Employer Concentration and Outside Options

Gregor Schubert, Anna Stansbury, and Bledi Taska *

January 18, 2021

Job Market Paper (Anna Stansbury)

Click here for most recent version

Abstract

This paper studies the effect of employer concentration on wages in the United States. We make two primary new contributions. First, we develop an instrument for employer concentration, based on differential local exposure to national firm-level trends. We use the instrument to estimate the effect of plausibly exogenous variation in employer concentration on wages across the large majority of U.S. occupations and metropolitan areas. Second, we adopt a flexible "probabilistic" approach to labor market definition, identifying relevant job options outside a worker's own occupation using new occupational mobility data constructed from 16 million resumes, and estimate the effect of these outside-occupation options on wages. We find that moving from the median to the 95th percentile of employer concentration reduces wages by 3%. But we also reveal substantial heterogeneity: the effect of employer concentration is at least four times higher for low outward mobility occupations than those with high outward mobility. Since the majority of U.S. workers are not in highly concentrated labor markets, the aggregate effects of concentration on wages do not appear large enough to have substantial explanatory power for income inequality or wage stagnation. Nonetheless, our estimates suggest that a material subset of workers experience meaningful negative wage effects from employer labor market power. Our findings imply that labor market regulatory agencies and antitrust authorities should take employer concentration seriously, but that measures of employer concentration – typically calculated for narrowly defined occupational labor markets – need to be adjusted to incorporate the availability and quality of job options outside the focal occupation.

*Schubert: Harvard University, gschubert@g.harvard.edu. Stansbury: Harvard University, annastansbury@g.harvard.edu. Taska: Burning Glass Technologies, btaska@burning-glass.com. Previous versions of this paper have been circulated under the titles "Monopsony and Outside Options", and "Getting Labor Markets Right". For their detailed comments and advice, we thank Justin Bloesch, Gabriel Chodorow-Reich, Karen Dynan, Richard Freeman, Ed Glaeser, Xavier Gabaix, Benny Goldman, Larry Katz, Larry Summers, Maya Sen, and Betsey Stevenson. For helpful discussions, we thank David Balan, John Coglianese, Oren Danieli, David Deming, Martin Feldstein, Claudia Goldin, Emma Harrington, Simon Jaeger, Robin Lee, Mike Lipsitz, Jeff Miron, Suresh Naidu, Nancy Rose, Isaac Sorkin, Elizabeth Weber Handwerker, Ron Yang and participants of the briq Workshop on Firms, Jobs and Inequality, the 2019 Federal Reserve System Community Development Research Conference, the IZA Summer School on Labor Economics 2019, the IZA/CAIS Workshop on Matching in Online Labor Markets, the Urban Economic Association Meetings 2019 and 2020, the Wharton People and Organizations Conference, the Asian and Australasian Society of Labour Economics 2019 Conference, the Oxford NuCamp Virtual PhD Workshop, the Harvard Center for International Development Growth Lab Seminar, the MIT IWER Seminar, and the Harvard Labor, Industrial Organization, and Macro lunches, Labor breakfast, and Multidisciplinary Seminar on Inequality and Social Policy. 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

Policy and academic debates have become increasingly focused in recent years on the issue of employer concentration. A lack of choice of job options for workers - as a result of a few large firms dominating their local labor market - has been posited as a possible explanation for inequality, low wages, and stagnant wage growth. Antitrust authorities have been called upon to consider employer concentration when reviewing mergers and acquisitions. Concerns have been raised that high employer concentration may facilitate (legal or illegal) restrictions to labor market competition, such as the use of no-poaching agreements or non-compete clauses. And, since employer concentration can be a source of monopsony power in labor markets,[1] concerns around high employer concentration have bolstered calls for higher minimum wages and for a strengthening of workers' collective bargaining power.[2]

But to assess whether – or in which cases – policy should respond to employer concentration, we need a deeper understanding of the nature and effects of employer concentration in the United States. Two major open issues remain. First, endogeneity: while there is a well-documented negative correlation between local employer concentration and wages, the extent to which this is causal – and the magnitude of any such causal effect – is unclear. Second, market definition: assessing the effect of local employer concentration on wages, and pinpointing the workers who are most affected by it, requires a good definition of the relevant local labor market for workers.[3]

Our paper addresses both of these issues, estimating the effect of employer concentration on wages across the majority of U.S. occupations and metropolitan areas. To address the endogeneity issue, we develop a new identification strategy based on differential local exposure to national firms' hiring growth. To address the market definition issue, we segment our analysis by the degree of outward mobility from different occupations. We also develop a new measure for the outside option value of local jobs in other occupations, proxying for workers' ability to find a job in another occupation

[1]See, for example, Robinson (1933) or Manning (2003) on monopsony power. Note that employer concentration is only one possible source of monopsony power in labor markets. Others include search frictions, switch costs, and worker and job heterogeneity.

[2]Authors making arguments referred to in this paragraph include, variously, Bahn (2018); Shambaugh, Nunn, Breitwieser and Liu (2018); Krueger and Posner (2018); Naidu, Posner and Weyl (2018); Marinescu and Hovenkamp (2019); Marinescu and Posner (2020); Naidu and Posner (2020).

[3]Prominent critiques of the emerging recent literature on employer concentration and wages focus on these two factors, including Rose (2019) and Berry, Gaynor and Scott Morton (2019).

using new occupational mobility data which we construct from 16 million U.S. workers' resumes, and use our new measure to estimate the joint effect of within-occupation employer concentration and outside-occupation options on wages.

Our baseline results suggest an important effect of employer concentration on wages: moving from the degree of employer concentration faced by the median worker to that faced by the worker at the 95th percentile results in a roughly 3% lower wage, holding all else equal. This average, however, masks substantial heterogeneity: the effect of employer concentration is at least six times higher for the least outwardly mobile than the most outwardly mobile occupations. A back-of-the-envelope calculation, using our coefficient estimates, suggests that over 10% of the 110 million workers covered in our data experience wage suppression of 2% or more as a result of employer concentration.

Overall, our findings point to a middle ground between two prominent views about the effects of employer concentration in the U.S. labor market. On the one hand, employer concentration is *not* a niche issue confined to a few factory towns: our results suggest that a material subset of U.S. workers see non-trivial effects of employer concentration on their wages. On the other hand, employer concentration *does not* seem to be an important determinant of wages for the majority of U.S. workers, and the effects of employer concentration 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.

The fact that employer concentration affects wages for several million American workers suggests that increased policy attention to this issue is appropriate, both from antitrust authorities and from labor market authorities who might promote countervailing worker power through labor market regulations, minimum wages, or collective bargaining. In making policy decisions which depend on employer concentration, however, the definition of the labor market is vitally important. Policymakers seeking to act on a (perceived) lack of labor market competition should take into consideration not only the degree of employer concentration within the local occupation, but also the degree of outward occupational mobility and the availability and quality of local job options in other feasible occupations.

We give a more detailed overview of our analysis in the rest of this section. In section 2, we outline the conceptual framework guiding our analysis. We outline our empirical approach in sections 3 and 4, and present our results in section 5. Section 6 discusses the implications of our findings, and section 7 concludes.

1.1 Overview of our analysis

Employer concentration and wages – theory: There are a number of models of the labor market in which employer concentration matters for wages. First, if there are few large employers in a given labor market, and each individual firm therefore faces an upward-sloping labor supply curve, firms will optimally pay a wage which is marked down below the marginal product of labor (one aspect of labor market monopsony power as initially described by Robinson (1933)). Second, the presence of a small number of firms in a labor market makes it easier for firms to collude to suppress wages. Third, in a labor market where rents or quasi-rents are present in employment relationships, and firms and workers bargain over these rents, employer concentration reduces the number of feasible outside job options workers have in a given labor market, reducing their relative bargaining position and therefore exerting downward pressure on wages.

A number of recent papers have demonstrated theoretically that higher employer concentration can reduce wages. Berger, Herkenhoff and Mongey (2019) develop a general equilibrium oligopsony model of the labor market, demonstrating that firms with higher market share have greater market power (due to a lower elasticity of labor supply to the firm), and showing that the wage-bill Herfindahl-Hirschmann Index ("HHI") is a relevant statistic for assessing the welfare effects of firms' labor market power. Azkarate-Askasua and Zerecero (2020) demonstrate that firms' employment shares affect the elasticity of labor supply to the firm, and therefore equilibrium wages, in a model with worker bargaining power as well as employer wage-setting power. Jarosch, Nimczik and Sorkin (2019) show that in a search model with wage bargaining, the presence of large employers worsens workers' outside option and so reduces wages, with the effect determined by a concentration index which is closely related to an HHI of firm employment shares. Arnold (2020) and Naidu and Posner (2020) show that an HHI of employer concentration partly determines the magnitude of the wage markdown in model of Cournot competition between employers, and Hershbein, Macaluso and Yeh (2019) show that under certain conditions, an HHI can be used to proxy for firms' elasticity of labor supply.

In this paper, we outline a brief conceptual framework which formalizes the intuition that employer concentration can exert downward pressure on workers' wages by reducing the value of workers' outside option in the wage bargain. Our framework has three core ingredients: (1) the expected value of workers' outside option plays a role in wage determination, as it creates a floor below which workers' wage cannot fall at any given employer; (2) the expected value of the outside option can be thought of as the weighted average of the wages offered by other employers, with the weights representing the likelihood that workers would be able to find jobs at these other employers; and (3) the likelihood of receiving job offers from other employers is a function of the share of vacancies in the labor market which are accounted for by firms *other than* the worker in question's own employer. With these three ingredients, higher employer concentration will always exert downward pressure on the wage, holding all else equal. While the primary focus of our paper is empirical, we see our conceptual framework as useful – and as complementary to the other work outlined above – in illustrating the generality of the possibility that employer concentration might depress wages, and in demonstrating that an employer HHI index can be a good proxy for the effects of employer concentration on wages.

Endogeneity: A growing empirical literature demonstrates a negative correlation between measures of employer concentration and wages across U.S. labor markets, including Azar, Marinescu, Steinbaum and Taska (2020a), Azar, Marinescu and Steinbaum (2020b), Rinz (2018), Lipsius (2018), Benmelech, Bergman and Kim (2018), Hershbein et al. (2019), and Qiu and Sojourner (2019). In practice, however, identifying a causal effect of employer concentration on wages is difficult. An increase in employer concentration could reflect a lack of local economic dynamism, with few new firms being created. This may be associated with slower wage growth because productivity growth is slower, rather than because of any increase in employer wage-setting power. On the other hand, an increase in employer concentration could reflect the expansion of a highly productive large firm, which may increase wages as average productivity in the labor market increases: this effect would bias towards zero estimates of the relationship between employer concentration and wages.

Responding to these concerns, some recent papers have demonstrated plausibly causal effects of changes in employer concentration on wages in certain specific cases. Arnold (2020) uses local merger and acquisition activity in the U.S. as an instrument which increases employer concentration in specific local labor markets, while being plausibly orthogonal to other local economic trends: he finds that M&A activity which increases employer concentration substantially depresses wages. Prager and Schmitt (2019) find that hospital mergers which increase employer concentration substantially reduce the wage growth of workers with highly industry-specific skills. A different strand of the literature instruments for changes in employer concentration in a given local occupation using changes in the inverse of the average number of firms in the same occupation in other localities (Azar et al., 2020b; Rinz, 2018; Qiu and Sojourner, 2019; Marinescu, Ouss and Pape, 2019; Gibbons, Greenman, Norlander and Sørensen, 2019).

Endogeneity – Our approach: In this paper we propose a new identification approach for the effects of employer concentration on wages. We note that changes in employer concentration in any given local labor market are primarily driven by changes in large firms' hiring behavior, and that in different localities, different firms are dominant. We take advantage of this differential local-level exposure to large national firms, instrumenting for the local change in employer concentration within a particular occupation with the *predicted* change in employer concentration based on locally-large employers' national hiring growth (and controlling for the predicted effect of this exposure on local labor demand). The intuition: when a large employer grows nationwide, this should increase employer concentration by more in cities which already had a large presence of that employer. This enables us to construct shocks to local concentration that are plausibly orthogonal to local changes in productivity or to any other local occupation-specific economic trends, with the key identifying assumption being that each large firm's decision to increase or decrease its hiring nationwide is exogenous with respect to the local economic conditions in any specific occupation-metropolitan area labor market.

Our approach derives from the "granular" instrumental variable approach (GIV) developed by Gabaix and Koijen (2020), which uses plausibly exogenous idiosyncratic firm-level variation to instrument for changes in market-level aggregates, and from the shift-share "Bartik" approach, which identifies causal effects based off differential local exposure to national trends. This approach enables us to disentangle the effects of local employer concentration from local occupation-specific economic trends, as well as from national occupational wage shocks or local metropolitan area wage shocks.[4] We see our approach as complementary to work which estimates the effect of employer concentration on wages using M&A activity, since our estimated wage effects are identified

[4]The common instrumental variable approach in prior literature, in contrast – which instruments for changes in employer concentration in a given local occupation using changes in the inverse of the average number of firms in the same occupation in other localities – cannot disentangle the effects of changes in local employer concentration from occupation-level shocks to wages.

from different variation in concentration (and we note that M&A activity accounts for less than 2% of changes in U.S. employer concentration over time (Arnold, 2020)). In addition, since our approach can be applied to the entire universe of U.S. labor markets, we hope this makes it useful in its broader applicability. We discuss our empirical approach, and the conditions under which it is valid, in more detail in section 4.

Market definition: A second difficulty in estimating the effect of employer concentration on wages is the question of market definition. Most theoretical models in which employer concentration affects wages rest on the concept of a clearly delineated labor market where all workers are equally good fits for all jobs in that market, and cannot work in any jobs outside that market. This approach is commonly used in antitrust policy in product markets and has been taken by almost all the research on employer concentration to date. Labor markets have typically been defined as a single occupation or industry within a given local area (commuting zone, metropolitan area, or county), and debate has focused on how narrow an occupational or industrial definition to draw. Jarosch et al. (2019) instead defines local labor markets as clusters of firms, where the clusters are inferred using worker flows, and Dodini, Lovenheim, Salvanes and Willén (2020) define local labor markets as clusters of jobs based on common skills.

In practice, however, labor markets are not clearly delineated from one another. Within a narrowly-defined occupation, it may be a good approximation that all workers and jobs are relatively substitutable – but, workers often move to jobs in different occupations, meaning that an employer concentration measure at the level of narrowly-defined occupations (such as those in Azar et al. (2020b), Azar et al. (2020a), and Hershbein et al. (2019)) may exaggerate the true degree of employer concentration faced. On the other hand, within a broader occupation group, it is highly unlikely that all jobs are equally good options for all workers, meaning that pockets of higher effective employer concentration may be obscured.[5]

Market definition – our approach: In this paper, rather than following the binary market definition approach in delineating a set of distinct labor markets, we propose a different approach. We measure employer concentration at the level of a narrowly-defined occupation (SOC 6-digit) within a given metropolitan area, but *also*

[5] While the firm clusters approach of Jarosch et al. (2019) improves upon administrative definitions of labor markets, it does not solve the problem: it still requires individual labor markets to be delineated from each other in a binary way.

identify a probabilistic cluster of other occupations which will comprise workers' likely labor market, identifying the relevance of other occupations for workers in the initial occupation based on empirical occupational mobility patterns.[6]

We use this probabilistic concept of a labor market to develop a measure of the value of workers' outside job options in other occupations: an "outside-occupation option index", defined as the weighted average of local wages in all occupations except the worker's own, with each weight the product of two factors: (i) national occupational mobility from the worker's occupation to each other occupation (to proxy for the likelihood that an occupation could be a relevant outside option), and (ii) the local employment share in each other occupation, relative to the national average (to proxy for the local availability of jobs in each other occupation). We show that such an index can be derived from our conceptual framework in which workers' wage depends on the expected value of their outside option.

To take into account workers' differential access to jobs outside their own occupation, we introduce two new factors into our baseline regressions of wages on employer concentration. First, we *segment our analysis* by occupations' degree of average outward mobility. This enables us to estimate different effects of employer concentration on wages for low-mobility occupations (for whom the SOC 6-digit occupation may be a good approximation to their true labor market) and for high-mobility occupations (for whom the SOC 6-digit occupation is almost certainly not a good approximation to their true labor market). Second, we *control for the availability* and quality of job options outside workers' own occupation, using our outside-occupation option index described above. This enables us to eliminate any bias introduced into wage-concentration regressions by failure to consider the quality of outside-occupation job options: empirically, it is the case that workers who are in high-concentration labor markets (*within* their local occupation) also tend to have poor local job options *outside* their occupation.

We also identify causal effects of changes in the value of outside-occupation job options on wages, instrumenting for demand shocks to workers' outside option occupations using the national leave-one-out mean wage in each of those occupations, and proxying for local exposure to these demand shocks using the initial local employment share in each of these occupations relative to the national average.

[6] Our strategy of using occupational transitions to identify outside options is most closely related to Arnold (2020), who uses worker transitions between industries to create a "flow-weighted" employer concentration measure for each local industry.

Data: Using the approach described above, we study the effects of employer concentration on wages across a large subset of all U.S. occupations (at the SOC 6-digit level) and metropolitan areas over 2013–2016: a total of over 100,000 occupation-metropolitan area cells. Our dependent variable of interest is the average hourly wage in each occupation-metropolitan area labor market in each year, which we obtain from the Bureau of Labor Statistics Occupational Employment Statistics (BLS OES).

To measure employer concentration, we construct a Herfindahl-Hirschman Index (HHI) of vacancy concentration by occupation and metropolitan area, using a database of the near-universe of online vacancy postings collected by Burning Glass Technologies (following Azar et al. (2020a) and Hershbein et al. (2019)).

To measure outside-occupation options, we use new occupational mobility data which we construct from a data set of 16 million resumes, also collected by Burning Glass Technologies. The large sample size of this resume data – an order of magnitude more than other data sources – enables us to reliably estimate occupational transitions between a large proportion of occupations in the U.S. We use this data, alongside annual wage and employment data at the level of individual occupations and metropolitan areas from the BLS OES, to construct our outside-occupation option index. In addition: we are making our new occupational mobility data set publicly available, since a highly granular data set of occupational transitions like this does not yet exist for the United States.

Results: In our baseline specification, controlling for differential availability of outside-occupation job options, we estimate a negative and significant effect of local within-occupation employer concentration on wages: moving from the median to the 95th percentile HHI as faced by workers (or roughly an HHI of 220 to an HHI of 2,200, measured at the occupation by metropolitan area level) is associated with 2.6 log points lower wages. Our instrumental variable estimates are about 50% larger than our OLS estimates, suggesting that some combination of omitted variables or measurement error bias the coefficient towards zero in simple regressions of wages on employer concentration.

The average effect of employer concentration on wages masks substantial heterogeneity: within-occupation employer concentration matters substantially less for workers who are easily able to find jobs in other occupations. For occupations in the bottom quartile of outward mobility, moving from the median to 95th percentile of concentration as faced by workers is associated with between 4 and 8 log points lower wages; for occupations in the highest quartile of outward mobility, our point estimate is zero and the confidence interval rules out any decrease in wages greater than 1.7 log points for an equivalent change in employer concentration.

As expected, we find that regressions of wages on employer HHI suffer from omitted variable bias if the availability of outside-occupation job options is not included in the analysis, with an upward bias in the coefficient size of about 30-40%. This occurs because workers with few local employers in their own occupation, and so a high HHI, also tend on average to have worse local outside-occupation job options. These results imply that simple regressions of wages on within-occupation HHI cannot accurately identify the effects of employer concentration without also taking into account heterogeneity of outward occupational mobility and the quality of local outside-occupation job options.

Finally, we find a positive and significant effect of an increase in the value of outside-occupation options themselves, holding constant within-occupation employer concentration: for the median occupation, moving from the 25th to the 75th percentile value of outside-occupation options across cities is associated with 3.7 log points higher wages. These magnitudes are large relative to the degree of geographic wage dispersion across cities: for the median occupation, moving from the 25th to the 75th percentile city by average wage is associated with a 21 log points higher wage. These results suggest that job options outside workers' own occupation play an important part in wage determination for the average worker; that empirical occupation transitions can be used effectively to proxy for the relevance of different occupations as outside job options; and that the differential availability of outside-occupation job options is an important determinant of pay for workers in different U.S. occupations and cities.

Relevant literature: Our paper makes a number of contributions to related literatures. First, our estimates of the effect of employer concentration build on a growing body of work demonstrating a relationship between wages and employer concentration, including Azar et al. (2020b), Azar et al. (2020a), Rinz (2018), Lipsius (2018), Benmelech et al. (2018), Hershbein et al. (2019), and Qiu and Sojourner (2019) in U.S. local occupations and/or industries, Gibbons et al. (2019) for guest workers in the U.S., Abel, Tenreyro and Thwaites (2018) in local industries in the UK, Marinescu et al. (2019) and Azkarate-Askasua and Zerecero (2020) in local occupations and industries in France, and Martins (2018) in local occupations in Portugal. Jarosch et al. (2019) develop a concentration index defined on clusters of firms, identified from worker flows, and estimate the relationship between this concentration index and pay in Austria. Prager and Schmitt (2019) and Arnold (2020) estimate the effect of increases in local employer concentration as a result of M&A activity on wages in the U.S.[7] We make two contributions to this literature. With a new instrument for changes in local employer concentration, based on employer "granularity", we can obtain plausibly causal and precisely estimated effects of employer concentration on wages across the majority of occupation-city labor markets in the U.S. And, in integrating this concentration literature with analysis of the availability of jobs outside workers' occupation, we show that if measures of employer concentration are to be used for labor market analysis (in, for example, antitrust screening as suggested by Marinescu and Hovenkamp (2019)), they must also take into account the availability of outside-occupation options. In contemporaneous work, Dodini et al. (2020) develop an employer concentration measure based on local skill clusters in Norway, which also demonstrates the importance of substitutability across occupations and industries in estimating the effects of employer concentration.

Second, our results add to a literature on the effect of workers' outside options in wage determination more broadly. Our identification for the effect of changes in outside-occupation options relates most closely to Beaudry, Green and Sand (2012), who show for the U.S. that local changes in the availability of high-wage jobs in some industries have spillover effects. We differ in using occupational mobility to identify *which* occupations are more or less affected by these spillovers. Our work is also similar to Caldwell and Danieli (2018), who construct an index of the value of workers' outside options in Germany based on the diversity of jobs and locations in which similar workers are observed, and find it is strongly associated with wages, Macaluso (2019), who studies how outcomes for laid-off workers in the U.S. vary depending on the similarity of available local jobs, and Alfaro-Urena, Manelici and Vasquez (2020), who estimate the outside option value of jobs at multinational corporations in Costa Rica. Our paper adds to this literature by developing a new way of measuring outside-occupation options based on observed occupational transitions.

[7] In addition, Azar, Huet-Vaughn, Marinescu, Taska and Von Wachter (2019b) find that minimum wage increases have smaller negative employment effects in local labor markets with higher employer concentration, consistent with monopsony power.

Third, in using empirical worker transitions to infer the extent of workers' labor markets, our work builds on other papers which use worker flows to identify the scope of workers' geographic labor markets (Manning and Petrongolo (2017) in the UK), to identify firm clusters representing labor markets (Nimczik (2018) in Austria), or to study similarity in skill requirements across occupations and industries (Shaw, 1987; Neffke, Otto and Weyh, 2017; Arnold, 2020).

Finally, in estimating the degree to which wages are sensitive to outside options, we contribute to a broader literature on imperfect competition in labor markets. First, our work relates to the broader literature on labor market monopsony and the elasticity of the labor supply curve to the firm, including the theoretical work in Boal and Ransom (1997), Manning (2003), Azar, Berry and Marinescu (2019a), Berger et al. (2019), and Azkarate-Askasua and Zerecero (2020), broad empirical estimates of the elasticity of labor supply to the firm in Webber (2015) and Sokolova and Sorensen (2020), and empirical analyses of specific industries, firms, or worker classes in Hirsch and Schumacher (2005), Staiger, Spetz and Phibbs (2010), Ransom and Sims (2010), Ashenfelter, Farber and Ransom (2013), Matsudaira (2014), Naidu, Nyarko and Wang (2016), Bassier, Dube and Naidu (2019), Goolsbee and Syverson (2019), and Dube, Jacobs, Naidu and Suri (2020). Second, in demonstrating the importance of outside options in wage determination, our work is relevant to the large search-and-matching literature, especially models featuring wage bargaining between workers and firms (e.g. Burdett and Mortensen, 1980; Pissarides, 2000). In particular, our findings complement those of Cahuc, Postel-Vinay and Robin (2006), who find that the degree of between-firm competition for workers is a key driver of wages.

Overall, our paper demonstrates – in data comprising over 100,000 U.S. occupation-city cells – that within-occupation local employer concentration matters substantially for the wages of a material subset of U.S. workers. The size of this effect, however, depends very much on the availability of job options outside workers' occupation. When studying the degree of competition in a labor market, whether for economic research or for policy purposes, analysts should focus both on the degree of employer concentration within an occupation *and* the availability of local job options outside the occupation.

2 Conceptual framework

Why might employer concentration reduce wages? As laid out in the introduction, employer concentration can exert downward pressure on wages in a number of different models of the labor market. In this paper, the conceptual framework guiding our empirical analysis is one where **employer concentration worsens workers' outside options**. The central premise is that increased employer concentration reduces the availability of outside job options to workers in a given labor market, reducing the value of this outside option in the wage bargain.[8] To develop this intuition, consider a market for retail workers in a small town where there are many mom-and-pop stores. For each worker at each individual store, the majority of other job options in their labor market are outside their own firm. They know, and their employer knows, that this gives them a high likelihood of being able to find another job outside their current firm. On the other hand, consider a comparable small town where there is one large supermarket and one small mom-and-pop store. The supermarket hires many workers, so any individual worker at the mom-and-pop store knows that if she quits her job, she will likely be able to get a job at the supermarket. But for the workers employed at the large supermarket, the option of leaving to get a job at the single mom-and-pop store is not particularly credible – this store is so small that the likelihood of it hiring at any given moment is also small – and there are no other options. This means that their outside option is not very valuable in the wage bargain.[9]

Below, we briefly outline a stylized framework where we formalize this intuition. We develop the framework in more detail in Appendix A. In this section, we will consider a clearly-defined labor market where all workers and all jobs are perfect substitutes. In section 3, we will relax this assumption, allowing for workers to have job options outside their current occupation.

Wage bargaining. At the start of each period, each employed worker Nash-bargains with her employer i over the wage. The outcome is a wage w_i, equal to the value of the worker's outside option if she leaves her job, oo_i, plus a share β – reflecting

[8] Jarosch et al. (2019) develop a search-and-matching model which formalizes a similar intuition: in their model, workers cannot re-encounter the same firm twice when searching, so the presence of large firms reduces workers' effective job offers and therefore depresses the wage. Our work was developed contemporaneously and independently.

[9] While this framework suggests that the outside options are different for workers at different firms in the same local labor market, we focus our analysis at the market level in this paper, estimating the effect of employer concentration in a given labor market on *average* wages in that labor market.

worker bargaining power – of the match surplus:

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

where p_i is the product of the match.[10] We assume all workers at the same firm i have the same outside option, such that all workers at firm i are paid the same wage.

Job search. After wage bargaining with incumbent workers has concluded, firms post vacancies to fill positions which have been vacated – either by worker exit from the labor market, or by the breakdown of wage bargaining with an existing worker. Posted vacancies contain a take-it-or-leave-it wage offer, with the posted wage equal to the wage the firm is paying to its other workers. Job seekers – workers whose previous wage bargain broke down, or workers who newly entered the labor force – are each paired randomly with a job vacancy within their labor market.[11] Random matching means that the chance of any given worker receiving a job offer from a particular firm j is equal to σ_j, the share of all vacancies in the labor market accounted for by firm j (in the spirit of Burdett and Mortensen (1980) and Jarosch et al. (2019)). If a worker does not receive any job offers, she moves to unemployment for the period and receives unemployment benefit b.[12]

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

---

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

[11]We consider workers' labor market, in this section, to be the set of local feasible jobs, where all jobs within the labor market are equally feasible and all jobs outside the labor market are infeasible. We will relax this assumption later when we discuss market definition.

[12]All workers who receive a job offer accept it, since the wage offered will always exceed unemployment benefit b and since the search process is the same for unemployed or recently employed job seekers.

weighted by the probability of being matched with each feasible firm σ_j, as well as the unemployment benefit b multiplied by the probability of receiving no job offers $\left(1 - \sum_{j \neq i}^{N} \sigma_j\right) = \sigma_i$.[13]

Equilibrium wage. The wage for workers at firm i therefore satisfies:

$$w_i = \beta p_i + (1 - \beta) \left(\sum_j \sigma_j \cdot w_j + \sigma_i \cdot b \right) \tag{2}$$

This expression suggests that the wage at firm i depends on the employment share of firm i in the labor market. Why? In a labor market with atomistic firms, every job seeker would receive a match from a feasible employer each period. However, in a labor market with some large employers, the chance that a worker is re-matched with her current firm is non-zero, and if this happens the large employer can threaten *not* to re-hire their former worker. This means that the probability that a worker at firm i receives a job offer from another firm if she leaves her job is proportional to the share of all jobs which are outside her firm, $(1 - \sigma_i)$, and the probability that she becomes unemployed if she leaves her job is proportional to her firm's share in the labor market, σ_i. In a slightly less stylized setting, one could instead represent this intuition as playing out with on-the-job search, where workers can only receive job offers from firms which are *not* their own firm: as they receive outside offers, their outside option value increases, but workers who are already at the largest employers in their labor market are less likely to receive outside offers because there are fewer vacancies outside their own firm.

Note that the outcome of the wage bargain reached by workers at firm i depends on the outcome of the wage bargain reached by workers at all other local firms j, and vice versa. To solve this "reflection problem", we iteratively substitute for w_j in wage expression (2). Since our empirics will focus on the *average* wage in the labor market, we then take the average wage across all firms in the labor market, $\bar{w} = \sum_i \sigma_i w_i$. This gives us an expression in increasing orders of employer concentration (shown in full in Appendix A). Taking a second order approximation in terms involving firm employment shares σ_j, we can reduce this complex expression to become a function of the squares of the employment shares – and, therefore, the commonly-used Herfindahl

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

Hirschmann Index ($HHI = \sum_i \sigma_i^2$):

$$\bar{w} = (1 - (1-\beta)HHI) \cdot \bar{p} + (1-\beta)HHI \cdot b - \beta(1-\beta)\sum_i \sigma_i^2 \tilde{p}_i \tag{3}$$

where $\bar{p} = \sum_i \sigma_i p_i$ is average productivity across firms, and $\tilde{p}_j = p_j - \bar{p}$ is the difference between firm j's productivity and the market average.

The first and second terms of our wage expression (3) illustrate that the wage declines as average employer concentration increases: workers are less likely to receive job offers from other firms outside their own, meaning that less weight in the wage equation falls on the option to receive competing job offers (represented by average productivity $\bar{p}$) and more weight falls on the unemployment benefit (b). The third term depends on the joint distribution of employment shares and productivity: if the correlation of market shares and firm productivity is small, it becomes small and the wage is simply a concentration- and bargaining power-weighted average of productivity p and unemployment benefit b.[14] There is also an interaction between employer concentration HHI and worker bargaining power β. The more bargaining power a worker has over the match surplus, the less the outside option matters in the wage bargain and therefore the less employer concentration matters for the wage.[15]

Note that while our framework provides us a particular lens through which to interpret the empirical results which form the main contribution of this paper, one *can interpret* our empirical analysis that follows in the context of alternate frameworks in which employer concentration affects outside options. The framework outlined in this section should be considered as providing an intuition for *why* concentration might affect wages, without constraining the interpretation of our empirical analysis to be solely within this framework.

[14] For example: If firms with higher employment shares are more productive, the average wage will rise as average productivity $\bar{p}$ is pushed up – but the third term suggests that the passthrough is less than complete because high-productivity large firms are in fewer workers' outside option, relative to a counterfactual with the same average productivity but evenly distributed across firms.

[15] Note: for simplicity the *only* way the worker can end up unemployed in this framework is they are randomly re-matched with their former employer in the job search process. Thus our framework generates a setting where the *only* reason there is a markdown of the wage from the marginal product is because there is both employer concentration and some non-zero firm bargaining power ($\beta < 1$). This is for clarity of exposition only: one could incorporate steady state unemployment, where higher employer concentration simply increases the probability of unemployment if a match breaks down.

3 Market definition and outside-occupation options

To estimate the effect of employer concentration on wages, one must define the relevant labor market. Most analysis of employer concentration defines a local labor market by occupation or industry. But any choice of *binary* labor market definition – which treats all jobs within the labor market as perfect substitutes, and all jobs outside it as irrelevant – ignores the fact that different jobs are differently valuable as outside options to workers.

In our paper, we follow the binary market definition approach in calculating employer concentration on a narrowly-defined "baseline" labor market. However, we *also* consider the outside option value of jobs outside that labor market. As our baseline labor market, we choose SOC 6-digit occupations within U.S. metropolitan areas.[16] To incorporate the outside option value of moving to jobs in other occupations into our analysis, we do two things: (1) we segment our analysis, estimating the effect of employer concentration on wages **separately by degree of outward occupational mobility**, and (2) we develop a measure of the outside option value of jobs in other local occupations (referred to in this paper as our **"outside-occupation option index"**) and use this as a control variable in our regressions to account for differential availability of local outside-occupation job options. This approach enables us to work with commonly-available data definitions and publicly-available data, and perform aggregate analysis across the large majority of the U.S. labor market – while also accounting flexibly for workers' differential mobility patterns. Note that in this paper, we focus on *local* labor markets – the set of jobs a worker can feasibly access without having to move – which we approximate as jobs within a worker's metropolitan area.[17]

[16]This is for two reasons. First, to ensure applicability to policy, we wanted to choose a baseline definition for which there is plenty of labor market data (suggesting that we choose either occupation or industry as our starting point). Second, research on human capital specificity suggests that occupations are a more accurate approximation than industries of the set of jobs open to workers (Kambourov and Manovskii, 2009; Sullivan, 2010).

[17]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. Our approach could be extended to incorporate job options outside workers' local area. Geographic mobility patterns suggest that for most workers the option to move geographically is substantially less important than the option to take a job in another occupations in their current city: only around 3% of U.S. workers move between metropolitan areas each year (according to IRS county-to-county migration data), over 80% of job applications by workers are sent to jobs within their metropolitan area Marinescu and Rathelot (2018), geographic mobility has declined over time (Molloy, Smith and Wozniak, 2011), and the proliferation of state-level occupational licensing has made geographic mobility more difficult for many workers (Johnson and Kleiner, 2020). There are, however, a subset of

In the rest of this section, we discuss our new occupational mobility data and describe the construction of our outside-occupation option index.

3.1 Measuring occupational mobility using resume data

To understand which sets of jobs comprise workers' true labor markets, we would like to see the extent to which workers move between different occupations.[18] There is, however, no existing U.S. data set on occupational mobility which has high enough granularity to study individual pairs of transitions between SOC 6-digit occupations. We therefore construct a new data set in which we can study occupational mobility patterns. To do this, we use a new proprietary data set of 16 million unique resumes with more than 80 million job observations over 2002–2018, provided by labor market analytics company Burning Glass Technologies ("BGT").[19] Resumes were sourced from a variety of BGT partners, including recruitment and staffing agencies, workforce agencies, and job boards. Since we have all data that people have listed on their resumes, we are able to observe individual workers' job histories and education up until the point where they submit their resume, effectively making it a longitudinal data set. We use this data to construct our baseline measure of occupational transitions $\pi_{o \to p}$:

$$\pi_{o \to p} = \frac{\#\text{ in occ } o \text{ in year } t \text{ observed in occ } p \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \text{ observed in a new job in year } t+1}$$
$$\approx Prob(\text{move from occ } o \text{ to occ } p | \text{leave job}) \tag{4}$$

more mobile workers – like highly educated professionals – for whom the failure to consider job options outside their own city may be more problematic.

[18]Why use occupational mobility data to identify workers' job options outside their occupations? The outside option value of a job in another occupation can be thought of as a function of both the feasibility and the desirability of a job in that occupation. Occupational transitions are a transparent, non-parametric way to capture a combination of *both* feasibility and desirability. Using mobility is, we think, better than using measures based on task/skill similarity: these cannot capture unobserved constraints that prevent moves in practice (e.g. regulation), or the desirability of moves between different occupations, and usually assume symmetry of similarity between occupation pairs (see Appendix E for a more detailed discussion).

[19]The largest publicly-available data set on which occupational transitions can be calculated, the CPS, has at least an order of magnitude fewer occupational transition observations than our data (aggregated over the same time period 2002–2018). This matters a great deal when calculating pairwise occupational transition shares between the 840 SOC 6-digit occupations: with 705,600 possible transition cells, data sets with even a few million observations are not big enough to capture many of the transition paths.

This measure seeks to approximate the **probability of a worker moving from occupation o to occupation p conditional on leaving her job**. Specifically, it captures the share of people we observe in a specific occupation o at some point in a specific year t, who are also observed in occupation p at some point in year $t+1$, expressed as a fraction of all those in occupation o in year t who are observed in a new job at some point in year $t+1$.[20] We also use this data to construct the 'occupation leave share', which approximates the **share of people who leave their occupation when they leave their job**:

$$\begin{aligned}\text{leave share}_o &= \frac{\#\text{ in occ } o \text{ in year } t \text{ \& no longer in occ } o \text{ in year } t+1}{\#\text{ in occ } o \text{ in year } t \text{ \& in a new job in year } t+1} \\ &\approx Prob(\text{leave occ } o|\text{leave job}) \end{aligned} \tag{5}$$

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

We estimate transition probabilities $\pi_{o\to p}$ for a large proportion of the possible pairs of SOC 6-digit occupations: we exclude the occupations for which we have fewer than 500 observations in the BGT data (roughly the bottom 10% of occupations), resulting in 786 origin SOC 6-digit occupations in our data. We average the observed annual occupation-to-occupation transitions over all observations in the data – across the whole of the U.S. and across all starting years 2002–2015[21] – to capture as much as possible the underlying degree of occupational similarity rather than transitory fluctuations from year to year.

The BGT resume data set is largely representative of the U.S. labor force in its

[20] We approximate the share of workers moving from occupation o to occupation p with the share of all workers observed in occupation o at any point in year t who are observed in occupation p at any point in year $t+1$ (since resumes don't always list specific dates at which workers were in a given job in a given year). We drop jobs lasting 6 months or less to exclude temporary work, summer jobs, and internships. Similarly, we approximate the share of workers in occupation o who take a new job as the share of all workers observed in a given job in occupation o at any point in year t who are observed in a different job at any point in year $t+1$. Note that our measure will capture not only sequential occupational mobility but also mobility between occupations in the form of working in jobs in two different occupations at the same time. Implicitly, we are assuming that taking up a secondary job in an occupation indicates its viability as an outside option.

[21] Where 2015 refers to the *starting* year, i.e. transitions observed between occupation o in 2015 and occupation p in 2016.

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

3.2 Descriptive evidence on occupational mobility

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

First, occupational mobility is high, and highly heterogeneous across occupations, suggesting that the narrowly-defined SOC 6-digit occupation fails to capture many workers' true labor markets (and is a worse approximation of labor markets for some occupations than for others). In our data, the average probability of a worker leaving her 6-digit occupation given that she leaves her job - the "occupation leave share" defined in expression (5) above - is 23%. The "leave share" at the 25th percentile of occupations is 19%, and at the 75th percentile at 28% (Table 1, Figure 2).[22] Second, aggregating up the SOC classification hierarchy - which groups occupations with ostensibly similar occupations - still fails to capture many transitions, suggesting that this is not sufficient to capture the full extent of workers' true labor markets. 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).[23]

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

[23]For example, only 39% of systems software developers leave their 2-digit occupation group when

Third, the occupational transition matrix is highly asymmetric, suggesting that the relevance of other occupations is not symmetric across occupation pairs (unlike in many task- and skill-based measures of occupational similarity). This partly reflects the fact workers in an occupation with specialized skills may be able to move to occupations which require generalist skills (e.g. retail salespersons) but the reverse flow is less feasible. Fourth, there are few observed transitions between most pairs of occupations – the transition matrix is sparse – suggesting that workers' relevant labor markets can be constructed from clusters of related occupations, as we do in this paper.[24]

Fifth, empirical occupational transitions reflect systematic similarities between occupations in terms of their task requirements, wages, amenities, and leadership responsibilities, suggesting that they do indeed reflect the underlying feasibility of an occupation as an outside option. To show this, we regress our measure of occupational transitions on a number of different occupational characteristics derived from the O*Net database: the vector difference in the importance scores for all "Skill" task content items (see Macaluso (2019)); task composites capturing the distinction between cognitive vs. manual, routine vs. non-routine task contents, and social skills, based on Autor, Levy and Murnane (2003) and Deming (2017); characteristics that proxy for flexibility on the job (Goldin, 2014), such as time pressure and the need for establishing and maintaining interpersonal relationships; and characteristics measuring leadership responsibilities. In every pairwise regression of occupational mobility on the absolute difference in characteristics (controlling for the difference in wages), the coefficients are significantly negative or statistically insignificant, as shown in Figure 3.[25]

3.3 A flexible approach to market definition

The analysis above suggests that the narrow SOC 6-digit occupation is not an accurate reflection of workers' true labor market, but that we can use occupational transitions to identify the cluster of occupations which *do* comprise workers' true labor markets. How can we incorporate workers' ability to switch occupations into an analysis of employer

they move across 6-digit occupations, compared to 95% of flight attendants. Note that management roles are often considered a separate 2-digit occupational group from non-management roles in the same field or specialty. If we exclude transitions to and from management occupations entirely, at the median 67% of SOC 6-digit occupational transitions still cross SOC 2-digit boundaries (see Table 1).

[24]See, for example, Appendix Figures A7, A9 and A10, and Appendix Table A5).

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

concentration? We suggest a flexible approach, which estimates the effect of employer concentration *within* workers' local occupation on wages, while (1) segmenting the analysis by degree of outward occupational mobility and (2) controlling for a measure of the outside option value of moving to jobs in other occupations.

We define our measure of the outside option value of jobs in other occupations - the **outside-occupation option index** oo^{occs} – as a weighted average of the wage in each alternative occupation, weighted by a measure of the likelihood that the worker will move to a job in each of those alternative occupations if she leaves her current job:

$$oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} Prob(\text{move to job in occ } p \text{ if leave job})_{o,k,t} \cdot \text{wage}_{p,k,t} \tag{6}$$

where the subscripts refers to the city (k), current occupation (o), possible destination occupations (p), and year (t). This measure can be conceived of as reflecting the expected value, to a given worker, of the option of moving to a job in a different occupation if she leaves her current job.[26]

To construct this index, we need an empirical analog for the probability that a worker who leaves her job in occupation o and city k will move to a job in occupation p and city k, denoted $Prob(\text{move to job in occ } p \text{ if leave job})_{o,k,t}$. We use the product of two variables: (1) the national average empirical transition share between occupation o and occupation p for workers that leave their job, $\pi_{o \to p}$, and (2) the relative employment share of occupation p in city k compared to the national average, $\frac{s_{p,k}}{s_p}$. The national occupation-to-occupation transition share is a proxy for the likelihood that, nationwide, the average worker's best job option outside her own occupation would be in each of these other occupations; the local relative employment share adjusts this for the local availability of job options in each destination occupation. Our empirical outside-occupation option index is therefore:

$$oo^{occs}_{o,k,t} = \sum_{p \neq o}^{N_{occs}} \pi_{o \to p} \cdot \frac{s_{p,k,t}}{s_{p,t}} \cdot \bar{w}_{p,k,t} \tag{7}$$

[26]Later in this section, we show that this measure can be derived from our stylized conceptual framework. We don't consider employer concentration in other occupations because the effect of employer concentration on the wage in outside-occupation options is already implicitly incorporated into our outside-occupation option index (through the local wage in those occupations).

We construct this outside-occupation option index at the annual level for as many SOC 6-digit occupations and U.S. cities over the years 1999–2016 as our data allows. We use data from the BLS Occupational Employment Statistics (OES) to obtain relative employment shares $\frac{s_{p,k,t}}{s_{p,t}}$ and average wages $\bar{w}_{p,k,t}$ by SOC 6-digit occupation, city, and year.[27] We use the Burning Glass Technologies resume data described above to construct occupational transition shares $\pi_{o\to p}$.

3.4 Conceptual framework: outside-occupation options

The outside-occupation option index defined above can be derived from the stylized conceptual framework we introduced in section 2. In this framework, the value of the outside option is a weighted average of the wage in each other local firm, where the weight was equal to the probability a job seeker would receive an offer from a given firm j (expression (2)). To incorporate workers' option to switch occupation, we delineate separate local labor markets based on individual occupations. The value of the outside option is still the weighted average of the wage in each other local firm, but the weight is this time a product of two factors: the probability that a worker from occupation o will receive *any* job offer from a firm in occupation p, which we denote $Prob(o \to p)$, and the vacancy share of firm j in occupation p, $\sigma_{j,p}$. This gives us the following wage expression for workers in firm i and occupation o:[28]

$$w_{i,o} = \beta p_{i,o} + (1-\beta)\underbrace{Prob(o \to o) \cdot \sum_{j \neq i} \sigma_{j,o} \cdot w_{j,o}}_{\text{own occupation options}} + (1-\beta)\underbrace{\sum_{p \neq o}^{N^{occs}} Prob(o \to p) \sum_{l}^{N_p} \sigma_{l,p} \cdot w_{l,p}}_{\text{outside occupation options } oo^{occs}}$$
$$+ (1-\beta)\underbrace{Prob(o \to o)\sigma_{i,o} \cdot b}_{\text{unemployment}} \tag{8}$$

[27] See summary stats of our index in Table 2. Note: we use "cities" to refer to the CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas) for which data is available in the BLS OES. Of the possible 786,335 occupation-city cells, wage data in the BLS OES only exists for approximately 115,000 each year. The missing occupations and cities are primarily the smaller ones. To create a consistent panel of occupations over time we crosswalk SOC classifications over time: see Appendix D.

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

The third term, reflecting outside occupation options, is a direct analog of our empirical outside-occupation option index under two assumptions. First, approximate local vacancy shares $\sigma_{l,p}$ in each occupation p with each firm's current employment share in occupation p. Then, $\sum_{l}^{N_p} \sigma_{l,p} \cdot w_{l,p}$ simply becomes the average wage in local occupation p.[29] Second, approximate the probability that a worker from occupation o will receive a job offer from a firm in occupation p with $\pi_{o \to p} \cdot \frac{s_{p,k,t}}{s_{p,t}}$, the product of the national occupational transition shares from occupation o to occupation p and the local relative availability of jobs in occupation p. Iteratively substituting for wages in other local employers j, taking an average of the local wage in occupation o, and taking a second order approximation in employer shares (as before) gives us the following expression for the average wage in occupation o:

$$\bar{w}_o = (1 - (1-\beta)Prob(o \to o)HHI_o)\left(\alpha \bar{p}_o + (1-\alpha)oo^{occs}\right) + (1-\beta)Prob(o \to o)HHI \cdot b - \beta(1-\beta)Prob(o \to o)\sum_i \sigma_i^2 \tilde{p}_{i,o} \qquad (9)$$

where $\alpha = \frac{\beta}{1 - Prob(o \to o)(1-\beta)}$.

Ignoring the final term, which is small if the average productivity of individual firms is not strongly correlated with their employment shares, this expression suggests that the average wage in occupation o is a weighted average of the average productivity in occupation o ($\bar{p}_o$), the value of jobs outside occupation o (oo_o^{occs}), and unemployment benefit (b). As before, employer concentration within workers' own occupation reduces their likelihood of receiving outside job offers from other firms, reducing the value of other jobs in the wage bargain and increases the weighting on b the unemployment benefit. In addition, there is now an interaction with $Prob(o \to o)$: the more likely a worker is to stay in her own occupation if she leaves her job (i.e. the less likely to be able to find a job in a different occupation), the more employer concentration in her own occupation matters for her wage.

This expression illustrates two problems that improper market definition could cause when estimating the effect of concentration on wages. First, since outward occupational mobility ($1 - Prob(o \to o)$) varies across occupations, one should expect

[29]Since we intend to study the effect of within-occupation outside options (proxied by employer concentration) and outside-occupation options separately, this eliminates the need for us to consider the reflection problem that wages in occupation o affect wages in occupation p and vice versa - instead, we can simply use data on the average wage in each local occupation p to construct our measure of the value of outside-occupation options.

the effect of within-occupation employer concentration on wages to be different for low-mobility vs. high-mobility occupations. Second, if the degree of employer concentration within a local occupation (HHI) is correlated with the quality of outside options *outside* the local occupation (oo^{occs}), then estimation of the effect of concentration on the wage may be biased without controlling for outside-occupation options.

4 Empirical Approach

Our analysis thus far suggests four testable predictions for the relationship between employer concentration, outside-occupation options, and wages. First, higher employer concentration (HHI) reduces wages. Second, better outside-occupation options increase wages. Third, the wage-HHI relationship should be stronger for occupations where workers have little ability to get jobs in other occupations. Fourth, the empirical relationship between wages and HHI may be biased if concentration within an occupation is correlated with the availability of outside-occupation job options. In this section, we lay out our empirical approach, which aims to test these predictions and to quantify as precisely as possible the magnitude of the effect of employer concentration on wages.

Before we do this, it is important to be clear about what question we are asking, when we ask "what is the effect of employer concentration on wages?" One might be asking "what is the effect of a particular labor market having become more concentrated, relative to the past?". The answer to this question would consider the effects of a given change in labor market structure on *both* outside options *and* average productivity. It is quite possible that a merger – for example – might increase productivity as well as concentration. On the other hand, one might be asking "what is the effect of employer concentration in a given labor market, in terms of reducing workers' wages relative to their productivity?". In that case, the answer to this question takes the degree of productivity in the labor market *as given* and looks only at the effect of employer concentration in reducing wages below that level of productivity by worsening outside options: isolating the effect of employer concentration on outside options. We focus on the latter question in this paper.

4.1 Measuring employer concentration

We follow Azar et al. (2020a); Hershbein et al. (2019) in using Burning Glass Technologies' ("BGT") database of online vacancy postings to measure employer concentration, calculating the Herfindahl-Hirschman Index (HHI) of the share of vacancy postings from each employer at the level of individual SOC 6-digit occupations within metropolitan areas, in each year 2013–2016. The BGT vacancy posting data covers the near-universe of online job postings, drawn from over 40,000 distinct online sources including company websites and online job boards, with no more than 5% of the vacancies coming from any one source (Hazell and Taska, 2019).[30]

We calculate the vacancy HHI at the level of each SOC 6-digit occupation by metropolitan area by year, as the sum of the squared vacancy posting shares for each unique employer:[31]

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

where $v_{i,o,k,t}$ denotes the number of vacancy postings from employer i in occupation o and metropolitan area k, during year t.

We construct our HHI indices from vacancy posting data rather than employment data for two reasons. First, pragmatically, there is no U.S. data which enables the calculation of firm-level employment by occupation and metropolitan area (which means we would not be able to estimate our instrumental variable for concentration). Second, one could argue that the vacancy posting HHI measure is more appropriate, given our focus on employer concentration's effect on outside options: if concentration reduces the wage by reducing the value of workers' outside options, vacancies – over the period for which the wage bargain will last – reflect the jobs workers could feasibly apply for.[32]

[30] Each vacancy observation contains the job title, company name, location, date, and various details about the job description. Using proprietary parsing technology, Burning Glass then imputes the relevant SOC 6-digit occupation code for each job posting. More details on the process by which BGT obtains, parses, and deduplicates this data can be found in Carnevale, Jayasundera and Repnikov (2014) and in Appendix B.

[31] Identifying an employer is not always simple: we largely group jobs by employer name, which means we may miss patterns of common ownership across employers with different names, but also means we count different establishments or franchises with the same name as the same employer.

[32] On the other hand, one could argue that the stock of employment is a better indicator of the underlying employment opportunities in a particular labor market than the flow of vacancies (and so are better suited for understanding the general equilibrium effects of employer concentration on wages)

Since the Burning Glass Technologies vacancy data covers the near-universe of online job postings, it is relatively representative of the vacancies which are advertised online. Of course, not all vacancies are posted online. Azar et al. (2020a) estimate that in 2016, the BGT vacancy database captured around 85% of all job vacancies both online and offline (as measured from the Help Wanted Online database). There are two important reasons, however, why one might have concerns about estimates of employer concentration based on online vacancy posting data. First, for occupations where a large share of jobs are advertised offline or informally, online vacancy posting data will capture only a small share of total vacancies. Second, for occupations where firms tend to hire many workers for each posted vacancy, our data will underestimate the degree of employer concentration as for a given number of vacancy postings, one might expect large employers to hire more workers per job posting than small employers.[33]

To understand the degree to which each of these might be an issue, we run representativeness checks by occupation, calculating the share of each SOC 6-digit occupation in the BGT vacancy database relative to the share of each occupation in total employment (as per the BLS OES). By this metric, occupations which are particularly underrepresented in our data include low-wage food service jobs, cleaners, home health aides, laborers, and cashiers. In our estimates of the effect of employer concentration on wages, we carry out a number of sensitivity checks based on representativeness of the vacancy data. We also control for occupation-by-year fixed effects, which should assuage concerns about the relative representativeness of the data for different occupations; however, when drawing conclusions about the degree of employer concentration affecting specific occupations, it will be important to bear in mind that our database may tend to disproportionately overestimate employer concentration for the occupa-

since vacancy posting decisions are endogenous to labor market conditions. In practice, the distinction between vacancy and employment HHIs may not matter much as, in equilibrium, one would expect the two to be highly correlated. Indeed, Marinescu et al. (2019) show that an HHI of employment flows – conceptually similar to posted vacancies, as these reflect filled vacancies – is highly positively correlated with the HHI of employment stocks in French administrative data.

[33]If there is a big difference in the ratio of hires per job posting between large and small firms, 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. 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. Ideally, we would have been able to calculate employer concentration at the level of true job openings/vacancies, rather than vacancy postings, but we are not aware of a data set that enables us to observe firm-level local vacancies by occupation in the U.S.

tions which are underrepresented in the data (and when singling out specific occupations we exclude those which are heavily underrepresented in the BGT vacancy data). For further discussion of the BGT vacancy data, see Appendix B.

4.2 Regression specification

Our baseline specification regresses the log of the average wage in a given occupation, metropolitan area, and year, on the log of the Herfindahl-Hirschman index of employer concentration within that occupation, metropolitan area, and year ($HHI_{o,k,t}$), the log of our outside-occupation option index ($oo^{occs}_{o,k,t}$), which measures the expected value of local jobs outside workers' own occupation, and a set of occupation-by-year and city-by-year fixed effects:

$$\ln \bar{w}_{o,k,t} = \alpha + \alpha_{o,t} + \alpha_{k,t} + \gamma_1 \ln HHI_{o,k,t} + \gamma_2 \ln oo^{occs}_{o,k,t} + \xi_{o,k,t} \tag{10}$$

where $\alpha_{o,t}$ and $\alpha_{k,t}$ are occupation-year and city-year fixed effects, and $\xi_{o,k,t}$ is the local occupational wage residual.

We also re-run the same regression, but allowing the coefficient on the HHI and outside-occupation job options to vary according to the occupation's degree of outward mobility ("leave share"), interacting $HHI_{o,k,t}$ and $oo^{occs}_{o,k,t}$ with an indicator variable for the applicable quartile of outward mobility of occupation o. In this specification, we therefore estimate different coefficients γ_1 and γ_2 for each quartile of outward occupational mobility.

We run these regressions across the largest possible subset of U.S. occupation-city-year cells for which we can obtain all our data: Our full data set for our baseline regressions over 2013–2016 comprises 212,417 occupation-city-year observations.[34] We use BLS OES data for average occupational wages by city and year for the dependent variable $\bar{w}_{o,k,t}$. As discussed earlier, we use BGT vacancy posting data to construct the HHI and we use a combination of wage and employment data from the BLS OES, and occupational mobility data constructed from the BGT resume data.

[34]This includes 387 cities and 715 SOC 6-digit occupations, with 94,809 occupation-city labor markets represented in at least one of the four years from 2013–2016. We have data on the wage, HHI, and outside-occupation option index (*but not* the instruments) for a larger set of occupation-city-year labor markets. We calculate summary statistics and counterfactuals on this larger set.

Note that our estimation of the effect of employer concentration on wages is to some extent model-agnostic: our results could be used to inform analyses based either on labor market models based on bargaining, or those where employer concentration and market shares affect the wage via the elasticity of labor supply to individual firms.[35] However, the construction of our outside-occupation option index is informed specifically by a bargaining model of the labor market.

4.3 Identification: employer concentration

Endogeneity issues may bias the coefficients on our HHI concentration measure. The direction of the bias is ambiguous: an increase in employer concentration could reflect the expansion of a highly productive large firm, which would result in higher employer concentration (expected to reduce wages) but also an increase in average productivity (expected to increase wages). On the other hand an increase in employer concentration could reflect a lack of local dynamism, with few new firms being created, which would lead to an increase in employer concentration and worsening of outside options alongside a general decrease in productivity.[36] We therefore instrument for local labor market concentration, creating an instrumental variable which leverages differential local occupation-level exposure to national firm-level hiring, in a strategy which builds on both the "granular" instrumental variable approach (GIV) of Gabaix and Koijen (2020) (which uses plausibly exogenous idiosyncratic firm-level variation to instrument for changes in market-level aggregates), and on the shift-share 'Bartik' approach.

First note that, mechanically, the growth in the local employer concentration of occupation o will be a function of the growth in local occupational employment for

[35] A bargaining framework, however, has different implications from a monopsonistic framework for employment outcomes: in the former case, changes in employer concentration affect only the distribution of surplus between workers and firms, whereas in the latter case, changes in employer concentration also affect firms' hiring decisions.

[36] Concerns like these are raised in many of the critiques of the empirical literature which finds a negative correlation between local employer concentration and wages, including Berry et al. (2019) and Rose (2019). Rose (2019) argues that empirical strategies attempting to identify a causal effect of employer concentration on wages must isolate the effect of employer concentration from changes in labor demand; our identification strategy attempts to do this. Hsieh and Rossi-Hansberg (2019) show that over recent decades there has been a substantial pattern of expansion of national productive firms into local labor market, reducing local employer concentration and likely increasing local productivity.

each employer j, $g_{j,o,k,t}$ (leaving aside firm entry):

$$\begin{aligned}\Delta HHI_{o,k,t} &= \sum_j \sigma^2_{j,o,k,t} - \sum_j \sigma^2_{j,o,k,t-1} \\ &= \sum_j \sigma^2_{j,o,k,t-1}\left(\frac{(1+g_{j,o,k,t})^2}{(1+g_{o,k,t})^2} - 1\right) \end{aligned} \tag{11}$$

That is, 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 $g_{o,k,t}$.

Our instrumental variable strategy uses this insight, using plausibly exogenous *firm-level* changes in vacancies as instruments for the overall change in local labor market concentration. For large firms operating in many local labor markets, the leave-one-out firm-level national growth in vacancies in occupation o (which we denote $\tilde{g}_{j,o,t}$) is likely to be correlated with growth in the firm's local employment in occupation o, but uncorrelated with determinants of occupation-specific productivity growth in any given local labor market k. Under this assumption, $\tilde{g}_{j,o,t}$ is a valid instrument for $g_{j,o,k,t}$. We therefore construct our instrument as:

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

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 local firm j.[37] Through the lens of shift-share instruments, our instrument features plausibly exogenous 'shocks' (a function of the leave-one-out occupational hiring growth of local firms j), and possibly endogenous exposure 'shares' (the last-period local occupational vacancy shares of each of those firms j), following the terminology and approach of Borusyak, Hull and Jaravel (2018).

One concern with this instrumental variable is that differential local exposure to national firms' growth may differentially affect total labor demand, not just employer concentration. In our model, the effect of a large firm's growth on local labor market

[37]Note that by taking the log of the instrument, we implicitly exclude observations where the predicted change in HHI based on national firm-level growth is negative. Note also that we are instrumenting for the local level of the HHI with an instrument derived from an expression for the change in the HHI.

concentration is quadratic, whereas the effect of a large firm's growth on local labor demand or productivity is linear. Thus, we control for (1) the growth rate of local vacancies in the occupation-city labor market ($g_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} g_{j,o,k,t}$), and (2) the predicted growth rate of local vacancies based on large firms' national growth (i.e. the direct linear analog to our concentration index: $\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} \tilde{g}_{j,o,t}$ as defined above). With these controls, we should be estimating the effect of a change in local labor market concentration due to changes in large firms' employment, holding constant any direct linear effect on local labor demand or productivity.[38] A second concern, raised in work on shift-share IVs by Borusyak et al. (2018), is that since our exposure 'shares' (the sum of squared employer shares) do not sum to one, a bias may be introduced by differential local occupation-level exposure to *any* possible hiring shocks. As such, following Borusyak et al. (2018) we also introduce a control we call the "exposure control": the sum of the squared vacancy shares of local firms j which have *any* growth in other metro areas $\sum_j \sigma^2_{j,o,k,t-1} \cdot \mathbb{1}[\tilde{g}_{j,o,t} \neq 0]$. This controls for the fact that different local occupations may have different initial shares of employment accounted for by large national firms. We elaborate on conditions for identification in Appendix G.[39]

Our approach is a novel instrument for local labor market concentration. Our instrument exploits the facts that (a) increases in local labor market concentration are often driven by specific, already-large, firms growing, (b) these large firms usually operate across many labor markets, (c) different local labor markets are differentially exposed to employer concentration arising from the growth of these large national firms, based on the initial share of employment with each firm, and (d) the employment growth of these large firms nationally is likely to be orthogonal to productivity changes or other conditions in a specific local occupational labor market. The intuition behind the instrument can be explained with a hypothetical example of insurance sales agents in Bloomington, Illinois and in Amarillo, Texas (two metropolitan areas of relatively similar size). In each city, there are several insurance companies who employ

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

[39]In robustness checks in Appendix Tables A7 and A9, we exclude the exposure control in column (c), exclude the controls for actual and predicted vacancy growth in column (a), and in column (b) also include a further control the equal-weighted growth rate of local firm-level vacancies $\frac{1}{N}\sum_j g_{j,o,k,t}$, following the Granular Instrumental Variable approach of Gabaix and Koijen (2020). This controls for any local increase in hiring which is common to all firms – i.e. conceivably caused by a common local demand shock.

insurance sales agents. Assume that in Bloomington, State Farm has a large share of local insurance sales agent employment (as Bloomington is their headquarters), while in Amarillo employment is more dispersed amongst a number of insurance companies. In years where State Farm grows substantially faster than other major insurance companies nationwide, under most assumptions about how that growth is allocated across occupation-city cells, employer concentration of insurance sales agents will grow by more in Bloomington IL than in Amarillo TX.

Our identification strategy has two weaknesses that should be considered. First is the assumption that, when large national firms grow, the effect of this growth on local occupational wages can be separated into a *linear* demand effect on vacancies (which should push wages up) and a *quadratic* employer concentration component (which should push wages down).[40] This concern may be partly assuaged by the fact that the introduction of various controls for changing local occupational demand does not seem to affect our baseline coefficient estimates much (as shown in Appendix Table A7). A secondary weakness is that, by construction, our strategy is unlikely to be a strong instrument for small changes in employer concentration in initially unconcentrated labor markets – if no firm makes up more than a trivial portion of local employment, even a large growth rate in that firm's hiring will not do much to change local employer concentration. To assuage this concern, we focus on applying our estimates of the effect of employer concentration on wages to local labor markets with above-median employer concentration.

We nonetheless see our approach as a step forward on the problem of estimating the effect of employer concentration on wages. Some recent empirical work on labor market concentration has instrumented for changes in concentration in a given occupation and city with changes in (the inverse of) the number of employers in the same occupation in other cities (e.g. Azar et al. (2020a,b)). This identification strategy is able to circumvent some endogeneity concerns, but remains subject to the concern

[40]The most plausible threat to our identification is that the growth of a large national firm, locally, pushes up wages by more in areas with inelastic labor supply to the local occupation than in areas with elastic labor supply to the local occupation, and that the elasticity of labor supply to the local occupation is correlated with the initial local employment share of the large national firm in that occupation. Specifically: if occupation-city labor markets with higher employment shares of the national firms which are growing over the period have more elastic labor supply to their local occupation, then it may appear as if shocks to local employer concentration driven by the growth of large national firms suppress wages when in fact what is happening is that areas with higher exposure to the large national firms which are growing see more muted wage responses to increases in labor demand than areas with lower exposure to the large national firms which are growing.

that national occupation trends in concentration may be correlated with unobservable national trends in occupational productivity, demand, or supply, which could be confounding the measured wage effects.[41] In contrast, our strategy allows us to control for national occupational trends because we are exploiting across-city variation in concentration due to differences in occupation-city labor markets' exposure to firm-level growth. Another recent approach is that of Arnold (2020) who uses cross-sectional variation in local exposure to M&A activity to generate plausibly exogenous variation in local labor market concentration. This approach is able to circumvent endogeneity concerns about the *cause* of the change in employer concentration, but cannot isolate the effects of employer concentration from other effects on local labor demand or employment caused by the M&A activity. Our approach is similar in using differential local exposure to large national firms' activity, but, by using our identification approach, we are not restricted only to studying M&A activity (which accounts for less than 2% of changes in local labor market concentration (Arnold, 2020)), we are able to examine the effects of employer concentration on wages across broad swathes of the U.S. labor market, and we are able to control at least somewhat for effects on local labor demand.[42] Ultimately, we believe that this set of different identification approaches – based off different variation, and with different strengths and weaknesses – can together provide a useful picture of the causal effects of employer concentration on wages.

4.4 Identification: outside-occupation options

Alongside the effects of (within-occupation) employer concentration, we also study the extent to which changes in the quality of outside-occupation job options affect wages. This is useful not only because it is interesting in and of itself, but also because finding large effects of outside-occupation job options would suggest that jobs outside workers' own occupation are relevant parts of their labor market, and therefore that an approach to labor market definition which ignores outside-occupation job options may be misleading in important ways.

However, we cannot simply regress local occupational wages on our outside-occupation

[41] Though the authors are able to address some of these concerns by controlling directly for variables like local occupational labor market tightness.

[42] A further complementary approach is that of Dodini et al. (2020), which demonstrates that workers laid-off in mass layoffs see larger wage losses in more concentrated labor markets in Norway.

index. Endogeneity issues may bias the coefficients on our outside-occupation option measure upwards: in a year where a worker's city experiences a positive demand shock for an occupation similar to her own, there may also be 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 therefore 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 instrument for local wages in each outside option occupation with plausibly exogenous national demand shocks to that occupation. Specifically, to instrument for wages in each outside option occupation p in city k, we use the leave-one-out national mean wage for occupation p, excluding the wage for occupation p in city k. To avoid endogeneity concerns over the local employment shares, we also instrument for the local relative employment share in each occupation, using the initial employment share in that occupation in 1999, the first year for which we have data.[43] Our instrument for the oo^{occs} index, $oo^{occs,inst}$, is therefore 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 city, $\frac{s_{p,k,1999}}{s_{p,1999}}$, and the national occupation transition shares from the worker's occupation o to each other occupation p, $\pi_{o\to p}$:

$$oo^{occs,inst}_{o,k,t} = \sum_{p}^{N_{occs}} \left(\pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\not{k},t} \right) \tag{12}$$

The identifying variation within a given occupation across different cities comes from differences in each city's initial exposure to outside option occupations. Identifying variation over time within the same occupation-city cell comes from national (leave-one-out) changes over time in wages of local outside-option occupations. That is, in a year when there is a national wage shock to one of occupation o's outside option occupations p, cities which had a higher proportion of their jobs in occupation p in 1999 should see bigger increases in the wage of occupation o (because they were more

[43] Or we use the first year the occupation-city cell is in the data, if it is not present in 1999.

exposed to the shock to their outside options).[44]

In Appendix G, we provide more details on the conditions required for our instrument to identify the effect of outside-occupation options on wages, using the approach in Borusyak et al. (2018). Intuitively, the assumptions require that the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p is correlated with the local wage of occupation p in location k (relevance condition), but does not affect the local wage in initial occupation o through a direct channel other than increasing the quality of local outside options $oo^{occs,inst}_{o,k,t}$. Note that this only needs to hold *conditional* on controlling for fixed effects which, in our main specification, include both wage effects common to all workers in occupation o itself nationwide in a given year, and wage effects common to all occupations in city k within a given year. The inclusion of these fixed effects means that differences in city-level trends or national productivity of different occupations do not represent an issue for our identification strategy.[45]

5 Empirical Results

Our specification allows us to test four empirical predictions about the effect of labor market concentration and outside-occupation options on wages: (1) that higher labor market concentration reduces wages, (2) that better outside-occupation options raise wages, (3) that the effect of labor market concentration is greater for occupations with less outward mobility, and (4) that regressions of wages on labor market concentration will be biased if they do not consider outside-occupation job options. Our empirical analysis in this section confirms all four predictions, and allows us to estimate with some precision the average effect of employer concentration on wages for occupations with different degrees of outward mobility.

[44] This instrumental variable strategy is closely related to that of Beaudry et al. (2012), who argue that they are able to avoid endogeneity and reflection problems in their index of cities' industrial composition by using national industry wage premia to substitute for city-level industry wages.

[45] 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 cities *within* occupations. An additional concern may be that groups of local occupations that share similar labor markets experience similar location-specific industry shocks. We show that our results are robust to controlling for common exposure to shocks to local industries (see Appendix Table A7).

5.1 Effect of employer concentration on wages

We start by exploring the correlation between log vacancy HHIs and log wages at the occupation-city level (illustrated in Figure 4). In the cross-section, occupation-city labor markets with high employer concentration tend to have low wages. This pattern also holds within cities (for a given city, workers in occupations with higher HHIs tend to have lower wages) and within occupations (for a given occupation, workers in cities with higher HHIs tend to have lower wages). Indeed, even when controlling for *both* occupation-by-year and city-by-year fixed effects, there is a negative relationship between the log of the employer HHI and log wages.[46]

When instrumenting for the HHI using our IV strategy outlined in section 4, the magnitude of the coefficient on the log HHI increases by around 40% relative to the OLS specification (Table 3, column (c)). This suggests some combination of omitted variable bias or measurement error biasing the coefficient toward zero in simple OLS regressions of wages on HHI.[47] We also introduce a control for our outside-occupation option index (Table 3, columns (b) and (d) for OLS and 2SLS respectively). The coefficient on the instrumented outside-occupation option index is positive and highly statistically significant, confirming that outside-occupation options matter for wages, and therefore that jobs outside workers' own occupation are relevant parts of workers' labor market. After introducing the outside-occupation option index the coefficient on the HHI falls by about a third in the OLS regressions and about a fifth in the IV regressions, consistent with omitted variable bias. This is because the vacancy HHI is negatively correlated with workers' outside-occupation options: workers with worse options *within* their occupation also have worse options *outside* their occupation (as

[46]The magnitude of the relationship between wages and concentration in our data, with a coefficient of -0.010 (Table 3 column (a)), is similar to that found in Hershbein et al. (2019) who estimate a coefficient of -0.014 when regressing wages on employer vacancy concentration at the level of SOC 6-digit occupation by metropolitan area with occupation, metropolitan area, and year fixed effects over 2010–2017. The estimates in other papers are not directly comparable: Rinz (2018) uses employer concentration measured at the 3- or 4-digit industry level, Azar et al. (2020b) use a measure of wages based on posted wages, and Azar et al. (2020a) estimate wage-concentration relationships for national industry averages.

[47]The first stage of this IV regression is shown in Table 5, column (a). Note that our ex ante expectation of the direction of the omitted variable bias was unclear: it was possible either that positive local productivity or demand shocks would increase both concentration and wages (biasing the coefficient in OLS regressions toward zero), or that negative local productivity or demand shocks would increase concentration but decrease wages (biasing the coefficient in OLS regressions away from zero).

illustrated in Appendix Figure A13).

How big is the average effect of employer concentration on wages? Our baseline coefficient estimates in Table 3 column (d) – when instrumenting for employer concentration and controlling for the value of outside-occupation job options – suggest that going from the HHI faced by the median worker to the HHI faced by the worker at the 95th percentile (from an HHI of 221 to 2,256) would be associated with a 2.6 log points lower hourly wage.[48] This is in the same ballpark as, but at the low end of, the existing estimates reviewed in Marinescu and Hovenkamp (2019). Reviewing existing evidence, they suggest that a 10% increase in employer concentration (at the SOC 6-digit occupation by commuting zone level) leads to a 0.3% to 1.3% decrease in wages. Our point estimates suggest a 10% increase in concentration (at the SOC 6-digit occupation by MSA level) leads to a 0.1% decrease in wages on average. However, this average effect conceals substantial heterogeneity as we discuss in the next section.

Heterogeneity by occupational outward mobility

We now re-run our baseline regression, but allowing the coefficients on the HHI and the outside-occupation option index to vary for occupations with different degrees of outward occupational mobility (segmenting our data into four quartiles of occupations' average "leave share" calculated from our BGT resume data). Why? Because, as outlined in section 3, different occupations have very different degrees of outward mobility – the likelihood that a teller leaves her occupation when she leaves her job is around four times greater than the likelihood that a nurse leaves her occupation when she leaves her job – suggesting that within-occupation employer concentration likely matters a lot more for workers in occupations with low outward mobility than high outward mobility.

As predicted, we find substantial heterogeneity (Table 4). For occupations in the lowest quartile of outward mobility, the effect of within-occupation employer concentration on wages is more than twice as large as the average effect. In contrast, we find no statistically significant effect (and a zero point estimate) for the occupations in the highest quartile of outward mobility. These results are visualized in Figure 5. For

[48]In Appendix Tables A12 and A13, we explore whether there is evidence for heterogeneity of the effect of concentration on wages for different quartiles of the wage distribution, or for different occupation groups, but do not find patterns which are clearly statistically significantly different across these groups.

the quartile of occupations with the lowest outward mobility, an increase in the HHI from the median to the 95th percentile as faced by workers would be associated with 5.8 log points lower wage, with the confidence interval suggesting an effect of between 3.7 and 8.0 log points. In contrast, for the quartile of occupations with the highest outward mobility, the point estimate is zero, and the confidence interval suggests that an increase in the HHI from the median to the 95th percentile as faced by workers would be associated with at most a 1.7 log point lower wage.

What are the implications of this heterogeneity for our analysis of labor markets – and for policy decisions which involve employer concentration? Our empirical results show that the binary market definition approach, by failing to consider workers' job options outside their occupation (or industry, or firm cluster), obscures substantial heterogeneity in the wage-concentration relationship and biases the size of the estimated effect of concentration on wages. Considered alone, the HHI defined at the level of a 6-digit SOC occupation is not therefore a sufficient indicator of the effects of employer concentration in any given labor market.[49]

Heterogeneity by level of HHI

One might expect that effects of employer concentration would be non-linear: a given percentage change in concentration from an already-high starting point might affect wages differently than a similar proportional change from a lower starting point. We do not, however, find evidence of different effects of concentration on wages when estimating the effect separately by quartile of initial employer HHI (Appendix Table A11). While we do not find evidence for nonlinearity, we emphasize that there is

[49] Our results are very consistent with the recent analysis of the effect of hospital mergers on the wages of hospital workers by Prager and Schmitt (2019). They find that mergers which induce large increases in local hospital concentration reduce the wages of nursing and pharmacy workers substantially relative to a no-merger counterfactual, have a somewhat suppressive effect on the wages of non-medical professionals in hospitals (working in employee benefits, administrative, human resource, and social service functions), but have no detectable effect on the wages of the remainder of hospital workers (in maintenance and repairs, operations, housekeeping, catering, and medical records). They interpret these differential results as reflecting the degree to which workers have industry-specific skills. Our estimates also suggest that nursing and pharmacy workers are some of the most affected by employer concentration across all U.S. occupation-city labor markets, since they are frequently in high-concentration occupation-city labor markets and have low outward occupational mobility (reflecting the limited availability of comparably good jobs outside their occupation). In contrast, workers in maintenance, housekeeping, and catering in our data tend to have very high outward occupational mobility (as well as tending to be employed across many industries), which according to our estimates would suggest a minimal effect of changes in employer concentration on wages.

reason to believe that we should not apply our baseline coefficient estimates to very low-concentration labor markets. Our instrument for the vacancy HHI in a given occupation-city cell relies on large firms' employment decisions having a significant effect on the local HHI. Note that – as outlined in section 4 – this is *only* likely to be a realistic assumption in labor markets that are already at least somewhat concentrated. If a labor market is made up of many tiny employers, no single employers' growth – even with a large growth rate from year-to-year – will make an appreciable difference to local employer concentration. This logic suggests that our instrumental variables strategy may not be appropriate to identify the effects of employer concentration in labor markets with extremely low levels of initial employer concentration.[50] We therefore caution against applying our effect size estimates to occupation-city labor markets with very low HHIs. Partly for this reason, when trying to quantify the effects of employer concentration on wages later in this paper, we will consider only the effect of reducing HHIs to the *median* as experienced by workers (roughly an HHI of 200).

Robustness checks: employer concentration

We explore a number of additional variations on our baseline analyses. The results are presented in Figure 6 and in Appendix Tables A7-A10.

First, we adjust the control variables used. In our baseline specification we control for local actual and predicted vacancy growth, as well as the "exposure control" for our HHI instrument (the sum of squared employer shares of *any* firm with hiring outside the metro area). In robustness checks in Tables A7 and A9, we show that our results are similar if we remove the controls for vacancy growth (column (a)), if we remove the exposure control (column (c)), or if we follow Gabaix and Koijen (2020) in adding an additional control for the equal-weighted vacancy growth of local firms, which could be construed to reflect common local occupation-specific shocks (column (b)).

Second, we control for an industry Bartik shock to alleviate concerns that local occupation-specific employment shocks may be driven by common exposure to national

[50]Figure A15 illustrates the correlation between our (log) HHI instrument and the (log) HHI for occupation-city labor markets in 2016: as can be seen, for occupation-city cells with a very low value of the HHI or of our HHI instrument, the linear relationship somewhat breaks down. A more formal analysis bears this out: regressing wages only on occupation-city cells with very low HHIs results in a much weaker instrument (with, for example, a Kleibergen-Paap F-statistic of 14 for occupation-city cells with HHIs less than 50) and no significant relationship between wages and HHI.

industry trends (Tables A7 and A9, column (d)). We discuss this industry Bartik more in our discussion of robustness checks for the outside-occupation option index.

Third, we re-run our baseline regressions with fixed effects for occupation-by-city and year (rather than occupation-by-year and city-by-year), with results shown in column (e) of Tables A7 and A9. In this specification, the identifying variation comes from year-to-year changes in employer concentration within the same occupation-city labor market over our four year period 2013–2016. In the full sample, we still see a statistically significant negative point estimate of the same order of magnitude as our baseline estimates.[51]

Fourth, we show that our results on the relationship between labor market concentration and wages are qualitatively similar – and in fact, the baseline coefficient estimates are larger – if we weight each occupation-city cell by employment or log employment (Tables A8 and A10, columns (a) and (b)). This assuages any concern that our results may have been solely driven by the dynamics of smaller occupations or cities which are not representative of the typical worker's experience.

Fifth, one might be concerned that our results are biased since our vacancy data from Burning Glass Technologies is differently representative for different occupations. For occupations which are substantially underrepresented in the BGT vacancy data, it is likely that we will overestimate employer concentration.[52] In column (c) of Tables A8 and A10, we show our baseline regression estimates *excluding* any occupations which are substantially underestimated in our BGT vacancy data (from which we calculate our employer concentration HHI).[53] The coefficient estimates are almost identical to those in our baseline regressions. We take an alternative approach in columns (d) and (e) of Tables A8 and A10 – rather than excluding occupations which are un-

[51]When breaking down the analysis by quartile of outward occupational mobility, the estimates are very noisy, which shoud not necessarily be surprising given the fact that employer concentration tends to change only relatively minimally across occupation-city labor markets in a short period.

[52]If our vacancies are sampled proportionately from every employer recruiting in these occupations, we will estimate employer concentration accurately. However, it seems plausible that the vacancies we are missing for heavily underrepresented occupations disproportionately come from small firms or households, or from self-employment, so adding these vacancies to the overall labor market would tend to reduce employer concentration in these occupations.

[53]We exclude occupations with a 'represented-ness' less than 0.5, where represented-ness by occupation is calculated as the share of all vacancies accounted for by a given occupation in the BGT vacancy data in a given year, divided by the share of employment accounted for by a given occupation in the BLS OES in that same year, averaged over 2013–2016. About one third of occupations in our data have occupation represented-ness of 0.5 or less in the BGT data.

derrepresented, we weight each occupation or metropolitan area respectively by its represented-ness in the BGT data. We still find large, negative, and significant effects of employer concentration on wages in these re-weighted regressions.

5.2 Effect of outside-occupation options on wages

We can also use our baseline regressions to ask: how big are the effects of outside-occupation options on wages? Our baseline coefficient estimate (Table 3 column (d)) suggests that a 10 log point higher outside-occupation option index is associated with roughly 1 log point higher wages in workers' own occupation.[54] This implies that moving from the 25th to the 75th percentile value of outside-occupation options across cities for the median occupation leads to 3.7 log points higher wages.[55] This estimated effect is quite large in the context of the geographic variation of occupational wages: for the median occupation, the interquartile range of average wages across cities within 2016 was 21 log points.

In addition, finding a large, significant, and positive effect of shocks to outside-occupation options on wages reinforces our conclusions that workers' true labor markets are broader than their narrow 6-digit SOC occupations, and that our "probabilistic" method of identifying relevant outside options can capture workers' true labor markets relatively well.

These estimates can be applied to explain patterns of pay variation for specific occupations in specific cities. Consider, for example, Baltimore, MD, and Houston, TX. They are a relatively similar size, with a very similar average hourly wage, but statisticians in Baltimore earned 26% more than statisticians in Houston in 2016. This may partly be explained by differential productivity of the occupations in the two cities – but the two cities also have very different values of our outside-occupation option

[54] Note that our outside-occupation option index include transitions to *all* SOC 6-digit occupations for which we have sufficient data, including management occupations. Transitions into management occupations may represent promotion rather than lateral moves, so may be less likely to represent relevant outside options. Our inclusion of management occupations likely attenuates our results.

[55] To calculate this statistic, we estimate the interquartile range of the outside-occupation option index for each occupation across cities in 2016, and apply our coefficient estimate of 0.95 to the median interquartile range estimate. Performing the same exercise with variation in employer concentration, we find that for the median occupation, moving from the 25th to the 75th percentile of employer concentration across cities, for the median occupation, would result in a wage increase of 1.4 log points.

index. Applying our baseline coefficient estimate on the effect of outside-occupation job options on wages would suggest that around 20% of this gap may be attributable to differential availability of outside-occupation job options.[56]

Robustness checks: outside-occupation options

There may be concerns that the regression coefficients on our instrumented outside-occupation option index are biased by differential geographic exposure to industry shocks which affect both a workers' own occupation and her outside option occupations.[57] To control for this possibility we construct a shift-share "industry Bartik" shock for each occupation-city-year cell, where the exposure of occupation o in city k to each industry ι is defined as the employment share of industry ι in occupation o nationwide, multiplied by the employment share of industry ι in city k. The industry Bartik shock has a significantly positive effect on wages, but our coefficients on the outside-occupation option index are only slightly attenuated relative to our baseline (Appendix Tables A7 and A9, column (d)).

In addition, while our HHI data only covers 2013–2016, we can estimate our outside-occupation option index for each year from 1999–2016. In Appendix Table A14, we regress log wages on outside-occupation options over 1999–2016, with various combinations of fixed effects. Even with both occupation-by-city and occupation-by-year fixed effects – and so identifying only off annual variation in outside-occupation options compared to their mean for each occupation-by-city and occupation-by-year unit – we find large, positive, and significant effects of outside-occupation options on wages. We also show that the results in our baseline regressions of the effect of outside-occupation

[56] While we do not consider the effects of outside-city options on wages in this paper, our methodology could easily be extended to do so. In practice, the wage effect of within-occupation employer concentration and outside-occupation options is limited by workers' option to move cities. Factors which increase the cost of geographic mobility, in turn, would be expected to increase the effect of local employer concentration and local outside-occupation options on wages.

[57] Imagine, for example, 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.

options on wages are qualitatively similar in a number of different scenarios: if we calculate the outside-occupation option index using occupational mobility at the SOC 2-digit or SOC 3-digit levels to address concerns about possible mismeasurement in occupational mobility at the level of detailed 6-digit SOC codes (Appendix Table A16); if we segment our analysis into different time periods to address concerns that our results may be driven by the idiosyncrasies of any of these periods (Appendix Table A17); or if we weight each occupation-city cell by employment (Appendix Table A18).

Finally, our construction of the outside-occupation option index was based on the concept of jobs in other occupations being a valuable outside option for workers in their wage bargain: higher wages in an outside-occupation job option give workers more bargaining leverage, leading to higher wages. But there is another channel by which outside-occupation job options can affect wages: mobility. As the wages in an outside option occupation p rise, some workers from occupation o will move to occupation p. The supply of workers in occupation o falls, and so the wage rises. (Both of these mechanisms imply that the alternative occupation p is a relevant outside option: the difference is simply that in the bargaining case, workers don't have to exercise their option to move). Even when controlling for local occupational employment, however, the value of outside-occupation options still matters for wages (Appendix Table A15), reinforcing our conclusion that the effects of outside-occupation options on wages largely reflect the value of that option to workers who *stay* in their own occupation in their wage bargain (and not only the exercised option of workers to *move* to that other occupation).

6 Discussion and Implications

Our results show that employer concentration is not a niche or irrelevant issue in U.S. labor markets: it has substantially broader relevance than the canonical example of the one-factory town, with large, negative, and significant effects of employer concentration on wages when estimated across data comprising the large majority of the U.S. labor market. Later in this section, we will apply our coefficient estimates in a back-of-the-envelope exercise which suggests that perhaps as many as 12 million workers are in occupation-city labor markets with wage effects of employer concentration of 2% or greater.

On the other hand, our results do not support the idea that employer concentration is a major factor in wage suppression for the majority of U.S. workers. (This does not imply that other sources of monopsony power are not at play for workers in unconcentrated labor markets, notably arising from search frictions, switching costs, or differentiated jobs). While a very large share of occupation-city *labor markets* are highly concentrated according to typical thresholds, the majority of U.S. *workers* are not in labor markets with high degrees of employer concentration as measured by typical thresholds.[58] This is because the most concentrated labor markets tend to be those with the fewest workers. So, while the effects of concentration on wages are non-trivial for the subset of workers in highly-affected labor markets, the *aggregate* effect of employer concentration on wages is unlikely to be very large, and employer concentration cannot explain more than a small share of aggregate income inequality.[59]

In this section, we carry out a back-of-the-envelope exercise to understand the degree to which different workers' labor markets might be affected by employer concentration, and discuss some policy implications for antitrust and other labor market policies. Overall, we see our study as giving strong support to an increased focus on employer concentration as an issue affecting wages for a material subset of U.S. workers, and a way to identify which labor markets should be prioritized in this effort. For a deeper understanding of the degree to which individual labor markets may be affected by specific policy changes to address this issue, labor-market-specific study is needed.[60]

[58]According to our estimates from the BGT vacancy data, as of 2016 the median U.S. private sector worker was in an occupation-city labor market with an HHI of 200, around one quarter of workers in the U.S. private sector were in occupation-city labor markets with an HHI of over 500, and a little less than 5% of workers were in occupation-city labor markets with an HHI of over 2,500. Our BLS OES data covers about 110 million workers: to the extent that our data disproportionately excludes workers who are in small occupations or in rural areas – and who might be likely to face a higher degree of employer concentration – our data will underestimate the total share of workers facing high levels of employer concentration.

[59]Similarly, while our work focuses only on the period 2013–2016, Rinz (2018), Berger et al. (2019), and Lipsius (2018) have shown that employer concentration appears to have fallen over recent decades in most local labor markets, suggesting that changing employer concentration appears unlikely to have been able to explain the trends of median pay stagnation or rising income inequality. It is quite possible, however, that the decline in countervailing worker power in the U.S. 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).

[60]This has parallels to the argument made in Schmalensee (1989), who argues that inter-industry research should be used to generate empirical regularities and stylized facts which in turn can be used to motivate in-depth research in specific markets.

6.1 Back-of-the-envelope quantification of the degree to which concentration affects U.S. workers

To what extent are workers' wages in the U.S. affected by employer concentration, and who is most affected? In this section, we use our coefficient estimates for a back-of-the-envelope quantification of the degree to which different workers might be affected by employer concentration. To do this, we evaluate the effect of employer concentration in occupation-city labor markets relative to a benchmark HHI of 200 – roughly the level of employer concentration experienced by the median worker in our data set in 2016. (As an example: an HHI of 200 might correpond to roughly 50 equal-sized employers being present in the labor market, or two large employers which each employ 10% of workers and an atomistic 'fringe' of firms employing the rest of the workers).[61]

Specifically, in this exercise we apply our regression coefficients to estimate the degree to which U.S. workers' wages might be higher if it were possible to reduce employer concentration in their labor market to 200 without affecting anything else. (Note that this rests on the assumption that we can apply our estimated coefficients linearly). We estimate a rough "wage effect" of employer concentration for each occupation-city in 2016 as follows:

$$\text{wage effect}_{o,k,t} = (log(HHI)_{o,k,t} - log(200)) \cdot \gamma_1^q$$

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

Note that this exercise considers the effect of changes in employer concentration *holding all else constant*, including local productivity. Therefore, it should be considered illustrative of the degree to which wages may be marked down from local occupational productivity as a result of employer concentration, rather than illustrating what would happen if a specific policy or business decision were to change local employer concentration (but might also change local productivity).

[61] This level of concentration is not typically thought to be a concern from an anti-competitive perspective in product markets. For example, the Department of Justice Horizontal Merger Guidelines considers only markets with an HHI of over 1,500 to be sufficiently concentrated to warrant further attention. While there may be monopsony or employer power in even relatively unconcentrated labor markets, it is more likely to derive from employer heterogeneity, search costs, or frictions than from employer size (Naidu and Posner, 2020).

Roughly 49 million of the 110 million workers in our data set are in occupation-city labor markets with an HHI greater than 200. Of these, our counterfactual wage exercise suggests that roughly 3 million workers would have wages which were 5% higher or more if employer concentration was 200 (and nothing else changed), and a further 9 million would have wage increases of between 2% and 5%. These 12 million workers represent a little over 10% of the workers in our data, and around 7.5% of the total U.S. civilian labor force in 2016 (and indeed, one might expect many of the workers *not* represented in our data – who include workers in non-metropolitan areas – to face wage suppression from employer concentration as well). That is: there are a large number of workers for whom employer concentration may be depressing wages by a not-insignificant amount.[62]

Our estimated counterfactual wage effect depends on only two factors for each occupation-city labor market: the labor market's employer HHI, and the occupation's outward mobility. Grouping occupation-city labor markets into categories based on these two variables, we show in Table 7 the average estimated wage effect for different combinations of employer concentration and outward occupational mobility.[63] Wage effects are strongest for low-mobility occupations with even low-to-moderate HHIs, and for medium-mobility occupations with high HHIs. In fact, our estimated coefficients would suggest that the 10 million workers in occupation-city labor markets with a relatively low HHI by product market standards (between 500 and 1,500), but with lower-than-median outward occupational mobility, see *larger* wage effects of employer concentration than the 1.4 million workers in highly concentrated occupation-city labor markets (HHI greater than 2,500) but where the occupation is in the highest quartile of outward mobility. This reiterates that within-occupation employer concentration alone can give a very misleading picture of the true degree to which workers' wages are affected by employer concentration.

[62]One may be concerned that our estimates of the number of workers affected by employer concentration in this exercise are overestimates since we include some occupations which are relatively underrepresented in our vacancy posting data (and for which we may therefore overestimate the true degree of employer concentration). We repeat this exercise excluding all occupations which have a 'represented-ness' of less than 0.5 in our BGT vacancy data (where represented-ness is calculated as the share of vacancies accounted for by occupation o in the vacancy data, divided by the share of all employment accounted for by occupation o nationwide according to the BLS OES). This would suggest 1.7 million workers with wage increases of 5% or more if employer concentration was 200, and 5.4 million workers with wage increases of between 2% and 5%.

[63]We replicate this exercise, but including only occupations with a represented-ness of 0.5 or greater in the BGT vacancy posting data, in Appendix Table A20.

Which occupations are most affected by employer concentration? Consider only occupation-city labor markets with an estimated wage effect of employer concentration of 2% or greater. In Table 8, we list the twenty-five largest occupations who are most affected by employer concentration, ranked according to the number of workers in that occupation who see an estimated wage effect of 2% or greater (but excluding occupations which are substantially under-represented in the BGT vacancy data, for whom we may have substantially over-estimated employer concentration).[64]

A large share of the most-affected occupations are in healthcare. More than one million registered nurses, licensed practical and vocational nurses, and nursing assistsants, and around 200,000 pharmacists and pharmacy technicians, for example, are in labor markets with an estimated wage effect of concentration of 2% or greater; the list of the most-affected occupations also includes radiologic technologists, lab technologists, phlebotomists, and respiratory therapists. 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, 2019), and a low elasticity of the labor supply of registered nurses to individual hospitals (Staiger et al., 2010). There are, however, other large occupations outside of healthcare where many workers appear to be substantially affected by employer concentration. According to our estimates, nearly a million security guards are in labor markets with an estimated wage effect of concentration of 2% or more, as large shares of local security guard labor markets are comprised of employment by a few large security services companies (although note that our BGT vacancy data is somewhat underrepresentative of security guard employment, so we may be overrepresenting the true degree of employer concentration faced by security guards). Hairdressers, hairstylists, and cosmetologists similarly are frequently in labor markets with 2% or greater wage effects of employer concentration according to our estimates, since large numbers of the hairdressing jobs in our database are accounted for by large salon chains (although, note that our data may miss small hair salons who advertise by word-of-mouth, and also that many of the large salon chains have franchise business models, so it is unclear whether it makes more sense to consider employer concentration at the level of the firm or at the level of individual franchised employers).

Do effects differ across the income distribution? Grouping occupation-city labor

[64] Our threshold is a 'represented-ness' of 0.5 in the BGT vacancy data, or around the bottom third of occupations.

markets by their position in the national hourly wage distribution, we show that the effects of employer concentration on wages appear to be relatively similar across the bottom three quartiles of the wage distribution (Figure 7). Around 10% of workers in the bottom, 2nd, and 3rd quartile of the wage distribution see wage effects of 2% or more, and around 3% of workers in each of these quartiles see wage effects of 5% or more arising from employer concentration. Employer concentration affects fewer workers in the top quartile of the wage distribution.[65] Similarly, we can estimate the share of workers in each metropolitan area who see wage effects of employer concentration of 2% or greater. The average degree to which workers are affected by employer concentration is substantially greater in smaller and lower-wage metropolitan areas as they tend to have fewer employers in any given occupation (as shown in Figure 8). These exercises suggest that the net effect of employer concentration is to slightly increase wage inequality across the income distribution and across U.S. regions (though the fact that most workers are not in highly concentrated labor markets means that differential employer concentration cannot explain more than a very small share of overall income inequality).

This is a back-of-the-envelope exercise and rests on the key assumption that our HHI coefficients can be applied linearly, so should be interpreted as a *rough* estimate of the degree to which employer concentration may be affecting the wages of American workers. It should be noted that we only estimate effects of employer concentration on average wages, and that employer concentration would likely have a similar affect on non-wage benefits and workplace amenities. Indeed, Qiu and Sojourner (2019) find that workers are less likely to receive employment-based health insurance in more concentrated labor markets, and Marinescu, Qiu and Sojourner (2020) find a greater prevalence of labor rights violations in more concentrated labor markets.[66]

6.2 Implications: antitrust

Our back-of-the-envelope exercise suggests that while employer concentration suppresses wages for a material subset of workers, the majority of American workers likely

[65]The large majority of the highly-affected workers in the top quartile of the wage distribution are in healthcare occupations.

[66]There may also be an interaction with the degree to which employers exert power over workers in other respects, for example by requiring them to sign mandatory arbitration clauses for workplace disputes, an increasingly common practice (Colvin, 2017).

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 mobility outside their occupation. One area where this can be done is in antitrust.

References to monopsony were introduced into the Department of Justice's and Federal Trade Commission's Horizontal Merger Guidelines in 1992 (Phillips, 2019). On the whole, however, antitrust authorities have paid relatively infrequent attention to employer concentration or anti-competitive practices in labor markets (Marinescu and Hovenkamp, 2019; Naidu and Posner, 2020).[67] In recent years, spurred partly by recent work on employer concentration, there have been increasing calls for antitrust authorities to take into account labor market impacts in merger reviews even if there is no impact on product markets and to scrutinize more closely possible anti-competitive behavior in labor markets (Marinescu and Hovenkamp, 2019; Naidu et al., 2018; Hemphill and Rose, 2017; Steinbaum and Stucke, 2020; Hovenkamp, 2018).[68] And indeed, a recent public comment by the FTC on a proposed hospital merger stated concerns not only about possible anti-competitive effects in product markets but also about the possibility that the merger would depress wage growth for registered nurses.[69]

Our findings would tend to support increased consideration of employer concentration and labor market power by antitrust authorities. However, it is important to note that our findings *do not* tell us that all increases in employer concentration reduce wages: rather, our findings should be interpreted as illustrating the degree to

[67] According to Naidu et al. (2018): "As far as we know, the DOJ and FTC have never challenged a merger because of its possible anticompetitive effects on labor markets, or even rigorously analyzed the labor market effects of mergers as they do for product market effects. Nor have we found a reported case in which a court found that a merger resulted in illegal labor market concentration."

[68] Indeed, issues of employer concentration featured in the Federal Trade Commission's hearings on "Monopsony and the State of Antitrust Law" in September 2018 and "Antitrust in Labor Markets" in October 2018, the Department of Justice's Public Workshop on "Competition in Labor Markets" in September 2019, and the House of Representatives' Committee on the Judiciary hearing on "Antitrust and Economic Opportunity: Competition in Labor Markets" in October 2019. The FTC has also announced plans to expand its retrospective review of mergers to address questions including whether these mergers created labor market monopsony power. There is disagreement between legal scholars as to whether antitrust authorities have sufficient authority to respond effectively to threats to labor market competition under existing legal frameworks, or whether a new standard ought to be adopted: see e.g. Hemphill and Rose (2017); Marinescu and Hovenkamp (2019); Steinbaum and Stucke (2020).

[69] Defining the relevant local labor market as the market for Registered Nurses in the Hendrick TX Commuting Zone, the FTC submission showed that the proposed merger would increase the local employment HHI from 5,449 to 8,598. They note that "the reduction in competition caused by the proposed Hendrick merger in the labor market for registered nurses is likely to be extraordinarily high" Federal Trade Commission (2020).

which employer concentration (by worsening worker outside options) reduces wages *holding constant* other firm-level and market-level factors like productivity. This has important implications for the appropriate antitrust approach to employer concentration. In some cases, seeking to reduce employer concentration (or to prevent it from increasing) may not be the best response to high employer concentration. While the presence of large firms in a certain local area may lead to high employer concentration, in some cases it may also make for substantially higher productivity than an equivalent labor market with many small firms – and, in some cases, may even *reduce* wages relative to a counterfactual with high concentration but high productivity (as is emphasized by Hovenkamp (2018), Berger et al. (2019), and Arnold (2020)). For example, a highly productive factory may be responsible for high employer concentration in a given occupation-city labor market, but its high productivity may mean that workers' wages are higher than they would be in an equivalent world with many small factories in their labor market. More broadly, the degree of employer concentration in a given local area is at least to some extent endogenous to local economic conditions.[70] As such, we see our work as demonstrating the practical importance of the potential for high employer concentration to reduce wages – but emphasize the need for close scrutiny of individual cases, and industry- and occupation-specific case studies (such as Prager and Schmitt (2019)) to understand whether antitrust action would be appropriate in any specific circumstance.[71]

With this caveat in mind: How should antitrust authorities take employer concentration into consideration when thinking about labor market power? Shapiro (2019) writes that "If the antitrust authorities seriously want to explore the possibility of challenging mergers on the basis of harm to competition in labor markets, developing a quick and efficient means of identifying mergers that involve a significant overlap

[70] This is one of the concerns which motivated industrial organization economists to move away from the structure-conduct-performance paradigm and focus less on market concentration statistics: see, for example, Berry et al. (2019) or Schmalensee (1989). Our empirical analysis – by isolating changes in employer concentration which are plausibly exogenous to local occupation-specific economic conditions – enabled us to isolate the effect of employer concentration holding other factors (like productivity) constant. It cannot, however, tell us what the effect *would be* of a business- or policy-induced change that increases concentration in a specific market would be (whether that is a merger, an antitrust action, or a set of regulations around firm entry).

[71] Naidu et al. (2018) argue that antitrust authorities should use a "worker welfare standard" to evaluate mergers alongside the consumer welfare standard. This standard would permit mergers where the incremental increase in workers' wages because of the productivity gains induced by the merger would *outweigh* the incremental decrease in workers' wages induced by the increase in employer concentration – so workers would be on net better off after the merger than before.

in plausible labor markets would be a good first step." In product markets, antitrust authorities use HHIs as a screen for potential anti-competitive effects of mergers. Marinescu and Hovenkamp (2019), Naidu et al. (2018), and Krueger and Posner (2018) argue that the same should be done in labor markets (with further detailed investigation into the possible labor market impact of the mergers which are flagged in the screening process).

Our analysis in this paper suggests one simple way to screen labor markets for potential anti-competitive effects of mergers using only two variables: the HHI at the level of an individual 6-digit SOC occupation by commuting zone (or metropolitan area), and the degree of outward occupational mobility from that occupation. This proposal differs slightly from that in Marinescu and Hovenkamp (2019), who argue that antitrust authorities should screen for anti-competitive effects of mergers based only on the HHI calculated at the level of a local SOC 6-digit occupation. Our work suggests that screening based only on local within-occupation concentration without considering the very differential degree of outward occupational mobility experienced by different workers will lead to some mergers being scrutinized which will have little effect on wages, while others which may have serious anti-competitive effects may go unnoticed. These distinctions are extremely practically important: our results suggest that the wage suppressive effect of a given increase in local employer concentration might be more than four times as high for occupations in the lowest quartile of outward mobility as for occupations in the highest quartile of outward mobility.[72] Of course, these screens should bear in mind that some mergers may increase employer concentration substantially not only in the occupation under scrutiny but also in the most relevant outside option occupations for workers – a particular concern for occupations whose outside options are predominantly in the same industry, like healthcare. In these cases, even high-outward-mobility occupations may see large wage effects from an increase in

[72] Marinescu and Hovenkamp (2019) and Naidu et al. (2018) argue that the appropriate labor market definition for concentration screens is one in which a hypothetical monopsonist could achieve a "SSNRW": Small Signifiant Non-transitory Reduction in Wages. Applying this concept alongside empirical estimates of labor supply elasticities, Marinescu and Hovenkamp (2019) argue that the local SOC 6-digit occupation is an appropriate labor market definition. However, our coefficient estimates would suggest that even large increases in employer concentration in some SOC 6-digit occupations may have only very small effects on wages for workers: moving from an HHI of 220 to 2,500, even when applying the maximum estimate from our baseline confidence interval for the quartile of occupations with the highest outward mobility, would only be associated with at most a 1.7% wage decrease. This suggests that even by the standards of the "SSNRW" test, the SOC 6-digit occupation is likely be too narrow a definition for a large share of workers.

employer concentration induced by a merger.

To illustrate the importance of considering occupational mobility in employer concentration screening, compare the examples of nursing assistants, bank tellers, and security guards in Duluth, a metropolitan area of a little under 300,000 people covering parts of both Minnesota and Wisconsin. Nursing assistants in Duluth in 2016 had an HHI of 1,571 in 2016, bank tellers had an HHI of 1,911, and security guards had an HHI of 580. A simple screen based on product market guidelines would consider nursing assistants and bank tellers to be in moderately concentrated labor markets, meriting further scrutiny (Marinescu and Hovenkamp, 2019), while the security guards' labor market would be considered to be unconcentrated. Nursing assistants, however, have an occupation leave share of 18%, whereas bank tellers have an occupation leave share of 32%. While the within-occupation HHI suggests that nursing assistants and bank tellers face similar labor market concentration, our regression estimates suggest that the negative effect of this employer concentration on wages could be very large for the nursing assistants, but may not exist at all for the bank tellers. Indeed, our estimates would suggest that the security guards in Duluth – with a much lower HHI of around 580, but a relatively low occupation leave share of 19% – experience a substantially larger effect of employer concentration on wages than do bank tellers, despite the substantially higher degree of within-occupation concentration faced by tellers. A labor market screen based only on the local within-occupation HHI would inaccurately deem the bank tellers as being in a labor market with substantial effects of employer concentration on wages, and would inaccurately consider the security guards to be in a labor market without substantial effects of employer concentration on wages.[73]

While increased antitrust scrutiny of labor markets is important, it is important

[73]It is also worth noting that any given merger is likely to affect different groups of workers differently. To illustrate this point, consider nurse practitioners and medical secretaries in Rochester, Minnesota. Due to the concentration of healthcare in this metro area, both occupations have an extremely high HHI (of over 8,000). Nurse practitioners, however, have an occupation leave share of 9%, putting them in the lowest quartile of outward occupational mobility, whereas medical secretaries have an occupation leave share of 26%, putting them in the third quartile of outward occupational mobility. While the within-occupation HHI suggests that both groups face similar labor market concentration, our estimates suggest that the negative effect of this employer concentration on wages could be around three times as large for the nurse practitioners as for the medical secretaries (who have substantially more comparable job options outside their occupation). This is consistent with Prager and Schmitt (2019)'s finding that hospital mergers disproportionately affects workers in nursing and pharmacy occupations. This suggests that different screens should be run for the different local occupational labor markets in which the firms in question operate, since the potential for anti-competitive effects is likely to be substantially greater in some occupations than in others.

to note that it is also unlikely to affect the majority of workers impacted by employer concentration.[74] Most changes in employer concentration are not caused by mergers and acquisitions, and it seems likely that most labor markets do not feature the use of illegal anti-competitive practices. As such, while an increased focus on labor markets from antitrust authorities would be welcome, other policy measures may be appropriate in the (many) circumstances where antitrust action is not.

6.3 Implications: policies to raise wages

In many cases, rather than seeking to change the level of employer concentration in a given labor market, it may be more appropriate to *recognize* the fact that employer concentration appears to 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 wage regulations or standards for benefit provision or occupational health and safety in local labor markets characterized by high employer concentration.

There is some suggestive evidence that labor markets with higher unionization rates see smaller effects of employer concentration on wages, as the theory of countervailing power would suggest.[75] When we re-run our baseline regression of the effect of employer concentration on wages with an interaction with states' right-to-work status – a proxy for the ease of forming a union – we find larger effects of both employer concentration and outside-occupation options on wages in right-to-work states.[76] Similarly, Prager and Schmitt (2019) find larger effects of hospital mergers on nursing wages when nursing unionization rates are lower and in right-to-work states, and Benmelech et al. (2018) find a stronger relationship between employer concentration and wages

[74] See Naidu and Posner (2020) for an in-depth discussion of the limits of antitrust as a response to employer concentration and other aspects of monopsony power.

[75] In our conceptual framework outlined in section 2, higher worker bargaining power β reduces the weight placed on the outside option in the wage bargain and therefore reduces the importance of employer concentration in wage determination.

[76] These results are shown in Appendix Table A19. Since there is no variation in states' right-to-work status over our short sample period, this is just a cross-sectional difference in effect between the states with and without right-to-work laws, and these states likely differ on other characteristics also. So this cannot be interpreted as a causal effect of right-to-work laws on the effect of concentration on wages, but rather is interesting in that it is consistent with what one might expect if unionization were to reduce the effects of employer concentration on wages.

in U.S. manufacturing firms where unionization rates are lower. Thus, in some cases it may be possible to address the effects of employer concentration not by reducing concentration but by strengthening workers' countervailing power. (This also suggests a possible interaction with the antitrust discussion in the prior section: mergers which increase employer concentration would be likely to have less damaging welfare impacts in labor markets where workers have countervailing power).

Moreover, policies which increases wages, like higher minimum wages, would be expected to have less of a negative effect on employment in labor markets where employers have monopsony power (and employer concentration is one possible source of monopsony power). Consistent with this, Azar et al. (2019b) find that occupation-city labor markets with higher employer concentration see smaller employment effects of minimum wage increases in the U.S. This suggests that when localities are considering the level at which to set the minimum wage, they should try to understand the likely degree of monopsony power in their local labor market. The degree of employer concentration and outward mobility in particular local occupations can help inform this calculation. The absence of employer concentration, however, does not imply absence of monopsony power: other sources of monopsony power include search frictions, switching costs, and worker and job heterogeneity.

6.4 Implications: policies to promote mobility

Our results illustrate that the wage suppressive effects of employer concentration within a local occupation depend to a large extent on whether workers are able to find similarly good jobs outside their occupation. Similarly, while we do not examine geographic mobility, it seems likely that the easier it is for workers to find jobs in other localities, the less of an effect employer concentration can have in their own local occupation. This suggests that policies which make it easier to switch occupation and/or to work in different geographic areas would – by expanding the scope of workers' labor markets and increasing their outside options – reduce the degree to which employer concentration can suppress wages.[77]

How might geographic mobility costs be reduced? One possible avenue is to in-

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

crease the reciprocal recognition of state-specific licenses and certifications. Johnson and Kleiner (2020) find that workers in occupations with state-specific licensure requirements move between states at a 7% lower rate than workers in occupations with national (or reciprocally recognized) licenses, suggesting that state-specific licensing acts as a barrier to geographic mobility. Note, though, that state-specific licensing may not be the binding constraint on mobility for some occupations: for example, DePasquale and Stange (2016) find that the rollout of reciprocal recognition of nursing licenses across states did not affect either nurses' geographic mobility or their wages.[78] A second avenue which would reduce mobility costs on the margin is to reduce constraints on housing supply in high-wage locations: Ganong and Shoag (2017) find that part of the decline in geographic mobility for low-income workers can be explained by high housing costs in high wage cities, which reduce the returns to migration.

How might occupational mobility costs be reduced? This is likely to be more difficult: the occupations with the lowest outward mobility tend to be occupations which require highly occupation-specific qualifications or training (like many medical occupations, hairdressers and cosmetologists, or fitness trainers), or which have highly occupation-specific task requirements or job characteristics (like security guards). This suggests that formal barriers to occupational mobility like licensing or certification requirements may not play a major role in limiting occupational mobility, since it is in some sense limited by the inherent specificity of individual occupations. Nonetheless, it seems likely that policies which increase workers' ability to move between occupations (including policies which reduce any disproportionate costs of acquiring training, licensing, or certification in different occupations) would increase outside options and reduce the effects of employer concentration on the margin.

Finally, it should be emphasized that restrictions on worker mobility *within* an occupation could also exacerbate the effects of employer concentration on wages by reducing the availability of valuable outside options for workers. As such, the proliferation of non-compete clauses (in industries where the protection of firm-specific intellectual property or relationships is not a concern) may exacerbate the effects of employer concentration on wages, particularly if such non-compete agreements have a broad scope across occupations or locations, or if they apply to workers with limited

[78] This is particularly interesting since nurses are one of the occupations which we find are most affected by local employer concentration, and therefore for whom an increase in the availability of outside options should be most valuable.

outward occupational mobility (Starr, Prescott and Bishara, 2019; Johnson, Lavetti and Lipsitz, 2020).

6.5 Incidence of effects of employer concentration

Finally, understanding the ultimate *incidence* of the effect of employer concentration is important to understand the welfare consequences of policies to tackle employer concentration. Our estimates suggest that increases in employer concentration ceteris paribus reduce local wages. But, our estimates are not informative as to whether these wage reductions are retained by firms in the form of higher profits, or are passed on to consumers in the form of lower prices. This is likely to depend on the nature of the product market competition faced by these firms: for example, local hospitals which are both monopsonists in the labor market and monopolists in the product market may retain profits made as a result of labor market power, whereas a local manufacturing firm competing in a highly competitive global market may pass on the lower wages in the form of lower consumer prices. Similarly, Kahn and Tracy (2019) argue that the ultimate incidence of local labor market concentration falls to a large extent on local landowners, arguing that lower local wages reduce local rents and house prices; they find that local land prices are lower in areas with higher employer concentration. Understanding the ultimate incidence of the effects of employer concentration for firms, workers, and localities is important to understand the welfare implications and determine any appropriate policy response.

7 Conclusion

In this paper, we use a new instrumental variable approach, based on differential local exposure to large national firms' growth, to estimate the causal effects of employer concentration on wages across the large majority of U.S. occupation-city labor markets. Documenting new facts about occupational mobility using a new data set of 16 million unique U.S. resumes, we show that existing occupational definitions are not appropriate measures of workers' labor markets, and that a failure to consider workers' job options outside their occupation when studying labor market concentration can lead to biased inference and obscure important heterogeneity. Informed by bargaining models of the labor market, where employer concentration reduces the availability of feasible outside options for workers, we develop a simple, tractable way to take into account workers' outside-occupation job options, using worker flows to identify relevant job options and using a shift-share approach to identify causal effects of changes in the value of these outside-occupation job options on wages.

Overall, we show that employer concentration in local occupational labor markets matters substantially for a material subset of U.S. workers. Moving from an occupation-city labor market with the median worker's HHI to the 95th percentile HHI is associated with a roughly 3 percent reduction in wages on average. This masks substantial heterogeneity: the same change in employer concentration is associated with a 6 percent reduction in wages for workers in the lowest quartile of occupations by outward mobility, but no significant reduction in wages for workers in the highest quartile of occupations by outward mobility.

These findings have important implications not only for our understanding of the nature of labor market competition in the U.S. but also for policy. Our findings show that employer concentration is *not* a niche issue confined to a few factory towns – but nor is it prevalent enough to be a major determinant of macroeconomic aggregates such as the overall wage level or degree of income inequality. With our estimates suggesting perhaps 10% of the U.S. private sector workforce experiences non-trivial wage effects of employer concentration, increased policy attention on these issues is justified.

When considering the effects of employer concentration, our work suggests the HHI (at the level of a local 6-digit SOC occupation) is only a good proxy if considered alongside the degree of outward occupational mobility (and, ideally the quality of outside-occupation job options in the local area). An only moderately concentrated local labor market where the occupation in question has low outward mobility should be much more of a concern than a highly concentrated local labor market where the occupation in question has high outward mobility. In helping identify the labor markets most affected by employer concentration, our findings can help inform the targeting of labor market screening in antitrust as well as the design of policies which seek to increase wages (like minimum wages) or increase workers' countervailing power in labor markets.

More generally, our paper suggests that labor market analysts can easily improve upon binary labor market definitions, instead constructing 'probabilistic' labor market measures using data on worker transitions between occupations, industries, or locations. Given the absence of publicly-available, fine-grained occupational mobility data, we have made the occupational transitions data set constructed from the Burning Glass Technologies data publicly available. We hope that the tools and insights provided in this paper enable other researchers to use, and improve upon, methods like ours to ensure that the labor markets they are researching are the ones that workers are experiencing.

References

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

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.

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

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

Ashenfelter, Orley, Henry Farber, and Michael Ransom, "Modern Models of Monopsony in Labor Markets: A Brief Survey," *IZA Discussion Paper*, 2013, (4915).

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

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

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.

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.

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

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

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

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

Benmelech, Efraim, Nittai Bergman, and Hyunseob Kim, "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?," 2018.

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

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

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

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," Technical Report, NBER 2018.

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

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

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

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

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

Colvin, Alexander JS, "The growing use of mandatory arbitration," *Economic Policy Institute*, 2017.

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.

DePasquale, Christina and Kevin Stange, "Labor supply effects of occupational regulation: Evidence from the nurse licensure compact," Technical Report, National Bureau of Economic Research 2016.

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

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

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang, "Imputing Missing Values in the US Census Bureau's County Business Patterns," 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.

Federal Trade Commission, "Federal Trade Commission Staff Submission to Texas Health and Human Services Commission Regarding the Certificate of Public Advantage Applications of Hendrick Health System and Shannon Health System," Technical Report, September 11th, 2020 2020.

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

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

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

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

Gibbons, Eric M., Allie Greenman, Peter Norlander, and Todd Sørensen, "Monopsony Power and Guest Worker Programs," 2019.

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

Goolsbee, Austan and Chad Syverson, "Monopsony Power in Higher Education: A Tale of Two Tracks," 2019.

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

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

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

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

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.

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

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

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

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

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

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

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

Johnson, Janna E and Morris M Kleiner, "Is Occupational Licensing a Barrier to Interstate Migration?," *AEJ: 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.

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

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

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

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

Krueger, Alan B and Orley Ashenfelter, "Theory and evidence on employer collusion in the franchise sector," Technical Report, National Bureau of Economic Research 2018.

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

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

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

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

— **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.

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

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

— **, Ivan Ouss, and Louis-Daniel Pape**, "Wages, Hires, and Labor Market Concentration," 2019.

— **, Yue Qiu, and Aaron Sojourner**, "Wage Inequality and Labor Rights Violations," *Available at SSRN*, 2020.

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

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

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

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

Naidu, Suresh and Eric A. Posner, "Labor Monopsony and the Limits of the Law," 2020.

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

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

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

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

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.

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

Pissarides, Christopher A, *Equilibrium Unemployment Theory*, MIT Press, 2000.

Prager, Elena and Matthew Schmitt, "Employer Consolidation and Wages: Evidence from Hospitals," 2019.

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

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

Rinz, Kevin, "Labor Market Concentration, Earnings Inequality, and Earnings Mobility," *U.S. Census Bureau Carra Working Paper Series*, 2018, *2018-10.*

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

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

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

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

Shapiro, Carl, "Protecting competition in the American economy: Merger control, tech titans, labor markets," *Journal of Economic Perspectives*, 2019, *33* (3), 69–93.

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

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

Sokolova, Anna and Todd Sorensen, "Monopsony in Labor Markets: A Meta-Analysis," 2020.

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

Stansbury, Anna and Lawrence H Summers, "The Declining Worker Power Hypothesis: An explanation for the recent evolution of the American Economy," *Brookings Papers on Economic Activity*, 2020.

Starr, Evan, James J Prescott, and Norman Bishara, "Noncompetes in the US labor force," *U of Michigan Law & Econ Research Paper*, 2019, (18-013).

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

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.

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

Weil, David, *The Fissured Workplace* 2014.

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

8 Figures and Tables

Table 1: Summary statistics: BGT occ. mobility data

Percentile (across occupations)	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)									
Obs.	0.6	1.1	1.6	4.9	20.8	112.3	466.8	853.9	3,471.9
Panel B: Share leaving job and occupation, by occ. (2002-2015)									
Share in diff. job	0.30	0.35	0.37	0.40	0.45	0.52	0.61	0.66	0.74
Share leaving 6d. occ.	0.047	0.062	0.074	0.90	0.10	0.12	0.14	0.18	0.29
Leave share	0.09	0.11	0.14	0.19	0.24	0.28	0.33	0.38	0.69
Panel C: Share of occupational transitions which cross SOC 2d boundary (2002-2015)									
All occ. transitions	0.55	0.65	0.70	0.79	0.87	0.93	0.97	0.98	1.00
Excl. management	0.40	0.48	0.51	0.59	0.67	0.75	0.80	0.83	0.87

Notes: We exclude occupations with <500 observations in the BGT resume data from this table and from all our analysis. In Panel A, an observation is a person-year unit, that is also observed in the data in the following year. Panel B shows the share of workers in our BGT occupational mobility data observed in a new job from one year to the next, the share in a new occupation from one year to the next, and the "leave share", which we define as the share leaving their occupation conditional on leaving their job (see section 3). Panel C shows the share – by origin occupation – of all SOC 6-digit occupational coincidences which also span SOC 2-digit boundaries. The percentiles refer to percentiles across occupations, such that the median occupation in our data has 20,800 observations (Panel A), the median occupation when sorted by the leave share has a leave share of 0.24 (Panel B), and the median occupation when sorted by the share of transitions which cross a SOC 2-digit boundary sees 87% of SOC 6-digit transitions also cross a SOC 2-digit boundary (Panel C).

Table 2: Summary statistics: main data set

Percentile	1	5	10	25	50	75	90	95	99
Panel A: Employer concentration HHI (2016)									
HHI	35	97	167	390	965	2,222	5,000	7,222	10,000
HHI (emp-wt)	10	22	33	81	221	536	1,357	2,256	5,556
Panel B: Outside-occupation option index oo^{occs} (2016)									
oo^{occs} (2016)	1.4	2.1	2.6	3.5	4.8	6.6	8.8	10.6	16.1
$\frac{oo^{occs}}{wage}$	0.03	0.06	0.09	0.15	0.23	0.34	0.45	0.53	0.74
$\frac{oo^{occs}}{wage}$, emp-wt	0.06	0.11	0.14	0.23	0.34	0.45	0.55	0.63	0.77
Panel C: Occupation-city wages and employment (2016)									
Employment	30	40	50	90	220	670	1,980	3,920	14,410
Mean hourly wage	9.05	10.50	11.94	15.40	21.42	31.08	44.07	53.68	90.50
Wage, emp-wt	8.97	9.94	10.99	13.39	18.33	30.28	45.10	56.42	80.50
Panel D: national hourly wage distribution (2016) from BLS OES									
Hourly wage	–	–	9.27	11.60	17.81	28.92	45.45	–	–

Notes: Panels A, B, and C show summary statistics for our main data set in 2016, calculated over all occupation-city-year cells for which we have wage data, a vacancy HHI, and an outside-occupation option index. This comprises 103,300 occupation-by-city labor markets and 109,366,900 workers (according to the BLS OES data). Panel D shows the national 10th, 25th, 50th, 75th, and 90th percentile of the hourly wage distribution according to the full BLS OES data set, for comparison. (This comparison suggests that our data set slightly overrepresents workers in the middle of the wage distribution relative to the tails: this underrepresentation of the tails is because the OES wage data by occupation-MSA excludes some of the smaller cells, which appear to disproportionately be drawn from the lower and higher tails of the wage distribution.)

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	-0.010***	-0.007***	-0.014***	-0.011***
	(0.001)	(0.001)	(0.003)	(0.003)
Log outside-occ. options		0.106***		0.095***
		(0.007)		(0.009)
Vacancy growth			-0.001	-0.001
			(0.001)	(0.001)
Predicted vacancy growth			-0.011	-0.013
			(0.010)	(0.010)
Exposure control			0.008	0.004
			(0.008)	(0.007)
Observations	184,411	184,411	184,411	184,411
F-Stat			705	379

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All regressions feature occupation-by-year and MSA-by-year fixed effects. Columns (a) and (b) show OLS regressions and columns (c) and (d) show 2SLS IV regressions (where both the log HHI and log outside-occ. option index are instrumented). The reported F-stat for the 2SLS IV regressions is the Kleibergen-Paap Wald F statistic. The vacancy growth and predicted vacancy growth variables are *rescaled* by dividing by 10 (so that 1% vacancy growth in a local area corresponds to a value of 0.001, rather than 0.01), such that the coefficient estimates can be seen in the table for most specifications. See text for detailed explanation of instruments and controls. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

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 X Q1 occ mobility	-0.015***	-0.014***	-0.024***	-0.025***
	(0.002)	(0.002)	(0.004)	(0.005)
Log HHI X Q2 occ mobility	-0.013***	-0.007***	-0.016***	-0.010***
	(0.001)	(0.001)	(0.003)	(0.003)
Log HHI X Q3 occ mobility	-0.011***	-0.005***	-0.013***	-0.007*
	(0.001)	(0.001)	(0.003)	(0.004)
Log HHI X Q4 occ mobility	-0.002*	0.002	-0.005	0.000
	(0.001)	(0.001)	(0.003)	(0.004)
Log outside-occ. options		0.083***		0.060***
X Q1 occ mobility		(0.009)		(0.011)
Log outside-occ. options		0.109***		0.100***
X Q2 occ mobility		(0.007)		(0.009)
Log outside-occ. options		0.115***		0.107***
X Q3 occ mobility		(0.007)		(0.010)
Log outside-occ. options		0.114***		0.109***
X Q4 occ mobility		(0.007)		(0.010)
Vacancy growth			-0.001	-0.001
			(0.001)	(0.001)
Predicted vacancy growth			-0.012	-0.016
			(0.010)	(0.010)
Exposure control			0.008	0.004
			(0.008)	(0.008)
Observations	184,411	184,411	184,411	184,411
F-stat			164	87

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All regressions feature occupation-by-year and MSA-by-year fixed effects. Columns (a) and (b) show OLS regressions and columns (c) and (d) show 2SLS IV regressions. The reported F-stat for the 2SLS IV regressions is the Kleibergen-Paap Wald F statistic. The vacancy growth and predicted vacancy growth variables are *rescaled* by dividing by 10 (so that 1% vacancy growth in a local area corresponds to a value of 0.001, rather than 0.01), such that the coefficient estimates can be seen in the table for most specifications. Independent variables labelled "X Qi outward mobility" show the coefficient on an interaction term between the HHI or outside-occupation option index (respectively) with an indicator variable which takes the value 1 if the occupation in question is in the ith quartile of outward occupational mobility (where "Q1" represents the least outwardly mobile occupations, and so on). See text for detailed explanation of variables. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 5: 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 vacancy HHI instrument	0.081***				
	(0.003)				
Log outside-occ. options instrument	-0.712***				
	(0.044)				
Log HHI instrument *X* Q1 occ mobility		0.078***			
		(0.004)			
Log outside-occ options instrument *X* Q1 occ mobility		-0.566***			
		(0.049)			
Log HHI instrument *X* Q2 occ mobility			0.084***		
			(0.003)		
Log outside-occ options instrument *X* Q2 occ mobility			-0.785***		
			(0.056)		
Log HHI instrument *X* Q3 occ mobility				0.074***	
				(0.003)	
Log outside-occ options instrument *X* Q3 occ mobility				-0.742***	
				(0.052)	
Log HHI instrument *X* Q4 occ mobility					0.085***
					(0.004)
Log outside-occ options instrument *X* Q4 occ mobility					-0.818***
					(0.050)
Vacancy growth	-0.061	-0.461***	-0.023*	-0.097	-0.588***
	(0.044)	(0.163)	(0.012)	(0.062)	(0.064)
Predicted vacancy growth	-0.038	-0.473	-0.023	-0.099	0.085
	(0.100)	(0.373)	(0.219)	(0.237)	(0.083)
Exposure control	1.424***	1.578***	1.335***	1.464***	1.432***
	(0.037)	(0.054)	(0.059)	(0.059)	(0.049)
Observations	184,411	46,300	46,189	46,080	45,840

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All regressions have occupation-year and CBSA-year fixed effects. 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. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 6: First-stage regressions: outside-occ. options instrument

Dependent variable: log outside-occ. options (segmented by quartile of occ mobility in cols (b)-(e))					
	Full sample	By quartile of occ mobility			
		Q1	Q2	Q3	Q4
	(a)	(b)	(c)	(d)	(e)
Log outside-occ. options instrument	0.783***				
	(0.018)				
Log HHI instrument	-0.000				
	(0.000)				
Log outside-occ options instrument *X* Q1 outward mobility		0.689***			
		(0.018)			
Log HHI instrument *X* Q1 outward mobility		-0.001**			
		(0.000)			
Log outside-occ options instrument *X* Q2 outward mobility			0.819***		
			(0.019)		
Log HHI instrument *X* Q2 outward mobility			0.000		
			(0.000)		
Log outside-occ options instrument *X* Q3 outward mobility				0.816***	
				(0.021)	
Log HHI instrument *X* Q3 outward mobility				-0.000	
				(0.000)	
Log outside-occ options instrument *X* Q4 outward mobility					0.810***
					(0.024)
Log HHI instrument *X* Q4 outward mobility					-0.000
					(0.000)
Vacancy growth	0.001	-0.007	0.003***	-0.004***	0.004
	(0.001)	(0.005)	(0.000)	(0.001)	(0.005)
Predicted vacancy growth	0.018	-0.010	-0.021	0.036	0.034
	(0.017)	(0.041)	(0.023)	(0.032)	(0.022)
Exposure control	0.001	0.010	0.008	-0.006	-0.004
	(0.005)	(0.010)	(0.008)	(0.009)	(0.005)
Observations	184,411	46,300	46,189	46,080	45,840

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All regressions have occupation-year and CBSA-year fixed effects. In column (a) we run a first-stage regression for our outside-occupation option index instrument. In columns (b) through (e) we run separate first-stage regressions for our outside-occupation option index instrument, segmenting our data into four quartiles by outward occupational mobility. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Counterfactual wage effects of setting HHI to 200
(predicted wage effect in each cell, with total workers affected in parentheses)

	2,500< HHI <10,000	1,500< HHI <2,500	500< HHI <1,500	200< HHI <500	0< HHI <200
Lowest mobility	7.8%	5.8%	3.6%	1.1%	0
	(1.2m)	(1.5m)	(5.2m)	(5.2m)	(8.0m)
Q2 mobility	3.1%	2.3%	1.4%	0.4%	0
	(1.3m)	(1.2m)	(5.2m)	(7.6m)	(18.3m)
Q3 mobility	2.2%	1.6%	1.0%	0.3%	0
	(1.1m)	(0.9m)	(4.3m)	(9.4m)	(19.0m)
Q4 mobility	0	0	0	0	0
	(1.4m)	(1.1m)	(4.6m)	(7.2m)	(5.6m)

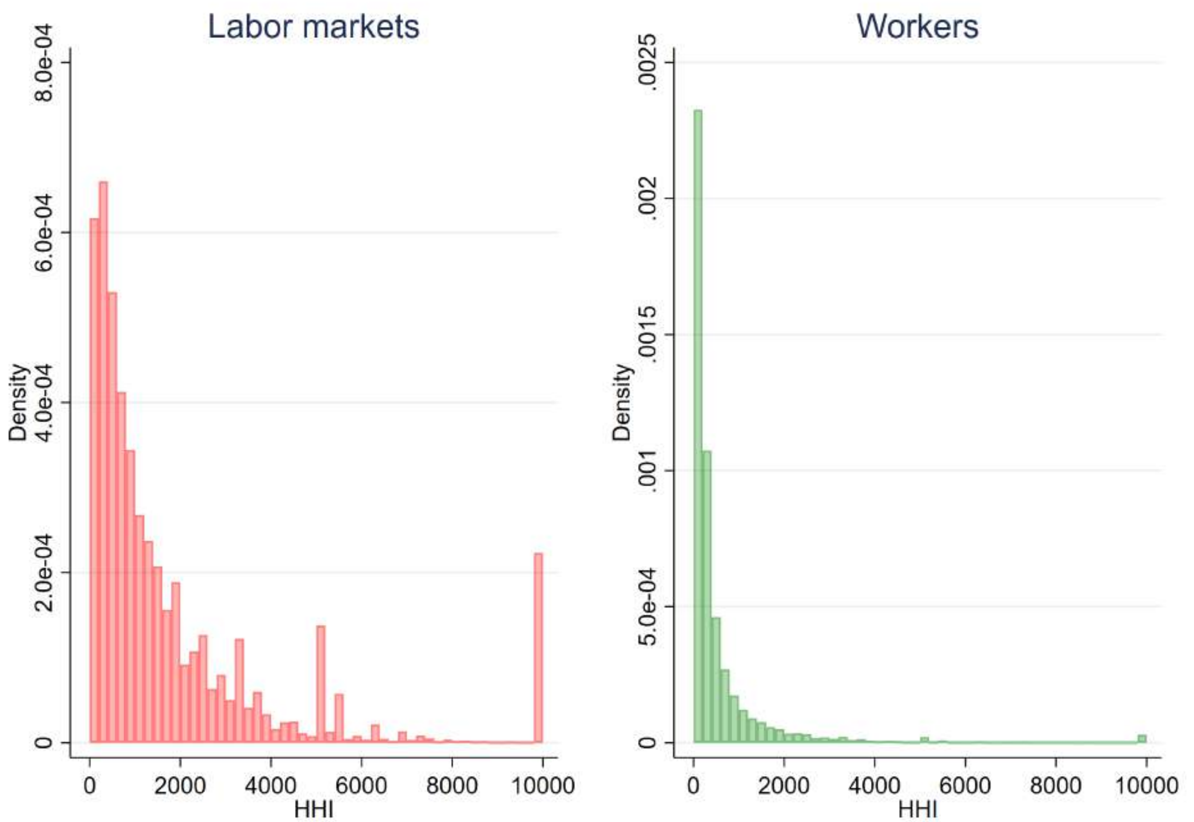
Notes: This table shows the estimated wage impact (and number of people affected in parentheses) of lowering the HHI to 200 in all occupation-city cells where it was greater than 200 in 2016. 200 is roughly the median HHI as experienced by workers. The estimated wage impact is calculated as the difference between the *actual* log HHI and the log of 200, multiplied by the estimated coefficient in our wage-HHI regressions (with the coefficient used corresponding to the appropriate quartile of occupational outward mobility, as estimated in Table 3 column (d)). The impact number in each cell in the table is the average impact across all workers in that cell: so, for example, for the 1.2 million workers in our data who are in occupations in the lowest quartile of outward mobility, and who are in occupation-city labor markets with an HHI greater than 2500, the *average* estimated wage impact of employer concentration on their wage is 7.8%. Note (1) this exercise implicitly holds productivity constant, and (2) our data set covers around 110 million workers in total, from the BLS OES occupation-by-metropolitan area employment and wage data.

Table 8: Twenty-five most-affected occupations (predicted wage effect of employer concentration 2% or greater)

Occupation	Employment in occupations w/ estimated effect 2% or greater	Share of national occ. w/ estimated effect 2% or greater	Represented-ness of occupation in BGT vacancy data
Security guards	969,230	.88	.63
Registered nurses	545,520	.19	2.1
Nursing assistants	488,040	.34	.75
Hairdressers, hairstylists, and cosmetologists	311,500	.88	.71
Nonfarm animal caretakers	161,540	.86	.69
Fitness trainers and aerobics instructors	158,310	.62	.61
Childcare workers	144,290	.25	1.3
Licensed practical and licensed vocational nurses	131,840	.19	1.4
Radiologic technologists	128,600	.64	.69
Pharmacists	117,360	.38	1.2
Emergency medical technicians and paramedics	112,290	.46	.59
Medical and clinical laboratory technologists	92,510	.55	.98
Phlebotomists	90,780	.75	1.4
Pharmacy technicians	84,360	.21	.91
Medical assistants	81,300	.13	.8
Respiratory therapists	78,360	.62	.99
Management analysts	68,230	.11	1.7
Lawyers	68,200	.11	.83
Librarians	66,730	.52	.5
Dentists, general	65,730	.62	1.1
Software developers, applications	64,990	.082	4.6
Bakers	64,300	.36	1.1
First-line supervisors of personal service workers	61,720	.32	.68
Real estate sales agents	60,190	.4	1.9
Massage therapists	58,820	.61	.83

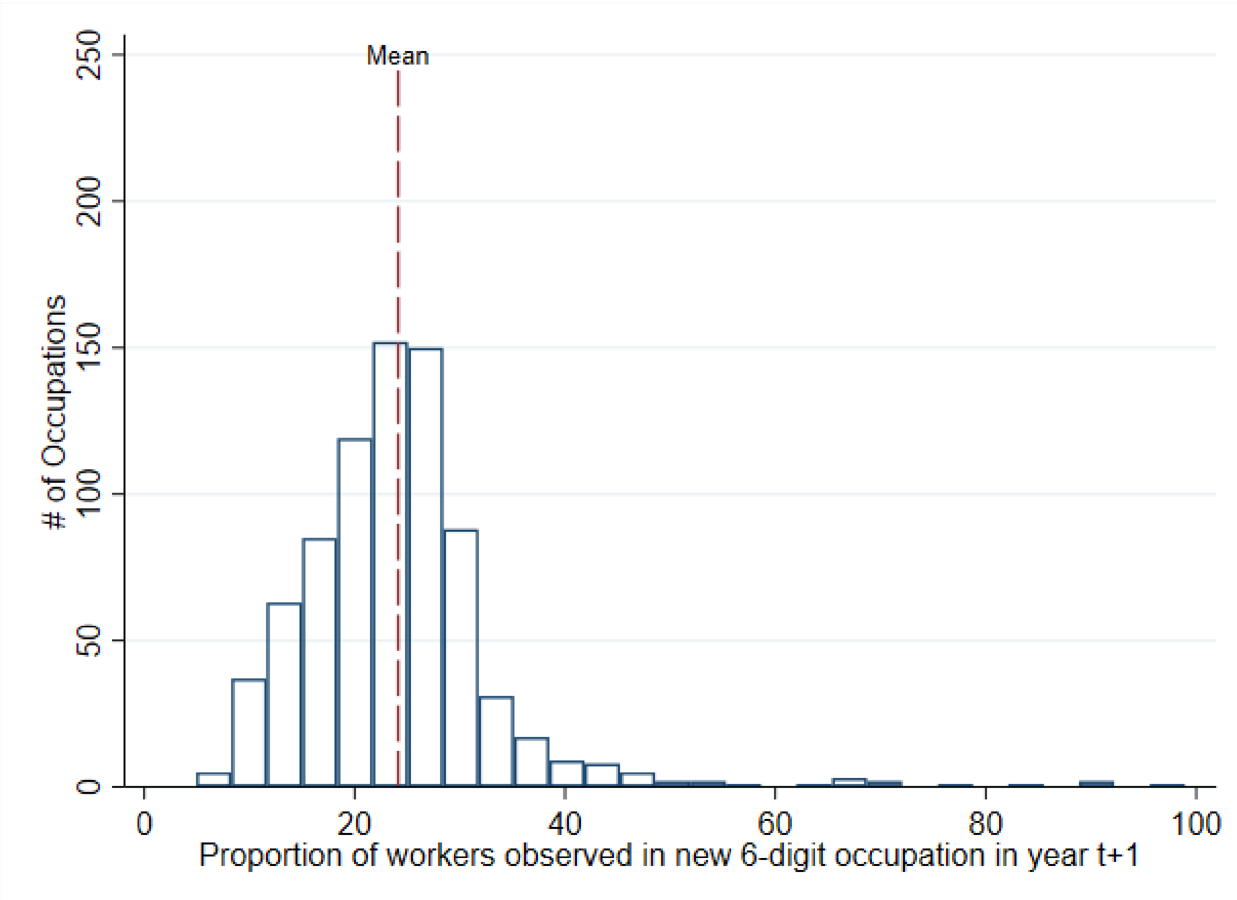
Notes: This table lists the largest occupations according to the number of workers in that occupation who experience an estimated wage impact of 2% or more as a result of employer concentration (see Table 7 for description of how this effect is calculated). We exclude "teachers and instructors, all other" since the classification of this group in the BGT data relative to OES may not be fully comparable, and since there is a large share of public sector employment of teachers. We also exclude occupations that are very under-represented in the BGT vacancy data relative to overall employment (with a cutoff with represented-ness<0.5, or around the 33rd percentile) – which, from this list, includes Personal care aides, Bartenders, Farmworkers and laborers (crop, nursery, and greenhouse), Dental hygienists, Lifeguards and ski patrol, Operating engineers and other construction equipment operators, Information and record clerks (all other), and Business operation specialists (all other), as well as some overwhelmingly public sector occupations (School bus drivers, Firefighters, Police officers, and Postal service mail carriers). We include the degree of represented-ness of each occupation in the BGT vacancy data in the table, because the more underrepresented an occupation is in the BGT vacancy data, the more likely we are overestimating the degree of employer concentration in these occupations and therefore overestimating the effect of concentration. On the other hand, in better-represented occupations we might be more confident that we are accurately prioritizing these occupations.

Figure 1: Histogram of employer HHI across occ-city labor markets and across workers, 2016



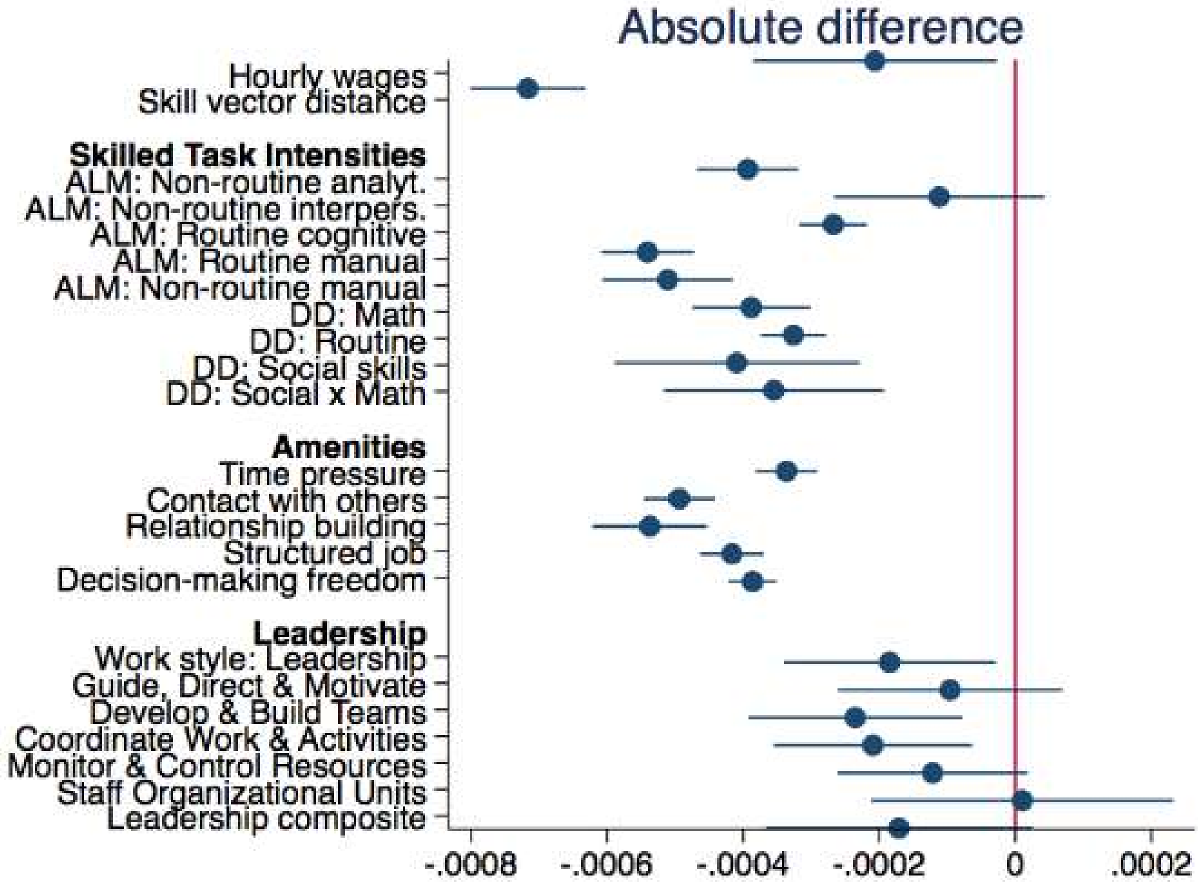
Note: HHI measured using Burning Glass Technologies vacancy data. HHI is calculated at the level of a labor market, defined as SOC 6-digit occupation by metro area by year. Our data covers occupation-metro area labor markets which include 110m of the 140m workers in the U.S. labor market in 2016.

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



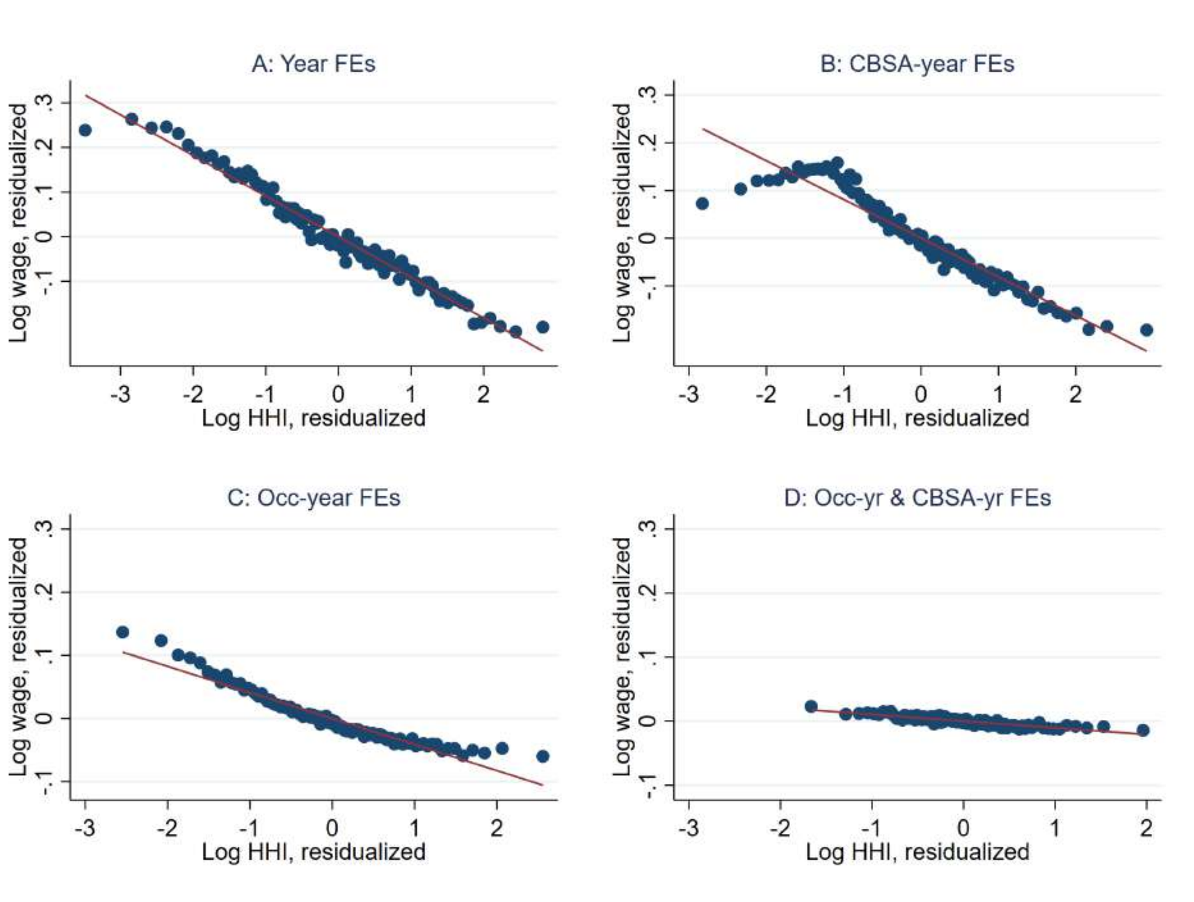
Distribution of the "occupation leave share" – the probability that a worker will leave their occupation conditional on leaving their job – by occupation. Occupation leave share is calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample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

Figure 3: Occupational transitions and occupational characteristic similarity



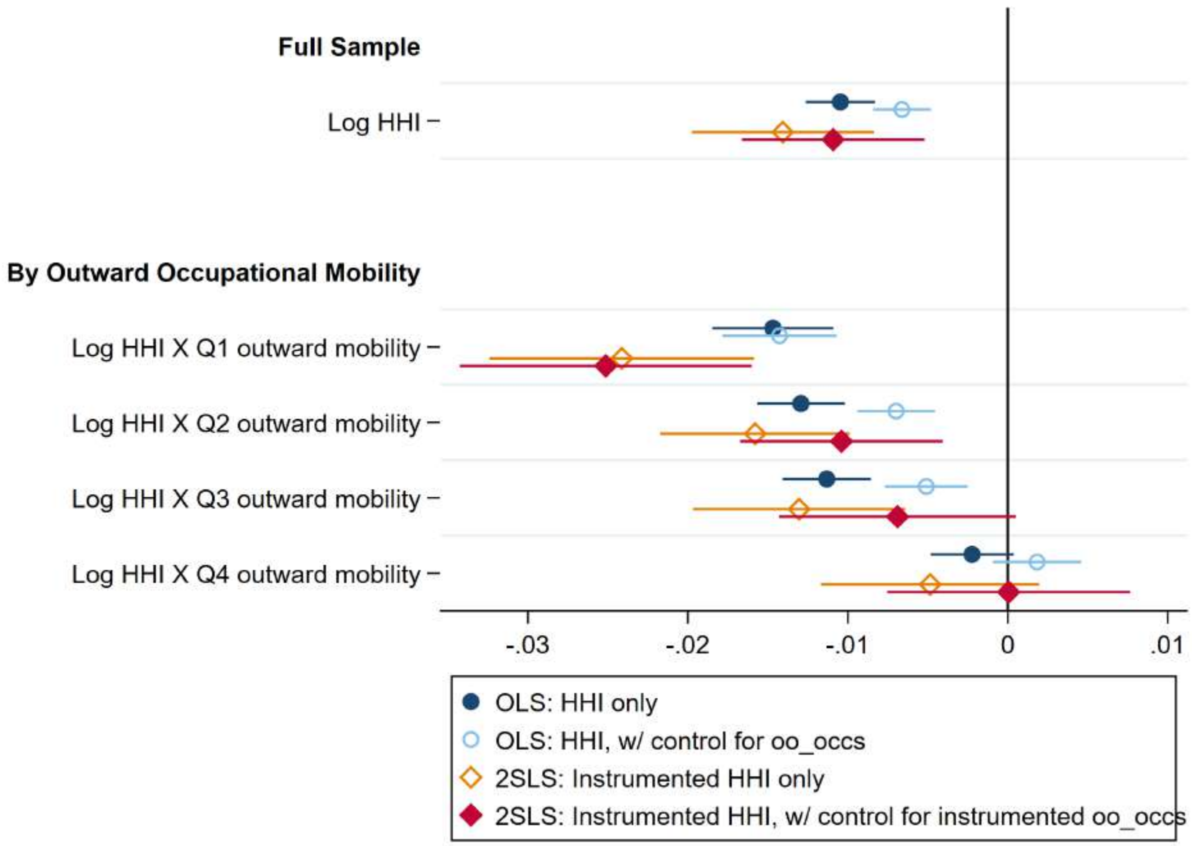
Note: This plot shows the coefficients and 95% confidence intervals from regressions of occupation transition shares $\pi_{o \to p}$, calculated from Burning Glass Technologies Resume data, on occupational characteristics: $\pi_{o \to p} = \alpha_o + \beta f(X_{occ\ o \to p}) + \gamma f(\Delta w_{o \to p}) + \epsilon_{op}$ where the function $f(\cdot)$ represents the absolute difference in characteristic between occupation o and p, and α_o is occupation o fixed effect. Regressions also include absolute avg. hourly wage differences (except for the amenities regressions). Standard errors are clustered at the origin occupation level. Appendix Figure A11, shows an analogous exercise: tghe coefficients on the regression of occupation transition shares on the *relative* difference in characteristics between the pairs of occupations.

Figure 4: Correlations between wage and HHI



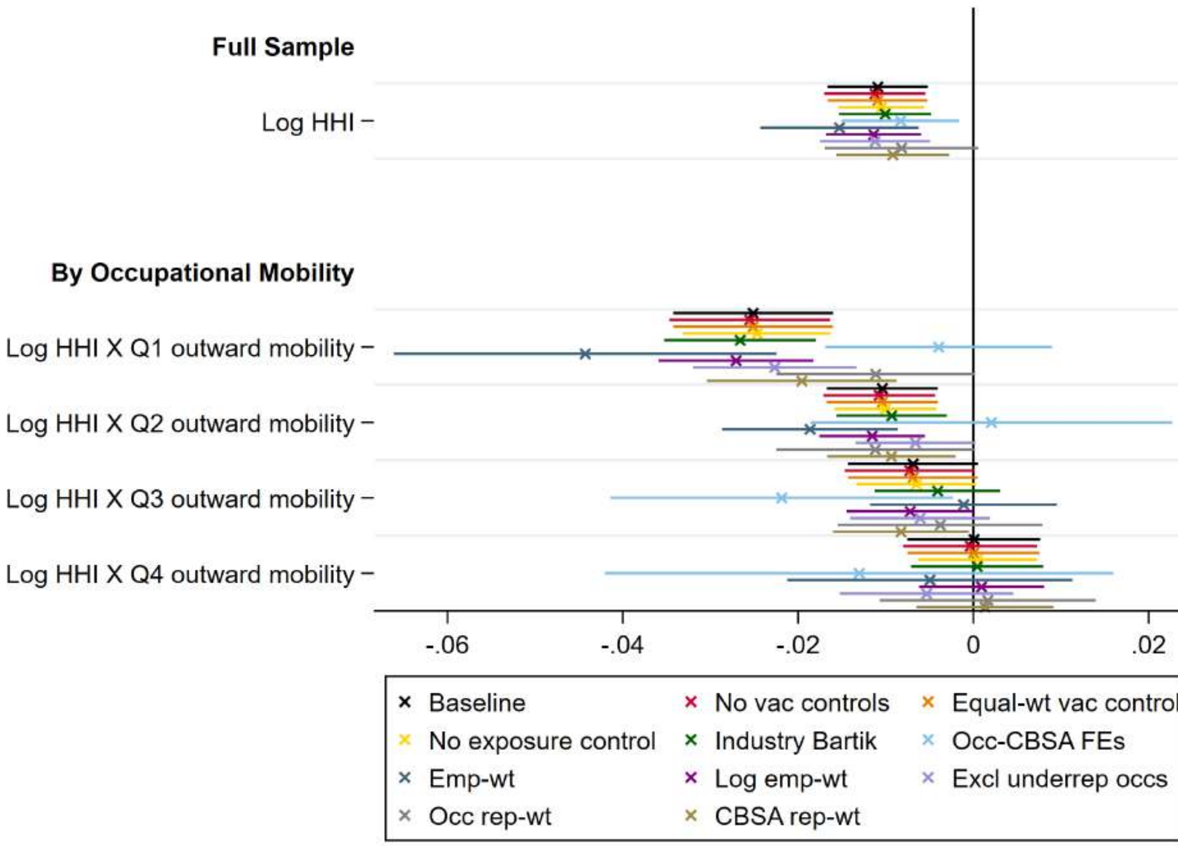
Note: Figure shows binned scatter plots of the correlation between average log wages and log within-occupation HHI index for occupation-CBSA cells over 2013–2016, residualized on different combinations of fixed effects. Regression coefficients for the line of best fit on each graph are: A: -0.09, B: -0.08, C: -0.04; D: -0.01.

Figure 5: Coefficients on wage-HHI regressions



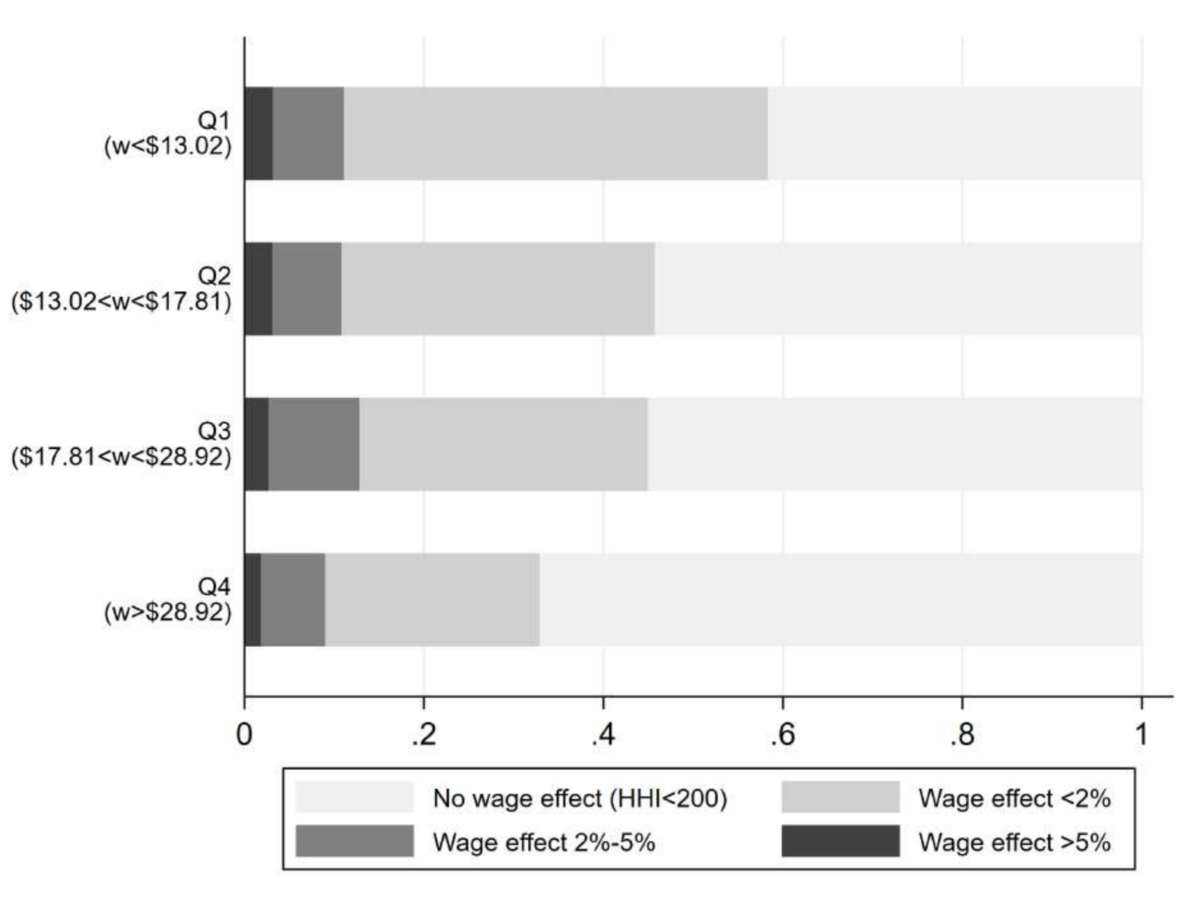
Note: Coefficients on log HHI and 95% confidence intervals from our baseline regressions of occupation-CBSA wages on employer HHI. Navy: OLS regression of wages on HHI; Light blue: OLS regression of wages on HHI with control for outside-occupation job options; Orange: 2SLS regression of wages on instrumented HHI; Red: 2SLS regression of wages on instrumented HHI, with control for (instrumented) outside-occupation job options. The top panel presents these coefficients for the full sample (estimates reported in Table **??**) and the bottom panels present the coefficients when estimated separately by quartile of outward occupational mobility (estimates reported in Table **??**). Regressions use annual data at the level of occupation-by-CBSA labor markets, span 2013-2016, and include occupation-year and CBSA-year fixed effects as well as controls as described in the main text. Standard errors are clustered at the CBSA level.

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



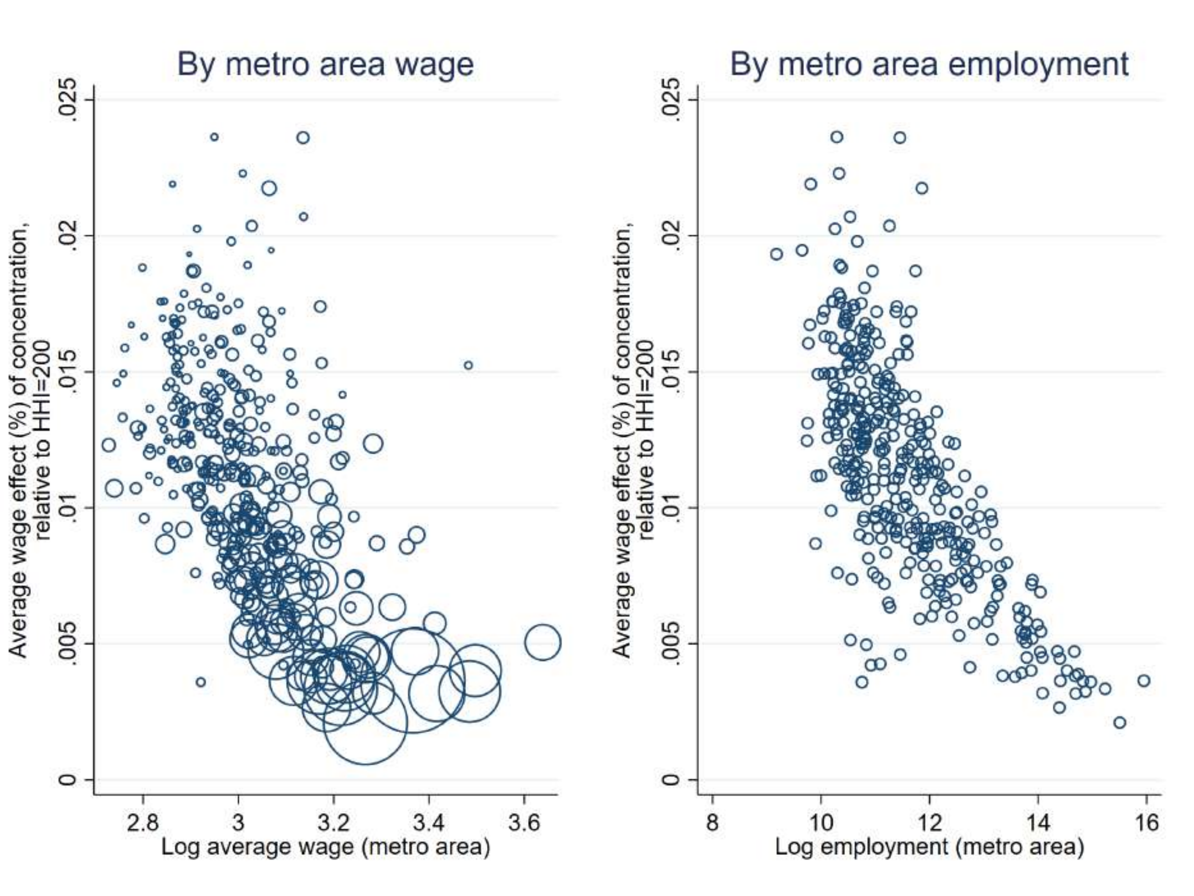
Note: Coefficients on log HHI and 95% confidence intervals from our baseline 2SLS IV regressions of occupation-CBSA wages on instrumented employer HHI, across various robustness checks. Specifications correspond to those in Tables A7-A10. The top panel presents estimated coefficients for the full sample and the bottom panels present the coefficients when estimated separately by quartile of outward occupational mobility. Regressions use annual data at the level of occupation-by-CBSA labor markets, span 2013-2016, and include occupation-year and CBSA-year fixed effects as well as controls as described in the main text, unless the robustness check specifies otherwise. Standard errors are clustered at the CBSA level.

Figure 7: Rough estimates of wage effects of concentration, across the income distribution, relative to a counterfactual with HHI=200



Note: This figure shows what share of workers may have experienced different degrees of wage suppression as a result of employer concentration, across the U.S. income distribution. We estimate this as described in section 6: we use our coefficient estimates for the effect of the HHI on the wage, by quartile of outward occupational mobility, to calculate a counterfactual wage for each occupation-city labor market *if* the HHI had been 200.

Figure 8: Average estimated effect of employer concentration relative to HHI=200, by metro area



Note: This figure shows the average estimated wage effect of concentration in each metro area relative to an HHI of 200 (holding all else constant), plotted against the average hourly wage in that metro area in 2016 (left panel) and the employment in that metro area in 2016 (right panel) according to our BLS OES data. Bubble size for the left hand graph represents metro area employment in 2016. We estimate the wage effect of employer concentration as described in section 6: we use our coefficient estimates for the effect of the HHI on the wage, by quartile of outward occupational mobility, to calculate a counterfactual wage for each occupation-city labor market *if* the HHI had been 200.

APPENDIX

A Appendix: Conceptual framework: more detail

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

Conceptual framework: effect of concentration on wages

Timing. Each period has two phases. During the hiring phase, workers exit or enter the labor market, employed workers bargain with their employer, and new workers are hired. During the production phase, workers remain employed at their firms and receive the wages determined during the hiring phase. At the start of the next period, workers may leave their firms and/or renegotiate their wage with their current employer. More details follow. To simplify exposition, we omit location subscripts, but all expressions in this section should be assumed to apply to a specific city k.

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

Wage bargaining. Each worker who is currently employed at the start of the period Nash-bargains with her employer i over the wage.[79] The outcome is a wage w_i, equal to the value of the worker's outside option if bargaining breaks down and she leaves her job, oo_i, plus a share β – reflecting worker bargaining power – of the match surplus:

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

If wage bargaining breaks down, the worker leaves her job and becomes a job seeker. We

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

assume that all workers at a given firm have the same set of outside options, such that all workers at any given firm will in equilibrium receive the same wage.

Vacancies. After labor market exit and wage bargaining have happened, firms post vacancies to fill the positions which have been vacated by either worker exit or the breakdown of wage bargaining. There is no cost to post vacancies to fill existing positions, but firms cannot post vacancies for new positions (i.e. there is no firm growth).[80] A firm can only post one vacancy for each position they wish to fill. Vacancies are posted as take-it-or-leave-it wage offers, with the posted wage equal to the wage the firm is paying to its other workers (which was decided on in the wage bargaining process described above). This constraint – that similar workers are paid similar amounts – is often observed in practice in firms and could be motivated by fairness concerns or explicit internal pay hierarchies or bargaining agreements.[81]

Job seekers. Job seekers are comprised of workers who have newly entered the labor market, workers who have left their previous job because their match broke down at the start of the period, and workers who were unemployed in the previous period. Job search can only take place at the start of the period, during the hiring phase. There is no on-the-job-search. Each job seeker applies to all *feasible* local employers $j \in N$, but search and matching frictions mean that each job seeker will receive exactly one job offer during each hiring phase. Specifically, a ball-urn matching procedure randomly matches pairs of job seekers and firms. For a given job seeker i, the probability of receiving an offer from each firm j is α_j, and the probability of receiving no offers is $1 - \sum_j \alpha_j$. If job seekers do not accept the job offer they receive, or if they receive no job offers, they remain unemployed for the period, receiving unemployment benefit b. We assume b is strictly lower than the wages offered by any feasible employer, such that a job seeker who receives an offer always accepts it. In our model this condition will always be satisfied as long as the productivity of all jobs is greater than the unemployment benefit: this is because the wage offered to each worker by firm j is equivalent to the wage bargained by the existing workers at firm j, which itself is strictly greater than unemployment benefit b. [82]

[80]This assumption is not necessary for our conclusions but keeps things simple.

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

[82]The analysis above assumes that workers care only about money. If we consider instead utility over

Outside option value for employed workers. The wage at each firm is determined by the bargain with employed workers, which in turn depends on the value of the outside option for these employed workers. What is this outside option value? The outside option for an employed worker bargaining with her employer is to leave her current job and become a job seeker. She does not know with certainty what her outcome will be as she will be matched with at most one feasible job if she leaves her current job. Her expected wage will be a weighted average of the wages paid by each feasible firm j, w_j, weighted by the probability of being matched with each feasible firm j, α_j, as well as the unemployment benefit b multiplied by the probability of receiving no job offers $1 - \sum_{j \neq i}^{N} \alpha_j$:

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

We assume that each firm-worker match has weakly positive surplus, such that the bargained wage is always weakly greater than the outside option value. This means that, in equilibrium, no bargaining session will break down.[83]

Equilibrium wage. Consider a clearly-defined labor market where all workers and all jobs are perfect substitutes. The bargained wage for workers at firm i (which is also paid to new hires at firm i) satisfies:

$$\begin{aligned} w_i &= \beta p_i + (1-\beta) oo_i & (15)\\ &= \beta p_i + (1-\beta)\left(\sum_j \alpha_j \cdot w_j + \left(1 - \sum^{j} \alpha_j\right) \cdot b\right) & (16) \end{aligned}$$

What do the probabilities of being matched with each feasible firm, α_j, correspond to? We assume that the probability of the offer a job seeker receives being from a particular firm j is proportional to the share of vacancies posted by that firm j, as a share of all vacancies in the labor market, σ_j. This is in the same spirit as Burdett and Mortensen (1980) and Jarosch et al. (2019) who use the employment share to proxy for the likelihood of getting a job offer from a given firm. (That is: in the urn-ball matching process, the probability of picking out a given firm is represented by the share of all job vacancies posted by this

pay and unemployment benefits, but also work and leisure, the argument still holds. In this case, the productivity of all jobs must be greater than the money-equivalent utility value of unemployment (including the unemployment benefit and any utility or disutility of unemployment relative to work). As long as all workers have the same utility/disutility of work and unemployment, the utility value of the wage bargained by existing workers at any firm j will be greater than the utility value of staying in unemployment such that all workers would accept a job offer from firm j.

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

firm). In a labor market with atomistic infinitesimally small firms, this procedure would lead to every job seeker receiving a match from a feasible employer each period. However, in a labor market with some large employers, the large employer can threaten *not* to re-employ a current worker if the worker leaves the firm but is re-matched with them in the job search process. That is, if bargaining breaks down between a given worker and her employer, firm j, and she chooses to become a job seeker, the random matching process may re-match that worker to her initial employer j, but in this case employer j refuses to re-hire her, as in Jarosch et al. (2019).[84] These assumptions imply the following wage equation for a worker at firm i:[85]

$$w_i = \beta p_i + (1-\beta)\left(\sum_{j \neq i} \sigma_j \cdot w_j + \sigma_i \cdot b\right) \tag{17}$$

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

$$w_i = \beta p_i + \beta(1-\beta)\sum_{j \neq i} \sigma_j p_j + \beta(1-\beta)^2 \sum_{j \neq i} \sigma_j \sum_{k \neq j} \sigma_k p_k + ... \tag{18}$$

$$+ (1-\beta)\sigma_i b + (1-\beta)^2 \sum_{j \neq i} \sigma_j^2 b + (1-\beta)^3 \sum_{j \neq i} \sigma_j \sum_{k \neq j} \sigma_k^2 b + ... \tag{19}$$

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

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

$$\bar{w} = \beta\left(\sum_i \sigma_i p_i + (1-\beta)\sum_i \sigma_i \sum_{j \neq i} \sigma_j p_j + (1-\beta)^2 \sum_i \sigma_i \sum_{j \neq i} \sigma_j \sum_{k \neq j} \sigma_k p_k + ...\right) \tag{20}$$

[84] In a single-period game, this strategy would not be time-consistent for the employer, as they are unable to fill the vacancy for the period and so lose the potential for positive surplus from re-hiring the worker. However, in a multiple period game with repeated interactions between firms and workers, and/or firm reputations, it can be in the firm's interest to avoid re-hiring a worker who has quit the firm in order to maintain the firm's reputation in future bargaining rounds with this or other workers - even if this comes at the cost of an unfilled vacancy in this period.

[85] Note that worker i's probability of becoming unemployed if she quits her current firm is identical to the her employer's share of the local labor market. We abstract away from other unemployment caused by search frictions, but this could be incorporated into the model at the cost of some simplicity without changing the substantive conclusions.

$$+ (1-\beta)b\left(\sum_i \sigma_i^2 + (1-\beta)\sum_i \sigma_i \sum_{j\neq i}\sigma_j^2 + (1-\beta)^2 \sum_i \sigma_i \sum_{j\neq i}\sigma_j \sum_{k\neq j}\sigma_k^2 + ...\right) \tag{21}$$

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

$$\Omega_r = \underbrace{\sum_i \sigma_i \sum_{j\neq i}\sigma_j \sum_{k\neq j}\sigma_k ... \sum_{m\neq n}\sigma_m \sum_{p\neq m}\sigma_p^2}_{\text{with } r \text{ summation terms or "steps" in the expression}} \tag{22}$$

such that the first order concentration index Ω_1 is the sum of the squared employer shares or HHI, $\Omega_1 = \sum_i \sigma_i^2$, the second order concentration index is $\Omega_2 = \sum_i \sigma_i \sum_{j\neq i}\sigma_j^2$, and so on. Define $\Omega_0 = 0$. We can then rewrite the average wage equation (23) as

$$\begin{aligned}\bar{w} = \beta\bar{p}\left(1 + \sum_{n=1}^{\infty}(1-\beta)^n\left(1 - \sum_{r=1}^{n}\Omega_r\right)\right) + (1-\beta)b\left(\sum_{n=1}^{\infty}(1-\beta)^{n-1}\Omega_n\right) \\ -\beta\left((1-\beta)\sum_i \sigma_i^2\tilde{p}_i + (1-\beta)^2\sum_i \sigma_i \sum_{j\neq i}\sigma_j^2\tilde{p}_j + (1-\beta)^3\sum_i \sigma_i \sum_{j\neq i}\sigma_j \sum_{k\neq j}\sigma_k^2\tilde{p}_k + ...\right)\end{aligned} \tag{23}$$

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

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

Note also that the relationship between employer shares σ_j and firm productivity p_j factor into the wage in the third term, but *also* in the first term in the sense that average productivity $\bar{p}$ is determined by the productivity of each employer, and the share of that employer in the labor market. If a particular firm j is more productive than average, and it grows to become a large share of the labor market, average productivity will rise, pushing the average wage up. However, the third term mitigates this effect somewhat: it reflects the fact that, if it is the largest firms which are the most productive, the passthrough of average productivity to average wages will be lower than if it is the smallest firms which are the most productive, because the largest firms are in fewer workers' outside option set.

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

$$\bar{w} = (1 - (1-\beta)HHI)\,\bar{p} + (1-\beta)HHIb - \beta(1-\beta)\sum_i \sigma_i^2 \tilde{p}_i \quad (24)$$

That is, this expression suggests that higher employer concentration increases the weighting on the unemployment benefit b, and decreases the weighting on the productivity of labor p, relative to a world with no employer concentration. If all firms have the same productivity $p_i = \bar{p} \forall i$, or if the correlation of market shares and firm productivity is small, the last term is close to zero and the wage is simply a concentration- and bargaining power-weighted average of productivity p and unemployment benefit b.[86]

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

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

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

Incorporating outside-occupation options

The conceptual framework explained above, however, assumes that all jobs and workers are perfectly substitutable in a clearly delineated labor market. As discussed extensively in this paper, this is rarely the case in practice. Instead, workers can switch between occupations – but differentially so across different occupations. Ideally, we would be able to delineate which firms in outside occupations are in a worker's feasible labor market for any given worker, and estimate the probability that a worker would receive a job in that outside occupation. In practice, we cannot. To work with commonly-available data definitions and publicly-available data, we must instead work with the occupational definitions from the SOC 6-digit classification scheme. We therefore extend our framework above by defining the primary labor market for workers as their local occupation o and incorporating an outside option term reflecting the value of moving to jobs in other occupations. Refer back to our equation (17) for the bargained wage in firm i,

$$w_i = \beta p_i + (1-\beta)\left(\sum_{j\neq i}\sigma_j \cdot w_j + \sigma_i \cdot b\right) \tag{25}$$

but now edit this equation to reflect that some feasible jobs are in workers' own occupation o whereas some are in other occupations p:[87]

$$\begin{aligned} w_{i,o} = \beta p_{i,o} + (1-\beta)\Bigg(& \underbrace{Prob(o\to o)\cdot\sum_{j\neq i}\sigma_{j,o}\cdot w_{j,o}}_{\text{own occupation options } oo^{own}} + \underbrace{\sum_{p\neq o}^{N^{occs}} Prob(o\to p)\sum_{l}^{N_p}\sigma_{l,p}\cdot w_{l,p}}_{\text{outside occupation options } oo^{occs}} \\ & + \underbrace{Prob(o\to o)\sigma_{i,o}\cdot b}_{\text{unemployment}}\Bigg) \end{aligned} \tag{26}$$

[87]Where employer share $\sigma_{j,o}$ now refers to firm j's share of vacancies *within* occupation o.

This expression states that the bargained wage for workers in firm i in occupation o is a function of their productivity $p_{i,o}$, the outside option value of moving to other firms j in their own occupation o, the outside option value of moving to other firms l in other occupations p, and the outside option value of moving to unemployment and receiving benefit b. Note that the expression for the probabilities of a worker from firm i in occupation o matching with firm l in occupation p if she becomes a job seeker is now the product of firm l's vacancy share in occupation p, $\sigma_{l,p}$, as well as the probability of the worker getting her job offer from some firm in occupation p conditional on having left her job in occupation o, $Prob(o \to p)$.[88]

To take this expression to the data, we first note that if employers' vacancy shares are relatively similar to their current employment shares, $\sum_l^{N_p} \sigma_{l,p} \cdot w_{l,p}$ can be approximated simply the average wage in local occupation p. This eliminates the need for us to consider the reflection problem that wages in occupation o affect wages in occupation p and vice versa - instead, we can use data on the average wage in each local occupation p to control directly for the effect of wages in occupation p on occupation o (as we do in our empirical implementation). Second, we approximate the probability that a job seeker from occupation o receives her job offer for a job in occupation p with the empirical transition shares of workers moving from occupation o to occupation p conditional on leaving their job, $\pi_{o \to p}$ (measured from the Burning Glass data). This gives us the wage expression

$$w_{i,o} = \beta p_{i,o} + (1-\beta)\left(\pi_{o\to o}\sum_{j\neq i}\sigma_{j,o}w_{j,o} + \sum_{p\neq o}^{N^{occs}}\pi_{o\to p}\bar{w}_p + \pi_{o\to o}\sigma_i b\right), \tag{27}$$

where $\sum_{p\neq o}^{N^{occs}} \pi_{o\to p}\bar{w}_p$ corresponds closely to our outside-occupation option index oo^{occs} (the transition-weighted average wage in local occupations p).[89] By the same procedure of iterative substitution of wages in other local firms j in the same occupation o, rearrangement, and substitution of our higher order concentration indices $\Omega_{r,o}$ (where subscript o denotes that this is the concentration index for employers in local occupation o), we can derive an expression for the average wage in occupation o which incorporates our outside-occupation

[88]For simplicity, we do not consider outside options in other cities and only consider options within the worker's own city k here, but this framework can easily be extended to consider options in other cities also. While these components of outside options certainly matter, the data suggests that for most workers they will be less important than the jobs in other occupations in workers' current city. Occupational mobility is substantially higher than geographic mobility: only around 3% of U.S. workers move between metropolitan areas each year. Note also that this expression assumes implicitly that each firm only employs workers of one occupation (or, alternatively, that the worker's own firm i can only refuse to re-employ that worker if she is re-matched with a job in her initial occupation o, but not if she is re-matched with firm i with a job in a new occupation p.

[89]Our empirical implementation of the outside-occupation option index also adjusts our national transition probabilities $\pi_{o\to p}$ by the local relative employment share in occupation p compared to the national average.

option index:

$$\bar{w} = (\beta\bar{p}_o + (1-\beta)oo_o^{occs})\left(1 + \sum_{n=1}^{\infty}(1-\beta)^n\pi_{o\to o}^n\left(1 - \sum_{r=1}^{n}\Omega_{r,o}\right)\right) + b\left(\sum_{n=1}^{\infty}(1-\beta)^n\pi_{o\to o}^n\Omega_{n,o}\right)$$
$$- \beta\left((1-\beta)\pi_{o\to o}\sum_i \sigma_i^2\tilde{p}_{i,o} + (1-\beta)^2\pi_{o\to o}^2\sum_i\sigma_i\sum_{j\neq i}\sigma_j^2\tilde{p}_{j,o} + (1-\beta)^3\pi_{o\to o}^3\sum_i\sigma_i\sum_{j\neq i}\sigma_j\sum_{k\neq j}\sigma_k^2\tilde{p}_{k,o} + \ldots\right) \quad (28)$$

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

$$\bar{w}_o = (1 - (1-\beta)Prob(o\to o)HHI_o)(\alpha\bar{p}_o + (1-\alpha)oo^{occs}) + (1-\beta)Prob(o\to o)HHI\cdot b$$
$$- \beta(1-\beta)Prob(o\to o)\sum_i\sigma_i^2\tilde{p}_{i,o} \quad (29)$$

where $\alpha = \frac{\beta}{1-Prob(o\to o)(1-\beta)}$.

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

As before, employer concentration within workers' own occupation increases the relative likelihood of workers ending up unemployed if they quit their job, increasing the weighting on b the unemployment benefit in the wage bargain and reducing the weighting on other jobs (productivity $\bar{p}_o$ and the value of moving to outside options outside workers' own occupation oo^{occs}). In addition, though, there is now an interaction with $\pi_{o\to o}$, the likelihood of the worker staying in her occupation if she leaves her job. The more likely she is to stay in her own occupation if she leaves her job – i.e., the less likely she is to be able to find a job in a different occupation – the more employer concentration in her own occupation matters for her wage. Finally, as before, note that there is an interaction with worker bargaining power β. The more bargaining power a worker has over the match surplus, the less the outside option matters in the wage bargain and therefore the less employer concentration matters for the wage.[90]

[90] Note that if there is no possibility of finding a job in another occupation (i.e. $\pi_{o\to o} = 1$), this expression

B Appendix: Burning Glass Technologies Vacancy Posting Data

This and the following section contains further information about the two data sets from Burning Glass Technologies ("BGT") which we use in this paper, discussed more briefly in Section 3.

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.

We use two data sets from BGT in this paper: vacancy data (online job postings) and resume data. The BGT vacancy data on online job postings has been used in several other academic papers, including Azar et al. (2020a) and Hazell and Taska (2019). We discuss this data further below. The BGT resume data set is a new proprietary data set of 16 million unique U.S. resumes spanning years over 2002–2018. We provide a detailed description of this data 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 U.S. job vacancies. They only measure *new* vacancy postings, so do not re-capture a given vacancy if it is left open for a long time.[91] They use proprietary algorithms to de-duplicate vacancies (for example if the same vacancy is posted on different websites).

We use BGT's vacancy data for the years 2013–2016. Over this four year period, we have data on 74 million vacancies, to which BGT has assigned the SOC 6-digit occupation and metropolitan area to which they apply. Of these, about one third or 24.8 million had no information about the employer, while 49.2 contained employer names (of a total of 1.02 million employers).

becomes identical to the expression for the average wage in the first part of this section, equation (24).

[91] 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 et al., 2014).

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

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

We do not capture relationships where one company owns another company but the names are not similar enough to identify this easily: this means that in some cases we will understate employer concentration by attributing vacancies to different employers. On the other hand, our employer categorization means that individual establishments of an employer – or even franchises of a brand – will be treated as the same employer, which may overstate employer concentration if pay decisions are made at the level of the establishment or franchise rather than the overall firm or brand group. It is not entirely conceptually clear whether employer concentration should be measured at the level of the establishment or the firm. On the one hand, individual establishments often have independent hiring policies; but on the other hand, multi-establishment firms often have common internal pay scales meaning they effectively operate as one employer across establishments. Similarly, it is not entirely conceptually clear whether franchises of the same brand should be considered as separate employers. One the one hand, they are independent businesses; on the other hand, franchisees' human resources policies are often at least partly dictated by the franchisor (Weil, 2014), and there have been a number of prominent cases where franchisors have required franchisees not to 'poach' each others' employees (with (Krueger and Ashenfelter, 2018) estimating that over half of major franchisors have no-poaching agreements in their franchise contract).

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. (2020a)). 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. (2020a) 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 49.2 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: 841 employers each posted more than 10,000 vacancies online over 2013–2016, and these 841 employers are responsible for a total of 32.4 million vacancies. While many of the small employers in our data are only present in the data for a subset of the four years 2013–2016, a subset of large employers are present for all four years (as shown in Table A1): as a result more than 75% of all vacancies in our database are listed by employers which are present in all four years of the sample. If employers hire a lot in any one year, they also tend to hire a lot in other years: the correlation of local occupation-specific vacancies by employer from one year to the next is 0.77.

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".[92] 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 later in the 'representativeness" section, 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 U.S.

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

[92]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.

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. (2020a), 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. (2020a) for more details. Note in particular that Azar et al. (2020a) 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 more systematically, we calculate the share of all vacancies in our data represented by each SOC 6-digit occupation, and compare this to the share of all employment in the BLS occupational employment statistics which is represented by each SOC 6-digit occupation. 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 U.S., 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.

Of the largest occupations in the data, sales occupations are relatively equally represented in BGT data as compared to the BLS OES; registered nurses, truck drivers, and computer and software occupations 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 underrepresentativeness 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-city labor markets even when controlling for occupation and fixed effects: that is, that within a given occupation, the lower-wage cities are underrepresented and within a given city, the lower-wage occupations are underrepresented. For our normative conclusions in terms of estimating the *aggregate number* of workers who are affected by employer concentration, and creating a ranking of which occupations are more or less affected, underrepresentativeness of the data is more of a concern: if some occupations are underrepresented in the BGT resume data, they may appear more concentrated when in fact, it is simply the case that online vacancy postings reflect fewer of the true vacancies available in the labor market for that occupation. As such, we take care when drawing these conclusions not to isolate specific occupations which appear to be severely underrepresented in our data.

C Appendix: Burning Glass Technologies Resume Data

The BGT resume data set is a new proprietary data set of 16 million unique U.S. 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.[93]
- Eliminate all jobs held before 2001.
- Eliminate all resumes with non-U.S. addresses.
- Eliminate any jobs which are listed as having lasted less than 6 months, to ensure that we are only capturing actual jobs rather than short-term internships, workshops etc.

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

From each of these resumes, we extract a separate observation for each job a worker was observed in, in each year they were observed in that job. (We define a 'job' as a unique job title-employer-occupation combination, meaning that a worker can in theory switch job

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

but remain at the same employer and/or in the same occupation.) For each job, we retain information on the SOC 6-digit occupation code. This gives us a data set of 80.2 million worker-job-occupation-year observations, where each worker might be observed in multiple jobs in the same year (either if jobs were held concurrently or the worker switched from one job to another within a given year).

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

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). We then aggregate them, reweighting by the average proportion of employment in each of these age categories in that occupation in the U.S. labor force over 2012–2017 (from the BLS Occupational Employment Statistics). Our aggregate occupational mobility matrix has therefore been reweighted to correspond to the empirical within-occupation age distribution in the labor force, reducing the potential for bias arising from the skewed age distribution of our sample.

Summary statistics

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

Job number and duration: The median number of jobs on a resume is 4, and more than 95% of resumes list 10 or fewer jobs (note that a change of job under our definition could include a change of job title or occupation under the same employer). The median length job was 2 years, with the 25th percentile just under 1 year and the 75th percentile 4 years. The median span of years we observe on a resume (from date started first job to date ended last job) is 12 years. Table A2 shows more information on the distribution of job incidences and job durations on our resumes.

Gender: BGT imputes gender to the resumes using a probabilistic algorithm based on the names of those submitting the resumes. Of our observations, 88% are on resumes where BGT was able to impute a gender probabilistically. According to this imputation, precisely 50% of our observations are imputed to be more likely to be male, and 50% are imputed to be more likely to be female. This suggests that relative to the employed labor force, women are very slightly over-represented in our data. According to the BLS, 46.9% of employed people were women in 2018.

Education: 141.3 million of our observations are on resumes containing some information about education. The breakdown of education in our data for these data points is as follows: the highest educational level is postgraduate for 25%, bachelor's degree for 48%, some college for 19%, high school for 8% and below high school for less than 1%. This substantially overrepresents bachelor's degree-holders and post-college qualifications: only 40% of the labor force in 2017 had a bachelor's degree or higher according to the BLS, compared to 73% in this sample (full comparisons to the labor force are shown in Figure A2). It is 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 therefore not be a reason for 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,[94] 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.[95]

Figure A4 compares the prevalence of occupations at the 2-digit SOC level in our BGT data to the share of employment in that occupation group in the labor force according to the BLS in 2017. As the figure shows, at a 2-digit SOC level, management occupations, business and finance, and computer-related occupations are substantially overweight in the BGT data relative to the labor force overall, while manual occupations, healthcare and education are substantially underrepresented. However, this does not bias our results, as we compute mobility at the occupation-level.

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. 115.4 million of our observations come from resumes that list an address in the 50 U.S. states or District of Columbia. Comparing the proportion of our data from different U.S. states to the proportion of workers in different U.S. states in the BLS OES data, we find that our data is broadly representative by geography. As shown in figure A5, New Jersey, Maryland and Delaware, for instance, are 1.5-2x as prevalent in our data as they are in the overall U.S. 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 U.S.

[94] As defined above, for our purposes, an observation is a person-occupation-year observation for which we also observe another occupation in the following year: i.e. the start of a year-to-year occupation coincidence sequence.

[95] 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.

labor force. However, the figure also suggests that the broad patterns of the demographic distribution of populations across the U.S. is reflected in our sample. Aggregating the state data to the Census region level, 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. This shows that our sample is very close to representative for the Midwest and South regions, and somewhat overweights the Northeast, while underweighting workers from the West region.

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 truly annual mobility (or 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). In addition, since individuals list the dates they worked in specific jobs on their resumes, we are able to observe occupational transitions at the desired frequency, whether that is annual or longer.[96] 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.

[96]Since many individuals list only the year in which they started or ended a job, rather than the specific date, measuring transitions at a sub-annual frequency is too noisy.

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.

Caveats and concerns

The BGT dataset 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 exexcutives in banking and finds that 51 responding executives are jointly aware of a total of 17 incidences of meaningfully falsified job titles, which, given the presumably large number of resumes and contained job listings that would have been processed under these executives seems small. All but one of the respondents estimated the incidence of falsification of *any* part of the resume to be below 20%, with most opting for lower estimates. Note that this study was done before online search made verification of basic resume information much faster and more affordable. More recently, Nosnik, Friedmann, Nagler and Dinlenc (2010) found that 7% of the publications listed by a sample of urology residency applicants on their resumes could not be verified.

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

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

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

Comparability with CPS occupational mobility

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

To what extent is our measure similar to measures of occupational mobility constructed from the CPS? Our measure is not strictly comparable to the concept of annual occupational mobility estimated from the CPS by Kambourov and Manovskii (2008) and Xu (2018) for two reasons. First, the occupation categorization is different: we use SOC 6-digit occupations (of which there are a total of 840 in U.S. data) and the CPS uses Census occupation codes, which are slightly broader. Second, because of the nature of our resume data, we cannot measure annual occupational mobility (share of workers whose main job was in occupation o on date d in year t whose main job was no longer in occupation o on date d in year $t+1$). Instead, our measure of the average share of workers leaving their occupation in any given year (11%) reflects the total number of workers who are observed in occupation o in year t who are *not* observed in occupation p at any point in year $t+1$. This makes it a slightly more conservative measure of occupational mobility than the annual occupational mobility concept commonly constructed from the CPS.

With these caveats in mind: our measure of occupational mobility – the share of workers leaving their occupation being 11% from one year to the next – is somewhat lower than the occupational mobility estimate from Kambourov and Manovskii (2008), who find occupational mobility of 0.20 at the Census 3-digit level in the CPS for the late 1990s. Our measure is, however, in a similar range to Xu (2018) who finds occupational mobility of 0.08 in 2014.

The fact that our measure is relatively low compared to Kambourov and Manovskii (2008) is interesting, since sample selection bias might be expected to *overstate* occupational mobility in our data set if the people applying for jobs (whose resumes we observe) are more mobile than average.

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

D Appendix: OES Occupational Code Crosswalk

To construct our data set of wages and employment at the occupation-CBSA level over 1999 - 2016 - as described in Section 4 - 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.[97] 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.[98]

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.[99]

Using these occupational crosswalks, we can stack the OES occupational employment and wage data by city provided by the BLS, creating an unbalanced panel of 2.5 million occupation-by-city-by-year data points of employment and mean hourly and annual wages for the years 1999-2016. Out of these data, the 2005-2016 panel that was originally coded by the BLS using the new CBSA definition has about 1.8 million data points.

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

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

[99]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.

E Appendix: Alternative approaches to estimating occupational similarity

What makes jobs in a given occupation a good outside option? Good outside option jobs should be both *feasible* in the sense that the worker can relatively easily become as productive as an average worker in that job, and should be at least somewhat *desirable* to work in (relative to the worker's current job). In Section 3 of this paper, we argue that using occupational transitions is a better way to identify workers' outside options than the two other commonly-used methods of estimating occupational similarity – skill-and task-based similarity measures, and demographic- and qualification-based similarity measures – because it better captures both feasibility and desirability of occupational transitions. We expand on this argument below.

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

This section expands on the results on determinants of occupational mobility discussed more concisely in section 3. To use worker transitions to infer the network of worker outside options, we must assume that the empirical occupational transitions we observe reflect the underlying feasibility and/or desirability of an occupation as an outside option. It is possible, however, that our empirical occupational transition shares simply reflect something idiosyncratic in our data, or short-run contractions or expansions of different occupations (though the size, time horizon, and relative representativeness of our data should do something to assuage these concerns). In this section, we describe in more detail our analysis on the degree to which our measure of occupational transitions reflects similarities between different occupations in terms of task requirements, wages, amenities, and leadership responsibilities.

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 et al. (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.[100]

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.[101] 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 re-

[100]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.

[101]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.

sponsibilities from the O*Net database, and create a new "leadership" composite measure defined at the level of each SOC 6-digit occupation. The measure incorporates the six characteristics most associated with leadership positions in the O*Net data, alongside the O*Net work style category for leadership. Since this is a new composite measure of an important occupational characteristic, we outline it in more detail here.

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

Occupational similarity and mobility

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

$$\pi_{o \to p} = \alpha_o + \beta^{abs} |X_{occ\, p} - X_{occ\, o}| + \gamma |\Delta w_{o \to p}| + \epsilon_{op}. \tag{30}$$

where $\pi_{o \to p}$ is the share of job changers in the origin occupation o that move into target occupation p, $|X_{occ\, p} - X_{occ\, o}|$ is the absolute difference between the target and the origin occupation in each of the occupational characteristics X_o defined above, and α_o are origin occupation fixed effects to control for differences in outward mobility across occupations.

We control for absolute wage differences between the occupations in all regressions except for those estimating the effect of wages or amenity differences on occupational mobility,[102] but note that the results are qualitatively similar without the wage controls.

We would expect the coefficient on the absolute difference in characteristics to be negative: the greater the difference between two occupations, the less likely we should be to observe the worker moving from one into the other. Our results bear this out: in every regression of pairwise occupational mobility on the absolute difference in characteristics, the coefficients are significantly negative or statistically insignificant, as shown in figure 3.[103]

The previous results impose symmetry on the likelihood of occupational transitions – but between many pairs of occupations, the probability of moving in one direction is likely to be different than the probability of moving in the other direction. To study whether differences in characteristics also predict the direction of occupational flows, we estimate a similar regression equation to that shown in equation (30), but now using the *relative* (target minus origin) difference in occupational characteristics as the independent variable:

$$\pi_{o \to p} = \alpha_o + \beta^{rel}(X_{occ\,p} - X_{occ\,o}) + \gamma \Delta w_{o \to p} + \epsilon_{op}. \tag{31}$$

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 (31) for the different measures are shown in Figure A11. (Note that this analysis involves directed relationships between occupations, so if the same share of moves in each direction is observed for an given occupation pair, the estimated effect of differences between them would be zero.)

A number of our predictions are borne out in the data: we find (1) that workers are more likely to move towards jobs with higher wages; (2) that workers transition on average *towards* jobs that require more leadership responsibility - as would be expected from moves up the career ladder; (3) that occupational transitions have on average been *towards* occupations that have higher analytical content and require more social skills, and out of occupations with more routine task requirements;[104] and (4) that workers have on average been moving

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

[103] 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.

[104] 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

into occupations that require more contact and working relationships with others (and so have less time flexibility).

While occupational transitions therefore do reflect similarity in tasks, temporal flexibility, and leadership requirements, we note that there is substantial variation in occupational transitions which is not captured by these other occupational similarity measures. Appendix Table A3 shows the adjusted R-squared statistics from regressions of $\pi_{o \to p}$ on our measures of skill distance, wage difference, amenity difference (temporal flexibility), leadership difference, and a composite skill measure. In all of these cases, while the correlation is strong and positive, the explanatory power is relatively low. This contrasts with results in Macaluso (2019), who shows that at a 2-digit level, skill distance can explain ∼23% of the variation in flows between occupational groups. The difference in these results shows that while skill distance may be a good predictor of mobility for more aggregate occupational groupings, for the more detailed analysis in this paper it cannot capture much of the variation the sparse matrix of mobility between 6-digit occupational pairs.

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 and Medical & clinical laboratory technologists; Carpenters and Dental assistants; Travel agents and 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 and 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.

---

and Autor (2011), and the increasing demand for social skills documented by Deming (2017).

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 et al. (2018), 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

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

where $\tilde{G}_{j,o,k,t} = \frac{(1+\tilde{g}_{j,o,t})^2}{(1+\tilde{g}_{o,k,t})^2} - 1$ is the predicted firm-level excess local vacancy growth relative to the average predicted local occupation vacancy growth - the time-varying shock - and $\sigma^2_{j,o,k,t-1}$ is the exposure of the local concentration index to that shock.[105]

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-city labor market ($g_{o,k,t} = \sum_j \sigma_{j,o,k,t-1} g_{j,o,k,t}$), and (2) the predicted growth rate of local vacancies based on large firms' national growth ($\tilde{g}_{o,k,t} = \sum_j \sigma_{j,o,k,t-1}\tilde{g}_{j,o,t}$). To control for differential initial exposure to non-local firms, we introduce our "exposure control" $e_{o,k,t} = \sum_j \sigma^2_{j,o,k,t-1} \cdot \mathbb{1}[\tilde{g}_{j,o,t} \neq 0]$.

In our fixed effects IV estimation of equation (10), the exclusion restriction for the instrument on the HHI concentration index is then equivalent to

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

[105]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.

Here, $\tilde{G}^{\perp}_{j,o,k,t}$ represents $\tilde{G}_{j,o,k,t}$ after it has been residualized with regard to city-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 $g_{o,k,t}, \tilde{g}_{o,k,t}, e_{o,k,t}$.

This orthogonality condition holds under two assumptions. First, we require that the national firm-level growth shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks $\xi_{o,k,t}$, the fixed effects Γ_{kt} and Γ_{ot}, actual and predicted average local vacancy growth $g_{o,k,t}$ and $\tilde{g}_{o,k,t}$, and initial exposure to non-local firms $e_{o,k,t}$. That is,

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

for some constant parameters τ_1^{HHI}, τ_2^{HHI}, τ_3^{HHI}, τ_4^{HHI}, and τ_5^{HHI}. 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.

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

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

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

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

To be concrete, note again the hypothetical example from the main text, which considered insurance sales agents in Bloomington, Illinois and in Amarillo, Texas. In each city, there are several insurance companies who employ insurance sales agents. Assume that in Bloomington, State Farm has a large share of local insurance sales agent employment (as Bloomington is their headquarters), while in Amarillo employment is more dispersed amongst a number of insurance companies. We noted before, that, in years where State Farm grows substantially faster than other major insurance companies nationwide, under most combinations of the distribution of that growth across cities and the initial distribution of employer shares in each city, 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 cities. Thus, it allows for each city to be exposed differently to overall trends in the demand for insurance sales agents. The identification only requires that once we account for overall city exposure to insurance sales agent demand, whether that demand was driven by the city'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 firm j grows nationally, local occupation-city labor markets with a higher share of vacancies accounted for by firm j in year $t-1$ see a larger increase in employer concentration. A sufficient condition for this to be the case under *most* initial employer share distributions is if firm j's new vacancies are allocated evenly across occupation-city labor markets, such that each occupation-city labor market sees the same growth rate in its firm j vacancies as the national average.[106] 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-city labor markets is higher than in high-initial-employment-share labor markets, as long as this relationship is not too strong. In our data, for a given employer, there is a negative relationship between the initial vacancy share in an occupation-city labor market and the next year's vacancy growth rate, but this relationship is not sufficiently strong to invalidate our first stage (for each one percentage point increase in the initial vacancy share, there is roughly a 1.4 percentage point lower vacancy growth rate from one year to the next). Empirically, our first stage holds for occupation-city labor markets with HHIs above all but very low levels, as shown in the main text.

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

[106] Note that this is not the case – i.e. the first stage might not hold – for *all* possible combinations of the distribution of employment growth and initial employer shares. For example, consider a world in which there is a labor market for where Employer X has 80% of the market in one city, and the rest of the market is comprised of atomistic firms; and Employer X has 65% of the market in another city, with the rest of the market comprised of atomistic firms. If Employer X grows by 10% in both locations in a given year, and the other firms do not grow at all, employer concentration will actually increase by more in the latter than the former market. This circumstance, however, only occurs when comparing two labor markets which both have extremely high levels of employer concentration already, and so is not relevant for the vast majority of the labor markets in our data (only 6% of which have HHIs of greater than 6,800 – and only 2% of workers in our data face HHIs of greater than 5,000). In practice our first stage regressions are positive and strongly significant even when segmenting our data to analyze only cells with high HHIs of 2,500 of more, meaning this concern is not hugely relevant in practice for our empirical analysis.

on national leave-one out mean wages (described in Section 4).

To avoid endogeneity concerns over the local employment shares, we instrument for the local relative employment share in each occupation using the initial employment share in that occupation in 1999, the first year in the data.[107] Our instrument for the oo^{occs} index, $oo^{occs,inst}$, therefore becomes 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 city, $\frac{s_{p,k,1999}}{s_{p,1999}}$, and the national occupation transition shares from the worker's occupation o to each of the other occupations, $\pi_{o\to p}$.

$$oo^{occs,inst}_{o,k,t} = \sum_{p}^{N_{occs}} \left(\pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\not{k},t} \right) \tag{32}$$

To make the assumptions transparent under which this wage instrument identifies the coefficient on our outside-occupation option index in equation (10), we follow the framework presented in Borusyak et al. (2018). 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. Note that we can write the instrument as

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

where $s_{okp} = \pi_{o\to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}}$ is a measure of predicted local exposure to the shock. In our fixed effects IV estimation of equation (10), the exclusion restriction for the instrument for outside-occupation options is then equivalent to

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

where $\bar{s}_{okp} = \mathbb{E}[s_{okp}]$ is the average exposure to occupation p, and $\phi^{oo}_{pt} \equiv \mathbb{E}[s_{okp}\xi_{o,k,t}]/\mathbb{E}[s_{okp}]$ is an exposure-weighted expectation of the structural wage residuals. Moreover, $w^{\perp}_{p,\not{k},t}$ represents $\bar{w}_{p,\not{k},t}$ after it has been residualized with regard to city-k-by-year-t fixed effects Γ_{kt} and occupation-o-by-year-t fixed effects Γ_{ot}, as well as the concentration index and control variables.

[107]Or we use the first year the occupation-city cell is in the data, if it is not present in 1999

Borusyak et al. (2018) show that this orthogonality condition holds under two assumptions. First, we require that the national occupation-level shocks are quasi-randomly assigned conditional on local exposure to structural wage shocks ϕ_{pt}, the fixed effects Γ_{kt} and Γ_{ot}, and the control variables. That is,

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

for some constant parameters τ_1 and τ_2. Second, there needs to be a large number of independent occupational shocks, that is,

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

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

The first assumption requires that the national leave-one-out mean wage $\bar{w}_{p,\not{k},t}$ in outside option occupation p is correlated with the local wage of occupation p in location k (relevance 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 $oo^{occs,inst}_{o,k,t}$. 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 city k. The inclusion of these fixed effects increases our confidence that the assumptions for instrument validity hold.

Industry Bartik shock

One possible concern with the identification assumptions required for our outside-occupation index – which may not entirely be picked up by our occupation-year, or city-year fixed effects – is that industry-level wage trends may differentially impact local occupations based on their city's direct exposure to those industries, rather than only based on indirect exposure through outside occupation job options. As discussed in the text, an example of this could be the following. Imagine that the finance industry and the tech industry employ both accountants and data scientists to a disproportionate degree relative to other occupations, and that San Francisco has a large share of employment in tech while New York has a large share of employment in finance. Imagine further that being a data scientist is a good outside option occupation for an accountant. In years where the tech industry is booming nationwide, this will impact San Francisco more than New York. Accountants in San Francisco will see wages rising by more than accountants in New York – partly driven by the increase in the outside option value of becoming a data scientist, but partly simply because more accountants in SF already work in the tech industry, as compared to accountants in NY, and so they will see their wages rise by more.

To control for this possible omitted variable bias, we incorporate an industry Bartik shock in a robustness check for our baseline regressions. We construct the industry Bartik shock for each occupation-city-year cell. The industry Bartik shock for occupation o in city k in year t is defined as

$$\sum_{\iota}^{industries} \frac{\frac{emp_{\iota,o,t-1}}{emp_{o,t-1}} \cdot \frac{emp_{\iota,k,t-1}}{emp_{k,t-1}} \cdot \bar{w}_{\iota,t}}{\frac{emp_{\iota,o,t-1}}{emp_{o,t-1}} \cdot \frac{emp_{\iota,k,t-1}}{emp_{k,t-1}}}$$

where ι denotes each NAICS 4-digit industry. The shock to each industry is the national industry wage $\bar{w}_{\iota,t}$, and the exposure of each local occupation to that shock is determined by (1) the national share of employment in that occupation which is in industry ι, $\frac{emp_{\iota,o,t-1}}{emp_{o,t-1}}$ and (2) the share of employment in that city which is industry ι, $\frac{emp_{\iota,k,t-1}}{emp_{k,t-1}}$. 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 Bartik instrument relies on the assumption that national industry-level wage shocks are uncorrelated with local occupation-level wage trends, except to the extent that the former causes the latter. We use data on 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 of each occupation by industry, and we construct data on industry employment shares by metropolitan statistical area ("city") from Eckert, Fort, Schott and Yang (2020)'s county-by-industry employment data, which was constructed from the County Business Patterns database. We report our baseline regression results, controlling for this industry Bartik shock, in Panel A of Table A7.

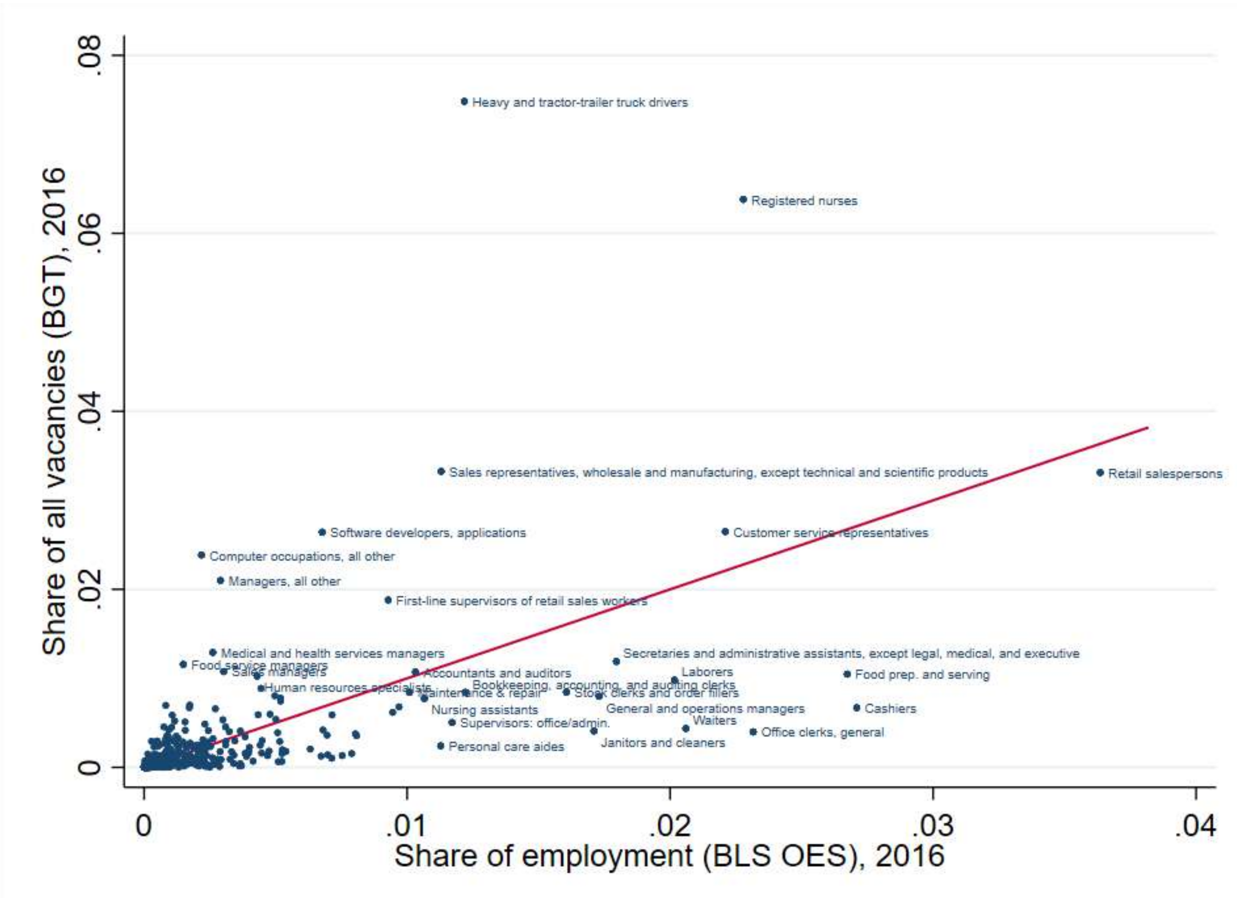
In other unreported specifications, we construct the industry Bartik shock with, variously, national wage growth in each industry ι, national employment growth in each industry ι, or the national wage bill growth in each industry ι. In all these specifications, our coefficients on the outside-occupation option index and HHI index (instrumented) remain very similar to those in the baseline specification.

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).

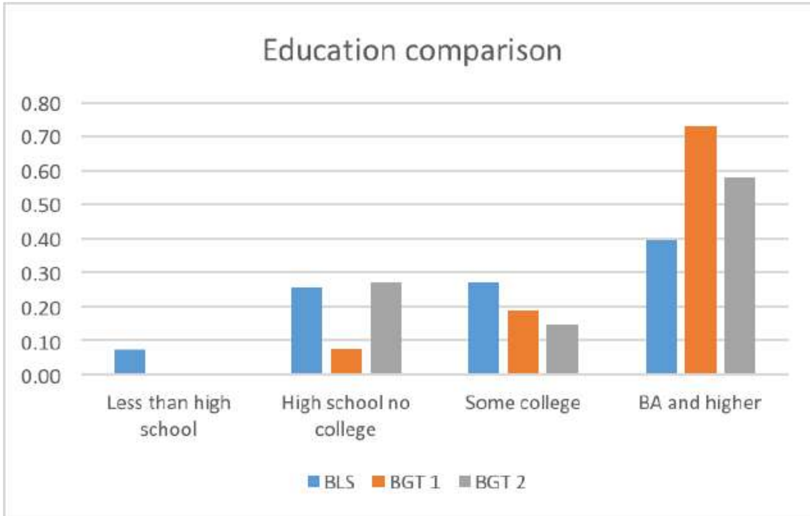
I Appendix: Figures

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



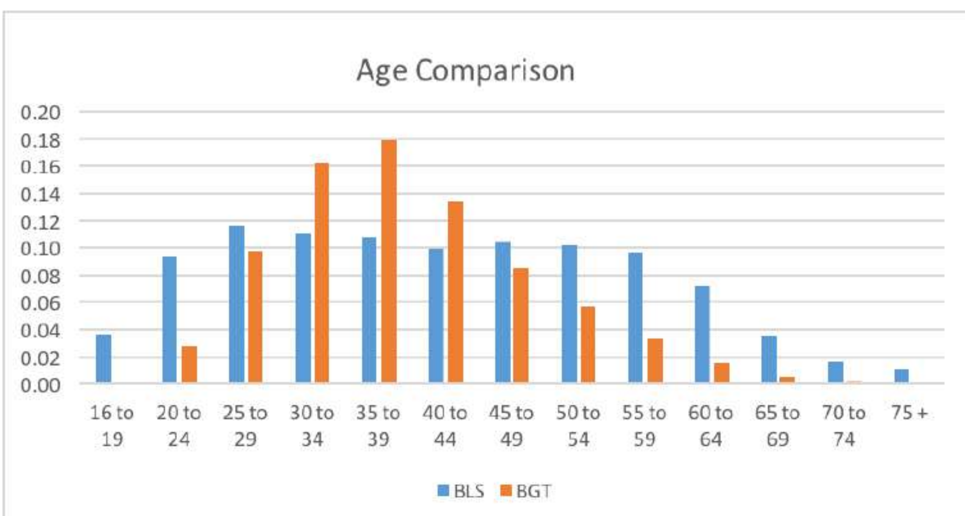
Comparison of distribution of share of each SOC 6-digit occupation in the BGT vacancy data, relative to its share in the BLS occupational employment statistics, with occupations comprising greater than 1% share of either data set labeled. Red line is the 45 degree line.

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



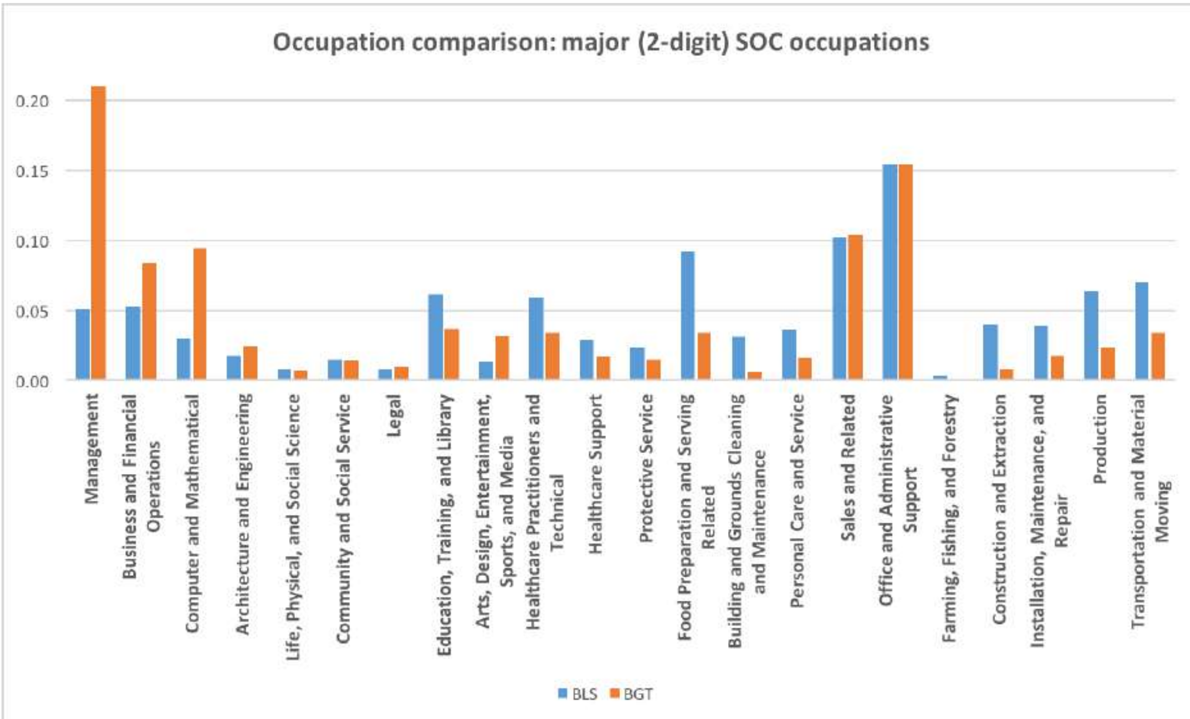
Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT resume data. Two versions are shown: BGT 1 excludes all resumes missing educational information, while BGT 2 assumes all resumes missing educational information have high school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

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



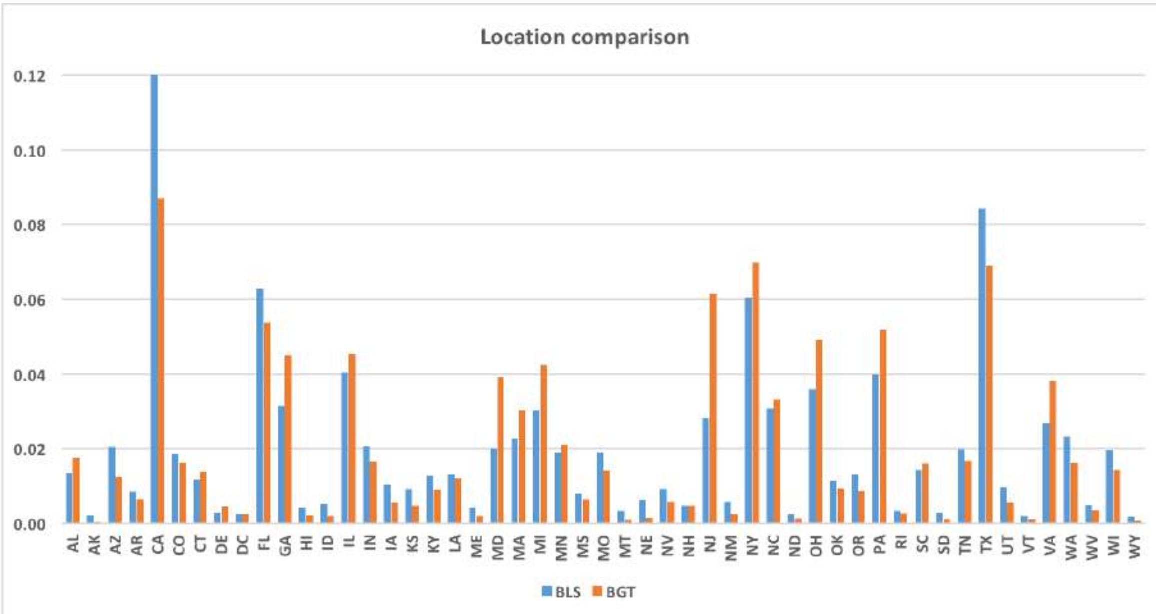
Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT resume data.

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



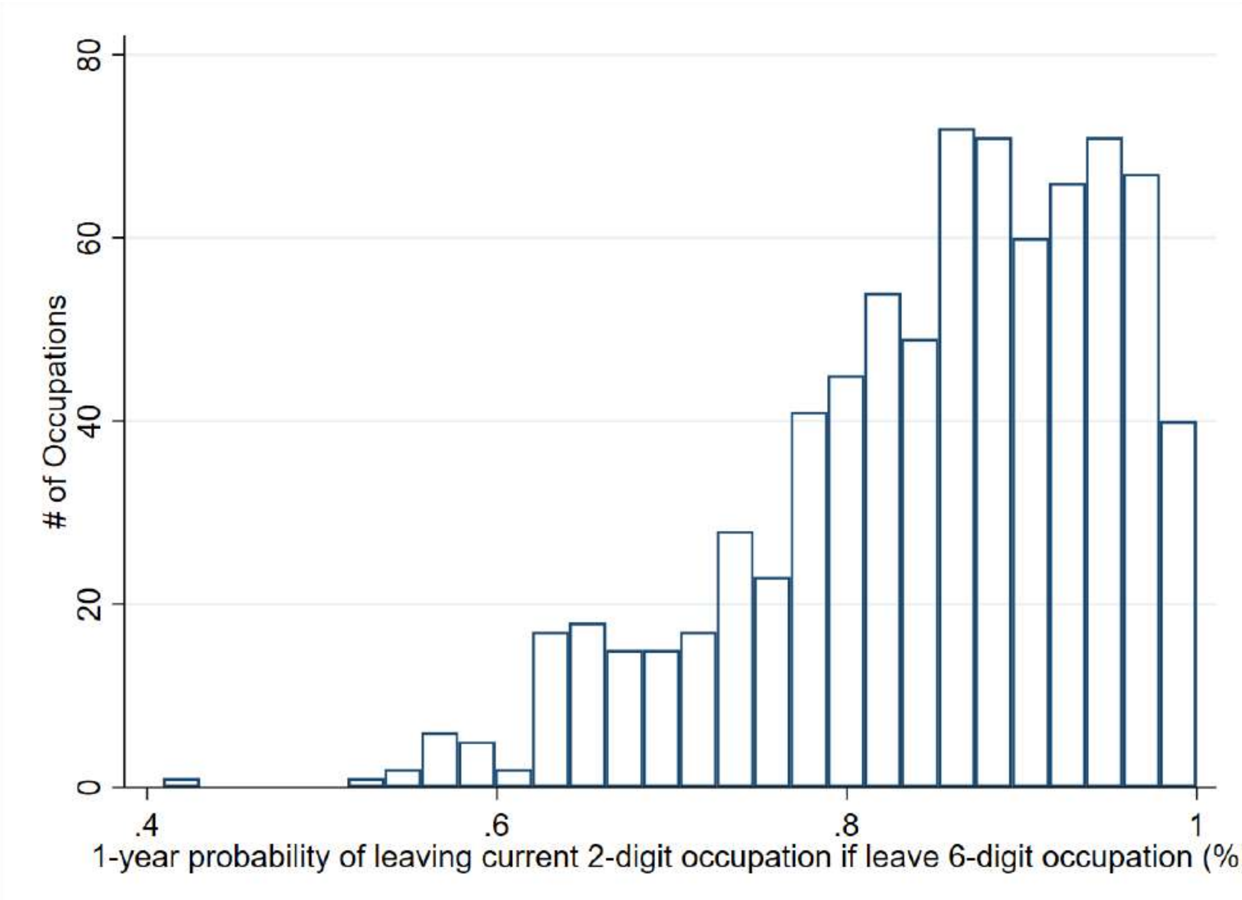
Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT job sequence data.

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



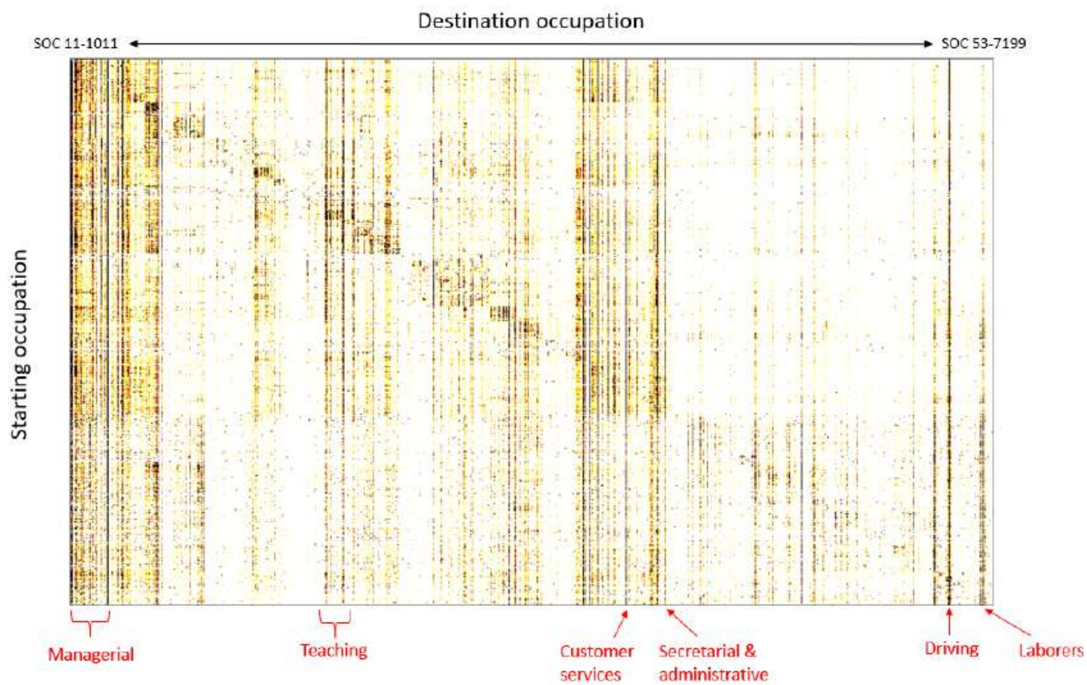
Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT job sequence data. Graph shows share of total in each state.

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



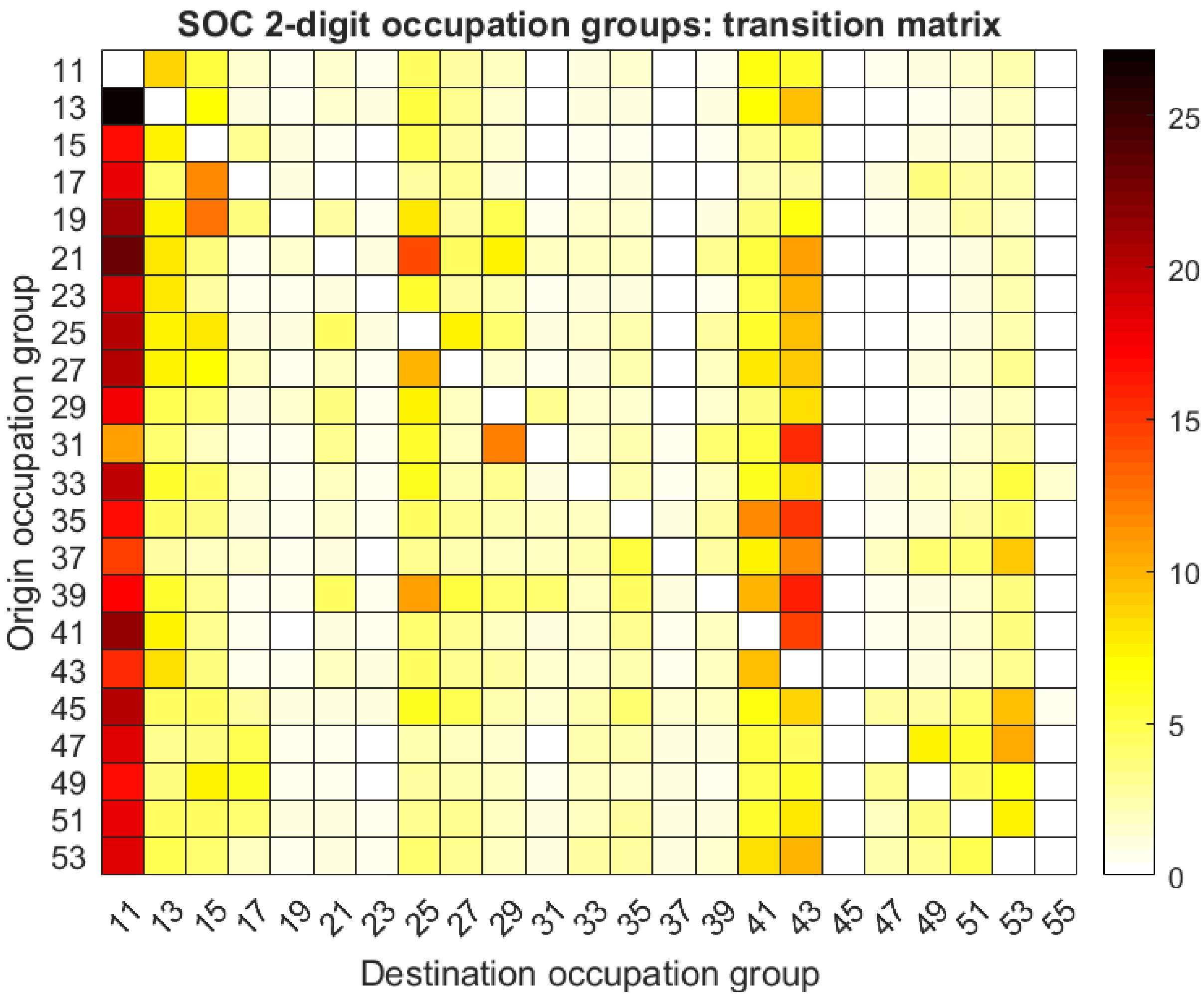
Distribution of the proportion of workers moving 6-digit SOC occupation who *also* move 2-digit SOC occupation, by occupation, calculated from BGT resume data for 2002-2015 period. Histogram shows 786 occupations.

Figure A7: 6-digit SOC occupational transition matrix



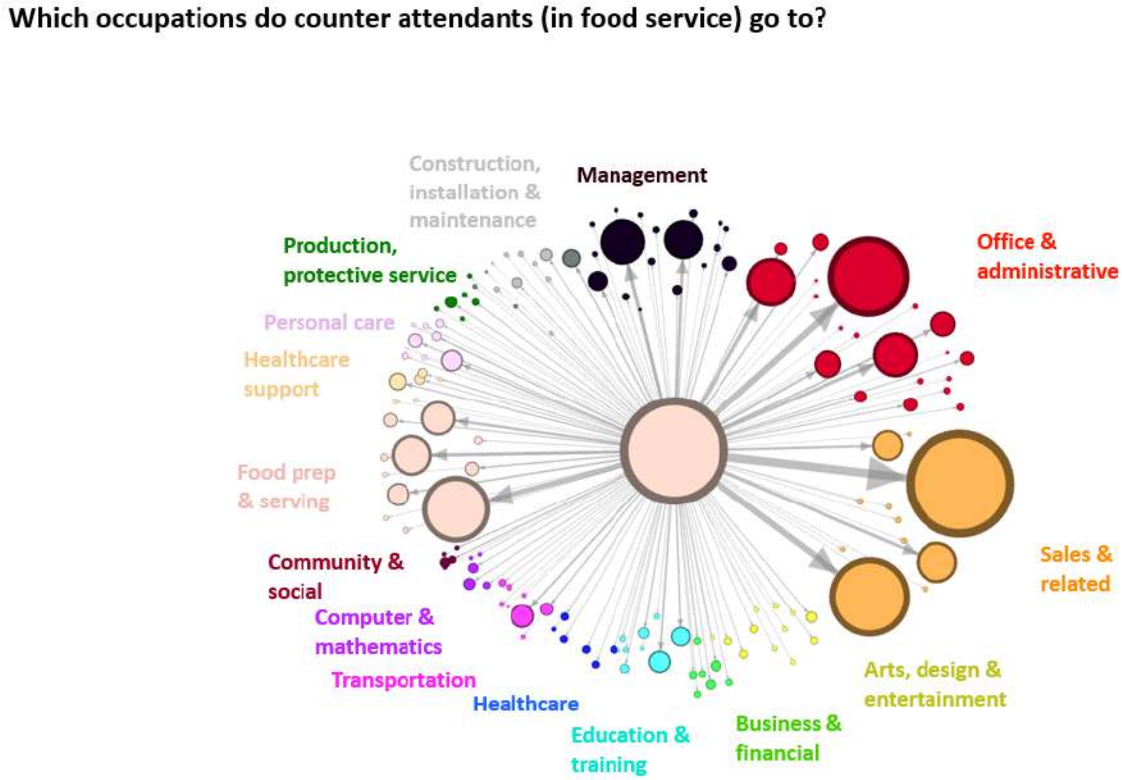
Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial job. Occupations are sorted in SOC numerical order. Cells colored black have a transition probability of 1% or greater conditional on leaving the initial job. Transitions to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

Figure A8: 2-digit SOC occupational transition matrix



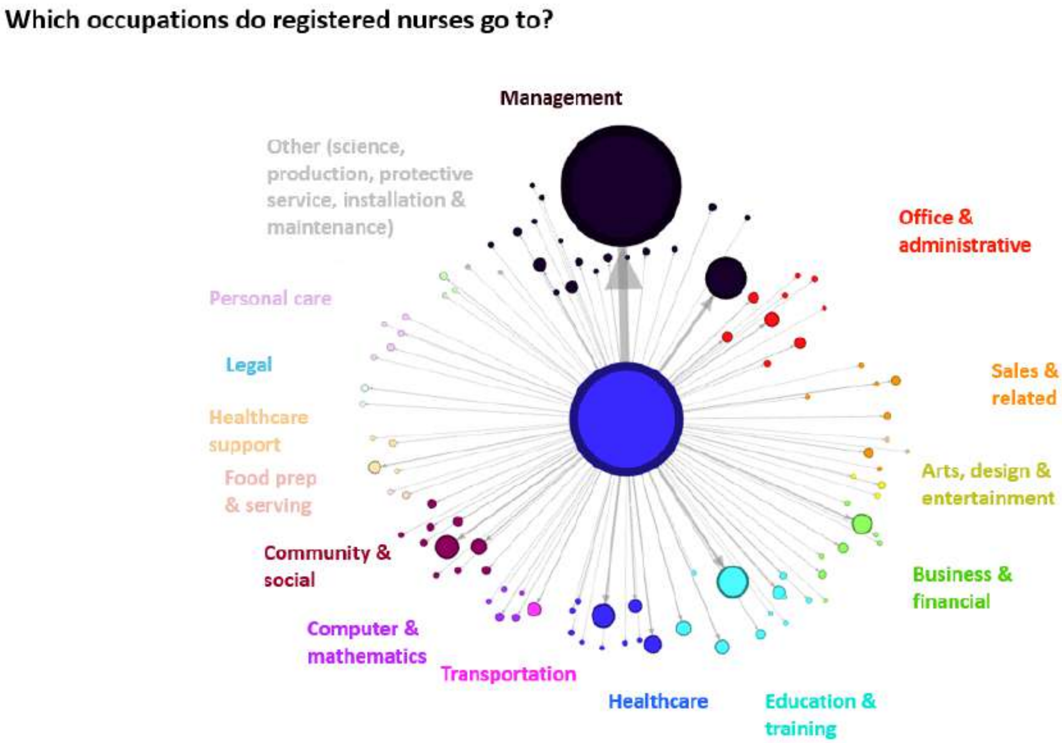
Occupational transition matrix showing transition probability between 2-digit SOC occupation groups conditional on leaving the initial job. Cells colored black have a transition probability of 25% or greater conditional on leaving the initial job. Job transitions within an occupation group are excluded.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

Figure A9: Examples of probabilistic labor markets: counter attendants



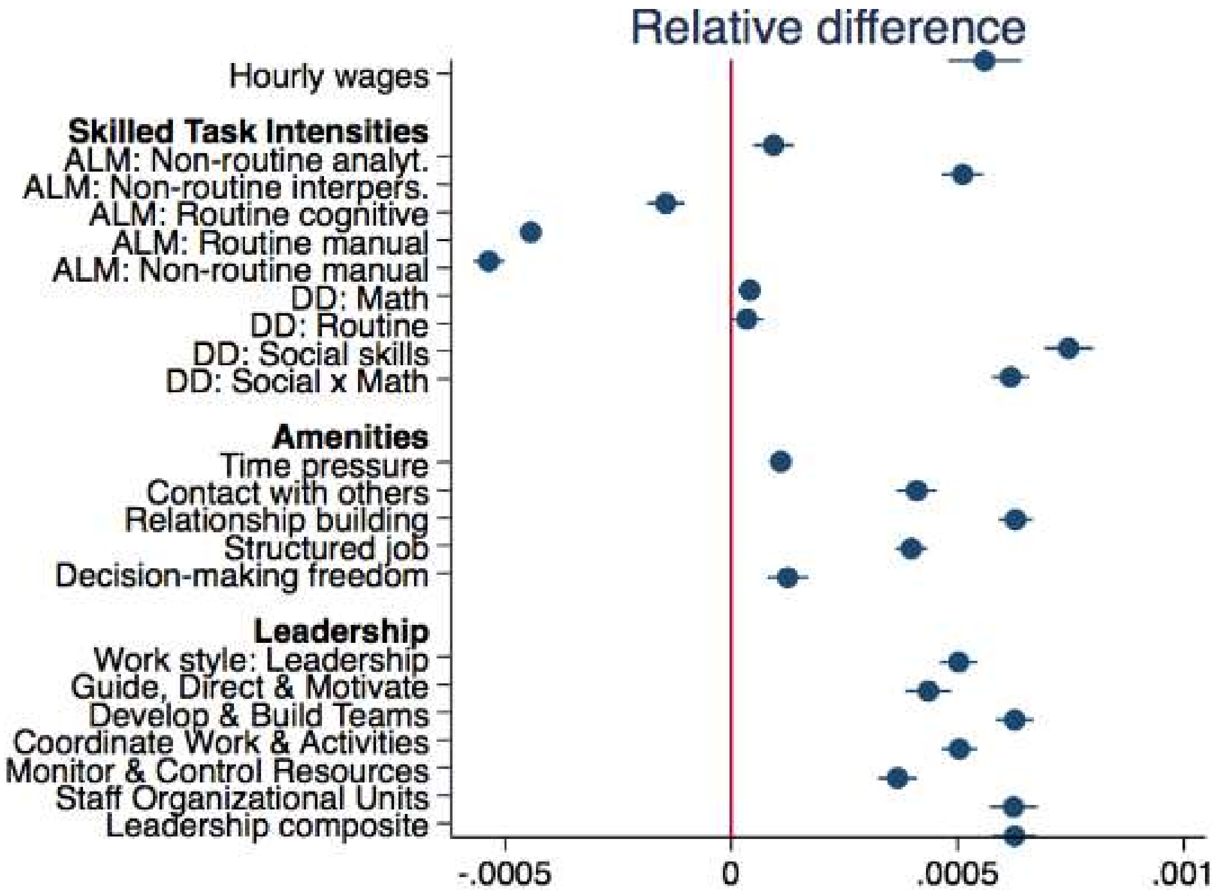
Occupational transitions for counter attendants in the food industry. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of counter attendants in the BGT data who are observed in each destination occupation in the following year.

Figure A10: Examples of probabilistic labor markets: registered nurses



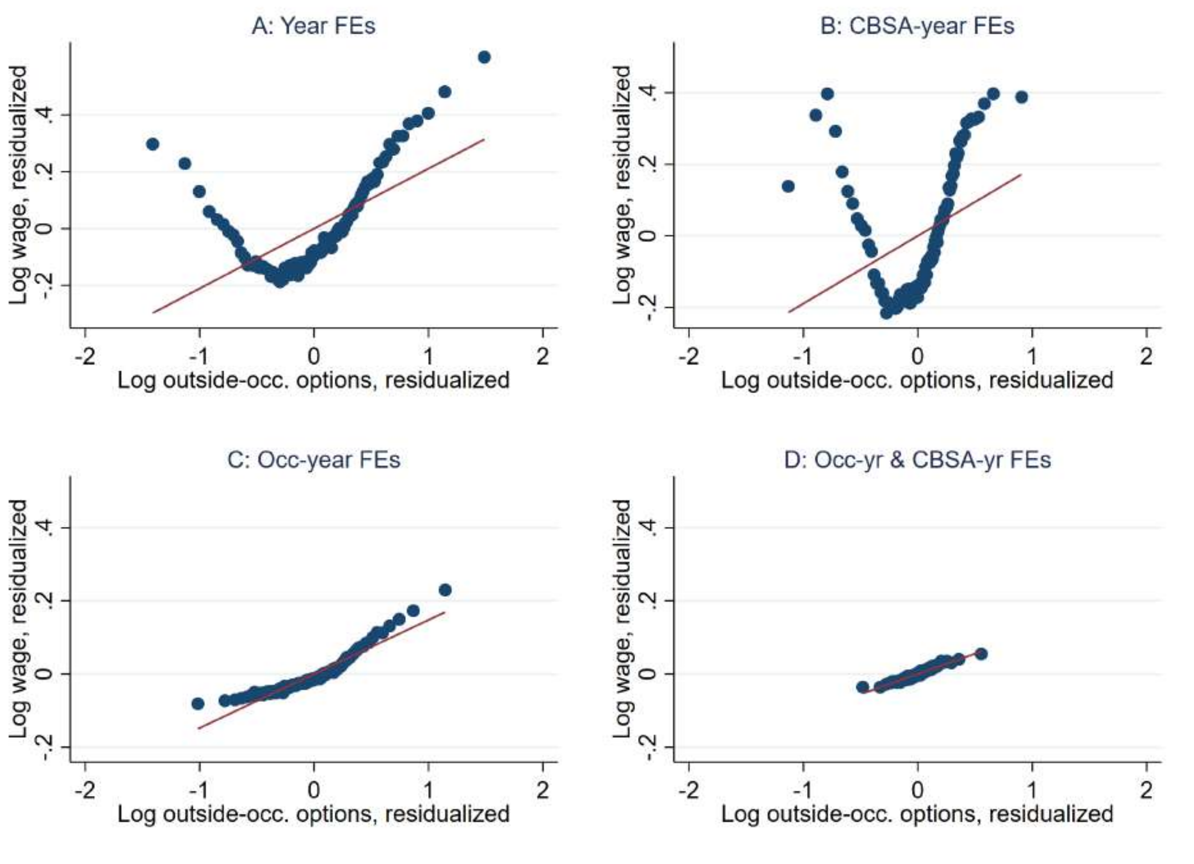
Occupational transitions for registered nurses. Each bubble is a SOC 6-digit occupation, and the colors represent SOC 2-digit occupational groups. The size of each bubble is proportional to the share of registered nurses in the BGT data who are observed in each destination occupation in the following year.

Figure A11: Determinants of occupational mobility



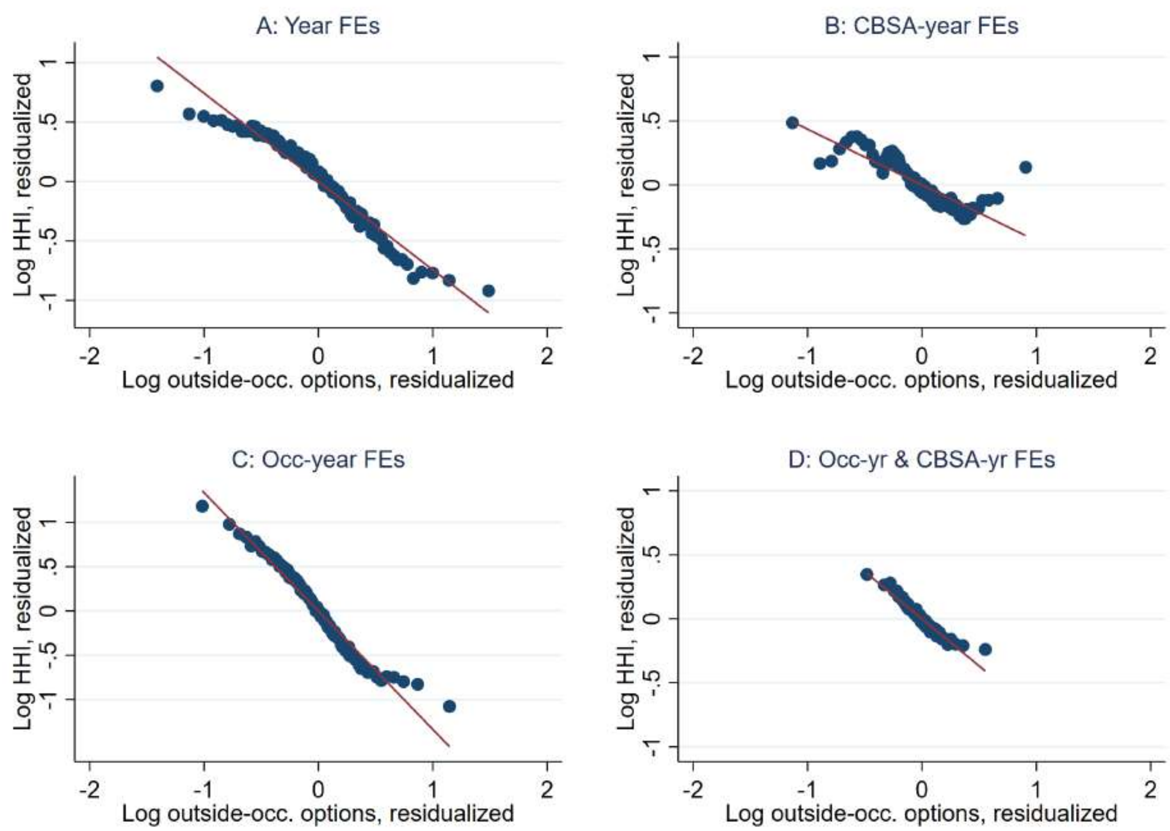
Note: This plot shows the coefficients and 95% confidence intervals from regressions of occupation transition shares $\pi_{o\to p}$, calculated from Burning Glass Technologies Resume data, on *relative* differences in occupational characteristics: $\pi_{o\to p} = \alpha_o + \beta f(X_{occ\ o\to p}) +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

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



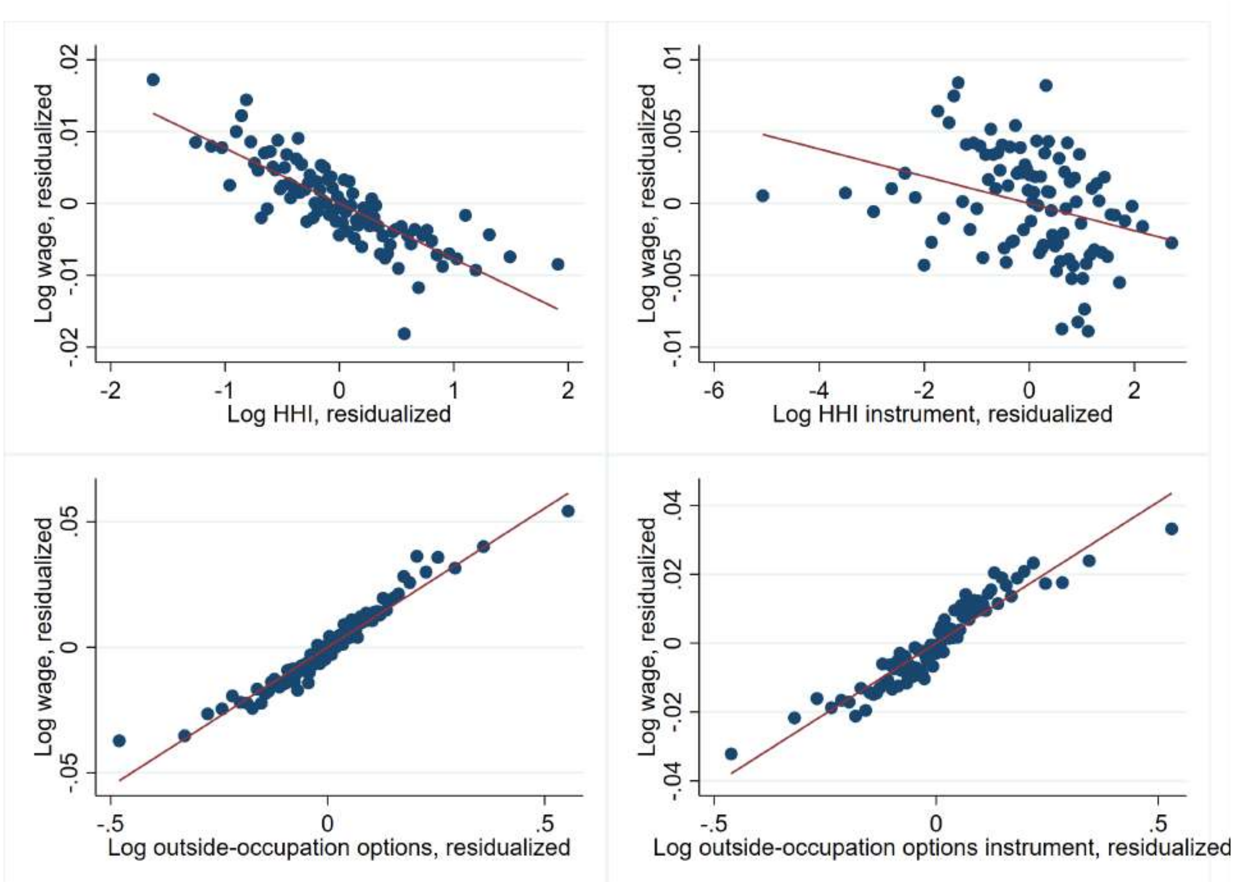
Note: Figure shows binned scatter plots of the correlation between average log wages and log outside-occupation option index for occupation-CBSA cells over 2013–2016, residualized on different combinations of fixed effects. Regression coefficients for the line of best fit on each graph are: A: 0.16, B: 0.12, C: 0.15; D: 0.11. The non-linear shape of the figures without occupation fixed effects (panels A and B) is explained by healthcare occupations which tend to have both low outward mobility and high pay.

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



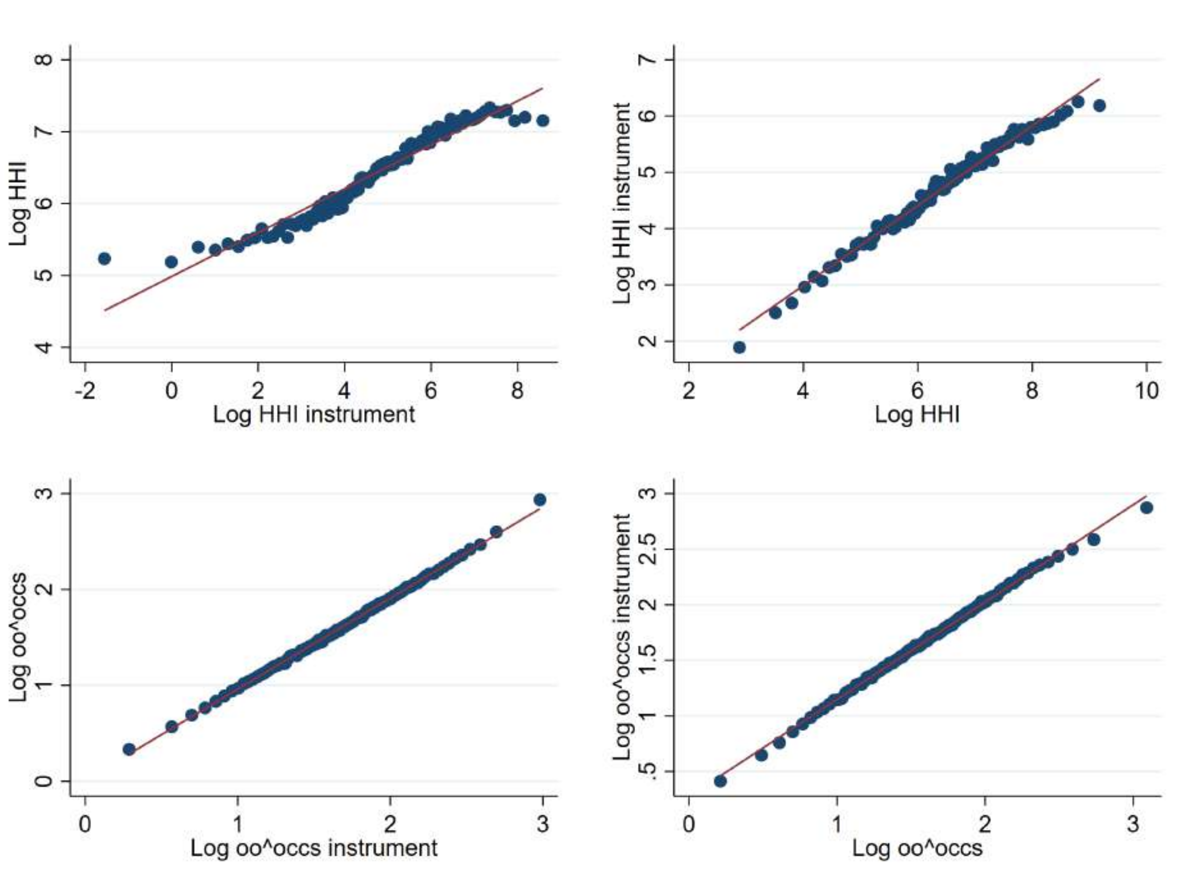
Note: Figure shows binned scatter plots of the correlation between average log HHI and log outside-occupation option index for occupation-CBSA cells over 2013–2016, residualized on different combinations of fixed effects. Regression coefficients for the line of best fit on each graph are: A: -0.74, B: -0.44, C: -1.34; D: -0.74.

Figure A14: Visualization of baseline regression results



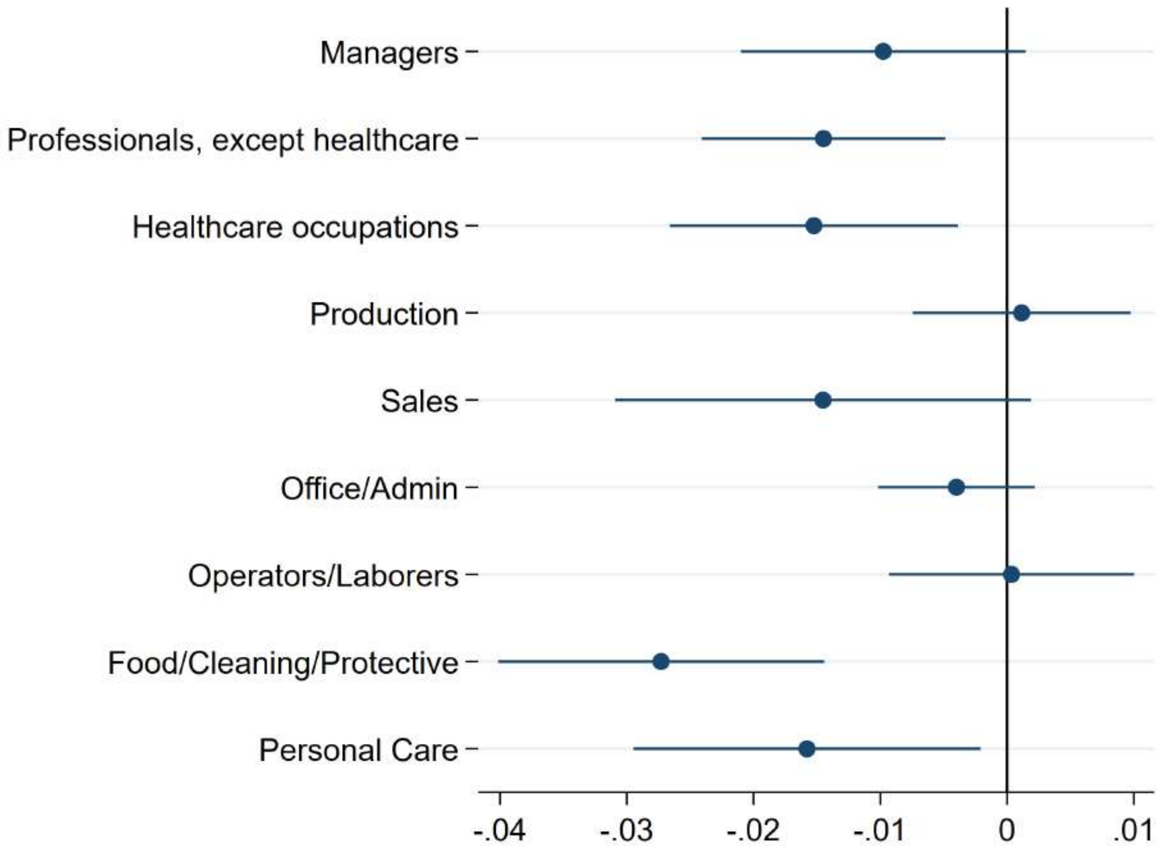
Note: Binned scatter plots of the log wage on log HHI and log outside-occupation options, and instrumented, including full controls as in baseline regression specification (i.e. the left panels correspond to coefficient estimates in Table 3 column (b), and the right panels correspond to the reduced form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

Figure A15: Visualization of first-stage regressions



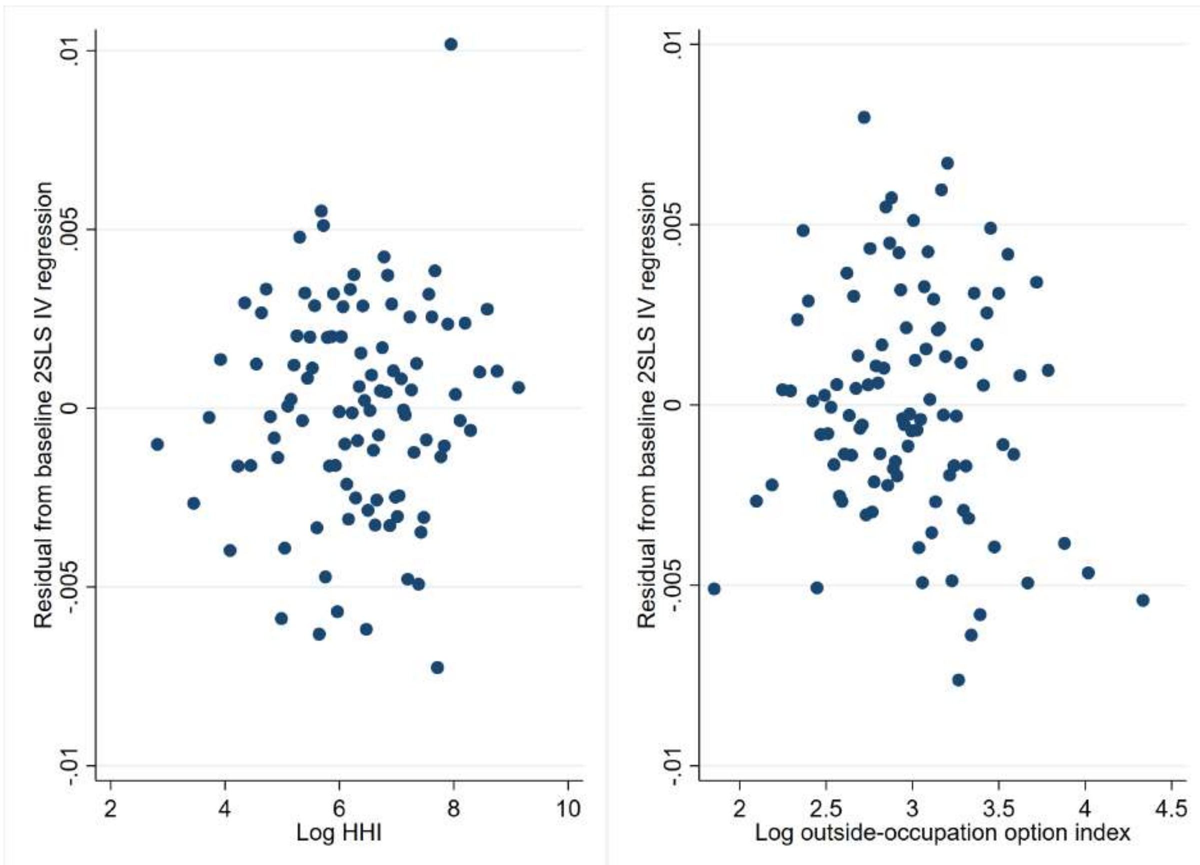
Note: Binned scatter plots of the correlation between the HHI instrument and raw variable (left side) and outside-occupation option index instrument and raw variable (right side) for occupation-CBSA cells in 2016.

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



Note: Coefficients on HHI and 95% confidence intervals from regressions of occupation-CBSA wages on instrumented local employer HHI, controlling for (instrumented) outside-occupation job options, with coefficient on HHI allowed to vary by occupation group. Regressions span 2013-2016 and include occupation-year and CBSA fixed effects. Standard errors are clustered at the CBSA level. Occupation groups are listed in descending order of average wage. Occupation groups map from SOC 2-digit occupations as defined in Appendix Table A6. Coefficient estimates correspond to those in Table A13 column (a).

Figure A17: Baseline regressions: residual plots



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

J Appendix: Tables

Table A1: Summary statistics in BGT vacancy data

	p5	p10	p25	p50	p75	p90	p95
Total vacancies posted by employer	1	1	1	2	5	19	58
No. of years employer present	1	1	1	1	2	3	4
No. of years employer present (vacancy-weight)	2	3	4	4	4	4	4
Occ. share relative to BLS OES	0.11	0.18	0.37	0.81	1.80	4.28	7.22
Occ. share relative to BLS OES (emp.-weight)	0.16	0.21	0.34	0.62	1.14	2.10	2.83
Metro area share relative to BLS OES	0.60	0.67	0.76	0.90	1.12	1.35	1.54
Metro area relative to BLS OES (emp.-weight)	0.60	0.63	0.75	0.90	1.03	1.17	1.34

This table shows some summary statistics from the BGT vacancy data for 2013–2016 inclusive. 'No. of years employer present' refers to the number of years in which a given employer posted a vacancy, with a maximum of 4. 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 2013-2016 then averaged across the four years). The employment-weighted version of this statistic weights each occupation-metro area cell by employment in that cell in 2016.

Table A2: Distribution of number of jobs on resume and duration of jobs in BGT resume data set.

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

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

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

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

Table shows adjusted R-squared from regressions of the form

$$\pi_{o \to p} = \kappa + \alpha_o + \beta \Delta X_{occ\ p-o} + \epsilon_{op}.$$

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

Table A4: Twenty large occupations with lowest leave shares and highest leave shares

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

This table shows the twenty large occupations with the lowest and the highest occupation leave shares - defined as the 1-year horizon probability of no longer working in their current occupation, conditional on leaving their job - in the BGT data over 2002-2015, as well as total national employment in that occupation in 2017 from the OES, the number of occupation-year observations in the BGT data (‘obs.’) and the most popular occupation that workers who leave the initial occupation move to (‘modal new occupation’). Large occupations are defined as those with national employment over 150,000 in 2017 (roughly the 75th percentile of occupations when ranked by nationwide employment). † Full occupation title is “Molding, coremaking, and casting machine setters, operators, and tenders, metal and plastic.”

Table A5: Forty thickest occupational transition paths for large occupations

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

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. Only occupations with at least 500 observations in the BGT data and 2017 OES employment data are shown.

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

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

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

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

Dependent variable:	Log wage				
	(a) No vac controls	(b) Equal-wt vac	(c) No exposure control	(d) Industry Bartik	(e) Occ-CBSA FEs
Log HHI, instrumented	-0.011*** (0.003)	-0.011*** (0.003)	-0.011*** (0.002)	-0.010*** (0.003)	-0.008** (0.003)
Log outside-occ. options, instrumented	0.095*** (0.009)	0.095*** (0.009)	0.095*** (0.009)	0.085*** (0.009)	-0.002 (0.007)
Exposure control	0.004 (0.007)	0.004 (0.007)		0.003 (0.008)	-0.001 (0.004)
Vacancy growth		-0.001 (0.001)	-0.001 (0.001)	-0.007 (0.004)	-0.003 (0.003)
Predicted vacancy growth		-0.013 (0.010)	-0.012 (0.010)	0.033 (0.030)	-0.008 (0.008)
Equal-weighted vacancy growth		-0.000 (0.000)			
Industry Bartik				0.146*** (0.014)	
Observations	184,411	184,411	184,411	169,341	158,393
F-Stat	386	373	491	620	221
Fixed effects	Occ-year CBSA-year	Occ-year CBSA-year	Occ-year CBSA-year	Occ-year CBSA-year	Occ-CBSA Year

Notes: These represent robustness checks for our baseline regression specification (reported in column (d) of Table 3). Each column is a different robustness check. All columns are 2SLS IV regressions estimated using our HHI and outside-occupation option instruments. Column (a) excludes our controls for local actual and predicted vacancy growth. Column (b) includes an additional control for equal-weighted vacancy growth of local firms in the relevant occupation. Column (c) excludes our HHI exposure control. Column (d) includes an industry Bartik shock to control for correlated industry shocks across occupation-metro area cells. All these specifications feature occupation-year and city-year fixed effects. Column (e) runs our baseline regression specification, but with occupation-city and year fixed effects. Other regression info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. F-Stat is Kleibergen-Paap Wald F statistic. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

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

Dependent variable:	Log wage				
	(a) Emp weight	(b) Log emp weight	(c) Drop low rep. occs	(d) Occ rep. weight	(e) CBSA rep. weight
Log HHI	-0.015***	-0.011***	-0.011***	-0.008*	-0.009***
	(0.005)	(0.003)	(0.003)	(0.004)	(0.003)
Log outside-occ. options	0.118***	0.103***	0.093***	0.108***	0.094***
	(0.027)	(0.009)	(0.009)	(0.011)	(0.009)
Vacancy growth	-0.000	-0.001	-0.001	-0.001	-0.001*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Predicted vacancy growth	-0.018**	-0.012	-0.018*	0.020	-0.020**
	(0.007)	(0.008)	(0.011)	(0.032)	(0.009)
Exposure control	0.003	0.007	0.008	0.018	0.003
	(0.017)	(0.008)	(0.009)	(0.013)	(0.008)
Observations	184,411	184,258	137,567	184,411	184,411
F-Stat	96	370	411	369	221

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) and (b) weight the regressions by employment and log employment of the occupation-MSA cell, respectively, with employment weights calculated as the average employment in that occupation-by-MSA labor market over 2013–2016 according to BLS OES. Column (c) drops all occupations with average represented-ness in the BGT vacancy data of 0.5 or less. Columns (d) and (e) weight the regressions by average represented-ness of the occupations, and MSAs, in the BGT vacancy data (respectively). Represented-ness by occupation (/CBSA) in the BGT vacancy data is calculated as the share of all vacancies accounted for by a given occupation (/CBSA) in the BGT vacancy data in a given year, divided by the share of employment accounted for by a given occupation (CBSA) in the BLS OES in that same year, averaged over 2013–2016. About one third of occupations in our data have occupation represented-ness of 0.5 or less in the BGT data. Other regression info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All specifications feature occupation-year and city-year fixed effects. F-Stat is Kleibergen-Paap Wald F statistic. 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 checks

Dependent variable:	Log wage				
	(a) No vac controls	(b) Equal-wt vac	(c) No exposure control	(d) Industry Bartik	(e) Occ-CBSA FEs
Log HHI X Q1 outward mobility	-0.026***	-0.025***	-0.025***	-0.027***	-0.004
	(0.005)	(0.005)	(0.004)	(0.004)	(0.007)
Log HHI X Q2 outward mobility	-0.011***	-0.010***	-0.010***	-0.009***	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)
Log HHI X Q3 outward mobility	-0.007*	-0.007*	-0.007*	-0.004	-0.022**
	(0.004)	(0.004)	(0.003)	(0.004)	(0.010)
Log HHI X Q4 outward mobility	-0.000	0.000	0.000	0.000	-0.013
	(0.004)	(0.004)	(0.003)	(0.004)	(0.015)
Log outside-occ. options X Q1 outward mobility	0.060***	0.060***	0.060***	0.051***	0.003
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
Log outside-occ. options X Q2 outward mobility	0.100***	0.100***	0.100***	0.090***	-0.020
	(0.009)	(0.009)	(0.009)	(0.010)	(0.017)
Log outside-occ. options X Q3 outward mobility	0.106***	0.107***	0.107***	0.100***	0.004
	(0.010)	(0.010)	(0.010)	(0.010)	(0.016)
Log outside-occ. options X Q4 outward mobility	0.108***	0.108***	0.109***	0.094***	0.013
	(0.010)	(0.010)	(0.010)	(0.011)	(0.024)
Exposure control	0.004	0.004		0.003	-0.002
	(0.008)	(0.008)		(0.008)	(0.004)
Vacancy growth		-0.001	-0.001	-0.007	-0.003
		(0.001)	(0.001)	(0.004)	(0.003)
Predicted vacancy growth		-0.015	-0.015	0.028	-0.009
		(0.010)	(0.010)	(0.030)	(0.009)
Equal-weighted vacancy growth		-0.000			
		(0.000)			
Industry Bartik				0.143***	
				(0.014)	
Observations	184,411	184,411	184,411	169,341	158,393
Fixed effects	Occ-year CBSA-year	Occ-year CBSA-year	Occ-year CBSA-year	Occ-year CBSA-year	Occ-CBSA Year

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. Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

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

Dependent variable:	Log wage				
	(a) Emp weight	(b) Log emp weight	(c) Drop low rep. occs	(d) Occ rep. weight	(e) CBSA rep. weight
Log HHI *X* Q1 outward mobility	-0.044***	-0.027***	-0.023***	-0.011*	-0.020***
	(0.011)	(0.004)	(0.005)	(0.006)	(0.006)
Log HHI *X* Q2 outward mobility	-0.019***	-0.012***	-0.007*	-0.011*	-0.009**
	(0.005)	(0.003)	(0.003)	(0.006)	(0.004)
Log HHI *X* Q3 outward mobility	-0.001	-0.007*	-0.006	-0.004	-0.008**
	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)
Log HHI *X* Q4 outward mobility	-0.005	0.001	-0.005	0.002	0.001
	(0.008)	(0.004)	(0.005)	(0.006)	(0.004)
Log outside-occ. options *X* Q1 outward mobility	0.093***	0.067***	0.059***	0.067***	0.069***
	(0.025)	(0.012)	(0.011)	(0.013)	(0.012)
Log outside-occ. options *X* Q2 outward mobility	0.135***	0.108***	0.105***	0.118***	0.098***
	(0.024)	(0.010)	(0.009)	(0.013)	(0.010)
Log outside-occ. options *X* Q3 outward mobility	0.125***	0.115***	0.105***	0.116***	0.099***
	(0.032)	(0.010)	(0.010)	(0.012)	(0.010)
Log outside-occ. options *X* Q4 outward mobility	0.101**	0.117***	0.102***	0.116***	0.105***
	(0.043)	(0.010)	(0.011)	(0.014)	(0.009)
Vacancy growth	0.000	-0.001	-0.001	-0.001	-0.001*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Predicted vacancy growth	-0.020**	-0.015**	-0.020*	0.016	-0.023**
	(0.008)	(0.008)	(0.011)	(0.031)	(0.009)
Exposure control	0.025	0.006	0.009	0.014	0.002
	(0.021)	(0.008)	(0.009)	(0.013)	(0.009)
Observations	184,411	184,258	137,567	184,411	184,411

Notes: This table repeats the robustness checks in Table A8, but allowing the coefficient on the HHI and outside-occupation option index to vary by quartile of outward occupational mobility. Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. All specifications feature occupation-year and city-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A11: Regression of wage on HHI and outside options: heterogeneity by quartile of HHI

Specification:	2SLS IV	First stage (by quartile of HHI)			
Dependent variable:	Log wage	Log HHI (by quartile of HHI)			
		Q1	Q2	Q3	Q4
Log HHI instrument	-0.015***	0.048***			
X Q1 HHI	(0.005)	(0.003)			
Log outside-occ. options instrument	0.102***	-0.608***			
X Q1 HHI	(0.010)	(0.057)			
Log HHI instrument	-0.010*		0.021***		
X Q2 HHI	(0.006)		(0.001)		
Log outside-occ. options instrument	0.087***		-0.137***		
X Q2 HHI	(0.009)		(0.016)		
Log HHI instrument	-0.010*			0.024***	
X Q3 HHI	(0.005)			(0.002)	
Log outside-occ. options instrument	0.088***			-0.082***	
X Q3 HHI	(0.008)			(0.014)	
Log HHI instrument	-0.011**				0.031***
X Q4 HHI	(0.005)				(0.003)
Log outside-occ. options instrument	0.091***				-0.177***
X Q4 HHI	(0.009)				(0.021)
Vacancy growth	-0.001	-0.891***	-0.013	-0.551***	-0.086
	(0.001)	(0.190)	(0.008)	(0.050)	(0.056)
Predicted vacancy growth	-0.013	0.455	-0.176	-0.086	0.213**
	(0.010)	(0.698)	(0.217)	(0.106)	(0.104)
Exposure control	-0.000	10.547***	1.725***	0.829***	0.785***
	(0.007)	(0.838)	(0.096)	(0.039)	(0.022)
Observations	184,411	45,827	45,884	45,916	45,897

Notes: This table repeats our baseline regression in column (a), but allows the coefficients on the HHI and outside-occupation option index to vary based on the average HHI of the occupation-city cell in question. Specifically, occupation-city labor markets are split into four quartiles based on their average HHI over the four year period in question 2013–2016. Quartile boundaries for the average HHI in the occupation-city labor market over 2013–2016 are 276 (p25), 622 (p50), and 1,340 (p75). The remaining columns show the first stage HHI regressions separately for each HHI quartile. Other info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. Regressions have occupation-year and city-year fixed effects. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A12: Regression of wage on HHI and outside options: heterogeneity by occupational wage quartile

Specification:	2SLS IV	First stage			
Dependent variable:	Log wage	Log HHI (by occ. wage quartile)			
		Q1	Q2	Q3	Q4
Log HHI, instrumented	-0.009**	0.082***			
X Q1 occ. wage	(0.004)	(0.004)			
Log outside-occ. options, instrumented	0.047***	-0.357***			
X Q1 occ. wage	(0.015)	(0.068)			
Log HHI, instrumented	-0.004		0.097***		
X Q2 occ. wage	(0.003)		(0.004)		
Log outside-occ. options, instrumented	0.091***		-0.459***		
X Q2 occ. wage	(0.012)		(0.048)		
Log HHI, instrumented	-0.011***			0.074***	
X Q3 occ. wage	(0.004)			(0.003)	
Log outside-occ. options, instrumented	0.083***			-0.647***	
X Q3 occ. wage	(0.010)			(0.041)	
Log HHI, instrumented	-0.014***				0.071***
X Q4 occ. wage	(0.004)				(0.003)
Log outside-occ. options, instrumented	0.089***				-0.648***
X Q4 occ. wage	(0.010)				(0.048)
Vacancy growth	-0.001	-0.350***	-0.560***	-0.032*	-0.511***
	(0.001)	(0.077)	(0.063)	(0.018)	(0.196)
Predicted vacancy growth	-0.013	0.175***	0.124	-0.392*	-0.258
	(0.010)	(0.054)	(0.125)	(0.208)	(0.233)
Exposure control	0.000	1.588***	1.515***	1.158***	1.768***
	(0.007)	(0.070)	(0.045)	(0.046)	(0.069)
Observations	184,411	19,428	40,181	64,770	60,031

Notes: This table repeats our baseline regression in column (a), but allows the coefficients on the HHI and outside-occupation option index to vary based on the average wage of the occupation in question. Specifically, each occupation's average wage is calculated in our BLS OES data sample as of 2016, and occupations are split by employment-weighted quartiles of this distribution (such that all observations from one occupation across cities and years are always in the same quartile). The cutoff for the 25th percentile is \$13.01, the median is \$18.49 and the 75th percentile is \$31.74. The remaining columns show the first stage HHI regressions separately for each occupation wage quartile. Other info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. Regressions have occupation-year and city-year fixed effects. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A13: Regression of wage on HHI and outside options: heterogeneity by occupation group

	2SLS IV	First stage (by occupation group)								
Log HHI, instrumented	-0.010*	0.057***								
X Managers	(0.006)	(0.003)								
Log outside-occ. options, instrumented	0.107***	-0.642***								
X Managers	(0.013)	(0.067)								
Log HHI, instrumented	-0.014***		0.071***							
X Professionals, excl managers	(0.005)		(0.004)							
Log outside-occ. options, instrumented	0.077***		-0.849***							
X Professionals, excl managers	(0.010)		(0.063)							
Log HHI, instrumented	-0.015***			0.058***						
X Healthcare	(0.006)			(0.003)						
Log outside-occ. options, instrumented	-0.008			-0.207***						
X Healthcare	(0.012)			(0.052)						
Log HHI, instrumented	0.001				0.085***					
X Production	(0.004)				(0.004)					
Log outside-occ. options, instrumented	0.081***				-0.817***					
X Production	(0.013)				(0.058)					
Log HHI, instrumented	-0.015*					0.074***				
X Sales	(0.008)					(0.006)				
Log outside-occ. options, instrumented	0.128***					-1.069***				
X Sales	(0.016)					(0.105)				
Log HHI, instrumented	-0.004						0.088***			
X Office/Admin	(0.003)						(0.005)			
Log outside-occ. options, instrumented	0.087***						-0.730***			
X Office/Admin	(0.011)						(0.085)			
Log HHI, instrumented	0.000							0.101***		
X Operators/Laborers	(0.005)							(0.007)		
Log outside-occ. options, instrumented	0.068***							-0.162		
X Operators/Laborers	(0.014)							(0.138)		
Log HHI, instrumented	-0.027***								0.068***	
X Food/Cleaning/Protective Service	(0.007)								(0.004)	
Log outside-occ. options, instrumented	0.022								-0.009	
X Food/Cleaning/Protective Service	(0.015)								(0.065)	
Log HHI, instrumented	-0.016**									0.091***
X Personal Care	(0.008)									(0.010)
Log outside-occ. options, instrumented	0.054***									-0.564***
X Personal Care	(0.015)									(0.109)
Vacancy growth	-0.001	-0.931***	-0.020**	-0.374*	-0.466***	0.344***	-0.051	-0.772***	-0.397***	-0.605***
	(0.001)	(0.082)	(0.009)	(0.213)	(0.069)	(0.122)	(0.034)	(0.106)	(0.047)	(0.111)
Predicted vacancy growth	-0.009	-0.543***	-0.516	0.017	-0.127	1.498*	0.247	4.493***	0.204***	0.158
	(0.010)	(0.182)	(0.474)	(0.091)	(0.356)	(0.762)	(0.319)	(1.341)	(0.050)	(0.548)
Exposure control	-0.002	1.947***	1.153***	1.533***	1.209***	1.654***	1.392***	1.637***	1.306***	1.458***
	(0.007)	(0.100)	(0.077)	(0.058)	(0.062)	(0.106)	(0.080)	(0.102)	(0.067)	(0.087)
Observations	184,411	27,955	25,917	28,326	26,802	11,924	20,977	9,181	18,373	6,611

Notes: This table repeats our baseline regression in the first column, but allows the coefficients on the HHI and outside-occupation option index to vary for different aggregated occupation groups. The remaining columns show the first stage HHI regressions separately for each occupation group. Occupation groups map from SOC 2-digit occupations as defined in Appendix Table A6. Other info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 2013–2016 inclusive. Regressions have occupation-year and city-year fixed effects. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A14: Regression of wage on outside-occupation options: 1999–2016

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: OLS regressions				
oo^{occs}	0.140***	0.082***	0.095***	0.044***
	(0.010)	(0.004)	(0.005)	(0.006)
Observations	*1,944,370*	*1,944,370*	*1,944,370*	*1,944,370*
Panel B: 2SLS IV regressions				
oo^{occs} , instrumented	0.122***	0.070***	0.076***	0.030***
	(0.010)	(0.005)	(0.006)	(0.006)
Observations	*1,944,370*	*1,944,230*	*1,944,370*	*1,944,114*
Fixed effects	Year	Occ-Year, City	Occ-Year, City-Year	Occ-Year, Occ-City

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 1999–2016 inclusive. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A15: Regressions of wages on outside-occupation options, incorporating employment share

Dependent var.:	Employment share	Log wage	
oo^{occs} , instrumented	-0.226***	0.030***	0.026***
	(0.027)	(0.006)	(0.006)
Empl. share			-0.015***
			(0.001)
Fixed effects	Occ-City, Occ-Year	Occ-City, Occ-Year	Occ-City, Occ-Year
Observations	*1,931,901*	*1,931,901*	*1,931,901*

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

Notes: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses. Units of observation are 6 digit SOC by city by year, for all observations with available data over 1999–2016.

Table A16: 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}	0.184***	0.091***	0.106***	0.071***
	(0.013)	(0.007)	(0.011)	(0.009)
IV: oo^{occs}, instrumented	0.143***	0.113***	0.105***	0.125***
	(0.018)	(0.012)	(0.015)	(0.014)
Observations	*486,487*	*486,481*	*486,487*	*485,815*
Panel B: Major SOC Group (2-digit) regressions:				
OLS: oo^{occs}	-0.063**	0.081***	0.002	0.080***
	(0.030)	(0.009)	(0.024)	(0.009)
IV: oo^{occs}, instrumented	-0.182***	0.079***	0.060**	0.327***
	(0.038)	(0.028)	(0.029)	(0.113)
Observations	*137,650*	*137,650*	*137,650*	*137,609*
Fixed effects	Year	Occ-Year City	City-Year Occ-Year	Occ-Year Occ-City

Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,*** $p < .01$. Units of observation are 2-digit or 3-digit SOC by city by year, for all observations with available data over 1999–2016 inclusive. As noted in the paper, 'cities' 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 A17: Regressions of wage on outside-occupation option index, sample split into three periods, with different combinations of fixed effects

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: 1999–2006:				
OLS: oo^{occs}	0.132***	0.071***	0.085***	0.029***
	(0.011)	(0.004)	(0.005)	(0.005)
IV: oo^{occs}, instrumented	0.100***	0.053***	0.060***	0.018***
	(0.011)	(0.005)	(0.006)	(0.004)
Observations	*788,519*	*788,463*	*788,519*	*772,025*
Panel B: 2007–2011:				
OLS: oo^{occs}	0.141***	0.091***	0.097***	0.035***
	(0.010)	(0.005)	(0.006)	(0.005)
IV: oo^{occs}, instrumented	0.130***	0.080***	0.084***	0.013***
	(0.010)	(0.006)	(0.007)	(0.005)
Observations	*579,283*	*579,242*	*579,283*	565,394
Panel C: 2012–2016:				
OLS: oo^{occs}	0.149***	0.096***	0.105***	0.026***
	(0.010)	(0.005)	(0.006)	(0.008)
IV: oo^{occs}, instrumented	0.145***	0.090***	0.094***	0.011
	(0.012)	(0.007)	(0.007)	(0.009)
Observations	*576,568*	*576,525*	*576,568*	*562,592*
Fixed effects	Year	Occ-Year City	City-Year Occ-Year	Occ-Year Occ-City

Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,*** $p < .01$. Units of observation are 6-digit SOC by city by year, for all observations with available data over 2002–2016 inclusive (split into three five-year periods). As noted in the paper, 'cities' refers to CBSAs (metropolitan and micropolitan statistical areas) or NECTAs (New England city and town areas). Each cell reports the coefficient for the variable of interest (outside-occupation option index) in one regression specification, with included fixed effects held constant within each column.

Table A18: Regressions of wage on outside-occ. option index, *employment-weighted*, with different combinations of fixed effects

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
Panel A: OLS				
oo^{occs}	0.382***	0.094***	0.132***	0.027***
	(0.026)	(0.010)	(0.009)	(0.008)
Panel B: 2SLS				
oo^{occs}, instrumented	0.396***	0.096***	0.106***	0.026***
	(0.025)	(0.013)	(0.015)	(0.010)
First stage:				
Coeff. on $oo^{occs,inst}_{o,k,t}$	1.106***	0.886***	0.874***	0.923***
	(0.052)	(0.020)	(0.020)	(0.081)
1st-stage F-Stat.	451	1929	1943	129
Fixed effects	Year	Occ-Year City	City-Year Occ-Year	Occ-Year Occ-City
Observations	*1,944,370*	*1,944,230*	*1,944,230*	*1,931,901*

Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,*** $p < .01$. Units of observation are 6 digit SOC by city by year, for all observations with available data over 1999–2016 inclusive. The instrumented outside-occupation option index uses the national leave-one-out mean wage in outside option occupations to instrument for the local (city-level) wage, and the initial local employment share in outside option occupations to instrument for the current local employment share. Observations are weighted by the average employment of their occupation-city over the sample period. Each cell reports the coefficient for the variable of interest (outside-occupation option index) in one regression specification, with included fixed effects held constant within each column.

Table A19: Regression of wage on HHI and oo^{occs}: Right-to-work and non right-to-work states

Dependent variable:	Log wage			
	(a) OLS	(b) OLS	(c) 2SLS IV	(d) 2SLS IV
Log HHI X Non right-to-work	-0.007*** (0.001)	-0.004*** (0.001)	-0.011*** (0.003)	-0.008*** (0.003)
Log HHI X right-to-work	-0.014*** (0.001)	-0.010*** (0.001)	-0.018*** (0.003)	-0.015*** (0.003)
Log outside-occ options X Non right-to-work		0.099*** (0.008)		0.086*** (0.009)
Log outside-occ options X right-to-work		0.114*** (0.008)		0.104*** (0.009)
Vacancy growth			-0.001 (0.001)	-0.001 (0.001)
Predicted vacancy growth			-0.012 (0.010)	-0.013 (0.010)
Exposure control			0.011 (0.008)	0.007 (0.007)
Observations	184,411	184,411	184,411	184,411
F-Stat			377	201

Notes: This table shows our baseline regression specifications as in Table 3 but allowing the coefficients on the HHI and outside-occupation options to differ in right-to-work states and non right-to-work states. States are classified as right-to-work or non-right-to-work according to data from the National Conference of State Legislatures. States which passed right-to-work laws in 2015 or 2016 (Wisconsin and West Virginia) are coded as non-right-to-work in our sample, under the assumption that it would take some time for the passage of right-to-work statutes to affect labor market behavior. Other info: Heteroskedasticity-robust standard errors clustered at the city level shown in parentheses: $^*p < .1$,$^{**}p < .05$,*** $p < .01$. Units of observation are 6 digit SOC occupation by city by year, for all observations with available data over 2013–2016 inclusive. All specifications include occupation-by-year and city-by-year fixed effects.

Table A20: Counterfactual wage effects of setting HHI to 200,
excluding occupations with a represented-ness<0.5 in our BGT vacancy data

	2,500< HHI <10,000	1,500< HHI <2,500	500< HHI <1,500	200< HHI <500	0< HHI <200
Lowest mobility	7.6%	5.7%	3.6%	1.1%	0
	(0.7m)	(0.9m)	(3.6m)	(4.1m)	(7.3m)
Q2 mobility	3.0%	2.3%	1.4%	0.5%	0
	(0.5m)	(0.6m)	(2.8m)	(3.7m)	(10.2m)
Q3 mobility	2.1%	1.6%	1.0%	0.3%	0
	(0.4m)	(0.4m)	(2.5m)	(7.0m)	(13.8m)
Q4 mobility	0	0	0	0	0
	(0.3m)	(0.4m)	(1.7m)	(1.5m)	(1.8m)

Notes: This table repeats the analysis in Table 7, but *dropping* any occupations with an average represented-ness in our BGT vacancy data of less than 0.5 (roughly the bottom third of occupations). This is because our concentration data on these occupations might be a substantial overestimate of the true degree of employer concentration, if our data is disproportionately sampled from large employers for these occupations.