer concentration on wages. This is an important innovation in a context where most evidence is subject to concerns about omitted variable bias aris-

[4]Measured as the median mobility-adjusted HHI across occupation-by-metro area labor markets in our data, which was around 150.

ing from local occupation-specific productivity or other economic conditions.[5] Second, we are the first to show that worker mobility leads to substantial heterogeneity in the effects of employer concentration on wages, and we propose a new mobility-adjusted labor market definition to solve this misspecification problem. We can then use our causal estimates based on these innovations to quantify the rough magnitude of the employer concentration problem across most US occupations.[6]

Second, in estimating the effect of outside-occupation options on wages, we add to a literature on outside options in the labor market (e.g. Beaudry, Green and Sand, 2012; Caldwell and Danieli, 2018), and in using occupational transitions to identify outside options we build on papers which use worker flows to identify the scope of workers' labor markets (e.g. Shaw, 1987; Manning and Petrongolo, 2017; Nimczik, 2018). Our paper links the two separate literatures on employer concentration effects and outside options, showing that causal estimates of the former need to account for the latter.

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

Roadmap. We lay out our empirical approach in section 2, outlining a simple partial equilibrium framework through which we can consider within-occupation employer concentration and outside-occupation options, explaining our data and independent variables of interest, and detailing our identification strategy. In section 3, we show our baseline results, explore their robustness to different specifications and estimation strategies, and show how labor markets can be adjusted to take mobility into account. We discuss the implications of our results for the aggregate economy and for policy in section 4, and conclude in section 5.

2 Empirical Approach

Why might employer concentration matter for wages? The core intuition is that employer concentration can reduce the availability of feasible outside job options for workers. This reduces the strength of workers' relative bargaining position with employers, and can therefore reduce their wage. For example, Jarosch, Nimczik and Sorkin (2019) show that employer

[5]Our work complements causal estimates from M&A activity (Arnold, 2020; Prager and Schmitt, 2021). These leverage a very different source of variation to that in our instrument – only 2% of changes in employer concentration come from M&A activity – and are subject to concerns about other simultaneous labor market effects of the M&A.

[6]Our approach addresses the key critiques of regressions of wages on local employer concentration (including Berry, Gaynor and Scott Morton (2019) and Rose (2019), which follow older critiques of the structure-conduct-performance paradigm (e.g. Schmalensee, 1989)), related to the concern that employer concentration is endogenously determined and dependent and that there is no appropriate definition of a market on which a meaningful concentration index can be calculated.

concentration reduces wages in a random search model with large ("granular") employers. In their model, workers bargaining with large firms have worse outside options: there are fewer other feasible job opportunities outside the firm they are currently bargaining with, because firms do not compete with their own vacancies. In models with wage posting rather than bargaining, employer concentration can generate upward-sloping labor supply curves to individual firms, leading to wage markdowns (Berger, Herkenhoff and Mongey, 2022; Azkarate-Askasua and Zerecero, 2020). Moreover, explicit collusion to suppress wages can be facilitated by the presence of a small number of firms. Several recent papers specifically demonstrate that a Herfindahl-Hirschmann Index ("HHI") across employers is a relevant statistic to measure labor market power arising from employer size. In Jarosch et al. (2019) the effect of concentration on wages is determined by an index which, to a second order approximation, reduces to an HHI.[7] In Berger et al. (2022), an HHI of the wage bill across employers is a relevant statistic for assessing the welfare effects of firms' labor market power. Naidu and Posner (2021) shows that an employer HHI is related to the size of the wage markdown under Cournot competition.

An HHI is by definition measured on a single labor market, where all jobs within the market are considered perfect substitutes and all jobs outside the market are irrelevant to the worker. This market might be a local occupation, industry, or cluster of firms. But this binary definition rarely captures workers' true labor markets. For example, using 6-digit SOC occupations as the base local labor market, we show in section 2.4 below that a local occupation does not capture workers' labor market well (and captures it differently poorly for different occupations), and that a local labor market is better described as a set of local occupations with differing relevance to the worker as an outside option.[8]

In this paper, we therefore consider within-labor-market employer concentration as part of workers' broader set of outside job options. Specifically we jointly consider the outside options available to workers in their own "base" local labor market - which we define as a local occupation - alongside the cluster of other occupations which may be relevant to them (their outside-occupation job options). Our approach captures three ideas. (1) Employer concentration within workers' local occupation reduces the availability of job options outside workers' own firm, which puts downward pressure on the wage. (2) Worker mobility across occupations makes other occupations relevant as outside options, meaning that higher wages in other relevant occupations improve the value of workers' outside options, putting upward pressure on the wage. (3) For workers for whom it is easy to switch occupation, within-

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

[8]This argument could also be applied to local industries or clusters of firms.

occupation employer concentration matters less for the wage, because their effective labor market is broader – a larger share of their outside options are outside their occupation. In the next section, we outline a simple partial equilibrium conceptual framework which captures these ideas, and use this to discipline the functional forms used in our estimation.

2.1 Conceptual framework

In this section, we outline a simple partial equilibrium framework designed to capture the intuition discussed above. First, we assume that wages w_i^o at each firm i in occupation o are a weighted average of labor productivity (p_i^o) and workers' outside option. We assume all workers at the same firm i and in the same occupation o have the same outside option in expectation (we will also focus on a single geography to simplify the notation.) That is,

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

The outside option is the expected wage if a worker leaves their current job:
$oo_i^o = E[w|\text{leave job at } i \text{ in } o]$. Upon leaving their current job, we assume a worker is matched to either their own occupation o or another occupation p (with some probability $\pi_{o\to p}$), and then randomly matched with a vacancy at a firm j within that occupation. Random matching means that the chance of receiving a job offer from a particular firm j within an occupation is equal to firm j's share of vacancies σ_j^o in that occupation o (in the spirit of Burdett and Mortensen (1980) and Jarosch et al. (2019)). The possible outcomes of the match are therefore (i) the worker is matched with some job in another occupation p with probability $\pi_{o\to p}$, earning in expectation the average wage w^p in that occupation; (ii) the worker is matched with a job at another firm j in their own occupation o with probability $\pi_{o\to o}\sigma_j^o$, earning wage w_j^o; or (iii) the worker may be randomly matched with the firm that they are leaving, in which case they are not re-hired and receive payoff of zero (following Jarosch et al. (2019)).[9] This means that the expected outside option of the worker is a weighted average of the wages in other firms in their own occupation, and the average wages in other occupations,

$$oo_i = \pi_{o\to o} \sum_{j\neq i} \sigma_j^o \cdot w_j^o + \underbrace{\sum_{p\neq o}^{N^{Occ}} \pi_{o\to p} w^p}_{oo_o^{occs}}, \tag{2}$$

[9]More realistically, a worker might expect a costly longer search time if they match with their current firm, but to simplify the intuition we set this payoff to zero. This formalizes the assumption that from a worker's perspective, jobs at their own firm are not a credible outside option, as in Jarosch et al. (2019). For simplicity, we do not consider unemployment in this framework - see, e.g. Jaeger, Schoefer, Young and Zweimueller (2020) that wages are empirically insensitive to unemployment benefit changes.

where j are indices over the set of firms in occupations o and p, respectively, σ_j^o is the share of all vacancies in occupation o represented by firm j, and $\pi_{o\to o}$ and $\pi_{o\to p}$ represent the probabilities of being matched with a vacancy in the current occupation o or a particular other occupation p (when coming from a job in occupation o). The second term represents what we will refer to as "outside-occupation options" oo_o^{occs}, which is the degree to which an occupation's wages are affected by wages in other occupations.

Taking the wages in other occupations as given, what would be the average wage w^o across firms in occupation o consistent with this method of wage setting? By substituting from equations 1 and 2, we can derive

$$\begin{aligned} w^o = \sum_j \sigma_j^o \cdot w_j^o &= \beta p^o + (1-\beta)\pi_{o\to o}\sum_j \sigma_j^o(1-\sigma_j^o)w_j^o + (1-\beta)\sum_{p\neq o}^{N^{Occ}} \pi_{o\to p} w^p \\ &\approx \left(\beta p^o + (1-\beta) oo_o^{occs}\right)\cdot\left(1+(1-\beta)\pi_{o\to o}(1-HHI^o)\right), \end{aligned} \tag{3}$$

where the last line takes a second order approximation and uses the definition of the commonly-used Herfindahl-Hirschman concentration index as $HHI^o = \sum_j (\sigma_j^o)^2$ and the definition of the outside-occupation options oo_o^{occs} shown above.[10] Intuitively, absent employer concentration, wages in occupation o would be a weighted average of the occupation's average productivity and the value of outside-occupation options. However, the higher is employer concentration HHI^o, the less valuable are workers' outside options within their own occupation - there are fewer credible other jobs they can move to, outside their own firm. This reduces the value of workers' outside option set and so reduces average wages in the occupation.[11] Log-linearizing equation 3, this implies the following log approximation

$$\widetilde{\ln w^o} = \gamma^o \widetilde{\ln p^o} + (1-\gamma^o)\widetilde{\ln oo_o^{occs}} - \zeta^o(\pi_{o\to o})\widetilde{lnHHI^o} \tag{4}$$

where the notation $\widetilde{x^o} = x^o - \overline{x^o}$ denotes deviations from equilibrium values, and $\zeta^o(\pi_{o\to o})$ is an increasing function of the likelihood of remaining in your own occupation $\pi_{o\to o}$.[12]

While this is not intended to be a full model of wage determination, this simple framework illustrates that accounting for the fact that workers in real-world labor markets are mobile across occupations is important when estimating effects of employer concentration. Typically, regressions of wages on employer concentration at the level of a local labor market take

[10] The expression is a second order approximation in a concentration index with higher order terms. p^o denotes average productivity in occupation o. We show the full derivation in the appendix.

[11] Note that in practice the wages in firms in other occupations p will depend on the wages set in occupation o, and vice versa – so there are general equilibrium effects and reverse causality concerns that are not captured in this simple framework because we assumed wages in other occupations to be exogenous. In our empirical analysis we will confront this endogeneity 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.

[12] Specifically, $\zeta^o(\pi_{o\to o}) = \frac{(1-\beta)\pi_{o\to o}\overline{HHI}^o}{1+(1-\beta)\pi_{o\to o}(1-\overline{HHI}^o)}$, and $\gamma^o = \frac{\beta\overline{p^o}}{\beta\overline{p^o}+(1-\beta)\overline{oo_o^{occs}}}$

the binary market definition approach: they do not take into account workers' (differential) ability to switch occupations.[13] The expression in equation 4 illustrates two problems that this could cause when estimating the effect of concentration on wages: (1) *Effect heterogeneity:* the effect of within-occupation employer concentration on wages should be stronger for occupations with low outward occupational mobility (high $\pi_{o\to o}$), because the occupation is a better reflection of their true labor market – as compared to occupations with high outward occupational mobility. (2) *Omitted variable bias:* if the degree of employer concentration within a local occupation (HHI^o) is correlated with the quality of outside options *outside* the local occupation (oo_o^{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 issues into account in the empirical analysis that follows and show that this leads to important new insights regarding the effect of employer concentration and outside options in labor markets.

While we use the occupation as a base labor market, we underscore that these issues will be present whatever base labor market the HHI is measured over – including when using more flexible measures like a cluster of related firms (e.g. Nimczik, 2018). This is because defining a single labor market and measuring an HHI over jobs in that labor market implicitly asssumes that all jobs within that labor market are equally good options for the worker and all jobs outside that labor market are irrelevant (the "binary" approach to market definition). In practice, there will always be jobs outside a plausibly-defined labor market which are relevant to the workers in that labor market, and to different degrees. The issues we identify thus cannot be solved simply by defining the base labor market more flexibly.[14]

Endogeneity. By taking wages in other occupations, the market structure, and the market shares of firms as given, the partial equilibrium framework above does not incorporate the fact that these local labor market characteristics are an endogenous outcome of productivity, demand, and supply conditions. This results in important identification concerns: employer concentration may be correlated with wages without having a causal effect on it. The quality of outside-occupation options (the wages in other occupations) suffers from a similar endogeneity issue. In the empirical implementation below, we therefore develop instruments for both the HHI and the outside-occupation option index, discussed further in sections 2.6 and 2.7.

[13]Labor markets have typically been defined as a single occupation or industry within a given local area (commuting zone, metro 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, Lovenheim, Salvanes and Willén (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.

[14]Though for settings where binary labor market measures are needed, we suggest defining a mobility-adjusted labor market or "MALMA" in section 3.3.

2.2 Empirical specification

Building on the conceptual framework outlined above, our baseline empirical specification is

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

where subscripts refers to the metro area (k), occupation (o), and year (t). $\alpha_{o,t}$ and $\alpha_{k,t}$ are a set of occupation-by-year and metro area-by-year fixed effects, which control for any national occupation-level economic shocks and any metro-area-level economic shocks which might affect both concentration and wages.

Our wage measure $\bar{w}_{o,k,t}$ is the average hourly wage for all workers in a given occupation, metro area, and year, from BLS OES data. We construct the HHI from Burning Glass Technologies' vacancy posting data (discussed in section 2.3), 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 in section 2.4).[15] We allow the coefficient γ_1 on the HHI to vary according to the occupation's degree of outward mobility (which we estimate using the "'leave share" calculated from resume data from Burning Glass Technologies and described in section 2.4). Specifically, we interact $HHI_{o,k,t}$ with an indicator variable for the applicable quartile of outward mobility of occupation o. We run our regressions across the largest possible subset of U.S. occupation-metro area-year cells for which we can obtain all our key variables: Our full data set for our baseline regressions over 2011–2019 comprises 443,493 occupation-metro area-year observations.[16]

2.3 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, Macaluso and Yeh (2020). The BGT vacancy posting data covers the near-universe of online job postings, drawn from over 40,000 sources including company websites and online job boards, with no

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

[16] This includes 367 metro areas and 660 occupations, with 93,005 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. Nonetheless, this larger data set is missing many occupations and metro areas. This is because, of the possible 786,335 occupation-metro area cells, wage data in the BLS OES only exists for approximately 115,000 each year. The missing occupations and metro areas are primarily the smaller ones. To create a consistent panel of occupations over time we crosswalk SOC classifications over time: see Appendix D. We use "metro areas" to refer to CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas).

more than 5% of vacancies from any one source (Hazell and Taska, 2019).[17] 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 metro area k in year t.

The Burning Glass Technologies vacancy data covers the near-universe of online job postings. There are, however, two possible concerns in terms of its representativeness of job vacancies. 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 in the Help Wanted Online database), but this is lower for occupations where a large share of jobs are advertised offline or informally. Second, in occupations where firms tend to hire many workers for each posted vacancy, our estimates of employer concentration will be biased to the degree that larger firms may hire more people per vacancy posting.[18] To understand the degree to which each of these are 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 (from BLS OES). Occupations which are particularly underrepresented include low-wage food service, cleaners, home health aides, laborers, and cashiers. In our estimates of the effect of employer concentration on wages, we include occupation-by-year fixed effects and also show that our results are robust to excluding underrepresented occupations. For further discussion of the BGT vacancy data, see Appendix B.

[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 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 use vacancies rather than employment to construct HHIs for two reasons. First, we are not able to obtain firm-level employment data within local occupations. Second, vacancies may be a better reflection of workers' feasible outside options. In equilibrium, vacancy and employment HHIs should be highly correlated (as illustrated in French data by Marinescu, Ouss and Pape (2021)).

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

2.4 Using occupational mobility to identify outside options

We use occupational mobility patterns to identify workers' job options outside their own occupation. 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. This resume data was collected by Burning Glass Technologies, who sourced the resumes from a variety of BGT partners, including recruitment and staffing agencies, workforce agencies, and job boards.[19]

We use this data to construct two objects. First, the **'occupation leave share'**, approximating the share of people who leave their occupation when they leave their job:[20]

$$\begin{aligned} leaveshare_o &= \frac{\#\text{ in occ } o \text{ in year } t \;\&\; \text{no longer in occ } o \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \;\&\; \text{in a new job in year } t+1} \\ &\approx Prob(\text{leave occ } o | \text{leave job}). \end{aligned} \tag{7}$$

We also construct the **'occupation transition share'** $\pi_{o \to p}$, which approximates the probability of a worker moving from occupation o to occupation p conditional on leaving her job:

$$\begin{aligned} \pi_{o \to p} &= leaveshare_o \cdot \frac{\#\text{ in occ } o \text{ in year } t \text{ observed in occ } p \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \text{ observed in any new occ in year } t+1} \\ &\approx Prob(\text{move from occ } o \text{ to occ } p | \text{leave job}). \end{aligned} \tag{8}$$

Intuitively, the leave share captures the probability that someone will go to any new occupation when they change job, while the transition share captures the probability that someone will go to a specific new occupation p when they change job. We calculate these as US national averages across all years in our data, to capture underlying occupational similarity rather than transitory fluctuations from year to year.[21]

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

[20] Specifically, this measure captures the share of people observed in occupation o in year t who are *no longer* observed in occupation o at any point in year $t+1$, as a share of those observed in occupation o in year t who are observed in some new job in year $t+1$. 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. This is supported by the finding in Lachowska, Mas, Saggio and Woodbury (2022) that secondary job wages can affect primary job wages for dual jobholders.

[21] We estimate transition shares 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

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 to have some degree of representativeness *within* each occupation. To correct for the over-representation of younger workers, we re-weight our observed occupational transitions to match the distribution of employment by age within each U.S. occupation, provided by the BLS for 2012-2017. We discuss the BGT resume data in more detail in Appendix C.

Using the BGT data, we document five stylized facts about occupational mobility:

1. Occupational mobility is high, suggesting that SOC 6-digit occupations fail to capture many workers' true labor markets: the median occupation's leave share is 24% (Table 1).
2. Mobility is heterogeneous across occupations, suggesting that SOC 6-digit occupations better capture the labor market for some occupations than others: a quarter have a leave share lower than 19%, and a quarter higher than 28% (Table 1, Figure 2).[22]
3. Aggregating up the SOC classification hierarchy - which groups ostensibly similar occupations - still fails to capture most occupational transitions, suggesting that this cannot solve the market definition problem.[23]
4. The occupational transition matrix is sparse, suggesting that workers' relevant labor markets are mostly comprised of only a few occupations, and is highly asymmetric, suggesting that the relevance of occupations as outside options is not symmetric across occupation pairs (unlike in many task- and skill-based measures of occupational similarity).[24]
5. Empirical occupation transitions reflect similarities between occupations in terms of their task requirements, wages, amenities, and leadership responsibilities, suggesting that occupational transitions reflect a combination of many different aspects of the underlying feasibility and desirability of an occupation as an outside option.[25]

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

[22]Occupations with low leave shares tend to be highly specialized, including medical, legal and educational occupations (see Appendix Table A2). In contrast, many high leave share occupations require more general skills, like restaurant hosts/hostesses, cashiers, tellers, counter attendants, and food preparation workers.

[23]For the median occupation, 87% of moves to a different 6-digit occupation are also to a different 2-digit occupation, but with substantial variation (see Table 1): e.g. 39% of systems software developers leave their 2-digit occupation when they move across 6-digit occupations, compared to 95% of flight attendants. Excluding transitions to and from management (which are often in a separate 2-digit occupation), at the median 67% of SOC 6-digit occupational transitions cross SOC 2-digit boundaries.

[24]See Appendix Figure A6 and Appendix Table A3. The asymmetry partly reflects the fact that workers in an occupation with specialized skills may be able to move to occupations which require generalist skills (e.g. retail salespersons) but the reverse flow is less feasible. This will be important for our construction of MALMA mobility-adjusted labor markets in section 3.3.

[25]We show this in regressions of occupational transitions on occupation-pair similarity in characteristics

Together, these stylized facts inform our empirical approach: they illustrate that (i) jobs outside workers' occupation are an important part of the labor market for many workers, (ii) this is differently true for different occupations, and (iii) occupational transitions can be used to identify these options.

2.5 Measuring outside-occupation options

To measure the value of workers' job options outside their occupation, we construct the empirical equivalent of the **outside-occupation option index** oo^{occs} that was defined in equation 2 and is motivated by the conceptual framework in section 2.1. This is a weighted average of the wage in each alternative occupation in a worker's metro area, $\mathrm{w}_{p,k,t}$, weighted by a measure of the relevance of each other local occupation in the worker's outside option set:

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

Specifically, we proxy for the relevance of other occupations in the worker's outside option set with the product of two variables: (1) $\pi_{o \to p}$, the likelihood of a worker moving to any job in occupation p if they leave their job in occupation o, constructed from our BGT resume data as described above, and (2) $\frac{s_{p,k}}{s_p}$, the relative employment share of occupation p in metro area k compared to the national average. The national occupation transition share $\pi_{o \to p}$ proxies for the likelihood that, nationwide, the average worker's best job option outside her current job would be a job in occupation p. The local relative employment share adjusts this for the local availability of jobs in each occupation p.

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

2.6 Identification: employer concentration

When estimating the effect of local occupational employer concentration on wages, identification requires that employer concentration varies independently of other shocks to local

derived from the O*Net database: specifically 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 job flexibility (Goldin, 2014); and characteristics measuring leadership responsibilities. In every pairwise regression of occupational mobility on the absolute difference in characteristics (controlling for the difference in wages), the coefficients are significantly negative or statistically insignificant, as shown in Figure A12. See Appendix F for more details.

occupation-specific labor demand, which might otherwise induce omitted variable bias.[26] These endogeneity issues may bias the estimated coefficients on the HHI. The direction of the bias is ambiguous: for example, the expansion of a highly productive large firm might increase both employer concentration and average productivity. Or, an increase in employer concentration could reflect a lack of local dynamism, with few new firms, leading to higher employer concentration alongside falling productivity. As Jarosch et al. (2019) and others illustrate, once we are able to identify exogenous changes in local labor market concentration, there is a clear conceptual channel by which increased concentration will, ceteris paribus, exert downward pressure on wages, as it reduces the value of workers' outside option set.

We therefore develop a novel instrument that induces exogenous variation in local labor market concentration by leveraging differential local occupation-level exposure to large national firms' hiring growth. We view our new instrument for employer concentration as a key contribution of our paper. Our strategy can be interpreted as a shift-share 'Bartik' instrument or through the lens of the "granular" instrumental variable approach (GIV) of Gabaix and Koijen (2020) (which uses plausibly exogenous idiosyncratic firm-level variation to instrument for changes in market-level aggregates). Our strategy is based on the facts that (a) increases in local employer concentration are often driven by individual large firms growing, (b) these firms usually operate across many labor markets, (c) local labor markets are differentially exposed to different large firms, and (d) the employment growth of these large firms nationally is likely orthogonal to economic conditions in a specific local occupation.

Specifically, we note that the growth in local employer concentration in occupation o is a function of the growth in local occupational employment for each employer j, $g_{j,o,k,t}$ (leaving aside firm entry): $\Delta HHI_{o,k,t} = \sum_j \sigma^2_{j,o,k,t} - \sum_j \sigma^2_{j,o,k,t-1} = \sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+g_{j,o,k,t})^2}{(1+g_{o,k,t})^2} - 1\right)$. The increase in local occupational employer concentration is a function both of initial concentration and of the growth rates of firm-level vacancies $g_{j,o,k,t}$ relative to overall vacancy growth in the labor market $g_{o,k,t}$. We instrument for the vacancy growth for each firm j in occupation o and metro area k with the national vacancy growth of that firm j in occupation o, leaving out the metro area in question k, (which we denote $\tilde{g}_{j,o,t}$).[27] Our instrument for

[26]Note that our occupation-by-year and metro area-by-year fixed effects in our baseline specifications already control for labor demand (or supply) shocks to the occupation or metro area respectively.

[27]We only use intensive margin shocks (excluding changes to or from zero). In our baseline specification, we use only positive shocks, but relax this in robustness analyses. Note that we are instrumenting for the local level of the HHI with an instrument derived from an expression for the change in the HHI, and by taking the log we exclude observations where the predicted change in HHI is negative. We formulate an alternative instrument based on the predicted level of the HHI in our robustness analyses.

the log HHI, $Z^{HHI}_{o,k,t}$, is therefore:

$$Z^{HHI}_{o,k,t} = \ln\left(\sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1\right)\right) \tag{9}$$

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.[28] Through the lens of shift-share instruments (Borusyak et al., 2022), our instrument features plausibly exogenous 'shocks' (a function of firms' national hiring growth), and possibly endogenous exposure 'shares' (the last-period local occupational vacancy shares of each of those firms). The key assumptions for our instrument to be valid are that the firm's national leave-one-out vacancy growth is (i) correlated with its local vacancy growth, but (ii) uncorrelated with the other determinants of occupation-specific wage growth in any given metro area k, conditional on our occupation-year and metro area-year fixed effects. Note that our inclusion of occupation-by-year fixed effects in all our baseline regressions assuages any concern around the GIV instrument being correlated with other occupation-level demand shocks.

The motivation behind our instrument is the existence of national firm-level labor demand shocks which are unrelated to the occupation-specific economic conditions in a particular local area. More specifically, assume a firm's labor demand in occupation o and local area k is a function of (i) occupation-level factors common to all firms, (ii) national firm-level factors – productivity (determined by national technology investments, IP, or production processes and systems), and product demand (perhaps a function of national brand recognition or advertising campaigns, or the specificity of a particular product or service), and (iii) local firm-specific productivity or product demand in local area k. Our occupation-year fixed effects hold constant component (i), the occupation-level factors common to all firms. Our identification comes from component (ii), the national firm-level shocks to labor demand, such as national changes to technology or production processes, new products or IP, or improved brand recognition or customer demand. These national firm-level shocks would be expected to increase firm i's hiring nationwide relative to other firms in the same occupation, and therefore to also increase firm i's hiring in any given local area k relative to other firms in the same occupation, for reasons that are orthogonal to specific labor supply, demand, or productivity changes in occupation o and metro area k, which will be captured in residual $\xi_{o,k,t}$.[29] For intuition, consider a hypothetical example: assume that in Bloomington, IL,

[28] The shock to firm i's hiring in occupation o and metro area k is estimated as firm i's national vacancy growth in occupation o leaving out growth in metro area k, to purge the shock of influences from any occ-o-metro-area-k specific economic conditions.

[29] The assumption that national firm demand shocks are a good proxy for local firm hiring decisions may

State Farm is the biggest employer of insurance sales agents, while in Amarillo, TX, employment is more concentrated in other insurance companies. If State Farm's brand improves nationwide such that demand for its products increases, it will grow faster than other insurance companies. Under most assumptions about how that growth is allocated geographically, employer concentration of insurance sales agents will grow by more in Bloomington, IL, than in Amarillo, TX.

The primary concern with this instrumental variable is that greater local exposure to fast-growing national firms may increase total local labor demand, as well as increasing employer concentration. Since increasing labor demand would be expected to increase wages, this would be expected to bias our estimated coefficient toward zero, which means that our (negative) estimates of the wage effect of labor market concentration would be, if anything, too small in magnitude. To address this, we control for two correlates of the change in local occupation-specific labor demand: (1) the growth rate of local vacancies in the occupation-metro area ($g_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} g_{j,o,k,t}$), and (2) the predicted growth rate of local vacancies based on large firms' national growth (i.e. the direct linear analog to our concentration index: $\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} \tilde{g}_{j,o,t}$ as defined above). With these controls, we are estimating the effect of a change in local labor market concentration due to changes in large firms' employment, holding constant any direct linear effect on local labor demand or productivity.[30] (In practice, we actually find that including these labor demand controls has little effect on our baseline coefficient estimates). Note that distinguishing labor demand and concentration effects in this way is not simply "identification by functional form" – the insight of our IV approach is precisely that the same average positive change in labor demand can affect concentration positively *or* negatively, depending on how the labor demand shocks are distributed across firms that start out with a small or a large share of local employment. It is this interaction of pre-existing employer market shares with national shocks that drives the local market variation in the instrument (and even nonlinear functions of mean labor demand changes would therefore be unable to eliminate market exposure to the instrument).

A second possible concern is that the firms which expand more nationally may be located in weaker labor markets where wage growth is lower: for firms' national hiring growth shocks to be "as good as randomly assigned", we must assume that the presence of a firm which is nationally fast-growing vs. nationally slower-growing in a local occupation is uncorrelated

be more justifiable in tradable occupations, so in a robustness check we show our results hold for tradables only. Even in non-tradables, however, there is evidence consistent with our identification strategy holding. For example, Giroud and Mueller (2019) find that in non-tradable industries, consumer demand shocks in the headquarters region of a given company affect establishment-level employment in even distant locations of that company.

[30] 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).

with local occupation-specific economic conditions, conditional on our fixed effects. Our baseline analyses include metro area-by-year fixed effects to capture any aggregate weakness in the local labor market, and a robustness check features occupation-by-metro area fixed effects to capture any occupation-specific labor market weakness in a given metro area. Robustness to both of these indicates that this is likely not the variation driving our results.

A third possible concern is bias due to the fact that the non-linear structure of our IV means that different local labor markets may be systematically differently exposed to the "shocks" of the growth of large national firms, and this pattern may be correlated with outcomes related to wages. This is a general concern in non-linear shift share instrumental variables as identified by Borusyak and Hull (2020). Following these authors' recommendation, we construct the expected value of the HHI instrument by generating 100 sets of counterfactual growth rates for the firms in our instrument (randomly re-assigning firms' national vacancy growth rates to other firms), constructing 100 counterfactual instruments, averaging over them and taking the log. We then control for this expected HHI instrument in all our regressions. This removes any omitted variable bias arising from non-random exposure to the national firm vacancy growth shocks.

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

We see our approach as a novel contribution with regard to the problem of estimating the causal 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 (2022); Qiu and Sojourner (2019); Marinescu et al. (2021); Gibbons, Greenman, Norlander and Sørensen (2019)). This circumvents some endogeneity issues, but a concern remains that national occupation trends in concentration may be correlated with unobservable national trends in occupational productivity, demand, or supply, which could confound estimated wage effects.[31] Our strategy based on local variation in concentration allows us to use occupation-year fixed effects to control for national occupation-level factors which affect wages. Other recent empirical work uses M&A activity to generate plausibly exogenous variation in local labor market concentration, including Arnold (2020) for all industries, and Prager and Schmitt (2021) for hospitals. This avoids endogeneity concerns

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

about the cause of the change in concentration, but reflects one specific source of concentration (M&A activity accounts for less than 2% of changes in local employer concentration (Arnold, 2020)) and cannot fully isolate the effects of employer concentration from other local economic effects of the M&A activity. Our approach allows us to examine the effects of various sources of variation across broad swathes of the US labor market, and to control at least somewhat for effects on local labor demand.[32] Ultimately, we believe that this set of complementary identification approaches – based off different variation, and with different strengths – can together provide a useful picture of the effects of employer concentration on wages.

2.7 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 in our data ($\frac{s_{p,k,1999}}{s_{p,1999}}$).[33] Our instrument for the log oo^{occs} is:

$$Z^{oo}_{o,k,t} = log\left(\sum_{p}^{N_{occs}}\left(\pi_{o\to p}\cdot\frac{s_{p,k,1999}}{s_{p,1999}}\cdot\bar{w}_{p,\not{k},t}\right)\right) \tag{10}$$

The identifying variation within a given occupation across different metro areas comes from differences in each metro area's initial occupational employment composition. Identifying variation over time within the same occupation-metro area cell comes from national (leave-one-out) changes over time in wages of local outside-option occupations.[34] For our instrument

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

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

[34] That is, in a year when there is a national wage shock to one of occupation o's outside option occupations p, metro areas which had a higher share of their jobs in occupation p in 1999 are more exposed to this shock and so should see bigger increases in the wage of occupation o. This instrumental variable strategy is closely

to be valid, the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p must be positively correlated with the local wage of occupation p in location k, but must not affect the local wage in initial occupation o other than by increasing the quality of local outside-occupation options (conditional on controlling for occupation-year and metro area-year fixed effects).[35] We discuss conditions for identification further in Appendix G, following the approach to shift-share IVs of Borusyak et al. (2022). Appendix Table A12 also contains answers to common questions about our empirical approach and identification strategy.

3 Results

Our results show that higher employer concentration causally reduces wages on average, and in particular that there is a large, robust, negative effect of employer concentration on wages in occupations with low outward occupational mobility – the occupations for which the local occupation may be a relatively good approximation to the actual labor market that workers face. Our results also show that better outside-occupation options increase wages. These results are robust to a range of alternate specifications and control variables. We discuss these findings further below.

3.1 Results: Employer concentration

In our data, there is a robust negative correlation between log vacancy HHIs and log wages at the occupation-metro area level (Figure 3). In a regression with occupation-year and metro area-year fixed effects, the OLS relationship is strongly statistically significant, with a coefficient of -0.015 (Table 3, column *a*). When instrumenting for the HHI, the coefficient magnitude nearly doubles, to -0.028 (column *c*).[36] This suggests that some combination of omitted variable bias or measurement error biases the coefficient toward zero in simple OLS regressions of wages on HHI. We also introduce a control for our outside-occupation option index (columns *b* and *d* for OLS and 2SLS respectively). The coefficient on the instrumented outside-occupation option index is positive and highly statistically significant, confirming that outside-occupation job options matter for wages, which we discuss further in the next section. After introducing the outside-occupation option index the coefficient on the HHI falls by four log points in the OLS regressions, and by three log points in the

related to Beaudry et al. (2012), who use national industry wage premia to substitute for metro area-level industry wages when estimating spillover effects of cities' industrial composition.

[35]The inclusion of these fixed effects means that differences in metro area-level economic conditions 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 industry shocks. We show that our results are robust to controlling for common exposure to industry shocks (see Appendix Table A6).

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

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 (Appendix Figure A13).

How big is the average effect of employer concentration on wages? Our baseline coefficient estimate of -0.025 in Table 3, column d – instrumenting for both employer concentration and outside-occupation job options – suggests that going from the HHI faced by the median worker to the HHI faced by the worker at the 95th percentile in 2019 (from an HHI of 137 to 1,845) would lead to a 6.5 log points lower hourly wage.[37]

Heterogeneity by occupational mobility. Re-running our baseline regression, but allowing the coefficients on the HHI to vary for occupations with different degrees of outward occupational mobility, we find that the average effect of within-occupation employer concentration on wages conceals substantial heterogeneity (Table 4, Figure 5). For the quartile of occupations with the lowest outward mobility, as proxied by our occupation "leave share", our coefficient estimate suggests that going from the median to the 95th percentile HHI faced by workers would be associated with 10.7 log points lower wages.[38] For the second quartile of outward mobility, an equivalent increase in the HHI would be associated with 7.5 log points lower wages, and for the third and fourth quartiles (the most outwardly mobile occupations), the predicted decline in wages would be 3.1 and 3.4 log points respectively.[39] Since occupations with low "leave shares" are more likely to be good approximations to workers' 'true' labor market, this suggests that high employer concentration – when calculated over workers' true effective labor market – can have a very large impact on wages. We list occupations with the highest and lowest "leave shares" in Appendix Table A2.

Robustness. We explore a number of variations on our baseline analyses, illustrated in Figure 6 (coefficient estimates and standard errors shown in Appendix Tables A6-A9).

First, we run our regressions with different specifications. Without employment weighting, our coefficient estimates remain significant but are smaller, suggesting that our large negative estimates of the effect of employer concentration on wages are not driven by small (unrepresentative) labor markets. Our coefficient estimates remain similar if we remove the controls for vacancy growth and the expected HHI instrument. In an additional robustness

[37] Calculated as $(\ln(1845) - \ln(137)) \cdot -0.025 = -0.065$. This is at the low end of the range of existing estimates presented in Marinescu and Hovenkamp (2019).

[38] Calculated as $(\ln(1845) - \ln(137)) \cdot -0.041 = -0.107$.

[39] This pattern is consistent with Prager and Schmitt (2021), who find that hospital mergers which induce large increases in concentration reduce nursing and pharmacy worker wages substantially, somewhat suppress wages of non-medical hospital professionals, and have no detectable effect on wages for workers in maintenance and repairs, operations, housekeeping, catering, and medical records. Nursing and pharmacy workers have much lower outward occupational mobility than the latter group.

check, we follow Gabaix and Koijen (2020) in adding a 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 *local* occupation-specific shocks that are common to all firms. These results are shown in Appendix Tables A6 and A8 , columns (a), (b), and (c) respectively.

Second, we run our regressions with different samples. Since the logic of our HHI instrument holds less well for very low HHI cells, we drop occ-metro area cells with very low HHIs (less than 50). To address concerns about representativeness in our BGT vacancy data, we exclude any occupations or metro areas which are substantially underrepresented in the vacancy data (those in the bottom third of 'represented-ness'). To address the concern that the logic of our instrument works less well for occupations producing non-tradable goods or services (since local hiring decisions may be driven by local occupation-specific economic conditions), we run a specification with only occupations which produce tradable goods or services.[40] In all these cases, our coefficient estimates are similar to the baseline both for the full sample and by occupational mobility quartile. These result are shown in Appendix Tables A6 and A8 , columns (d), (e), and (f) respectively.

Third, we run our regressions with alternate versions of our HHI instrument. One might be concerned that the exclusion restriction in our instrument is weak if our results are driven by firms which are only present in a few metro areas. To address this, we create a version of our instrument which only uses the national leave-one-out occupational vacancy growth of *large* firms (firms with vacancies in at least five metro areas in an occupation o in year t) to instrument for local firm-level vacancy growth $g_{j,o,k,t}$ in occupation o. Another concern might be that by excluding negative shocks (i.e. where a firm's national occupation-specific hiring contracts from one year to the next), we exclude useful information. We therefore create a version of our instrument which uses all hiring shocks, positive and negative. A further concern might be around using a expression for a change in HHI to instrument for the level of the HHI. We therefore formulate an alternative instrument based on the predicted level rather than change in HHI, $Z^{HHI,alt}_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \ln\left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2}\right)$.[41] The magnitude and significance of coefficients, as well as their pattern across occupational mobility quartiles, remains similar to our baseline with each of these three altered HHI instruments. These results are shown in Appendix Tables A7 and A9, columns (a), (b), and (c) respectively.

In further robustness analyses, we control for an industry Bartik shock to proxy for local

[40] We define tradable occupations as all occupations in manufacturing or production, extraction, or farming, fishing, and forestry (SOC codes starting in "45", "47-5", and "51"), as well as all occupations defined as at least somewhat teleworkable by Dingel and Neiman (2020) (i.e. those with a teleworkability index > 0).

[41] This alternative formulation is also a different way to reduce the bias introduced from the log transformation in our baseline instrument, as opposed to our approach in our baseline specifications of controlling for the expected HHI instrument (Borusyak and Hull, 2020).

metro area occupation exposure to common national industry trends (Appendix Tables A7 and A9 column (d)), and re-run our regressions by occupation mobility quartile using a measure of outward occupational mobility we construct from the CPS rather than from the BGT resume data (Appendix Table A9 column (f)).[42]

Finally, we run our baseline regressions with fixed effects for occupation-metro area and year, rather than occupation-year and metro area-year (Appendix Tables A7 and A9 column (e)). Here, identifying variation comes from differential year-to-year national growth rates of the large employers present in a given occupation-metro area labor market over the period 2011–2019, rather than from differential exposure to the fast-growing national employers in a local occupation relative to other localities and occupations. The coefficient estimate for the average effect, at -0.07 is smaller than our baseline estimate but still statistically significant at the 10% level. However, in line with our insight that average concentration effects conceal substantial heterogeneity, when we allow the estimates to vary by quartile of outward occupational mobility, the coefficient estimate for the lowest-mobility quartile is -0.017 and statistically significantly different from zero at the 1% level (and coefficient estimates for other quartiles follow a decreasing pattern as in our main results, but are not statistically significantly different from zero).[43]

These additional analyses consistently show the same key pattern as our baseline results: the average effect of employer concentration on wages is statistically significant but of modest size, while there is a consistently large, economically and statistically significant negative effect for the least mobile occupations (those for which the occupation is a good measure of their true labor market). This effect then declines in magnitude for more mobile occupations, which is a robust pattern across all the different specifications.

3.2 Results: outside-occupation options

While not the primary focus of our paper, we can also use our baseline regressions to ask: how big are the effects of changes in the value of outside-occupation options on wages? Finding a positive and significant coefficient on our outside-occupation option index is a joint test of two aspects: first, that job options outside workers' occupation matter for the wages inside their occupation, and second, that occupational mobility flows (from our BGT resume data)

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

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

can be used to infer what those relevant outside occupations are.[44] Our baseline 2SLS IV coefficient estimate (Table 3 column d) is indeed large, positive, and significant, and suggests that a 10 log point higher outside-occupation option index leads to 1.02 log points higher wages in workers' own occupation. Since the outside-occupation option index approximates a weighted average of wages in other local occupations, this implies roughly that 10 percentage point higher wages in workers' outside option occupations is associated with 1 percent higher wages in their own occupation in the same year. Applying the coefficient to understand spatial wage differences, we see that moving from the 25th to the 75th percentile value of outside-occupation options across metro areas for the median occupation leads to 4.1 log points higher wages.[45] 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.[46] The magnitude and statistical significance of the effect of these outside-occupation option shocks on wages are highly robust to a number of alternate specifications, shown in Appendix Tables A6-A11 and discussed in Appendix G.

Finding a large, significant, and positive effect of shocks to outside-occupation options on wages thus demonstrates both that workers' true labor markets are broader than their narrow 6-digit SOC occupations, and that our "probabilistic" method of identifying relevant outside options using observed occupation switches can capture workers' true labor markets relatively well.[47]

3.3 Integrating mobility into labor market analysis

Our results above suggest that conventional regressions of wages on labor market concentration are misspecified for occupations which do not represent workers' true labor markets – and more so, the more mobile workers are. However, there is increasing demand by policymakers to consider labor market power in antitrust enforcement, and also by researchers to study monopsony power as a consequence of, and determinant of, of firm-level decisions.

[44] This is important for the interpretation of our concentration results above: our interpretation of our results for low-outward-mobility occupations relies on the hypothesis that outside-occupation job options, as proxied by mobility flows, are an important part of workers' labor markets.

[45] Calculated using the median occupation's interquartile range of the log outside-occupation option index (across metro areas) in 2019, which was 0.40, and applying our coefficient estimate of 0.102. For employer concentration, for the median occupation moving from the 25th to the 75th percentile HHI across metro areas would result in a wage increase of 3.6 log points.

[46] For a specific example, 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 half of this difference may be attributable to differential availability of outside-occupation job options.

[47] 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.

Both of these efforts require measures of employer concentration over clearly-defined labor market boundaries. While our findings caution against simply using administrative definitions of occupational codes or industries to delineate labor markets in these applications, they also raise the question if there is a pragmatic way to improve these analyses in a way that accounts for worker mobility.

In this section, we suggest that worker mobility flows can be used to construct "mobility-adjusted labor markets" (MALMAs) that better capture the effective range of options that workers are likely to consider in their job search. First, we use our measure of the likelihood of transition to other occupations $\pi_{o \to p}$ defined before to capture the "relevance" of other occupational labor markets for a worker's job search. Intuitively, a $\pi_{o \to p}$ of 1%, for example, means that, of all the workers in occupation o who change job between years t and $t+1$, 1% of them go to jobs in occupation p. (Note that most workers, when they change job, stay in the same occupation, and $\pi_{o \to p}$ captures this for $p = o$).

Then, the MALMA for a worker in occupation o is defined as the set of all job vacancies in metro area k, in the occupations p that have relevance above a chosen threshold x:

$$MALMA_{o,k,t} = \bigcup_{p:\pi_{o \to p} > x} v_{i,p,k,t}, \tag{11}$$

where $v_{i,p,k,t}$ is the set of all vacancies of firms i in occupation p in metro k at time t. Note that in all empirically relevant settings this means that the local MALMA for workers in an occupation always includes all jobs in their current occupation, and also jobs in other occupations that are sufficiently likely to be destinations for jobseekers.

While our motivation is similar to that in other papers which construct labor markets based on mobility data (e.g Schmutte (2015); Nimczik (2018)), a key difference, which we believe is essential for antitrust applications and the study of labor market power, is the asymmetry of these clusters. As we illustrate in Section 2 and Appendix Section E, mobility between occupations can be highly asymmetric due to feasibility or desirability of moves being more limited in one direction than in another. Labor market definitions should allow for this asymmetry, such that occupation B being part of the relevant labor market for workers currently in occupation A does *not* imply the converse (A being relevant for workers in B) – and this asymmetry is permitted by the definition in equation 11. While mobility-adjusted labor markets can be symmetrical they do not have to be - and in practice often are not. This distinguishes this labor market definition from approaches that simply cluster occupations or industries into new broader labor markets that are still implicitly symmetric: instead, we define a new MALMA from the perspective of each initial six-digit SOC occupation.

The corresponding mobility-adjusted labor market concentration is then constructed over the broader MALMA rather than narrowly within the worker's current occupation and is given by

$$HHI_{o,k,t}^{\text{MALMA}} = \sum_{i=1}^{N} \left(\frac{v_{i,p,k,t} \times \mathbb{1}[\pi_{o \to p} > x]}{MALMA_{o,k,t}} \right)^2. \tag{12}$$

This adjusted employer concentration measure is in most cases smaller than the equivalent concentration computed for the worker's current occupation only, since expanding the set of occupations workers can consider typically reduces the concentration of any single employer. To the degree that workers are able to take up the jobs in these other occupations to which we frequently observe such moves, this lower employer concentration more closely reflects the worker's actual labor market options. For a non-trivial minority of occupation-metro area cells, however, the HHI^{MALMA} is actually larger than the HHI calculated only on their local occupation: this can be because the large employers in the initial occupation o are also (even larger) employers in the transitioned-to occupations p, or because the transitioned-to occupations are highly concentrated. There will also be local variation in the degree to which measured employer concentration changes when MALMAs are used: even if a worker *could* in theory move into jobs in other occupations, this only expands their labor market if there are jobs in those other occupations available in their particular geographic location.

To show the degree to which MALMAs differ from narrow occupational labor markets, we compute for each major occupation group the average employment-weighted local employer concentration for the included 6-digit occupations. We compare the average employer concentration obtained for within-occupation labor markets and for MALMAs calculated with a 1% $\pi_{o \to p}$ relevance threshold in Figure A15. The figure shows that employer concentration in MALMAs is generally smaller than in single-occupation labor markets, but that the adjustment differs substantially across major occupation groups. In particular, for low mobility occupation groups like health care or teaching, MALMA HHIs are more similar to single-occupation HHIs, while for relatively high mobility occupation groups like food preparation & serving, or office & administrative support, the difference is substantial.

Mobility-adjusted concentration effects. How does the use of a mobility-adjusted labor market definition change our estimates of the effect of employer concentration on wages? Table 5 shows the estimated effects of employer concentration on wages, using the same instrument as in the previous sections to identify exogenous movements in concentration. Columns (a) and (b) show the results for a more inclusive MALMA with a threshold of 1% relevance for inclusion of other occupations, while columns (c) and (d) apply a 2% relevance threshold for constructing the MALMA. The table shows two results: on the one hand, the estimated effects of employment concentration are larger than the effects estimated for single occupations, suggesting in column (a) that a change from the median to the 95th percentile of this mobility-adjusted HHI faced by workers (with a 1% relevance threshold for the MALMA) would be associated with an 14 log point decline in wages in the labor market - more than twice the baseline effect for more narrow occupations.[48] Intuitively, the single occupation labor markets *over*-estimate the changes in the level of employer concentration in a worker's labor market, as they draw the boundaries too narrowly. As a result of overestimating the change in the independent variable, the size of the coefficient in the baseline estimation is too small and is corrected upwards when labor markets are adjusted for mobility.

On the other hand, the coefficient estimates across different mobility quartiles are all statistically significantly different from zero, but no longer statistically significantly different *from one another* or from the average effect estimate when we consider the more inclusive definition of MALMA in column (b). This suggests that the mobility-adjusted labor markets do an adequate job at capturing the variation in labor market size and employer concentration impacts due to different occupational mobility. This analysis shows that our simple method of adjusting labor markets for mobility can produce employer concentration effect estimates that better account for heterogeneity in the extent of workers' labor markets. Furthermore, once labor markets have been adjusted to account for mobility, the average effect estimate across workers does a reasonable job of capturing employer concentration effects for *all* workers and it does not seem necessary to additionally allow for heterogeneity in the effect estimates.[49] Overall, we find that adjusting labor market definitions using mobility is an easily implemented correction when analyzing the labor market effects of changes in monopsony power, or for antitrust applications. Moreover, this adjustment matters: the estimated wage effects of concentration are substantively larger when labor markets account for mobility.

4 Discussion and Implications

What do 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-concentration occupation-by-metro area labor market in 2019, relative to a scenario where their mobility-adjusted HHI^{MALMA} is reduced

[48] This is calculated as $(ln(565) - ln(37) * -0.053 = -0.144$.

[49] Columns (c) and (d) of Table 5 show that applying a higher threshold for relevance of other occupations results in MALMAs and employer concentration effect estimates that are only partially adjusted: the effect estimates are larger and more similar across mobility quartiles than in our baseline estimates, but this smaller labor market adjustment does not fully eliminate differences in effect estimates across mobility quartiles. However, even this partial adjustment likely represents a closer approximation to comparable labor markets across occupations with different mobility–relative to using only the current occupation.

to 150 (which is roughly that of the median labor market in 2019),[50] as follows:

$$\text{wage effect}_{o,k,t} = \left(log(HHI^{MALMA}_{o,k,t}) - log(150)\right) \cdot \gamma_1^q \tag{13}$$

where γ_1^q denotes the estimated coefficient on the $\log(HHI^{MALMA})$ in our regression specification in Table 5 column (b), for the appropriate quartile q of outward occupational mobility. Note that this exercise considers the effect of changes in employer concentration *holding all else constant*, including local productivity. It can illustrate the degree to which wages may be marked down from local occupational productivity as a result of employer concentration, but cannot necessarily illustrate what would happen if a specific policy or business decision were to change local employer concentration (as it might also change local productivity). It also rests on the assumption that we can apply our estimated coefficients linearly.

Of the 117 million workers in labor markets covered by our wage, employment, and concentration data in 2019, our counterfactual wage exercise suggests that roughly 14% had wages which were at least 2% lower as a result of above-median employer concentration.[51] Our estimates suggest that employer concentration slightly widens inequality: the share of workers affected by employer concentration is larger in low-wage occupations and in low-wage cities (Figures 7 and 8).

Which occupations are most affected by employer concentration? In Table 6, we list the twenty occupations with the largest number (in Panel A) or share (in panel B) 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 the most affected occupations are in healthcare. This includes an estimated 1.4 million registered nurses, 700,000 nursing assistants, 300,000 licensed practical and vocational nurses, and 250,000 pharmacists, with wage effects of 2% or greater.[52] This is because healthcare workers often have (i) high employer concentration within their occupations, (ii) low occupational mobility, and (iii) what mobility they do have is often to other

[50] 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 market power concern in product markets (though of course factors other than concentration can still generate market power (Naidu and Posner, 2021)). Note also that while our estimates focus on wages, note that employer concentration may also affect non-wage benefits and workplace amenities (Qiu and Sojourner, 2019; Marinescu, Qiu and Sojourner, 2020).

[51] We estimate a very similar share excluding occupations which are highly underrepresented in the BGT vacancy data (a 'represented-ness' of less than 0.5). Our estimate of 14% is likely an underestimate of the effect of employer concentration across the full U.S. labor market since our data lacks coverage of non-metropolitan areas and of some small occupations, where one would expect employer concentration to be higher than average.

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

healthcare occupations, which often share the same set of employers (such as large hospitals). We also see several large low-wage occupations where a non-trivial minority of workers are affected by employer concentration: for example, an estimated 125,000 retail salespersons (employed in small metropolitan areas with high concentration).[53] Importantly, the list of most-affected occupations is very different using our HHI^{MALMA} estimates measured over mobility-adjusted labor markets: simply applying the *average* estimated effect for an unadjusted HHI on a single occupation, without accounting for workers' differential moblity across occupations, leads to substantial overestimation of the effect of employer concentration for high-mobility occupations with a diverse range of possible employers, like bank tellers, retail salespersons, or restaurant cooks, and substantial underestimation of the effects of employer concentration for low-mobility occupations with a limited range of possible employers, like registered nurses, nursing assistants, or pharmacists.

While we estimate that employer concentration suppresses wages for several million workers, we underscore also that the majority of American workers do not work in labor markets with even moderate degrees of employer concentration – and therefore likely do not experience significant wage suppression as a result of employer concentration. Thus, policymakers should focus attention on the subset of workers who face both concentrated labor markets within their occupation and limited opportunities for occupational mobility.[54]

Implications: antitrust. One area where this analysis can be applied is antitrust.[55] 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 it is crucial this screen takes into account occupational mobility. Section

[53]According to our estimates, large numbers of hairdressers, hairstylists, and cosmetologists are also affected by employer concentration. Note that we consider a salon chain, many of which are franchised, to be a single employer. It is an open question as to whether it is better to consider employer concentration at the level of individual franchised salons. Covenants not to compete within a franchise may often make this distinction moot in practice.

[54]While our estimates show that increases in employer concentration reduce local wages, they cannot tell us whether the ultimate incidence of these wage reductions falls on firms in the form of higher profits, or consumers in the form of lower prices. The balance likely depends on the nature of product market competition. Kahn and Tracy (2019) argue that the ultimate incidence of local labor market concentration falls to a large extent on local landowners as lower local wages reduce local rents and house prices. Understanding the ultimate incidence of these effects is important to determine the appropriate policy response.

[55]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, in the FTC retrospective merger review process, and in FTC public comments (e.g. regarding a hospital merger in Hendrick Texas in September 2020).

3.3 suggests one potential approach to incorporating mobility into the analysis of employer concentration: by adjusting labor markets to (asymmetrically) account for the mobility-related relevance of other occupations, policymakers and researchers can better capture relevant labor markets and obtain estimates of concentration changes and predicted wage effects that are stable across workers with differential mobility.[56]

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.[57] 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. An alternative might be strengthening minimum wages or benefits standards in local labor markets characterized by high employer concentration.[58]

[56]These adjustments are easy to implement using mobility data that can be obtained from the authors' websites or computed from public survey data. Our recommendation differs from for example 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. Screening based only on local within-occupation HHI without considering outward occupational mobility would likely lead to some mergers being scrutinized which may have little effect on wages, while others which may have anti-competitive effects may go unnoticed.

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

[58]In our conceptual framework or in Jarosch et al. (2019) for example, higher worker bargaining power reduces the weight placed on the outside option in the wage bargain and therefore reduces the importance of employer concentration in wage determination. Prager and Schmitt (2021) and Benmelech et al. (2022) find evidence consistent with higher unionization reducing the strength of the wage-concentration relationship. Higher minimum wages would be expected to have less of a negative effect on employment in labor markets where employers have monopsony power, and Azar, Huet-Vaughn, Marinescu, Taska and Von Wachter (2019) find evidence of this.

Implications: promoting mobility. Our results suggest that employer concentration within a local occupation matters substantially less if workers can find similarly good jobs outside their occupation. By the same logic, this would also be true if workers can easily move geographically. This suggests that policies which make it easier to switch occupation and/or to work in different geographic areas may – by increasing workers' outside options – reduce the degree to which employer concentration can suppress wages. 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.[59]

Implications: labor market definition. Our analysis of occupational mobility revealed large and heterogeneous rates of workers moving out of their current occupation every year, showing that their effective labor market transcends the boundaries of their current occupation. Moreover, our results suggest that ignoring the effective boundaries of dynamic labor markets may lead to biased estimates of labor market impacts, for instance by neglecting the role that simultaneous changes in outside options play in worker outcomes. This means that academic or policy-oriented analyses of the impact of shocks on workers should take into account both the larger scope of which jobs are relevant for workers, and the potential for spillovers of labor market changes between occupations. Concretely, research on labor markets should explicitly grapple with, and justify, why a particular administrative definition (like SOC 6-digit occupations, or NAICS 4-digit industries) captures the relevant labor market for the issue at hand, and should consider probabilistically incorporating other occupations or industries that may affect a worker's outcomes.[60] While studies of product markets have long recognized the need to define product markets well, the same care is needed when discussing labor markets.

5 Conclusion

Our findings point to a middle ground between two prominent views about the effects of employer concentration in the US labor market. On the one hand, employer concentration is *not* a niche issue confined to a few factory towns: we find large, negative, and significant effects of employer concentration on wages when estimated using nuanced market definitions and plausibly exogenous variation across the majority of the US labor market, and our back-of-the-envelope calculations suggest that one in six U.S. private sector workers experience

[59]See Johnson and Kleiner (2020) on the mobility effect of state licensing standards and Ganong and Shoag (2017) on the mobility effect of housing costs. Restrictions on worker mobility *within* an occupation like non-compete clauses (Starr, Prescott and Bishara, 2021; Johnson et al., 2020) could also exacerbate the effects of employer concentration on wages.

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

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).[61] The fact that employer concentration affects wages for several million American workers suggests that increased policy attention to this issue is appropriate, in terms of antitrust, policies to raise wages, and policies to increase worker mobility. For these policy decisions, our work underscores that the definition of the labor market is vitally important, that the local occupation is a good definition of the local labor market for occupations with low outward mobility but broader concepts of the labor market should be used for more outwardly mobile occupations, and that when estimated within well-defined labor markets, employer concentration can have large effects on workers' wages.

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

References

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

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

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

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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, "Monopsony in the US labor market," *Federal Reserve Bank of Richmond Working Paper*, 2020.

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

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

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.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A Woodbury, "Wage posting or wage bargaining? a test using dual jobholders," *Journal of Labor Economics*, 2022, *40* (S1), S469–S493.

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

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

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

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

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

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

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

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

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

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.

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.

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

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

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

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

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

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.

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.

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	0.09	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 occ. transitions which cross SOC 2d boundary, by occ. (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: Excludes occupations with <500 observations in BGT resume data. In Panel A, an observation is a person-year that is also observed in the data the following year. Panel B shows, by occupation, the share of workers observed in a new job or new occupation from one year to the next, and the "leave share", defined in section 2.4. Panel C shows, by occupation, the share of SOC 6-digit occupational transitions which cross SOC 2-digit boundaries. Percentiles are 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	79	138	343	882	2,143	4,763	8,025	10,000
HHI, emp-wt	6	14	20	53	137	404	1,035	1,845	5,038
HHI^{MALMA}	8	14	22	51	145	386	1,022	1,955	5,400
HHI^{MALMA}, emp-wt	5	7	9	16	37	110	313	565	1,828
				Panel B: Outside-occupation option index oo^{occs} (2019)					
oo^{occs}	1.43	2.19	2.68	3.63	4.94	6.73	8.98	10.76	16.39
oo^{occs}, emp-wt	2.15	3.13	3.72	4.88	6.39	8.46	11.33	13.49	20.87
$\frac{oo^{occs}}{wage}$	0.03	0.06	0.09	0.14	0.22	0.32	0.43	0.50	0.70
$\frac{oo^{occs}}{wage}$, emp-wt	0.06	0.11	0.14	0.22	0.32	0.42	0.52	0.59	0.74
				Panel C: Occupation-metro area wages and employment (2019)					
Employment	30	40	50	90	220	670	2,000	3,990	15,010
Employment, emp-wt	80	240	470	1,532	5,820	21,020	55,270	96,500	184,050
Mean hourly wage	9.73	11.64	13.22	16.78	22.98	32.64	45.68	55.33	88.64
Wage, emp-wt	9.54	10.86	12.23	14.76	19.88	32.21	47.94	59.48	84.00
				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. In 2019 this comprised 107,695 unique occupation-by-metro area labor markets, which collectively contained 117,022,999 workers. Rows labeled "emp-wt" show the percentiles weighted by occupation-metro area employment. Panel D shows the national 10th, 25th, 50th, 75th, and 90th percentile of the hourly wage distribution according to the full BLS OES data set, for comparison. The HHI^{MALMA} in Panel A is defined in section 3.3.

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	-0.015***	-0.011***	-0.028***	-0.025***
	(0.001)	(0.001)	(0.005)	(0.005)
Log outside-occ. options		0.133***		0.102***
		(0.006)		(0.011)
Vacancy growth			-0.462***	-0.407***
			(0.123)	(0.123)
Predicted vacancy growth			0.217*	0.125
			(0.117)	(0.118)
Expected HHI instrument			0.003**	0.003**
			(0.001)	(0.001)
Observations	443,233	443,233	443,233	443,233
F-stat			364	175
Fixed effects				
Occ-year	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y

Notes: Heteroskedasticity-robust standard errors clustered at the metro area-by-year 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). See text for detailed explanation of instruments and controls.

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

Dependent variable:	Log wage			
	(a) OLS	(b) OLS	(c) 2SLS IV	(d) 2SLS IV
Log HHI	-0.025***	-0.021***	-0.045***	-0.041***
X Q1 occ mobility	(0.002)	(0.002)	(0.006)	(0.006)
Log HHI	-0.020***	-0.015***	-0.034***	-0.029***
X Q2 occ mobility	(0.001)	(0.001)	(0.005)	(0.005)
Log HHI	-0.008***	-0.004***	-0.016***	-0.012**
X Q3 occ mobility	(0.001)	(0.001)	(0.005)	(0.005)
Log HHI	-0.002*	-0.001	-0.015***	-0.013**
X Q4 occ mobility	(0.001)	(0.001)	(0.005)	(0.005)
Log outside-occ. options		0.130***		0.103***
		(0.006)		(0.011)
Vacancy growth			-0.452***	-0.396***
			(0.123)	(0.123)
Predicted vacancy growth			0.222*	0.129
			(0.119)	(0.120)
Expected HHI instrument			0.003*	0.003**
			(0.001)	(0.001)
Observations	443,233	443,233	443,233	443,233
F-stat			91	70
Fixed effects				
Occ-year	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y

Notes: Heteroskedasticity-robust standard errors clustered at the metro area-by-year 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). Independent variables labelled "X Qi outward mobility" show the coefficient on an interaction term between the HHI or outside-occupation option index (respectively) with an indicator variable which takes the value 1 if the occupation in question is in the ith quartile of outward occupational mobility (where "Q1" represents the least outwardly mobile occupations, and so on). See text for detailed explanation of variables.

Table 5: Wage effects of employer concentration: mobility-adjusted labor markets

Dependent variable:	Log wage			
Labor market:	MALMA (1 pp threshold)		MALMA (2 pp threshold)	
	(a) 2SLS IV	(b) 2SLS IV	(c) 2SLS IV	(d) 2SLS IV
Log HHI^{MALMA}	-0.053***		-0.030***	
	(0.011)		(0.006)	
Log HHI^{MALMA} X Q1 occ mobility		-0.063***		-0.044***
		(0.011)		(0.007)
Log HHI^{MALMA} X Q2 occ mobility		-0.055***		-0.034***
		(0.010)		(0.006)
Log HHI^{MALMA} X Q3 occ mobility		-0.037***		-0.016***
		(0.011)		(0.006)
Log HHI^{MALMA} X Q4 occ mobility		-0.042***		-0.017***
		(0.013)		(0.006)
Log outside-occ. options	0.095***	0.098***	0.104***	0.105***
	(0.012)	(0.012)	(0.010)	(0.010)
Vacancy growth	-0.343***	-0.342***	-0.395***	-0.394***
	(0.121)	(0.121)	(0.122)	(0.123)
Predicted vacancy growth	0.163	0.169	0.131	0.136
	(0.127)	(0.129)	(0.119)	(0.121)
Expected HHI instrument	0.004**	0.004**	0.003**	0.003**
	(0.002)	(0.002)	(0.001)	(0.001)
Observations	443,233	443,233	443,233	443,233
F-stat	80	33	151	61
Fixed effects				
Occ-year	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y

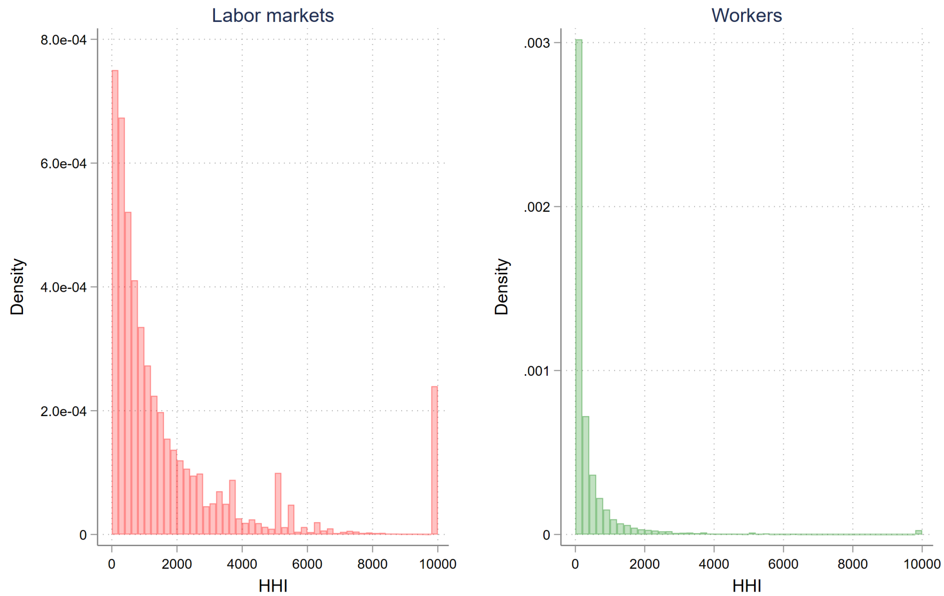
Notes: Heteroskedasticity-robust standard errors clustered at the metro area-by-year level shown in parentheses. Units of observation are 6 digit SOC occupation by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions feature occupation-by-year and metro 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). Independent variables labelled "X Qi outward mobility" show the coefficient on an interaction term between the MALMA HHI and an indicator variable which takes the value 1 if the occupation in question is in the ith quartile of outward occupational mobility (where "Q1" represents the least outwardly mobile occupations, and so on). See text for detailed explanation of variables.

Table 6: Occupations highly affected by employer concentration

Occupation	National employment	Wage effect 2% (Share)	(Number)
Panel A: Twenty occupations with largest number of workers affected by employer concentration			
Registered nurses	2,982,280	.46	1,384,070
Nursing assistants	1,419,920	.5	711,400
Hairdressers, hairstylists, and cosmetologists	385,960	.92	353,710
Licensed practical and licensed vocational nurses	697,510	.43	298,860
Pharmacists	311,200	.82	255,270
Secretaries and administrative assistants*	2,038,340	.11	227,180
Fitness trainers and aerobics instructors	325,500	.63	203,620
Heavy and tractor-trailer truck drivers	1,856,130	.11	199,090
General and operations managers	2,400,280	.068	164,000
Emergency medical technicians and paramedics	521,200	.31	162,430
Educational, guidance, school, and vocational counselors	296,460	.53	158,050
Cooks, restaurant	1,401,890	.11	150,180
Radiologic technologists	207,360	.7	144,870
Retail salespersons	4,317,950	.029	125,780
Lawyers	657,170	.19	124,070
Medical assistants	712,430	.17	118,550
Bookkeeping, accounting, and auditing clerks	1,512,660	.073	110,690
Preschool teachers, except special education	431,350	.24	103,980
Nurse practitioners	200,600	.51	102,470
First-line supervisors of food preparation and serving workers	1,011,100	.099	100,110
Panel B: Twenty large occupations with largest share of workers affected by employer concentration			
Hairdressers, hairstylists, and cosmetologists	385,960	.92	353,710
Pharmacists	311,200	.82	255,270
Radiologic technologists	207,360	.7	144,870
Physician assistants	120,090	.63	75,640
Fitness trainers and aerobics instructors	325,500	.63	203,620
Educational, guidance, school, and vocational counselors	296,460	.53	158,050
Nurse practitioners	200,600	.51	102,470
Nursing assistants	1,419,920	.5	711,400
Registered nurses	2,982,280	.46	1,384,070
Phlebotomists	128,290	.45	58,190
Licensed practical and licensed vocational nurses	697,510	.43	298,860
Healthcare social workers	174,890	.4	69,750
Clinical, counseling, and school psychologists	113,270	.35	39,900
Emergency medical technicians and paramedics	521,200	.31	162,430
Occupational therapists	133,570	.29	38,340
Physical therapists	233,350	.27	62,600
Speech-language pathologists	154,360	.26	40,320
Preschool teachers, except special education	431,350	.24	103,980
Physicians and surgeons, all other	390,680	.24	93,320
Medical and health services managers	394,910	.19	75,150

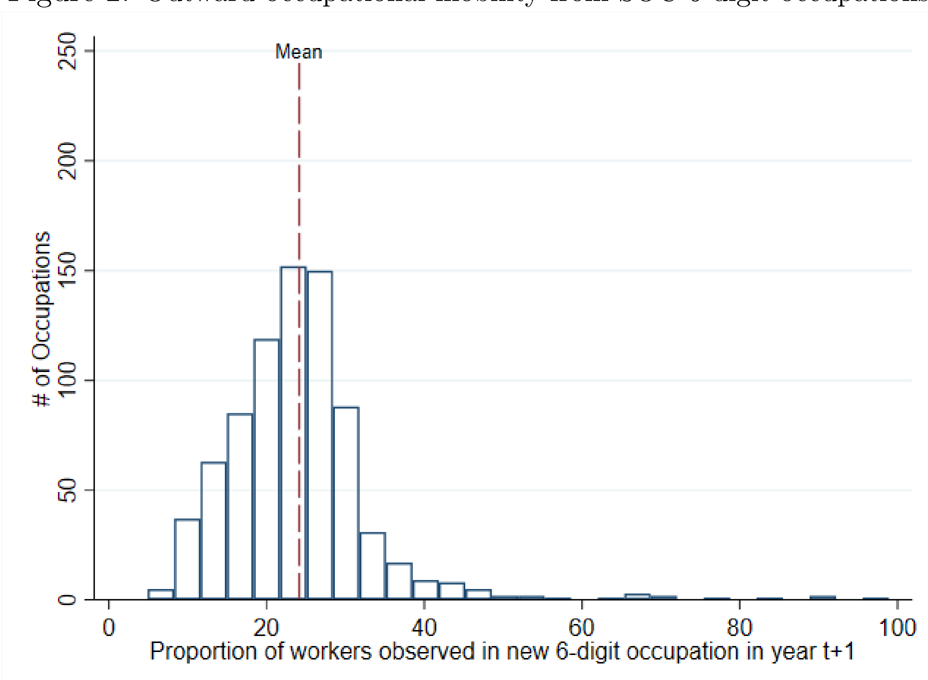
Notes: Shows occupations with the largest number (panel A) or share (panel B) of workers who experience an estimated wage impact of 2% or more as a result of above-median employer concentration in 2019. Panel B sample limited to occupations with national employment 100,000 in 2019. The columns list, 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). We exclude occupations from this list that are highly under-represented in the BGT vacancy data relative to overall employment. Excluded occupations include several large low-wage occupations for which we have low representativeness in the BGT vacancy data, such as Personal care aides, Waiters and waitresses,Janitors and cleaners, Cashiers. We also exclude Maids and housekeeping cleaners, and Nonfarm animal caretakers, since the large employers identified by the BGT vacancy data for this occupation are disproportionately postings on online listing sites. *: Full title is *Secretaries and administrative assistants, except legal, medical, and executive.*

Figure 1: Histogram of employer HHI across labor markets and across workers, 2019



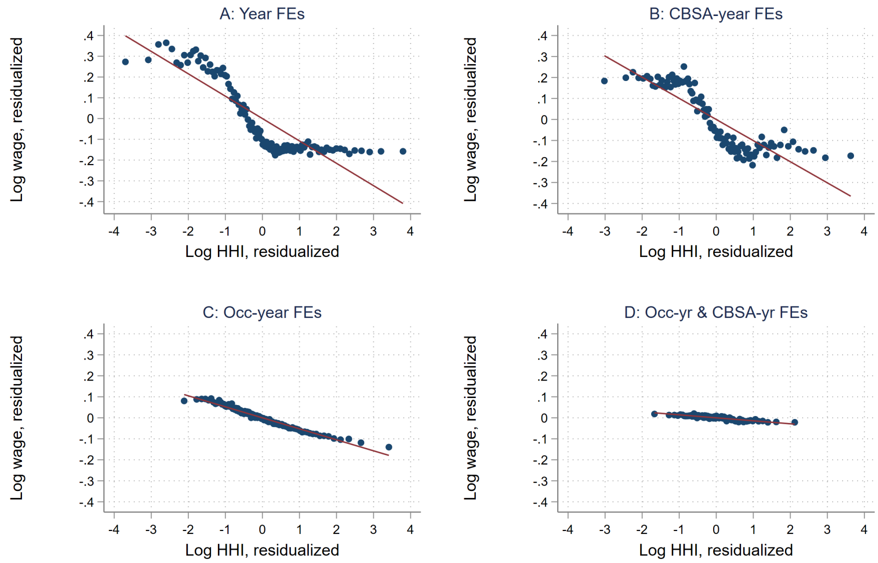
Note: HHI is measured using Burning Glass Technologies vacancy data, at the level of a SOC 6-digit occupation by metro area labor market. Our data covers occupation-metro area labor markets which include 117m of the 151m workers in the U.S. labor market in 2019. Left panel shows the distribution of HHIs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

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



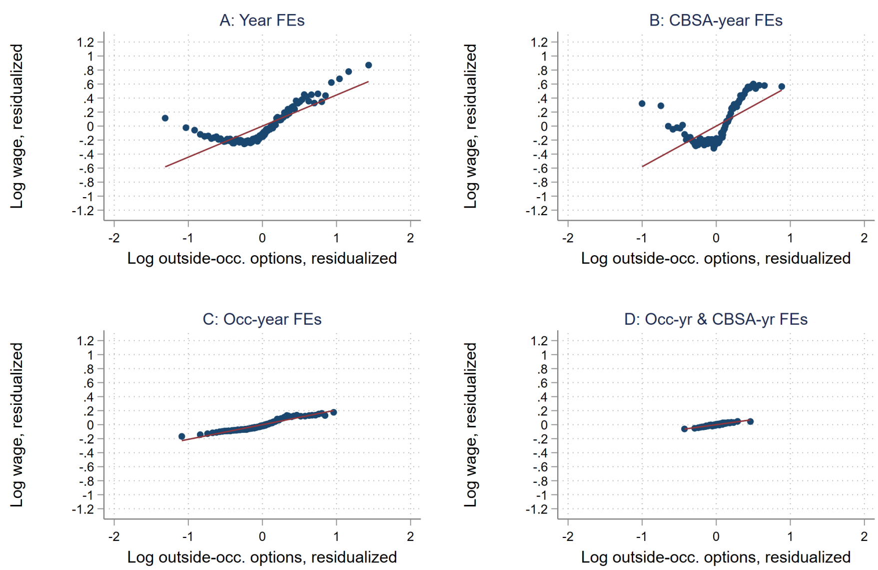
Distribution of the "occupation leave share" – the probability that a worker will leave their occupation conditional on leaving their job – by occupation. Occupation leave share is calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample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

Figure 3: Correlations between wage and HHI



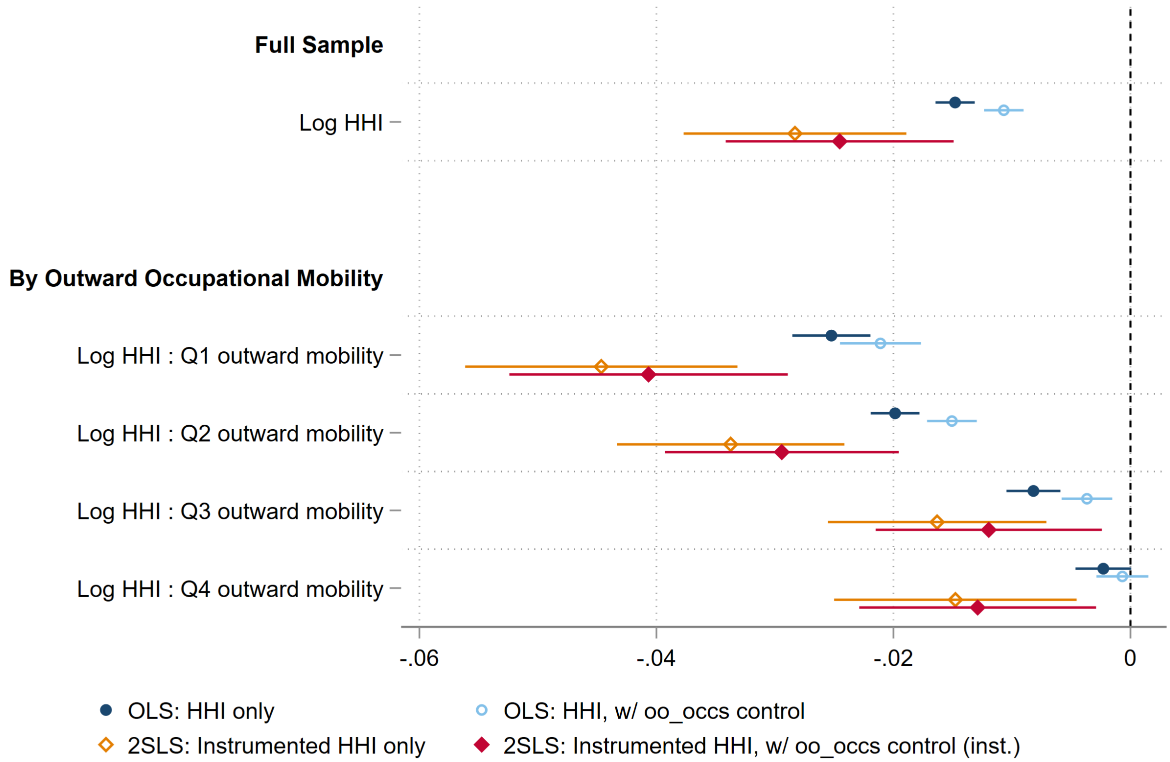
Note: Figure shows binned scatter plots of the relationship between log wages and log employer HHI for occupation-metro area cells over 2011–2019, weighted by occupation-metro area average employment and residualized on different combinations of fixed effects. Slopes for the line of best fit on each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

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



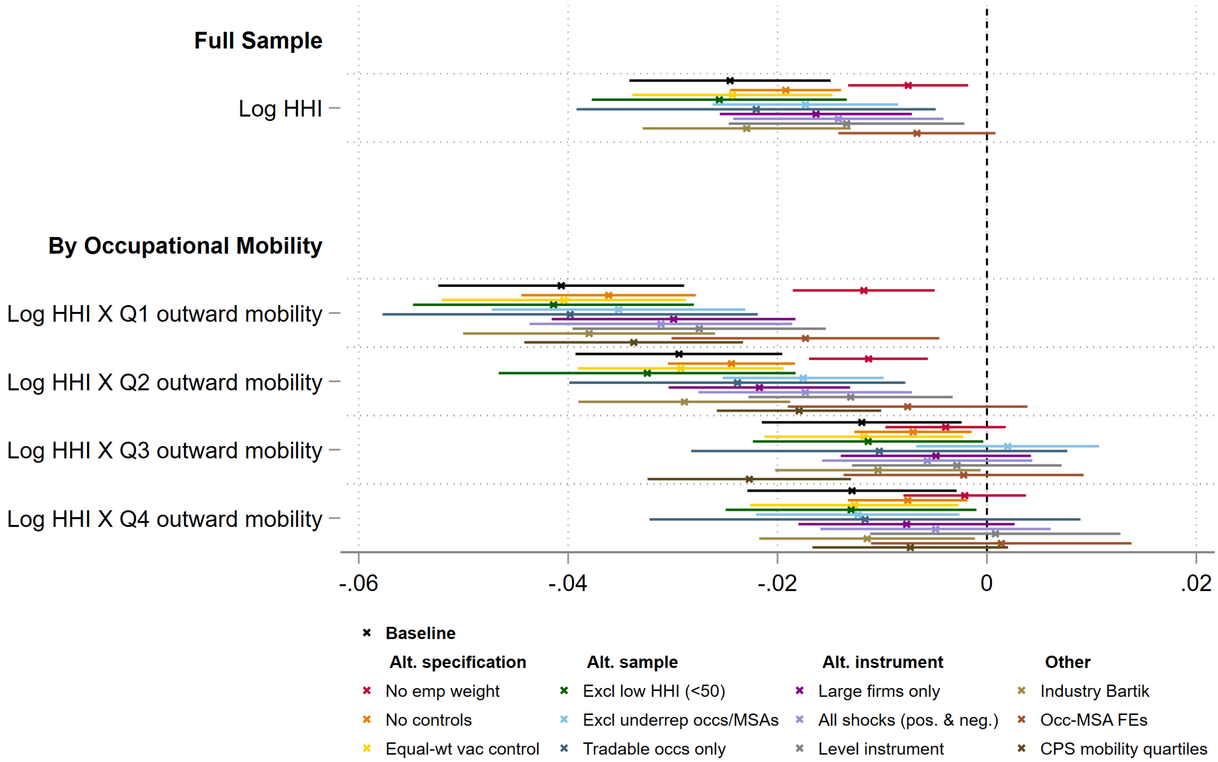
Note: Figure shows binned scatter plots of the employment-weighted relationship between average log wages and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on different combinations of fixed effects (as described by the panel titles). Regression coefficients for the line of best fit on each graph are: A: 0.44, B: 0.58, C: 0.21; D: 0.15. The non-linear shape of the figures without occupation-year fixed effects (panels A and B) is explained by healthcare occupations which tend to have both low outward mobility and high pay.

Figure 5: Coefficients on wage-HHI regressions



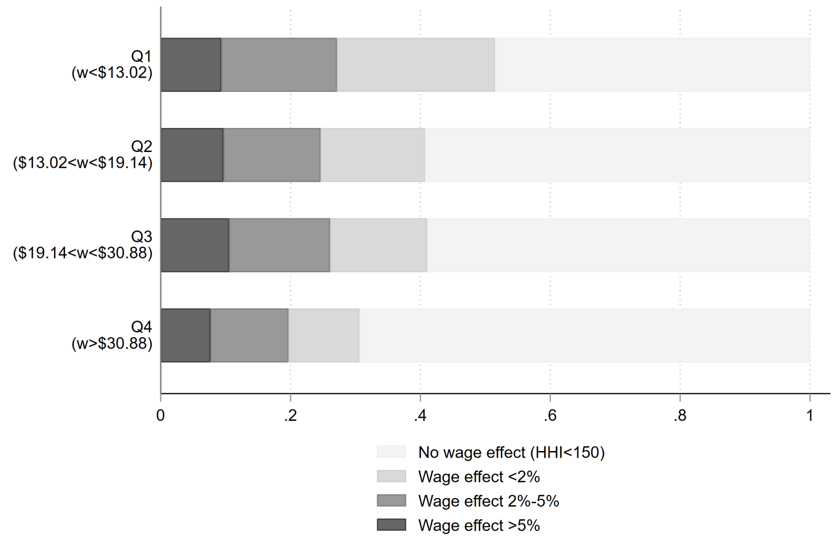
Note: Coefficients on log HHI and 95% confidence intervals from our baseline regressions of occupation-metro area log hourly wage on employer HHI. Navy and light blue represent OLS regression coefficient of wages on HHI, without (navy) and without (light blue) control for outside-occupation job options.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

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



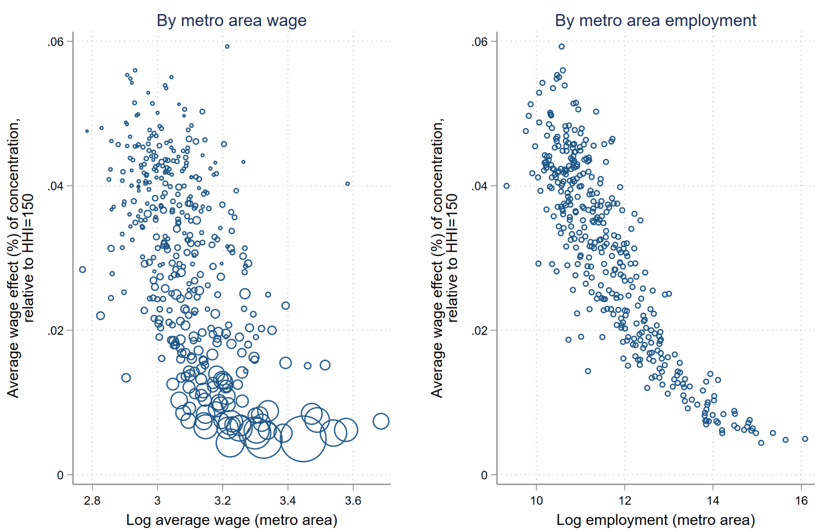
Note: Coefficients on log HHI and 95% confidence intervals from our baseline 2SLS IV regressions of occupation-metro area log hourly wage on instrumented employer HHI, across various robustness checks (as in Appendix Tables A6-A9). Black shows the baseline estimates from Figure 5.

Figure 7: Average estimated wage effect of employer concentration relative to HHI^{MALMA}=150, by hourly wage



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

Figure 8: Average estimated wage effect of employer concentration relative to HHI^{MALMA}=150, by metro area



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

APPENDIX

A Appendix: Conceptual framework: more detail

This section contains the full derivation of the average wage from our conceptual framework in section 2.1. Taking our expression for the wage as the weighted average of within firm productivity and outside options, the average wage in occupation o is given by:

$$\begin{aligned} w^o &= \sum_i \sigma_i^o w_i^o \\ &= \sum_i \sigma_i^o \beta p_i^o + \sum_i \sigma_i^o (1-\beta)\pi^{oo} \sum_{j\neq i} \sigma_j^o w_j^o + \sum_i \sigma_i^o (1-\beta) \sum_{p\neq o}^{N^{Occs}} \pi^{op} w^p \\ &= \beta p^o + (1-\beta) oo_o^{occs} + (1-\beta)\pi^{oo} \sum_i \sigma_i^o \sum_{j\neq i} \sigma_j^o \left(\beta p_j^o + (1-\beta) oo_o^{occs} + (1-\beta)\pi^{oo} \sum_{k\neq j} \sigma_k^o w_k^o \right) \end{aligned} \tag{14}$$

where $p^o = \sum_i \sigma_i^o p_i^o$ is the average productivity across firms in occupation o, and $oo_o^{occs} = \sum_{p\neq o}^{N^{Occs}} \pi^{op} w^p$ is our outside-occupation option index.

To simplify notation, define concentration index Ω_r^o as the sum of employer shares index with r "steps" as in the expression

$$\Omega_r = \underbrace{\sum_i \sigma_i^o \sum_{j\neq i} \sigma_j^o \sum_{k\neq j} \sigma_k^o ... \sum_{y\neq x} \sigma_y^o \sum_{z\neq y} \sigma_z^o}_{\text{with } r \text{ summation terms or "steps" in the expression}} \tag{15}$$

such that the second order index Ω_2^o is one minus the sum of the squared employer shares or HHI ($\Omega_2^o = \sum_i \sigma_i^o \sum_{j\neq i} \sigma_j^o = 1 - HHI^o$), and $\Omega_1^o = 1$. Denote each firm i's deviation from average occupational productivity p^o as $\tilde{p}_i^o = p_i^o - p^o$. Iteratively substituting in for the wage in other firms in occupation o, we can then rewrite the average wage equation 14 as

$$w^o = (\beta p^o + (1-\beta) oo_o^{occs}) \left(\sum_{n=1}^{\infty} \left((1-\beta)\pi^{oo}\right)^{(n-1)} \Omega_n \right) + \beta \sum_{n=1}^{\infty} \left((1-\beta)\pi^{oo}\right)^{(n-1)} \Pi_n \tag{16}$$

where Π_n is a function of the deviations of each firm's productivity from average productivity, and is defined analogously to Ω_r as the index with r "steps":
$\Pi_r = \sum_i \sigma_i^o \sum_{j\neq i} \sigma_j^o \sum_{k\neq j} \sigma_k^o ... \sum_{y\neq x} \sigma_y^o \sum_{z\neq y} \sigma_z^o \tilde{p}_z^o$,
where $\Pi_1 = \sum_i \sigma_i^o \tilde{p}_i^o = 0$, $\Pi_2 = \sum_i \sigma_i^o \sum_{j\neq i} \sigma_j^o \tilde{p}_j^o$, and so on. Taking a second order approximation and assuming that the term in Π_2 is sufficiently small, this can be written as

$$w^o \approx (\beta p^o + (1-\beta) oo_o^{occs}) \left(1 + (1-\beta)\pi^{oo}(1 - HHI^o)\right) \tag{17}$$

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 2.4). Burning Glass Technologies is an analytics software company that provides real-time data on job growth, skills in demand, and labor market trends. They frequently collaborate with academic researchers by providing data. The BGT vacancy data on online job postings has been used in several other academic papers, including Azar, Marinescu, Steinbaum and Taska (2020) and Hazell and Taska (2019). (We also use a different data set from BGT – the resume data set – to construct our measures of occupational mobility. We discuss this further in Appendix C.)

Vacancy posting data overview

Burning Glass Technologies constructs its vacancy database by collecting online job postings from about 40,000 websites, capturing the near-universe of online US job vacancies. They only measure *new* vacancy postings. To capture vacancies which firms keep online to hire workers continually for a given job, BGT consider a vacancy to be "new" if the identical vacancy is still online after 60 days (Carnevale, Jayasundera and Repnikov, 2014). BGT use proprietary algorithms to de-duplicate vacancies (for example if the same vacancy is posted on different websites).

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

Defining the employer and calculating the HHI

A key aspect for our purposes is how an "employer" is defined in the data. BGT's algorithm attempts to group together name variants for employers into a standard set, counting for example "Lowe's" or "Lowes" as the same employer. However, there may be some instances where employers which are in reality the same have not been detected by the algorithm due to large differences in spelling, punctuation, or naming conventions. We therefore carry out an additional layer of grouping by removing punctuation, spacing, and capitalization, and adjusting for common spelling differences or acronyms. We also used the Agency for

Healthcare Research Quality's "Compendium of US Health Systems" database for 2016 to link hospitals to the health systems which own them where possible, treating a health system as a single employer rather than a specific hospital. This match was not always perfect: there are several cases where we have not necessarily succeeded in matching all hospitals to their owner, because of the presence of multiple hospitals in our database with the same name. We also manually scanned several thousand of the largest employers in the database to group together different employer names which were evidently part of the same ultimate employer.

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

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

How do we treat the one third of vacancies which do not include an employer name? When we calculate our HHI statistics for each occupation-metropolitan area-year cell we assume that each vacancy listing by an employer with no name information in the database is a *separate employer* (as do Azar et al. (2020)). This will lead us to mechanically underestimate the HHI, as it is likely that at least some of these different vacancy postings where no name information is available come from the same employer in practice (Azar et al. (2020) note that the vacancy postings without employer name information are often due to staffing companies not disclosing on whose behalf they are posting a given job).

Summary statistics

Here, we provide summary statistics for the roughly 175 million vacancies which contain employer names. As one might expect given the skewed distribution of employment, the large majority of these vacancies are accounted for by a small group of large employers: 2,118 employers each posted more than 10,000 vacancies online over 2011–2019, and these 2,118 employers are responsible for a total of 45.7 million vacancies. On the other hand, the median employer in our dataset posted only 2 vacancies over the entirety of 2011–2019 (Table A1). While many of the small employers in our data are only present in the data for a subset of the 2011–2019 period, many large employers are present for all nine years (as shown in Table A1): as a result more than 50% of all vacancies in our database are listed by employers which are present in all nine years of the sample, and more than 75% are listed by employers which are present in at least eight of the nine years. If employers hire a lot in any one year, they also tend to hire a lot in other years: the correlation of vacancies within a given 6-digit SOC occupation and metropolitan area, by employer, from one year to the next is 0.76.

Vacancy postings, job vacancies, and employment

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

BGT estimates that its vacancy data covers the near-universe of online job postings. The Bureau of Labor Statistics' JOLTS database (Job Openings and Labor Turnover Survey) collects data on job *openings*, where each opening represents a specific position that the firm is actively recruiting to fill. The conceptual difference between a job posting and a job opening is that one job posting (a job advertisement) could be used to fill multiple job openings, if the firm needs to hire several people for a job with the same title, job description, and location at the same time. This may be a particular concern when measuring employer concentration, as a large employer may hire more workers per job posting than a small employer, and so we would systematically underestimate concentration in labor markets with a highly skewed distribution of employer size, relative to labor markets with more symmetric distributions of employer size. For example, when hiring for warehouse laborers, a large warehousing company like Amazon might hire several workers under a job ad for a "Warehouse Associate".[62] 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.

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

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

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 to Carnevale et al. (2014), Hershbein and Kahn (2018), and Azar et al. (2020) for more details. Note in particular that Azar et al. (2020) use the BGT vacancy data for the same purposes as we do: to calculate employer HHI concentration indices at the level of local SOC 6-digit occupations.

Hershbein and Kahn (2018) compare the distribution of BGT vacancies across major industry groups to the distribution of job vacancies in the Bureau of Labor Statistics' JOLTS database. While BGT is overrepresented in health care and social assistance, finance and insurance, and education, and underrepresented in accommodation and food services, public administration/government, and construction, the differences are mostly small in magnitude. Hershbein and Kahn (2018) also compare the distribution of BGT vacancies by occupation to both the stock and flow of employment in the United States, showing that BGT vacancy data has a much larger than average representation of computer and mathematical occupations, management, healthcare, and business and financial operations, and lower representation in transportation, food preparation and serving, production, and construction. This degree of representativeness does not change much over time in the BGT sample.

To analyze representativeness by occupation systematically, we calculate a measure we call 'represented-ness': the share of all vacancies in our data represented by each SOC 6-digit occupation in a given year, divided by the share of all employment in the BLS occupational employment statistics database which is represented by each SOC 6-digit occupation in that year. Note that our 'represented-ness' measure captures three dimensions: one is the degree to which the BGT vacancy *posting* data is representative of the totality of vacancy postings in the US, one is the degree to which vacancy *postings* are representative of true vacancies (job openings), and one is the degree to which individual occupations have high or low turnover (and as a result, a high or low ratio of vacancies to employment). We are interested primarily in the first two of these three, and would ideally compare the representativeness of our BGT vacancy data to a data set of the universe of online *and* offline vacancies by occupation, but this is not available. We show a scatter plot of the share of vacancies each occupation accounts for in our data, relative to the share of employment that occupation accounts for in the BLS OES, in Appendix Figure A1.

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

Similarly, zooming in on the next tier of occupations by size, we see overrepresentation of financial, information, management, and healthcare occupations, relatively even representation of sales, delivery, and mechanical occupations, and underrepresentation of workers in occupations with a large share of self-employment (construction, plumbing, landscaping), employment by individual households (maids and housekeeping cleaners, home health aides), or employment where firms may run single job ads for many workers, or which may advertise informally (dishwashers, cooks, food preparation workers, receptionists).

For our purposes, we have two potential representativeness concerns. One concern might be that the representativeness of our data is correlated in some way with factors which would affect both employer concentration and the wage. This concern is only relevant for the *estimated effect of concentration in our regressions* if our database systematically underrepresents low-wage occupation-metro area labor markets even when controlling for occupation and metro area fixed effects: that is, that within a given occupation, the lower-wage metro areas are underrepresented and within a given metro area, the lower-wage occupations are underrepresented. For our normative conclusions in terms of estimating the *aggregate num-*

ber 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.[63]
- 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.

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

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

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

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

Illustrative example of a resume.

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

We consider the switch from being a purchasing manager to being a compliance officier as *both* a change of job and a change of occupation, and the switch from being a compliance officer at Stansbury Inc to being a compliance officer at Taska Ltd as being a change of job but not a change of occupation. For the period 2004-2006, we would therefore record the sequential job holding patterns as follows:

Illustrative example of sequential job holding data.

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

This matching of sequential job-year coincidence pairs results in 178.5 million observations (including year-to-year pairs where workers are observed in the same occupation in both years). We use these sequential job-year coincidence pairs to construct our measures of occupational mobility as follows. For each pair of (different) occupations o to p, we count the total number of sequential job-year coincidence pairs where the worker is observed in occupation o at any point in year t and is observed in occupation p at any point in year $t+1$. We then divide this by the total number of workers in occupation o in year t who are still observed in the sample in the following year $t+1$.

Since our data is not fully representative on age within occupations, we compute these occupation transition shares separately for different age categories (24 and under, 25 to 34, 35 to 44, 45 to 54, and 55 and over).[64] We then aggregate them, reweighting by the average proportion of employment in each of these age categories in that occupation in the US labor force over 2012–2017 (from the BLS Occupational Employment Statistics). Our aggregate occupational mobility matrix has therefore been reweighted to correspond to the empirical within-occupation age distribution in the labor force, reducing the potential for bias arising from the skewed age distribution of our sample.

Summary statistics

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

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

Gender: BGT imputes gender to the resumes using a probabilistic algorithm based on the names of those submitting the resumes. Of our observations, 88% are on resumes where BGT was able to impute a gender probabilistically. According to this imputation, precisely 50% of our observations are imputed to be more likely to be male, and 50% are imputed to be more likely to be female. This suggests that relative to the employed labor force, women

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

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.[65]

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

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

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.

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.[66]

Caveats and concerns

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

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

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

We are only aware of a very limited number of studies directly trying to estimate the incidence of misrepresentations on resumes. For instance, Sloane (1991) surveys HR executives in banking and finds that 51 responding executives were jointly aware of a total of 17 instances of meaningfully falsified job titles, which seems small given the presumably

[66] 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).

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

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

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

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

Comparability with CPS occupational mobility

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

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

First, the occupation categorization is different: we use SOC 6-digit occupations (of which there are a total of 840 in US data) and the CPS uses Census occupation codes, which are broader.

Second, because of the nature of our resume data, we cannot measure annual occupational mobility (share of workers whose main job was in occupation o on date d in year t whose main job was no longer in occupation o on date d in year $t+1$). Instead, our measure of the average share of workers leaving their occupation in any given year (11%) reflects the total number of workers who are observed in occupation o in year t who are *not* observed in occupation p at any point in year $t+1$. This makes it a more conservative measure of occupational mobility than the annual occupational mobility concept commonly constructed from the CPS.

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

We can also compare how occupational mobility *by occupation* differs between our BGT measure and the CPS. To calculate outward occupational mobility in the CPS, we first map the CPS occupation codes to 6-digit SOC occupation codes using the crosswalk provided by the BLS. We then calculate outward occupational mobility in the CPS ASEC using the self-reported "occupation last year" variable. Specifically, we calculate outward mobility for each occupation o as the number of people who report that their occupation last year was o

and their occupation this year was not o (but they were employed), divided by the number of people who report that their occupation last year was o and who are still employed this year. An employment-weighted binned scatter plot of the 722 occupations for which we are able to calculate both a BGT and a CPS outward mobility measure is shown in Appendix Figure A9. As can be seen, there is a very strong positive correlation between the measures. As would be expected (given the discussion above), CPS outward mobility is higher than BGT outward mobility for any given occupation.

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

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.[67] 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.[68]

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.[69]

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 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 2.4 of this paper, we define workers' baseline labor market as a SOC 6-digit occupation within a metropolitan area.[70] 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

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

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

[69] 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.

[70] 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.

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 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 2.4 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.[71]

[71] 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.

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.[72] 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.

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;

[72] 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.

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 relationship between the likelihood of moving between occupation pairs o and p and their task and characteristic similarity.

Specifically, we estimate the following regression:

$$\pi_{o\to p} = \alpha_o + \beta^{abs}|X_{occ\,p} - X_{occ\,o}| + \gamma|\Delta w_{o\to p}| + \epsilon_{op}. \tag{18}$$

where $\pi_{o\to p}$ is the transition share as defined in 8 (the share of people who leave 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,[73] but note that the results are qualitatively similar without the wage controls.

We would expect the coefficient on the absolute difference in characteristics to be negative: the greater the difference between two occupations, the less likely we should be to observe the worker moving from one into the other. Our results bear this out: in every regression of pairwise occupational mobility on the absolute difference in characteristics, the coefficients are significantly negative or statistically insignificant, as shown in Figure A12.[74]

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 and how

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

[74] 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.

differences in characteristics also predict the direction of occupational flows, we estimate a similar regression equation to that shown in equation (18), but now using the *relative* (target minus origin) difference in occupational characteristics as the independent variable:

$$\pi_{o \to p} = \alpha_o + \beta^{rel}(X_{occ\ p} - X_{occ\ o}) + \gamma \Delta w_{o \to p} + \epsilon_{op}. \tag{19}$$

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 (19) for the different measures are also shown in Figure A12. (Note that this analysis involves directed relationships between occupations, so if the same share of moves in each direction is observed for an given occupation pair, the estimated effect of differences between them would be zero.)

A number of our predictions are borne out in the data: we find (1) that workers are more likely to move towards jobs with higher wages; (2) that workers transition on average *towards* jobs that require more leadership responsibility - as would be expected from moves up the career ladder; (3) that occupational transitions have on average been *towards* occupations that have higher analytical content and require more social skills, and out of occupations with more routine task requirements;[75] and (4) that workers have on average been moving into occupations that require more contact and working relationships with others (and so have less time flexibility).

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

The failure of skill similarity measures to explain many occupational transitions can be illustrated by a few cases from our data. First, consider some occupation pairs that are very similar on a skill distance metric (in the lowest distance decile), but where our data shows almost no (less than 0.01%) chance of moving from one to the other when switching jobs, in either direction: Surveyors vs. Medical & clinical laboratory technologists; Carpenters

[75] 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).

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 (2022), with exogenous 'shocks' in the form of differential national hiring patterns for large firms, and initial squared employer shares of each firm in a given local labor market determining the exposure to those shocks.

We can rewrite the concentration instrument as

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

where $\tilde{G}_{j,o,k,t} = \frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1$ is the predicted firm-level excess local vacancy growth relative to the average predicted local occupation vacancy growth - the time-varying shock - and $\sigma^2_{j,o,k,t-1}$ is the exposure of the local concentration index to that shock.[76]

As noted in the main text, we use only intensive margin shocks (i.e. we do not incorporate growth to and from zero vacancies), since the logic of our instrument holds for expansions or

[76]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.

contractions of existing establishments, and less well for expansions into new markets. In our baseline specification, we use only positive vacancy growth shocks $\tilde{g}_{j,o,t} > 0$. In robustness checks, we include all shocks (positive and negative), and we also create a version of the instrument only using shocks to large firms, where large firms are defined as firms which have vacancies in the relevant occupation o and pair of consecutive years $t-1$ and t in at least five different metropolitan areas k (and $\tilde{g}_{j,o,t}$ is set to zero for all other firms). In all cases, we use only firms which are present in at least two metro areas in consecutive years within a given occupation and which have at least ten vacancy postings in the occupation and year in question.

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 ($\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} \tilde{g}_{j,o,t}$).

We explain our third control in more detail here. The motivation for this control is that, as Borusyak and Hull (2020) demonstrate, nonlinear transformations in shift-share IVs risk introducing omitted variable bias. In our case, a log transformation of a shift-share IV – a concave transformation – can introduce omitted variable bias since the expected instrument becomes systematically higher for units of observation where the shift-share sum has lower variance. We therefore follow the recommendation of Borusyak and Hull (2020) in incorporating a control for the expected value of the HHI instrument in all our baseline regressions. To construct this expected HHI instrument control, we first construct 100 counterfactual HHI instruments by randomly re-assigning the actual observed firm-occupation vacancy growth rates to the other firms in the data, and using these counterfactual firm-occupation vacancy growth rates $\hat{g}_{j,o,t}$ to construct a counterfactual HHI instrument $e_{o,k,t} = \left(\sum_j \sigma^2_{j,o,k,t-1} \left(\frac{(1+\hat{g}_{j,o,t})^2}{(1+\hat{g}_{o,k,t})^2} - 1 \right) \right)$. We then construct our expected HHI instrument by taking 100 of these counterfactual HHI instruments, averaging over them, and taking the log. When we control for the expected HHI instrument in our regressions, we identify only off the difference between the actual firm-level vacancy growth shocks experienced, as compared to a counterfactual where the firm distribution was the same but vacancy growth shocks were different.

In a robustness check, we also use a different strategy which corrects for the bias introduced by the nonlinear IV specification, following Borusyak and Hull (2020) to construct an alternative instrument defined as $Z^{HHI,alt}_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \log \left(\frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} \right) \approx \sum_j \sigma^2_{j,o,k,t-1} \left(2(\tilde{g}_{j,o,t} - \tilde{g}_{o,k,t}) \right)$. That is, the instrument is constructed by expressing the log HHI in period t as a function of growth rates and employer shares in $t-1$, $HHI_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \left(\frac{(1+g_{j,o,t})^2}{(1+g_{o,k,t})^2} \right)$, defining the relative squared growth rate of each employer j as $\hat{g}_{j,o,t} = \frac{(1+g_{j,o,t})^2}{(1+g_{o,k,t})^2}$ and taking a log-linear approximation around $\hat{g}_{j,o,t} = 1$ (i.e. a log-linear approximation around the point at which the vacancy growth rate of each employer j is identical to the local labor market vacancy growth rate, so the HHI does not change from one year to the next).

With our three baseline controls and our fixed effects, in the IV estimation of equation (5) the exclusion restriction for the instrument on the HHI concentration index is equivalent to

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

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

This orthogonality condition holds under two assumptions. First, we require that the national firm-level growth shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks $\xi_{o,k,t}$, the fixed effects Γ_{kt} and Γ_{ot}, actual and predicted average local vacancy growth $g_{o,k,t}$ and $\tilde{g}_{o,k,t}$, and expected HHI instrument $e_{o,k,t}$. That is, once we account for the control variables, expected local squared exposure to excess national firm-level growth needs to be random in expectation.[77] Second, there needs to be a large enough number of independent firm-level shocks to drive sufficient variation in the instrument to identify the effect of interest.

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

To be concrete, note the hypothetical example from the main text, which considered insurance sales agents in Bloomington, Illinois and in Amarillo, Texas. In each metro area, there are several insurance companies who employ insurance sales agents. Assume that in

[77]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).

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 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. First, this requires that the occupation vacancy growth rate of firm j in metro areas outside metro area k is positively correlated with the occupation vacancy growth rate of firm j in metro area k. This is the case in our data. A regression of the latter on the former gives a coefficient of 0.02 (standard error 0.0001); when restricting to firms with vacancies in 5 or more metro areas in the same occupation, the coefficient is 0.35 (standard error 0.0006). Next, this requires that these local level increases in firm j hiring lead to higher employer concentration. A sufficient condition for this, under *most* initial employer share distributions, is if firm j's new vacancies are allocated proportionally across occupation-metro area labor markets, such that each occupation-metro area labor market sees the same growth rate in its firm j vacancies as the national average.[78] 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

[78]Note that this is not necessarily the case – i.e. the first stage might not hold – for *all* possible combinations of the distribution of employment growth and initial employer shares. For example, consider a world in which there is a labor market for where Employer X has 80% of the market in one metro area, and the rest of the market is comprised of atomistic firms (HHI=6,400); and Employer X has 65% of the market in another metro area, with the rest of the market comprised of atomistic firms (HHI=4,200). If Employer X grows by 10% in both locations in a given year, and the other firms do not grow at all, employer concentration will actually increase by more in the latter than the former market. This circumstance, however, only occurs when comparing two labor markets which both have extremely high levels of employer concentration already, and so is not relevant for the vast majority of the labor markets in our data (for example only 10% of occupation-metro area labor markets, containing only 1% of workers, had an HHI greater than 5,000 in our data in 2019).

our data there is a negative relationship between an employer's initial vacancy share in an occupation-metro area labor market and its next year's vacancy growth rate, but this relationship is not sufficiently strong to invalidate our first stage. Empirically, our first stage holds strongly as shown in Appendix Table A5.

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

Identification assumptions for outside-occupation option index instrument

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

As described in Section 2, our instrument for the oo^{occs} index, Z^{oo}, is the weighted average of national leave-one out mean wages in occupation p, $\bar{w}_{p,\not{k},t}$, 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, $\frac{s_{p,k,1999}}{s_{p,1999}}$,[79] and the national probability of a worker moving to a job in occupation p conditional on leaving a job in occupation o, $\pi_{o\to p}$:

$$Z^{oo}_{o,k,t} = \log\left(\sum_{p}^{N_{occs}} \pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\not{k},t}\right) \tag{20}$$

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. (2022).[80] Note that we can write the instrument as

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

where $s_{okp} = \pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}}$ is a measure of predicted local exposure to the shock. In our

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

[80]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.

fixed effects IV estimation of equation (5), the exclusion restriction for the instrument for outside-occupation options is then equivalent to

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

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

Borusyak et al. (2022) show that this orthogonality condition holds under two assumptions. First, we require that the national occupation-level shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks ϕ_{pt}, the fixed effects Γ_{kt} and Γ_{ot}, and the control variables. Second, there needs to be a large number of independent occupational shocks. The first assumption requires that the national leave-one-out mean wage $\bar{w}_{p,\cancel{k},t}$ in outside option occupation p is correlated with the local wage of occupation p in location k (relevance condition), but does not affect the local wage in initial occupation o through a direct channel other than increasing the quality of local outside options $Z^{oo}_{o,k,t}$. However, this lack of a direct effect only needs to hold *conditional* on controlling for fixed effects that include the national wage trend in occupation o itself and wage trends that are common to all occupations in metro area k.[81] The inclusion of these fixed effects increases our confidence that the assumptions for instrument validity hold.

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

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

We show that our outside-occupation option results are robust to this alternate specification for the instrument (Appendix Tables A7 and A9, column c).

Outside-occupation options: Robustness analysis

Industry Bartik: One possible concern with the identification assumptions required for our outside-occupation index – which may not entirely be picked up by our occupation-year,

[81] 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.

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. For example, 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.

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

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

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

We use data on national employment by NAICS 4-digit industry and SOC 6-digit occupation from the Bureau of Labor Statistics Occupational Employment Statistics to construct the employment shares in each industry by occupation, and we use Quarterly Census of Employment and Wages data from the BLS to construct industry employment shares by metropolitan statistical area, national industry employment shares, and leave-one-out national industry wage growth.

We report our baseline regression results, controlling for this industry Bartik shock, in column d of Tables A7 and A9. Coefficients on the outside-occupation option index remain robust to its inclusion.

Alternate specifications: We also explore alternate specifications. We find large, statistically significant, and positive effects of outside-occupation options on wages without employment weighting (Appendix Tables A6 and A8, column *a*), with occupation-metro area and year fixed effects (Appendix Tables A7 and A9, column *e*).

Alternate instrument formulation: Following the recommendation of Borusyak and Hull (2020), we construct an alternate specification of the outside-occupation option instrument which addresses the possibility for bias introduced by the log transformation. Specifically, our alternate instrument takes the log of the wage inside the sum, instead of taking the log of the sum of the transition-weighted wages: $Z^{oo,alt}_{o,k,t} = \sum_p^{N_{occs}} \pi_{o \to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \log \bar{w}_{p,\not{k},t}$. Our results are robust using this alternative instrument, as shown in Appendix Tables A7 and A9, column *c*.

Alternate time-frame or level of aggregation: 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 A10). Finally, 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 A11, 1999–2016).

H Appendix: Stata commands

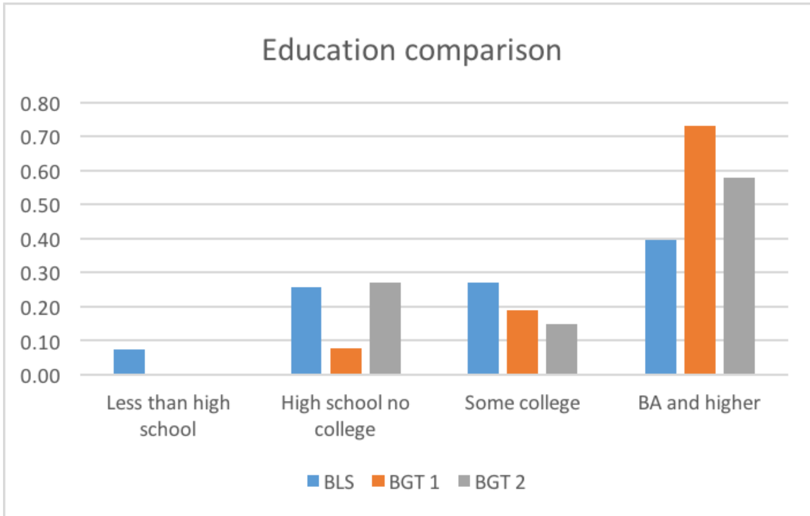
In our estimation, we used a number of user-written Stata commands: *reg2hdfe* (Guimaraes and Portugal, 2010), *reghdfe* (Correia, 2016), *ivreg2hdfe* (Bahar, 2014), *binscatter* (Stepner, 2013), *binscatter2* (Droste, 2019), and *coefplot* (Jann, 2013).

I Appendix: Figures

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

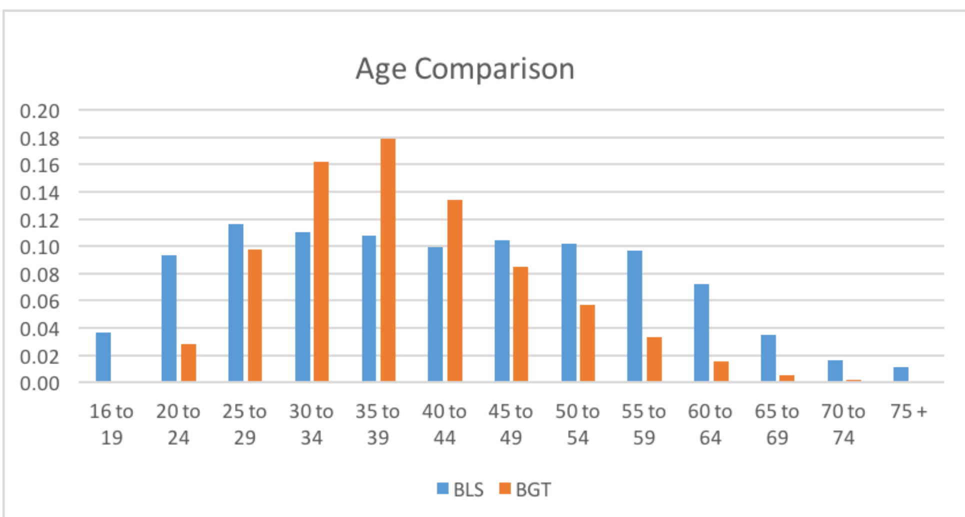
Note: Comparison of distribution of share of vacancies accounted for by each SOC 6-digit occupation in the BGT vacancy data in 2019, relative to that occupation's share of total employment in 2019 in the BLS occupational employment statistics. Occupations comprising greater than 1% share of either data 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



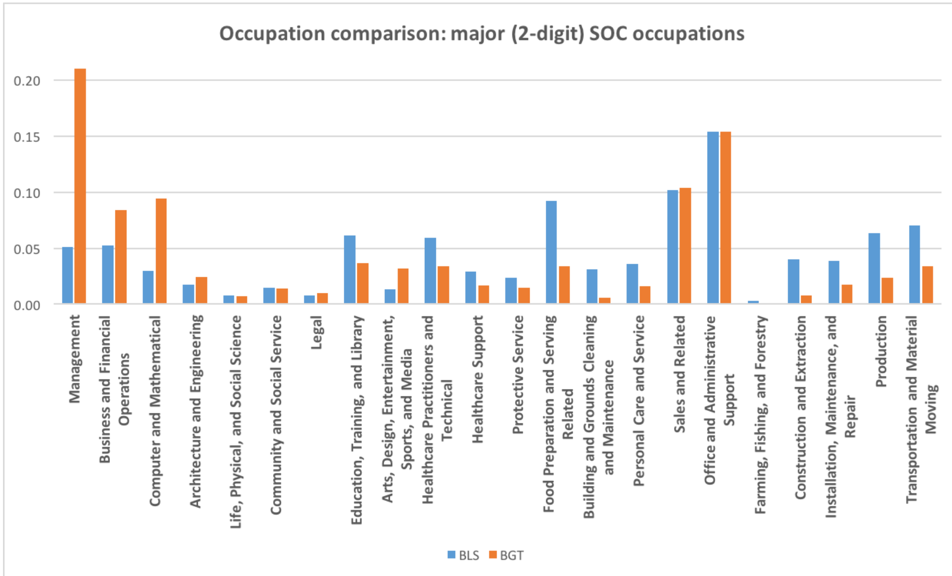
Note: Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT resume data. Two versions are shown: BGT 1 excludes all resumes missing educational information, while BGT 2 assumes all resumes missing educational information have high 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



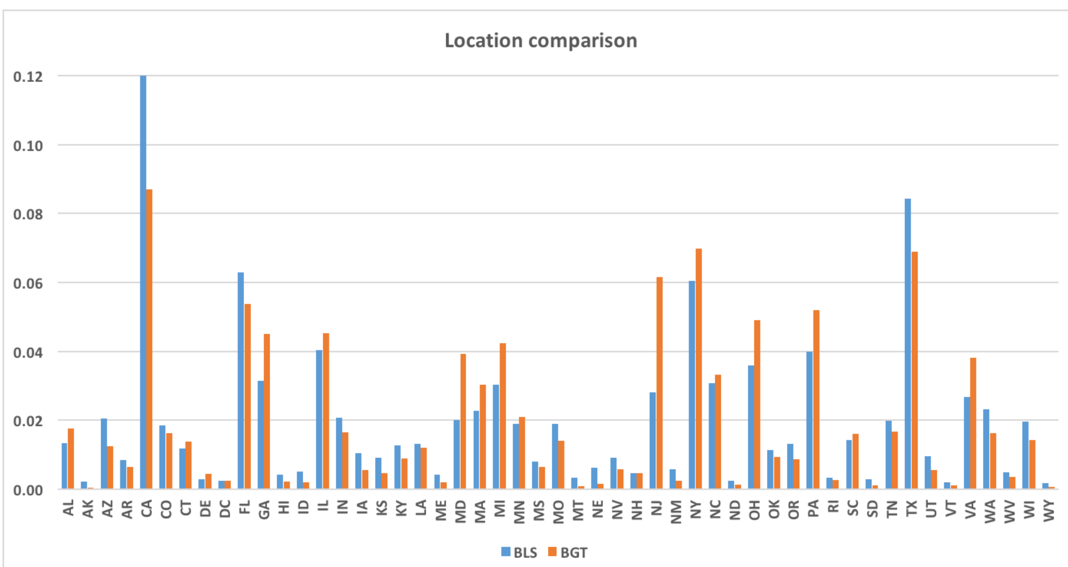
Note: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT resume data. The resume data is discussed in detail in Appendix C.

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



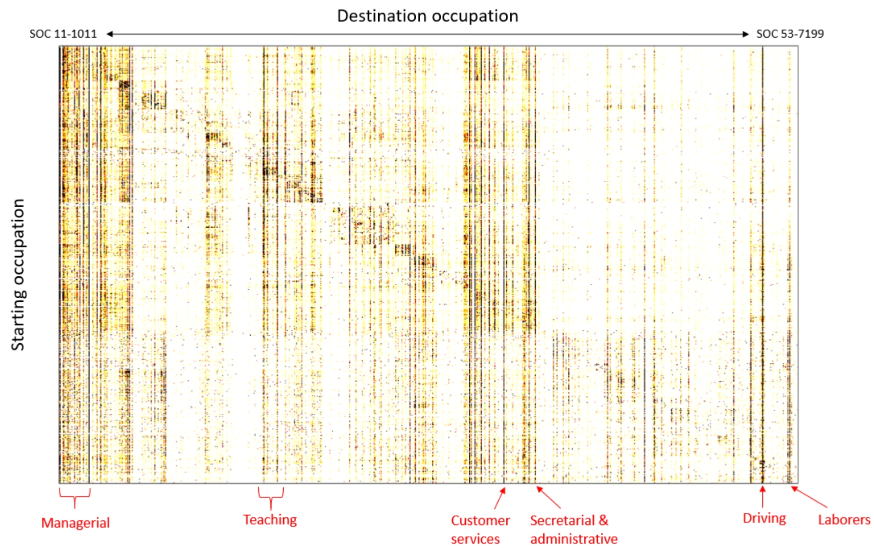
Note: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT resume data. The resume data is discussed in detail in Appendix C.

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



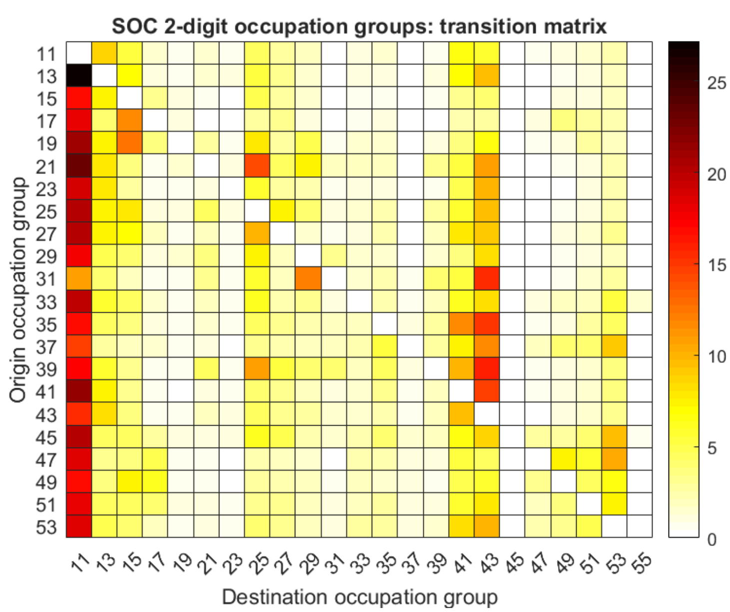
Note: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT resume data. Graph shows share of total in each state. The resume data is discussed in detail in Appendix C.

Figure A6: 6-digit SOC occupational transition matrix



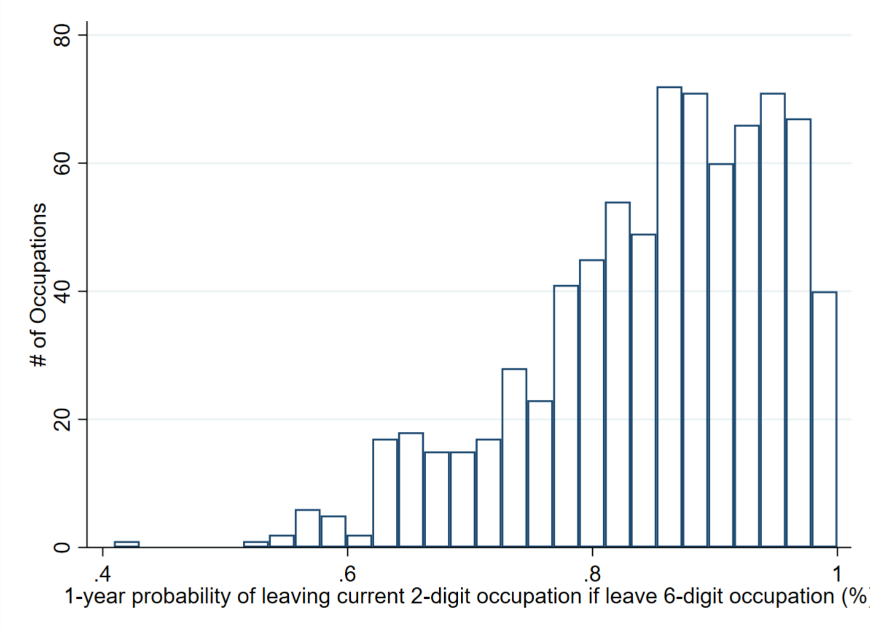
Note: Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial job. Occupations are sorted in SOC numerical order. Cells colored black have a transition probability of 1% or greater conditional on leaving the initial job. Transitions 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 A7: 2-digit SOC occupational transition matrix



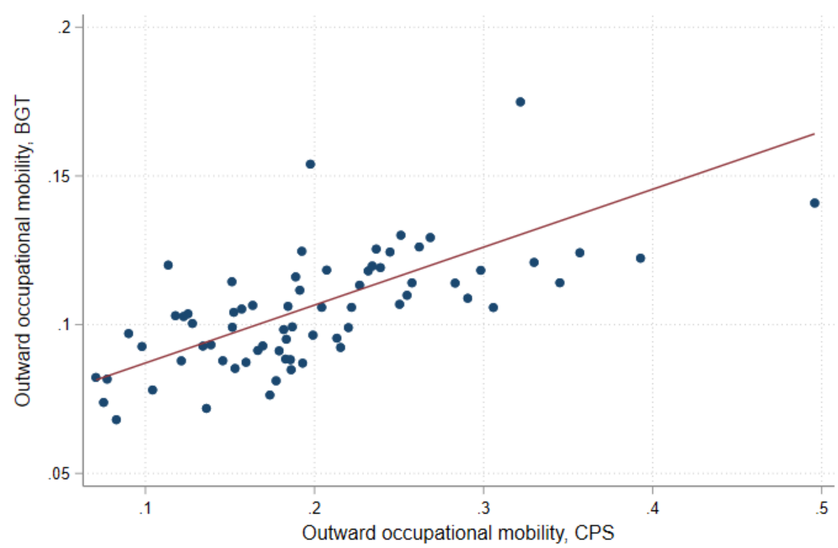
Note: Occupational transition matrix showing transition probability between 2-digit SOC occupation groups conditional on leaving the initial job. Cells colored black have a transition probability of 25% or greater conditional on leaving the initial job. Job transitions within an occupation group are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

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



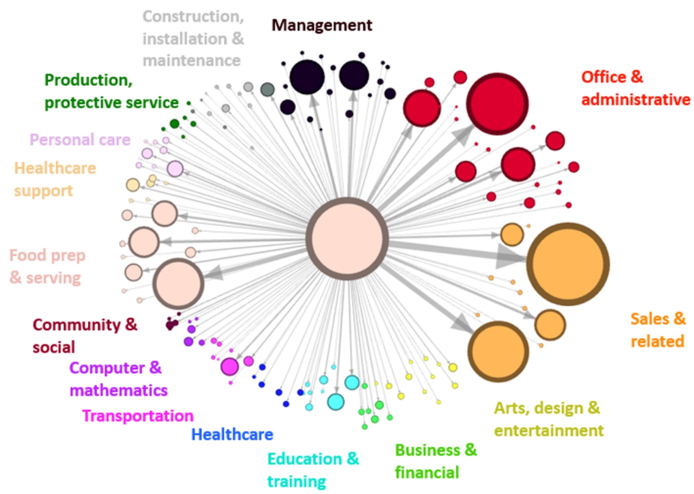
Note: Distribution of the proportion of workers moving 6-digit SOC occupation who *also* move 2-digit SOC occupation, by occupation, calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations. The resume data is discussed in detail in Appendix C.

Figure A9: Occupational mobility: comparing CPS and BGT measures



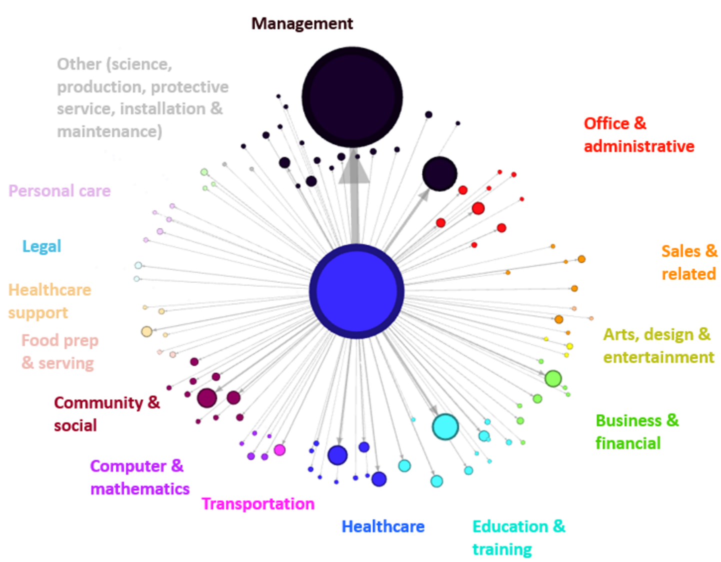
Note: Binned scatter plot of outward occupational mobility by occupation as calculated in our BGT resume data vs. in the CPS. Details of calculation in Appendix C.

Figure A10: Examples of probabilistic labor markets: counter attendants



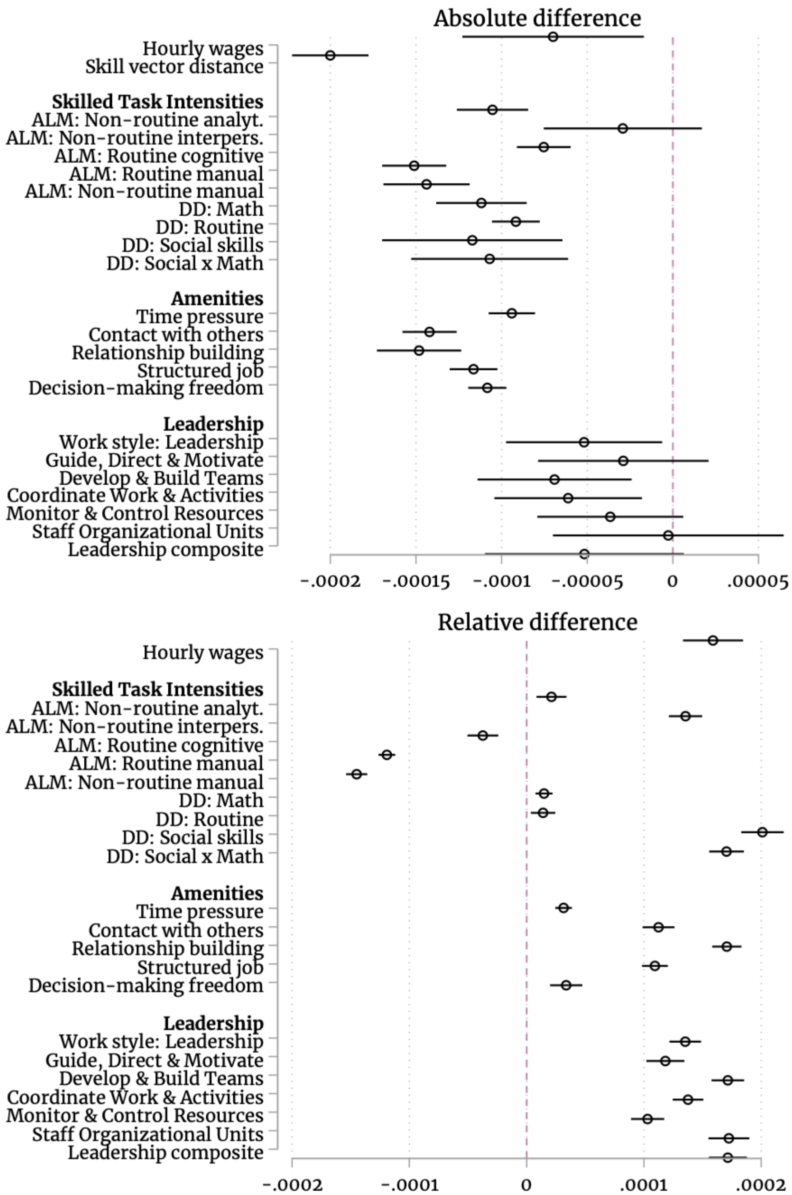
Note: Example visualization of occupational transitions for counter attendants in the food industry. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of counter attendants in the BGT data who switch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

Figure A11: Examples of probabilistic labor markets: registered nurses



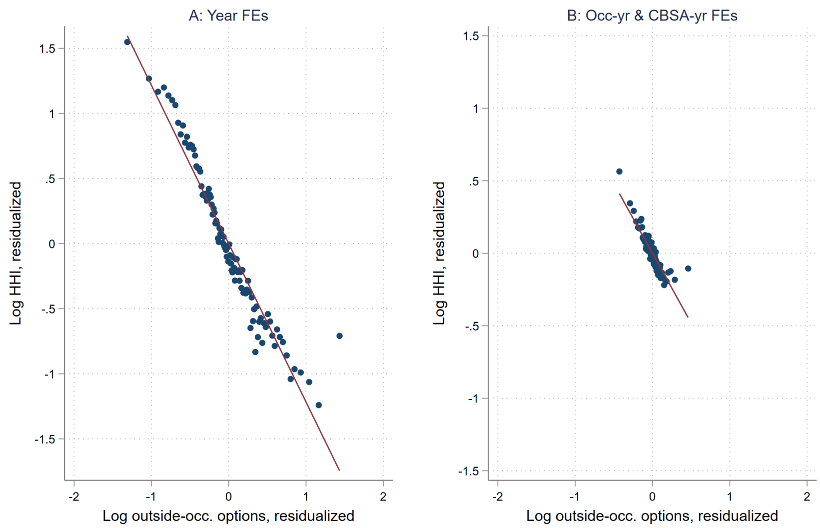
Note: Example visualization of occupational transitions for registered nurses. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of registered nurses in the BGT data who switch occupation, who are observed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

Figure A12: Occupational transitions and occupational characteristic similarity



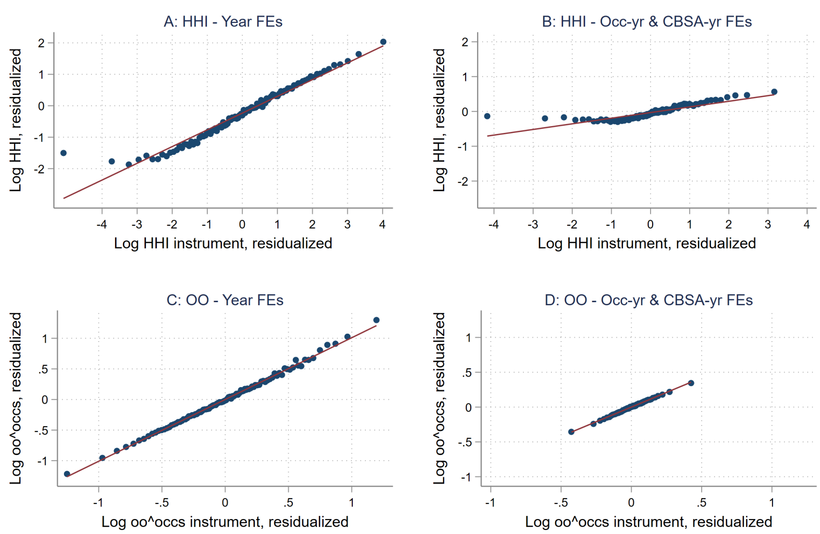
Note: This plot shows coefficients and 95% confidence intervals from regressions of occupation transition shares on occupational characteristics: $\pi^{op} = \alpha_o + \beta f(X_{occ\ o \to p}) + \gamma f(\Delta w_{o \to p}) + \epsilon_{op}$, where α_o is an occupation o fixed effect. In panel "Absolute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

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



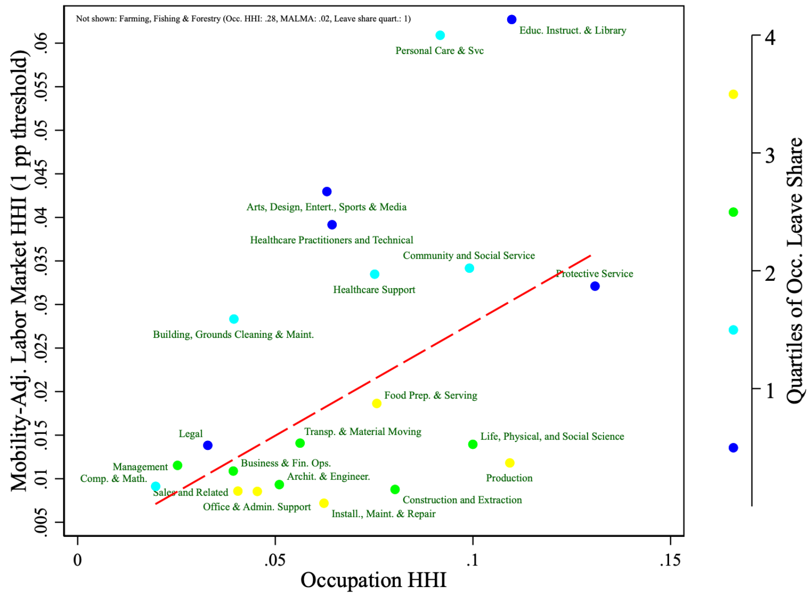
Note: Figure shows binned scatter plots of the employment-weighted relationship between log HHI and log outside-occupation option index for occupation-metro area cells over 2011–2019, residualized on year FEs in panel A and occupation-year and metro area-year FEs in panel B.

Figure A14: Correlation between instruments and independent variables



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

Figure A15: Employer concentration in mobility-adjusted and within-occupation labor markets in 2015



Note: The graph shows for each major occupation group the average employment-weighted local employer concentration for the included 6-digit occupations, computed from vacancy data for the years 2011-2019. The horizontal axis shows the within-occupation HHI, while the vertical axis shows the HHI for mobility-adjusted labor markets created by including all other occupation with at least 1% relevance (transition likelihood) for the worker's current 6-digit occupation.

J Appendix: Tables

Table A1: Summary statistics for BGT vacancy data

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

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

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

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

Note: This table shows the twenty large occupations with the lowest and the highest occupation leave shares - defined as share of workers observed in one occupation in one year but not in the following year, divided by the share that leave their job over that period (see Section 2.4) - in the BGT data over 2002-2015, as well as total national employment in that occupation in 2017 from the OES, the number of occupation-year observations in the BGT data ('obs.') and the most popular occupation that workers who leave the initial occupation move to ('modal new occupation'). Large occupations are defined as those with national employment over 150,000 in 2017 (roughly the 75th percentile of occupations when ranked by nationwide employment). † Full occupation title is "Molding, coremaking, and casting machine setters, operators, and tenders, metal and plastic."

Table A3: Forty thickest occupational transition paths for large occupations

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

Note: This table shows the 'thickest' occupational transition paths from large occupations (defined as those with national employment greater than 150,000 in 2017). The transition share from occupation o to occupation p is defined as the share of all occupation leavers from the initial occupation o who move into that particular new occupation p (as in Section 2.4). Only occupations with at least 500 observations in the BGT data and 2017 OES employment data are shown.

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

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

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

Table A5: First stage regressions: HHI instrument

Dependent variable: log vacancy HHI (segmented by quartile of occ mobility in cols (b)-(e))	Full sample	By quartile of occ mobility			
		Q1	Q2	Q3	Q4
	(a)	(b)	(c)	(d)	(e)
Log HHI instrument	0.105***	0.118***	0.102***	0.093***	0.098***
	(0.006)	(0.009)	(0.008)	(0.010)	(0.006)
Log outside-occ. options instrument	-0.761***	-0.493***	-0.785***	-0.724***	-0.889***
	(0.035)	(0.063)	(0.053)	(0.041)	(0.050)
Observations	443,233	110,404	111,193	127,573	93,993
Controls					
Vacancy growth	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y
Expected HHI instrument	Y	Y	Y	Y	Y
Fixed effects					
Occ-year	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y

Note: In column (a) we run a first-stage regression for our HHI instrument. In columns (b) through (e) we run separate first-stage regressions for our HHI instrument, segmenting our data into four quartiles by outward occupational mobility (the occupation "leave share" as defined in Section 2.4). Heteroskedasticity-robust standard errors clustered at the metro metro area-by-year level shown in parentheses. Units of observation are 6 digit SOC by metro area by year, for all observations with available data over 2011–2019 inclusive. All regressions have occupation-year and metro area-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A6: Regression of wage on HHI and outside-occupation options: robustness

Dependent variable:	Log wage					
	(a) No emp. weight	(b) No controls	(c) Equal-wt. control	(d) Drop low HHI	(e) Drop low rep.	(f) Tradable occs
Log HHI, instrumented	-0.008**	-0.019***	-0.024***	-0.026***	-0.017***	-0.022**
	(0.003)	(0.003)	(0.005)	(0.006)	(0.005)	(0.009)
Log outside-occ. options, instrumented	0.098***	0.105***	0.102***	0.105***	0.120***	0.140***
	(0.004)	(0.009)	(0.011)	(0.009)	(0.009)	(0.010)
Observations	443,233	502,609	443,233	432,733	241,890	213,643
F-stat	1,221	401	176	252	176	83
Controls						
Vacancy growth	Y		Y	Y	Y	Y
Predicted vac. growth	Y		Y	Y	Y	Y
Expected HHI instrument	Y		Y	Y	Y	Y
Equal-wt. vac. growth			Y			
Fixed effects						
Occ-year	Y	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y	Y

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

Table A7: Regression of wage on HHI and outside-occupation options: robustness (2)

Dependent variable:	Log wage				
	(a) Large Firm HHI Inst.	(b) Pos. & Neg. Shock HHI Inst.	(c) Level HHI Inst.	(d) Industry Bartik	(e) Occ-MSA FEs
Log HHI, instrumented	-0.016*** (0.005)	-0.014*** (0.005)	-0.013** (0.006)	-0.023*** (0.005)	-0.007* (0.004)
Log outside-occ. options, instrumented	0.114*** (0.010)	0.107*** (0.010)	0.135*** (0.010)	0.106*** (0.011)	0.021 (0.014)
Observations	382,389	329,855	443,233	405,873	427,472
F-stat	205	266	221	168	86
Controls					
Vacancy growth	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y
Expected HHI instrument	Y	Y	Y	Y	Y
Industry Bartik				Y	
Fixed effects					
Occ-year	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	
Occ-metro					Y

Notes: These represent robustness checks for our baseline regression specification (reported in column (d) of Table 3). Each column is a different robustness check. All columns are 2SLS IV regressions estimated using our HHI and outside-occupation option instruments. Columns (a), (b), and (c) use alternative formulations of our HHI instrument, with column (a) instrumenting only with national hiring growth for large firms (with vacancies in at least 5 metro areas in that occupation), column (b) using both national hiring growth and declines, and column (c) uses an HHI instrument based on the predicted level rather than change in HHI. Column (d) introduces a control for an industry Bartik shock to control for correlated industry shocks across occupation-metro area cells. Column (e) has occupation-metro area and occupation-year fixed effects instead of occupation-year and metro-year fixed effects. Other regression info: Heteroskedasticity-robust standard errors clustered at the metro area-by-year level shown in parentheses (except for column (e), where standard errors are clustered at the metro area-by-occupation level consistent with the fixed effects). 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, except column (e) which has occupation-metro area and occupation-year fixed effects. All regressions are employment-weighted. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A8: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness

Dependent variable:	Log wage					
	(a) No emp. weight	(b) No controls	(c) Equal-wt. control	(d) Drop low HHI	(e) Drop low rep.	(f) Tradable occs
Log HHI	-0.012***	-0.036***	-0.040***	-0.041***	-0.035***	-0.040***
X Q1 outward mobility	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)	(0.009)
Log HHI	-0.011***	-0.024***	-0.029***	-0.032***	-0.018***	-0.024***
X Q2 occ mobility	(0.003)	(0.003)	(0.005)	(0.007)	(0.004)	(0.008)
Log HHI	-0.004	-0.007**	-0.012**	-0.011**	0.002	-0.010
X Q3 occ mobility	(0.003)	(0.003)	(0.005)	(0.006)	(0.004)	(0.009)
Log HHI	-0.002	-0.008***	-0.013**	-0.013**	-0.012**	-0.012
X Q4 occ mobility	(0.003)	(0.003)	(0.005)	(0.006)	(0.005)	(0.010)
Log outside-occ. options	0.097***	0.106***	0.103***	0.106***	0.126***	0.139***
	(0.004)	(0.009)	(0.011)	(0.009)	(0.008)	(0.009)
Observations	443,233	502,609	443,233	432,733	241,890	213,643
F-stat	480	160	70	102	71	33
Controls						
Vacancy growth	Y		Y	Y	Y	Y
Predicted vac. growth	Y		Y	Y	Y	Y
Expected HHI instrument	Y		Y	Y	Y	Y
Equal-wt. vac. growth			Y			
Fixed effects						
Occ-year	Y	Y	Y	Y	Y	Y
Metro-year	Y	Y	Y	Y	Y	Y

Note: This table repeats the robustness checks in Table A6, but allowing the coefficient on the HHI and outside-occupation option index to vary by quartile of outward occupational mobility (the occupation "leave share" defined as in Section 2.4). Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A9: Regression of wage on HHI and outside-occupation options, by quartile of outward mobility: robustness (2)

Dependent variable:	Log wage					
	(a) Large Firm HHI Inst.	(b) Pos. & Neg. Shock HHI Inst.	(c) Level HHI Inst.	(d) Industry Bartik	(e) Occ-MSA FEs	(f) CPS mobility
Log HHI	-0.030***	-0.031***	-0.027***	-0.038***	-0.017***	-0.034***
X Q1 occ mobility	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.005)
Log HHI	-0.022***	-0.017***	-0.013***	-0.029***	-0.008	-0.018***
X Q2 occ mobility	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)
Log HHI	-0.005	-0.006	-0.003	-0.010**	-0.002	-0.023***
X Q3 occ mobility	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Log HHI	-0.008	-0.005	0.001	-0.011**	0.001	-0.007
X Q4 occ mobility	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Log outside-occ. options	0.115***	0.108***	0.141***	0.107***	0.019	0.099***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.014)	(0.010)
Observations	382,389	329,855	443,233	405,873	427,472	443,233
F-stat	72	109	92	67	30	72
Controls						
Vacancy growth	Y	Y	Y	Y	Y	Y
Predicted vac. growth	Y	Y	Y	Y	Y	Y
Expected HHI instrument	Y	Y	Y	Y	Y	Y
Industry Bartik				Y		
Fixed effects						
Occ-year	Y	Y	Y	Y		Y
Metro-year	Y	Y	Y	Y		Y
Occ-metro					Y	
Year					Y	

Notes: This table repeats the robustness checks in Table A7, but allowing the coefficient on the HHI and outside-occupation option index to vary by quartile of outward occupational mobility (the occupation "leave share" defined as in Section 2.4). It also includes an additional robustness check in column (f), which allows the coefficients to vary by the occupation's quartile of outward mobility as calculated in the CPS rather than in the BGT resume data. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A10: Regression of wage on outside-occupation options: 1999–2019

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: OLS regressions, employment-weighted				
oo^{occs}	0.392***	0.089***	0.130***	0.028***
	(0.007)	(0.005)	(0.003)	(0.005)
Panel B: 2SLS IV regressions, employment-weighted				
oo^{occs} , instrumented	0.408***	0.099***	0.106***	0.030***
	(0.007)	(0.003)	(0.004)	(0.006)
Panel C: OLS regressions, non employment-weighted				
oo^{occs}	0.129***	0.084***	0.094***	0.047***
	(0.002)	(0.001)	(0.001)	(0.002)
Panel D: 2SLS IV regressions, non employment-weighted				
oo^{occs} , instrumented	0.114***	0.072***	0.077***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)
Panel E: First stage, employment-weighted				
oo^{occs}	1.072***	0.862***	0.869***	0.823***
	(0.016)	(0.009)	(0.005)	(0.047)
Panel F: First stage, non employment-weighted				
oo^{occs}	0.982***	0.850***	0.804***	0.969***
	(0.004)	(0.005)	(0.003)	(0.016)
Observations	2,268,006	2,267,839	2,267,839	2,255,447
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-by-year 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 A11: 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 metro area-by-year 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 A12: Selected Challenges to the Empirical Approach and How They Are Addressed

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

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

Table A12: (continued)

Concern	How it is addressed
2. How does your measure of outside options compare to that of Danieli & Caldwell (2018)?	Caldwell and Danieli (2018)'s index of the value of workers' outside options captures the degree to which workers of a given type are distributed across different job types. Intuitively, a worker has more outside options if other workers of the same type are distributed across a wider range of types of job, where job type is captured empirically by a combination of establishment characteristics, industry and occupation task requirements, and geographic location. (Worker type is captured with gender, age, secondary education qualification, citizenship status, and occupation at labor market entry.) Their index reflects an equilibrium distribution of workers across job types and space. This assumes that workers' outside option set is best captured by the outside option set of others who look similar to them in terms of underlying characteristics (rather than current occupation), and that the outside option set of these other workers is in turn best captured by the types of firms and task requirements in the jobs worked in by these other workers. In contrast, our outside option index captures the degree to which workers in the same occupation move to different occupations. Our primary assumption is therefore that workers' outside option set (outside their occupation) is best captured by the outside option set of others in the same occupation, which in turn is best captured by the revealed occupational switches of others in the same occupation.
3. To account for mobility, could we just define a labor market to be a cluster of more than one occupation?	While this kind of clustering may more closely approximate a worker's labor market than administrative classifications, such as SOC codes, it still represents a form of "binary" market definition that does not allow for differential weighting of a worker's options but rather limits the weights to 0% (out of the market) and 100% (in the market). Moreover, the asymmetry of empirical movements between occupations (due to differential feasibility or desirability of reverse moves) is not accounted for by simply grouping occupations into new clusters. In Section 3.3 we suggest a compromise by creating labor markets that are mobility-adjusted and binary from the perspective of a particular occupation, but not symmetric.

Table A12: (continued)

Concern	How it is addressed
4. Instead of looking at both occupational mobility and concentration, could you construct a better HHI that combines both of these dimensions into one index that better measures concentration in probabilistic labor markets?	While we think the approach considering the interaction of both mobility and concentration that we outline in this paper is more transparent and model-agnostic, some contexts (e.g. legal thresholds for antitrust purposes) may require a labor market definition that allows for measuring a single comparable index of concentration across occupations. Our heterogeneity results suggest that monopsony power has the highest negative effect on wages in low mobility occupations. Therefore, an index that represents the product or interaction of an occupational mobility measure (e.g. our *"leave share"*) and a conventional concentration measure like the HHI may be more suitable for those purposes. Our paper's contribution is to establish that these factors should be considered jointly. While the exact functional form and relevant thresholds will vary based on the context and usage, we show in Section 3.3 that an approach that creates "mobility-adjusted labor markets" based on mobility data can be used to construct mobility-adjusted concentration indices which no longer show differential effects on different occupations sorted by mobility. Such mobility-adjusted labor markets can therefore be used for comparing changes in concentration across different labor markets in a way that implies similar wage effects for similar concentration changes. We encourage other researchers and practitioners to adapt this general approach to their specific needs and further study the properties of such a combined index.
Identification strategy	
5. What if an employer's expansion in *other* cities predicts an employment *contraction* in the focal city?	While this is certainly possible in individual cases, on average this is not true in our data: a regression of $g_{j,o,k,t}$ on $g_{j,o,\not{k},t}$ gives a coefficient of 0.02 (standard error 0.0001); when restricting to firms with vacancies in 5 or more metro areas in the same occupation, the coefficient is 0.35 (standard error 0.0006) (where $g_{j,o,k,t}$ is the growth in vacancy postings of firm j in occupation o and metro area k between years $t-1$ and t, and where $g_{j,o,\not{k},t}$ is the growth in vacancy postings of firm j in occupation o in all metro areas nationwide *except* k between years $t-1$ and t).
6. Does the fact that some cities have booming local economies while others do not - and that this affects wages - represent an issue for your identification approach?	The inclusion of metro-by-year fixed effects means that differences in metro area-level trends or national productivity of different occupations do not represent an issue for our identification strategy. Additionally, we show that our results are robust to controlling for common exposure to national industry shocks at the metro level (see Appendix Table A6).

Table A12: (continued)

Concern	How it is addressed
7. What if the expansion of a national firm in a local labor market changes wages through a simple labor demand channel rather than through its effect on concentration?	We address this concern in two ways. (1) Mathematically, the effect of a large firm's growth on local labor market concentration is quadratic, whereas the effect of a large firm's growth on local labor demand is likely linear (i.e. a 10% increase in local vacancies by national firms increases labor demand twice as much as a 5% increase). In addition, growth of a national firm can either increase or decrease local employment concentration: if the firm was a small player in the local labor market, its growth might reduce local concentration, whereas if it was already large, its growth will increase concentration. We control for the potential direct effect of higher labor demand on wages by including the expected change in local vacancy postings in the regressions. Controlling for the exposure of local labor demand to national employment trends directly to prevent it from confounding a nonlinear IV is similar to the "double Bartik" approach in Chodorow-Reich and Wieland (2020). Note that the inclusion of these demand terms does not substantially affect our baseline coefficient estimates (Appendix Tables A6 and A8). (2) The most likely direction of correlation between an increase in concentration due to an increase in hiring by national firms and overall labor demand is positive. Note that, if the labor supply curve slopes upward, this means that our instrument should be positively correlated with wages, which would bias the negative estimates of the effect of concentration on wages towards zero. That is, if our attempts to control for labor demand are insufficient, this most likely means that our estimates are *too conservative*. Consistent with this, when we exclude our labor demand controls our coefficient estimates on the wage-HHI relationship get smaller in absolute magnitude (closer to zero), although the differences are not large.
8. Does the fact that some occupations have recently seen an increase in labor demand and wages - and that this might correlate with increasing concentration in particular industries, such as the tech sector - represent an issue?	The inclusion of occupation-by-year fixed effects means that differences in occupation-level trends at the national level do not represent an issue for our identification strategy. Our estimates are identified off variation across metro areas within occupations, due to local shocks that differ from those experienced by the average labor market for that occupation.
9. Does your IV approach require exposure of local labor markets to large firms to be randomly assigned?	No. In the conventional framework of thinking of shift-share shocks as consisting of an "exposure" (here, the initial local employment concentration of different firms) and a "shifter" (here: national time-varying shocks to hiring by particular large firms), our causal effects are identified off the exogenous variation in the shifter, conditional on the included controls and fixed effects. We discuss conditions for identification further in Appendix G, following the approach to shift-share IVs of Borusyak et al. (2022).

Table A12: (continued)

Concern	How it is addressed
10. Could your outside options results be driven by the fact that connected local occupations are affected by common *industry* shocks?	We explicitly try to control for this confounding factor by constructing a shock that reflects the common predicted impact of national industry wage trends for each occupation-metro area-year cell, similar to an approach to controlling for common trends in Chodorow-Reich and Wieland (2020). When we control for these common industry shocks, the coeffcient on the outside options index barely changes, suggesting that common industry variation is likely not a major driver of our results. See Appendix G and Tables A7 and A9 for details.

Appendix References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

_ and _ , "A Cautionary Note on Using (March) Current Population Survey and Panel Study of Income

Dynamics Data to Study Worker Mobility," *Macroeconomic Dynamics*, 2013, *17*, 172–194.

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

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

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

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

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

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

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

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

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

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

Weil, David, *The Fissured Workplace* 2014.

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