Getting labor markets right: outside options and occupational mobility [*]

Gregor Schubert, Anna Stansbury and Bledi Taska[†]

DRAFT. COMMENTS WELCOME.

June 2019

Abstract

Many analyses of important questions in labor economics use occupations as proxies for workers' labor markets. Yet workers often switch occupations, suggesting that workers' true labor markets rarely coincide with occupational boundaries. In this paper, we use a large novel dataset on occupational mobility to infer workers' job options outside their current occupation. Informed by labor market search models, we construct a measure of the value of workers' outside-occupation job options as the weighted average wage across other local occupations, weighted by occupational transition shares. We show that workers in cities with better outside-occupation job options have higher wages, and that plausibly exogenous Bartik-style shocks to the wages of workers' outside option occupations have a large, positive, and significant effect on wages in workers' own occupation. We then re-evaluate the recent literature on local labor market concentration, showing that failing to consider job options outside workers' occupations biases the estimated relationship of concentration and wages upwards and obscures important heterogeneity. We propose a measure of labor market concentration that takes into account workers' ability to change occupation and show that this measure has a stronger relationship with wages than a conventional single-occupation HHI. Overall, our work suggests that outside-occupation job options are important for workers' labor market outcomes, and that occupational mobility data provides a tractable way to incorporate them easily into labor market analyses.

[†]Schubert & Stansbury: Harvard University. Taska: Burning Glass Technologies.

1 Introduction

All labor market analysis requires a definition of the worker's labor market: the jobs that the worker could feasibly take outside her current job, or her outside options. It is common to take a simple binary approach to labor market definition, defining a local labor market as an occupation or industry within a geography: all jobs within that labor market are considered accessible to workers, and all jobs outside are not.

If the boundaries of an occupation, industry or geographic area were impermeable, so that workers could never (or rarely) switch, this approach would be appropriate. Workers' true empirical labor markets, however, include jobs in multiple different occupations, industries and geographic areas. Ignoring workers' ability to switch occupation or move location when defining labor markets will therefore underestimate workers' true outside options. At the same time, adopting too broad a definition of a labor market is likely to overestimate workers' outside options and their 'true' labor markets by including a number of jobs which are not feasible for workers to take.

We argue that this problem can be avoided by moving beyond the binary approach to labor market definition, and instead taking a probabilistic approach. In this paper, we focus on the occupational dimension. We identify workers' likely job options outside their own occupation using observed worker transitions between pairs of occupations. Using observed occupational transitions is a simple, non-parametric way to identify workers' 'revealed' labor market. It captures the job options outside workers' current occupation which are both sufficiently feasible and sufficiently desirable to be meaningful outside options – as revealed by workers' actual job moves. Importantly, empirical transitions capture a number of factors that are not visible in the most common alternative approach to measuring occupational similarity – using skill or task data – such as differences in amenities between occupations, and explicit labor market barriers like occupational licensing requirements. While not the focus of this paper, our method can also easily be adapted to identify outside options based on industrial or geographic mobility.

We obtain our occupational transition data from a new and unique dataset of resumes, collected by Burning Glass Technologies ("BGT"). The data captures 23 million workers in more than 100 million jobs during the years 2002–2018. Since resumes describe workers' career histories, this data gives us longitudinal excerpts from workers' lives and allows us to observe their job transitions. The large sample size enables us to document average transition shares between almost all of the 840 x 840 pairs of 6-digit SOC occupations in the U.S. with a high degree of confidence in their representativeness. We first use these data to show that occupational mobility is high (the probability of a worker changing her

1

SOC 6-digit occupation when she changes her job is greater than 20%[1]), highly heterogeneous across occupations, and poorly captured by aggregating up the SOC occupational hierarchy. We then use the network structure of the data - directed pairwise occupational transition shares - to create a new high-dimensional measure of 'revealed' pairwise occupational similarity. We show that these occupational transitions are plausible measures of occupational similarity: they capture underlying similarity between occupation pairs in tasks, leadership responsibilities, and amenities, and the direction of the transitions we observe intuitively corresponds to a worker's tendency to move up the career ladder and is consistent with documented changes in the structure of the labor market during the last two decades[2].

We use this occupational transition data, alongside data from the BLS Occupational Employment Statistics, to create an index of the average value of workers' outside-occupation job options within their local area for over 116,000 occupation-by-city units in the US[3]. This index is constructed as the weighted average of local wages in all occupations except the worker's own, where each weight is the product of the empirical occupation-to-occupation transition share (which proxies for the likelihood that the average worker's best job option outside her own occupation is in each of these other occupations), and the relative employment share of jobs in each destination occupation in the city (which proxies for the local availability of job options in the destination occupation).

Conceiving of the value of outside-occupation options as a transition-weighted average across local wages in different occupations is intuitively plausible. It can also be rationalized with a simple labor market search model. We present a framework in which employers offer employed workers a wage which depends on the ex ante expected value of their outside option each period. If workers reject this offer, they search in the labor market. All workers in a given occupation and city are identical and have an identical set of outside options, but because of labor market frictions each worker only receives offers from a subset of her outside options each period. The ex ante expected value of workers' options outside their current occupation or city are therefore the wages offered in those jobs, multiplied by the probability of moving into them, which can be proxied by observed occupational transitions and the local relative

---

[1]Note that since our measure only captures transitions from "steady" jobs (held longer than 6 months) to other "steady" jobs, and does not include short-term or part-time moves into other occupations while continuing to work in an old occupation, it likely underestimates actual occupational mobility. Our measure is comparable to measures in other work. For example, Kambourov and Manovskii (2009) find that annual occupational mobility in the PSID is 13%-18% and annual industry mobility is 10%-12%. For Austria, Nimczik (2018) shows that about three quarters of job movers leave their 2-digit industry annually. While not the focus of our paper, geographic mobility is also high: Molloy et al. (2011) find that 13% of US workers move to a different commuting zone within 5 years.

[2]In a similar vein, Macaluso (2019) finds that mobility between SOC 2-digit occupation groups in the US is highly correlated with task similarity, and Nimczik (2018) finds that most job moves in Austria involve moving up the career ladder.

[3]Specifically, we use almost the entire set of occupation-by-city units for which the BLS OES provides wage and employment data. We use "city" as shorthand for CBSAs (metropolitan and micropolitan statistical areas) and NECTAs (New England city and town areas).

availability of jobs in those occupations.

This framework gives structure to the way in which job options outside workers' own occupation can be expected to affect their wages. The greater the likelihood a worker will be able to transfer into a different occupation, the greater the number of jobs available in that other occupation, and/or the higher the relative wage in that other occupation, the more valuable an outside job option in that occupation is to the worker's current wage. This enables us to estimate the extent to which outside options outside workers' own occupation matter, and for which occupations they matter more or less.

In regressions at the occupation-city-year level over 1999–2016, we find that our indices of outside-occupation job options are significantly and positively related to wages. This relationship is both economically and statistically significant, and exists both cross-sectionally within occupations and within cities, and over time within the same occupation and city.

There is a concern, however, that this relationship could be driven by common shocks to similar occupations in the same city. We therefore generate quasi-exogenous shocks to workers' outside options: we instrument for local demand shocks to workers' outside-option occupations using the national leave-one-out mean wage in those occupations (analogous to Bartik-style shocks to outside occupation wages, in a method similar to Beaudry et al. (2012)). That is, we examine the effect of a nation-wide increase in the wage of outside option occupations p on the local average wage of occupation o. The positive and significant results persist: after a nation-wide increase in the wage of outside option occupation p, workers in occupation o who are in cities with more jobs in outside option occupation p see bigger wage increases. A one standard deviation[4] increase in the value of outside-occupation options is associated with 1.4-3.4 log points higher wages.

These results show that – in data that comprises almost the entire set of U.S. occupations and cities for which wage data is available, over 17 years – workers' wages respond to the value of their job options *outside* their own occupation. This in turn suggests (1) that the commonly-used labor market definitions of occupation-by-city are too narrow to reflect workers' true labor markets, (2) that revealed occupational mobility patterns can be used to infer workers' relevant job options outside their occupations, and (3) that differential availability of outside-occupation job options in different cities affects workers' wages.

We show the relevance of these insights for labor market research by applying our method to the study of the effect of labor market concentration on wages. A recent literature has documented a negative empirical correlation between local labor market employer concentration and wages in the U.S.[5] In theory,

---

[4]Within a given SOC 6-digit occupation and year, across different cities

[5]See, e.g. Azar et al. (2017, 2018); Rinz (2018); Lipsius (2018); Benmelech et al. (2018); Hershbein et al. (2019)

higher levels of employer concentration can create monopsony power for employers, enabling them to suppress wages below the competitive level since workers have few outside job options. However, all of the empirical analyses have been conducted only at the narrow local industry or occupation level. Our results above suggest that this is too narrow a labor market definition to appropriately identify employer concentration.

We therefore use our metrics of workers' outside-occupation job options to revisit these recent analyses of local labor market concentration and wages. We show that the coefficient on the HHI in wage regressions is biased upward (in absolute value) when workers' options outside their occupation are not considered: this occurs because workers with few local employers in their own occupation also tend to have worse local outside-occupation job options as measured by our index. We also show that the coefficient on the HHI in wage regressions is substantially higher for occupations with low outward mobility than for occupations with high outward mobility, which is consistent with the hypothesis that a lack of job options outside the occupation compounds the effects of labor market concentration within the occupation. Together, these suggest that a simple HHI calculated at the level of a local 6-digit occupation is inappropriate to identify local labor market concentration.

As an alternative, we suggest a probabilistic HHI measure which better captures the availability of job options outside workers' own occupation. In horse races with a conventional single-occupation HHI, we find that only the probabilistic measure has a negative association with wages. Moreover, the relationship between wages and the probabilistic HHI are significantly larger than that with the single-occupation concentration measure - even when controlling for outside-occupation job options.

Overall, our theory and empirical results suggest that a broader concept of local labor markets and worker outside options, taking into account occupational mobility, is important to labor market analysis. The differential availability of outside-occupation job options affects workers' wages, and ignoring this can result in misleading inferences about local labor market dynamics. Our probabilistic framework provides a simple, tractable way for researchers to incorporate workers' job options outside their occupations into labor market analysis.

1.1 Related work

The results in our paper build on a substantial literature on labor market definition, occupational similarity, and worker outside options.

Labor market definition and worker flows: Our paper contributes to a small body of work using worker mobility to estimate the extent of workers' labor markets. Manning and Petrongolo (2017) use

unemployment and vacancy flows among U.K. census wards to infer that workers search across spatially proximate areas, which leads to interdependent effects in response to local shocks. Nimczik (2018) uses the job mobility network among Austrian firms to identify clusters of firms which do not align well with traditional geographic units, and which predict the pattern of spillovers from local labor market shocks. Our work is related in using worker occupational mobility to estimate the extent of workers' local labor markets, and applies this to our knowledge for the first time to the U.S. context.

Occupational similarity: Our work also relates to the literature estimating the similarity in job requirements between pairs of occupations (or industries), using skill and task data (Macaluso, 2019; Gathmann and Schönberg, 2010), worker demographic similarity (Caldwell and Danieli, 2018), or – most closely – worker mobility flows between occupations or industries (Shaw, 1987; Neffke et al., 2017). Our new and unique large dataset of U.S. worker resumes provides estimates of 'revealed' occupational similarity through workers' occupational mobility. We demonstrate that occupational mobility reflects many dimensions of occupational task, skill and amenity similarity, as well as other aspects of occupational similarity not captured in these measures.

Outside options: Imperfect competition models of the labor market – where a degree of firm monopsony power arises either from search frictions or from firm size – suggest a role for outside options in wage determination (Boal and Ransom, 1997; Ashenfelter et al., 2013; Manning, 2003). Our paper therefore informs a large literature on imperfect competition and outside options. In particular, a range of papers identify the effects of plausibly exogenous empirical shocks on outside options[6]. These include Caldwell and Harmon (2018), who use exogenous variation in information about outside options through changes in workers' coworker networks in Denmark, showing that higher labor demand at other firms in a worker's information network leads to higher wages at her current firm, and Macaluso (2019), who shows that displaced workers whose skills are a worse match for the other available jobs in their city face worse post-layoff labor market outcomes.

Within this broad literature, our paper relates most directly to Beaudry et al. (2012) and Caldwell and Danieli (2018). Beaudry et al. (2012) demonstrate that the industrial structure in workers' local labor markets affects their wages through workers' outside options[7]: they show that local changes in the

---

[6]In a local setting, empirical analyses on this topic often need to contend with a version of the "reflection problem" identified by Manski (1993): a worker's outside option in bargaining (e.g. another job's wage) may be affected by the worker's own outcomes, thus creating a circular causal chain. As a result, some source of exogenous variation in outside options is necessary to identify causal effects. A theoretical resolution of this issue is provided by Talamàs (2018), who notes that if there is an unambiguously best match where neither of the parties has a credible outside option, all the other matches and bargaining outcomes can be determined from that.

[7]Bidner and Sand (2018) also show that the local industrial structure affects the gender wage gap, and that declines in the quality of outside job options for men have contributed substantially to the decline in the gender wage gap in the U.S.

availability of high-wage jobs in some industries have spillover effects on wages in all other local industries, as would be expected if those jobs represented relevant worker outside options in a Nash bargaining setting. In their empirical estimation, differential industrial composition across cities generates differential local exposure to national wage changes (uncorrelated with local unobservable trends), allowing them to estimate the general equilibrium effects and spillovers of plausibly exogenous wage changes. We use a similar IV strategy in our wage regressions, instrumenting for shocks to workers' local outside-occupation job options with the interaction of national wage shocks and local occupational composition: however, we differ in estimating the scope of workers' labor markets based on empirical estimates of occupational mobility, rather than assuming that *all* industries in a city matter equally to all workers.

Caldwell and Danieli (2018) is the first empirical paper that we know of to construct an outside options index at the aggregate level. They construct a measure of the quality of a given worker's outside options based on the diversity of jobs and locations in which other similar workers are observed: the more other jobs that workers similar to the initial worker do, the more outside options the initial worker is assumed to have. They find that diversity of outside options in German workers' labor market is strongly and significantly associated with higher wages. Our index differs methodologically from their index: specifically, the dynamic nature of our occupational mobility data allows us to incorporate the directed nature and asymmetry of job moves, as well as the dependence of individual's next job on their current job (because of on-the-job learning, network acquisition, or the development of some task-specific human capital). Our paper is also - to our knowledge - the first to study empirically validated outside options for the full set of occupations in the U.S. and their relationship with wages.

Employer concentration and wages: Recent work has found a large, negative and significant relationship between employer concentration and wages for occupations or industries within given geographic areas (Azar et al., 2017, 2018; Rinz, 2018; Lipsius, 2018; Benmelech et al., 2018; Hershbein et al., 2019). Our work suggests that performing aggregate analyses across all occupation-by-city labor markets without considering the differential degree to which occupations actually represent workers' true labor markets can lead to bias and obscure important heterogeneity. This suggests that if antitrust analyses are to use a measure of labor market concentration like an HHI (as recommended by Marinescu and Hovenkamp (forthcoming)), the HHI calculated at the level of a simple 6-digit occupation can be misleading, and an alternate HHI reflecting a better approximation of workers' true labor market should be used.

The remainder of the paper proceeds as follows: Section 2 discusses the BGT data set and presents descriptive findings on occupational mobility patterns and their determinants. Section 3 provides a sim-

ple search framework that motivates our empirical measure of outside-occupation job options. Section 4 estimates the effect of outside-occupation job options on wages. Section 5 re-evaluates the relationship between labor market concentration and wages in light of the importance of outside-occupation job options, and suggests an alternative HHI measure to capture this. The last section concludes.

2   Using occupational mobility to identify outside-occupation options

A worker's labor market is the set of jobs which are realistic options for her to work in: this includes both her current job and her outside options. For each worker, the labor market is likely to be slightly different, determined by many factors which vary across workers: the skills and qualifications required, the location, and the worker's individual preferences and constraints (for example around family responsibilities or commuting). Ideally, labor market analysis could define each individual worker's relevant labor market appropriately.

For more aggregate analysis however, it is not possible to define different labor markets for each individual worker. Instead, a relevant labor market must be defined at the desired level of analysis. We focus in this paper on occupations. We ask: on average, how valuable are jobs in occupation p as outside options for workers in occupation o? Alternatively put, how likely are they to be in these workers' relevant labor market?

2.1   Three approaches to estimate occupational similarity

The outside option value of jobs in occupations other than the worker's current occupation can be thought of on a two-dimensional spectrum. One dimension is feasibility: the likelihood that the worker could easily become a typically productive worker in the new occupation (the distance from the worker's current skill set). The other dimension is desirability: the degree to which the worker would like to do a job in the new occupation, compared to a job in their current occupation. A typical job in the new occupation is a more valuable outside option to the worker, the more feasible it is and the more desirable it is.

There are three plausible ways of estimating the relevance of one occupation as an outside option for another occupation:

1. Skill and task similarity

2. Demographic & qualification similarity

3. Occupational transitions

7

Skill- and task-based measure: Skill- and task-based occupational similarity measures define two occupations as more similar, the more similar the skills and tasks are that they require. A number of authors create measures of skill- and task-based occupational similarity. Macaluso (2019) for example measures occupational similarity as the vector difference of occupational skill content. Gathmann and Schönberg (2010) measure occupational similarity as the angular separation of occupations' task content vectors. A skill- and task-based occupational similarity measure is likely to capture many aspects of the feasibility of moving from one occupation to another, but cannot capture non-skill-related aspects of feasibility such as occupational licensing barriers. It also does not 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 - so the kind of people that work in one occupation may not want to work in the other. In addition, skill- or task-based similarity measures require substantial assumptions as to how skill and task data should be combined to create a similarity measure.

Demographic- and qualification-based measure: 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. Moreover, 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.

Transition-based measure: A transition-based measure defines occupation p as a better outside option for occupation o, the more workers move from jobs in occupation o to jobs in occupation p - using workers' mobility to reveal their true labor markets. This measure captures some combination of feasibility and desirability. By definition the occupational transitions that actually occur were feasible for the individuals making those transitions. In addition, in most cases since occupational transitions involve some element of choice, presumably the new occupation is on average similarly or more desirable

8

than the old occupation (incorporating the value of amenities such as work-life balance, status, or career concerns). Unlike a skill/task-based approach, a transition-based approach does not require the imposition of symmetry: 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. Finally, a transition-based measure has the advantage of being non-parametric, allowing 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 et al., 2007).

The transitions-based measure has a problem in that 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. 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[8]. 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.

For these reasons, in this paper we adopt the third, 'revealed labor market' approach, identifying outside-occupation options using occupational transitions. Our measure uses observed empirical occupational transitions as a proxy for the likelihood of occupation p being a feasible and desirable outside option for a worker in occupation o. Our preferred measure of occupational transitions $\pi_{o \to p}$ is the probability of a worker moving from occupation o to occupation p conditional on leaving her job (as defined formally in equation 1), since this explicitly captures workers' decisions between jobs in their own occupation and in other occupations. The higher is the proportion of workers of occupation o who transition

---

[8]This is similar to the way that choice probabilities map to expected value functions in discrete choice models with i.i.d. preference shocks (McFadden, 1974)

to work in occupation p when they leave their job, the more relevant we consider jobs in occupation p as outside options for workers in occupation o. Formally, we define:

$$
\begin{aligned}
\pi_{o \to p} &= Pr(\text{move from occ } o \text{ to occ } p | \text{leave job}) \\
&= \frac{Pr(\text{move from occ } o \text{ to occ } p \cap \text{leave job})}{Pr(\text{leave job})} \\
&= \frac{\text{Share moving from occ } o \text{ to occ } p}{\text{Share leaving job}}
\end{aligned}
\tag{1}
$$

2.2 Resume data from Burning Glass Technologies

Our data on occupational and job transitions is from a new proprietary data set of 23 million unique resumes covering 100 million jobs over 2002–2018, provided by labor market analytics company Burning Glass Technologies ("BGT"). 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 dataset.

We would ideally use this data to estimate annual transition probabilities between full-time jobs from one occupation to the next. Unfortunately, resumes often do not list the months in which jobs started and ended, and do not always indicate if jobs were part-time or full-time. To describe occupational mobility conditional on workers leaving their job, we therefore 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 $t+1$[9]. 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 this measure will also capture mobility between occupations in the form of working in two different occupations at the same time, as well as job mobility that consists of taking a new job while continuing to work in an old job[10].

Our measure of the probability of transition from occupation o to occupation p, $\pi_{o \to p}$, is therefore

---

[9]We drops jobs which are listed as lasting for 6 months or less to exclude temporary work, summer jobs and internships.

[10]Implicitly, we are assuming that taking up a secondary job in an occupation indicates its viability as an outside option to the same degree as moving primary occupations. Under this assumption, our inclusive measure is more appropriate than one that focuses only on job or occupation moves that involve abandoning a previous job or occupation entirely.

constructed empirically as follows:

$$\pi_{o \to p} = \frac{\text{Share from occ } o \text{ moving into occ } p}{\text{Share from occ } o \text{ moving into a new job}}$$

$$= \frac{\text{number working in occ } o \text{ in year } t \text{ who are observed in occ } p \text{ in year } t+1}{\text{number working in occ } o \text{ in year } t \text{ who are observed in a new job in year } t+1} \quad (2)$$

We estimate these occupation transition probabilities at the national level between 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 and 285,494 non-empty occupation-to-occupation transition cells out of a total 705,600 possible transition cells (840 x 840). We average the observed occupation-to-occupation transitions over all observations in years 2002–2015,[11] to capture as much as possible the underlying degree of occupational similarity rather than transitory fluctuations in mobility.

We calculate one further statistic from this data: the 'occupation leave share', which we use as an approximation to the share of people leaving their jobs who also leave their SOC 6-digit occupation:

$$\text{leave share}_o = \frac{\text{Share from occ } o \text{ leaving occ } o}{\text{Share from occ } o \text{ moving into a new job}}$$

$$= \frac{\text{number working in occ } o \text{ in year } t \text{ who are no longer observed occ } o \text{ in year } t+1}{\text{number working in occ } o \text{ in year } t \text{ who are observed in a new job in year } t+1} \quad (3)$$

The BGT resume data set is largely representative of the U.S. labor force in its distribution by gender and location. However, it over-represents younger workers and white-collar occupations. Since we are estimating occupational transition probabilities within each occupation, 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. We correct for the over-representation by age by re-weighting the observed occupational transitions by the U.S. population age shares by occupation, provided by the BLS for 2012-2017[12].

2.3 Occupational mobility: high and heterogeneous across occupations

We have argued that we can use occupational mobility to infer workers' latent likelihood of moving between two occupations. If, however, this latent likelihood is small or very homogeneous across oc-

---

[11]Most resumes in our data have observations up to 2017 or 2018. We exclude transitions in the most recent years to avoid bias: if we observe someone applying for a job in 2017, for instance, who has changed job in 2017 or 2016, they are not likely to be representative of the average worker (who stays in their job for 2 years on average). Therefore, the last year t for which we compute observed mobility is 2015 (where the 2015 values reflect workers in their initial occupation in 2015 who are observed in a different occupation in 2016).

[12]Further discussion on the data representativeness, including on sample selection concerns, is in the Data Appendix.

cupations, then job options outside of a worker's occupation are likely to matter in theory but not in practice. In this section we present descriptive statistics on occupational mobility over 2002-2015 in the BGT resume database,[13] showing that occupational transitions are in fact frequent, highly heterogeneous across different SOC 6-digit occupations in terms of both magnitude and directions, and poorly captured by aggregating up the SOC occupational hierarchy.

The average share of workers leaving their occupation in our data - or the probability of no longer being observed in their initial SOC 6-digit occupation from one year to the next - is 0.11 in our data.[14] The average share of workers moving to a new job - the probability of being observed in a different job in year $T+1$ from the one you are observed in in year T - is about 0.46, consistent with the average length of a job in our data being 2 years.[15] Combining these statistics, the average probability of a worker leaving her 6-digit occupation given that she leaves her job - the "occupation leave share" - is 0.23. The full distribution of the occupation leave share is shown for all 6-digit occupations in Figure 1 and in Table 2. As these indicate, there is fairly large variation in the average share of workers leaving their occupation when they leave their job. Ranking occupations by their occupation leave share, the median occupation has a leave share of 0.24, with the 25th percentile at 0.19 and the 75th percentile at 0.28[16].

Almost all of the occupations with low leave shares are highly specialized: mostly, various medical, legal and educational occupations, and people with specific skills such as firefighters or graphic designers (see Table 4). In contrast, many of the occupations with high leave shares require mostly generalizable skills, such as restaurant hosts and hostesses, cashiers, tellers, counter attendants, and food preparation workers. The difference in mobility can be substantial: over 30% of telemarketers or hosts and hostesses leave their occupation when they leave their job – around three times greater than the occupation leave share for pharmacists, lawyers or licensed practical and vocational nurses. This suggests that the SOC 6-digit occupation is a substantially better measure of the true labor market for some occupations than

---

[13]Note that these averages overweight more recent years, since we have more observations in those years.

[14]The average leave share for all people (*not* just conditional on those leaving their job) in our data is 0.11. This is somewhat lower than the occupational mobility estimate from Kambourov and Manovskii (2009), who use the CPS to find occupational mobility of 0.20 at the Census 3-digit level for the late 1990s; it conversely is somewhat higher than Xu (2018) who finds annual occupational mobility of 0.1 in 1994 falling to 0.08 by 2014. The fact that our measure is relatively low compared to Kambourov and Manovskii (2009) is interesting, since sample selection bias would be expected to *overstate* occupational mobility in our data set if the people applying for jobs (whose resumes we observe) are more likely to be mobile than those not applying for jobs. However, our measure of outward occupational mobility is not strictly comparable to the concept of annual occupational mobility used by Kambourov and Manovskii (2009) and Xu (2018) because of the nature of our resume data: we count people who are observed in occupation o in year t at any point, but not observed in the same occupation o at any point in the following year $t + 1$, whereas conventional measures of annual occupational mobility count people who are no longer working in their same occupation o at the same time the following year.

[15]Note that leaving your job does not necessarily entail leaving your firm. The CPS reports that median employee tenure in 2018 was 4.2 years, so the average duration of a job at 2 years is consistent with workers working on average 2 jobs at their same employer.

[16]The spread is even wider in the tails: the 5th percentile is 0.11 and the 95th percentile is 0.38.

for others.

The SOC hierarchy structure groups occupations with other ostensibly similar occupations. However, mobility to a different SOC 6-digit occupation is not substantially lower than mobility to a different SOC 2-digit occupation. For the median occupation, 86% of moves to a different SOC 6-digit occupation are also to a different SOC 2-digit occupation, but this is highly heterogeneous by occupation: it is only 69% at the 10th percentile and is 96% at the 90th percentile (see Table 3 and Figure 2). For example at the low end, only 39% of systems software developers who move to a different 6-digit occupation also move to a different 2-digit occupation: most move to other computer-related occupations within the same 2-digit SOC occupation group. In contrast, 95% of flight attendants and 87% of counter attendants who leave their 6-digit SOC occupation also leave their 2-digit SOC occupation group. Flight attendants rarely move to other transportation occupations, instead moving to other white-collar jobs in office and administrative work or sales; counter attendants do often move to other food preparation and service occupations, but more often move to jobs in administrative work or retail[17]. This suggests that inferring occupational similarity or mobility by aggregating up the SOC classification structure does not capture workers' true occupational labor markets, and captures them differentially well or poorly for different occupations.

As would be expected, there are few observed transitions between most pairs of occupations: the occupational transition matrix is sparse (as shown in Figure 3). While many people are observed moving into a new occupation each year, there are only a few 'thick' occupational transition paths, where the transition probability is greater than negligible. For example, conditional on moving into a new occupation, there are only 189 pairs of 6-digit occupations which have a transition probability of 10% or greater (out of 284,797 observed non-zero occupational transition cells)[18].

Finally, the occupational transition matrix is highly asymmetric. Many occupational transition paths

---

[17]It is interesting to note that the 2-digit SOC occupation group fails to capture two different types of move: lateral moves to different types of occupations/industries, and promotions within the same type of occupation/industry. This is illustrated well by a comparison of counter attendants and registered nurses. Figures 5 and 6 illustrate the most common occupation transition paths for counter attendants and registered nurses respectively. For both of these occupations, the majority of people who leave their SOC 6-digit occupation also leave their SOC 2-digit occupation group, but the pattern is very different. Counter attendants' outside-occupation job options are very diverse, and dispersed across several different occupational groups: principally, counter attendants' occupation moves are lateral moves into jobs in sales, office & administrative work, and food preparation and service. In contrast, almost all registered nurses who leave their occupation do so through a promotion to become medical and health service managers, with very few transitions to any other occupations.

[18]Many of these thick transition paths are within close occupational families. For instance, 30% of licensed practical and vocational nurses who leave their occupation become registered nurses, 15% of short order cooks who leave their occupation become restaurant cooks, and 13% of light truck drivers become heavy truck drivers. Another set largely represent career progressions: 16% of human resources specialists become human resources managers, and 18% of legal secretaries who leave their occupation become paralegals or legal assistants the following year. Finally, some thick transition cells demonstrate occupational similarity across conventionally defined occupational boundaries: 21% of biological scientists who leave their occupation become operations research or management analysts, and 13% of meat, poultry and fish cutters who leave their occupation become truck drivers.

are thick one way and thin the other: the correlation between the transition share of occupation o to occupation p and the transition share of occupation p to occupation o is only 0.02[19]. This partly appears to reflect career progression; it also reflects the fact that some occupations appear to be fall-back job options for many different other occupations, particularly for transitions where workers in an occupation with specialized skills move to one which requires generalist skills (for example, some commonly transitioned-to occupations include retail salespersons, cashiers and secretaries and administrative assistants).

Taken together, these facts suggest that: (1) the SOC 6-digit occupation tends to be a better proxy for workers' true labor market for occupations requiring highly specialized skills, than for those requiring generalist skills; (2) there is a very large difference across occupations in the degree to which the SOC 6-digit occupation is an appropriate definition of workers' true labor market; (3) aggregating to a higher level of SOC code for occupations is not an appropriate way to fix this issue of labor market definition, (4) the sparse nature of the occupation-to-occupation transition matrix suggests that for many occupations, workers' true labor markets can be constructed out of relatively small clusters of similar occupations (as we do in this paper), and (5) the directed nature of the occupation-to-occupation transition matrix suggests that outside-occupation job options should not be considered symmetric across occupations.

These facts inform the approach that we take in this paper: imputing workers' outside options outside their local occupational labor market from occupation-to-occupation transition shares.

2.4 Determinants of occupational mobility

To use worker transitions to infer the network of worker outside options, we must assume that two occupations with more frequent transitions between them are more similar to each other in feasibility and/or desirability. It is possible, however, that our empirical occupational transition shares reflect something idiosyncratic in our data, or short-run contractions or expansions of different occupations, rather than latent similarities between occupations. Though the size, time horizon, and relative representativeness of our data should do something to assuage these concerns, as with any finite sample, our data may contain spurious variation in occupational flows that does not represent underlying occupational simliarity. To allay these concerns, we explore the degree to which occupational mobility measured using our resume database reflects similarities between different occupations in terms of task requirements, wages, job amenities, and leadership responsibilities.

---

[19]The correlation between the absolute size of the flows, on the other hand, is large: 0.93. This reflects the fact that transitions from a small occupation to a large one are much more likely than transitions from a large occupation to a small one.

2.4.1 Job characteristic measures

Task requirements. To measure similarity in terms of tasks required, we use two different approaches from prior literature. Our first approach, proposed by Macaluso (2019), is to use the vector difference between the importance scores for "Skill" task content items provided by the O*Net database of occupational characteristics.[20] 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[21].

Job amenities. Another potential factor in determining moves between jobs - and a source of non-monetary benefits - may be job amenities offered. We focus on amenities in the form of "temporal flexibility" of jobs[22]. To measure similarity in the temporal flexibility offered by different occupations, we use 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.[23].

Leadership responsibility. Another reason for observing occupational transitions may be career advancement, as workers move into positions of increasing responsibility or seniority (which is often reflected in a change of occupation). To study whether this appears in our data, we identify occupational

---

[20]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).

[21]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.

[22]These 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.

[23]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. 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.

characteristics measuring leadership responsibilities from the O*Net database, and create a new "leadership" composite measure defined at the level of each SOC 6-digit occupation. 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.[24] The final list obtained from this selection algorithm comprises 7 different occupational characteristics[25]. We use the mean score across these 7 characteristics as our "leadership" composite.[26]

2.4.2 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:

$$\pi_{o \to p} = \alpha_o + \beta^{abs} |X_{occ\,p} - X_{occ\,o}| + \gamma |\Delta w_{o \to p}| + \epsilon_{op}. \tag{4}$$

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[27], but note that the results are qualitatively similar without the wage controls.

If our occupational transitions measure does capture the feasibility and desirability of an occupational move, we would expect the coefficient on the absolute difference in characteristics to be negative: the

---

[24] 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.

[25] 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.

[26] 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.

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

greater the difference between two occupations, the less likely we should be to observe the worker moving from one into the other when she leaves her job. 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 7. 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.

2.4.3 Directed occupational transitions

The previous results impose symmetry on the likelihood of occupational transitions. However, 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. For example, it is more likely that a worker will be able to move from a job requiring specialized skills to a job requiring generalized skills than in the reverse direction; it is more likely that a worker will move up rather than down the career/leadership hierarchy; and it is more likely that workers will move towards structurally growing than structurally declining occupations.

To study whether differences in characteristics also predict the direction of occupational flows, we estimate a similar regression equation to that shown in equation 4, but now using the *relative* (target minus origin) difference in occupational characteristics as the independent variable:

$$\pi_{o \to p} = \alpha_o + \beta^{rel}(X_{occ\ p} - X_{occ\ o}) + \gamma \Delta w_{o \to p} + \epsilon_{op}. \tag{5}$$

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 5 for the different measures are shown in Figure 8[28].

A number of our predictions are borne out in the data: First, we find that workers are more likely to move towards jobs with higher wages. This suggests that the deliberate exercise of outside options likely plays a substantial role in the transitions that we observe, and alleviates some concern that occupational transitions may be caused to a large degree by layoffs or other negative shocks.

Second, we find that workers transition on average *towards* jobs that require more leadership re-

---

[28]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.

sponsibility - as would be expected from moves up the career ladder. This suggests that our transition probabilities do not just capture lateral moves, but also the outside option of moving up to a job with greater responsibility.

Third, the results show that occupational transitions have on average been *towards* occupations that have higher analytical content and out of occupations with more routine task requirements, as well as towards ocupations that require social skills or both social and analytical skills. 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 Autor et al. (2006), and the increasing demand for social skills documented by Deming (2017).

Fourth, our results show that workers have on average been moving into occupations that require more contact and working relationships with others (and so have less time flexibility). Once again this could reflect both career progression for individual workers towards managerial occupations, and/or an aggregate trend towards less time flexibility in the labor market[29].

2.4.4 Explanatory power of characteristic-based measures for mobility

Since similarity in tasks, temporal flexibility, and leadership requirements are strongly correlated with occupational transitions, it is worth asking whether our occupational mobility measure actually goes beyond the variation captured by these characteristic-based similarity measures. One way to do this is to examine the explanatory power of the different characteristic-based measures for the likelihood of an occupational transition, $\pi_{o \to p}$. Table 18 in the Appendix 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: by itself, skill distance explains only 1.1% of the observed variation in occupational mobility conditional on job changes at the 6-digit level[30], and relative and absolute differences in wages, job amenities, leadership composites, and other skill composites, explain 0.3%, 2.1%, 1.7%, and 3.5%, respectively.[31] This suggests that characteristic-based measures of occupational similarity fail to capture a number of dimensions that are important for

---

[29]To the degree that less time flexibility exacerbates the gender gap in wages (Goldin (2014)), this trend could contribute to the slow-down in the closing of the gender wage gap in recent decades as documented, for instance, by Blau and Kahn (2017).

[30]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.

[31]In regressions not reported here, we also find that all of these variables have incremental explanatory power of a similar magnitude when included *together* with skill distance in the regression.

predicting occupational mobility.

This 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.

Overall, these results support our approach of using observed occupational transitions to proxy for the feasibility and desirability of a given occupation as an outside option, since (1) occupational mobility is correlated with other characteristic-based measures of occupational similarity, (2) occupational mobility appears to be capture a number of factors which are unobserved in current measures of characteristic-based similarity, (3) occupational mobility is able to capture the asymmetry of occupational transitions more easily than characteristic-based measures, and (4) occupational mobility measures do not require the enumeration of all relevant occupational characteristics and the construction of a parametric model of the determinants of occupational similarity.

3 Theoretical framework

The degree of occupational mobility we document in section 2 suggests that many of workers' outside job options are outside their own occupations. We use our empirical measures of pairwise, directed occupational mobility to construct a measure of the value of these outside-occupation job options, and study its relationship with workers' own wages.

An intuitive measure of the value of workers' job options outside their own occupation is a weighted average of the wage in each alternative occupation, weighted by some measure of the likelihood that the

worker's best outside job offer will be in each of those alternative occupations:

$$\text{value of outside-occupation option}_{o,k} = \sum_{p}^{occs} Pr(\text{job in occ } p \text{ is best outside option}) \cdot \text{wage}_p \quad (6)$$

In the next section, we outline a simple search model which lays out more formally the assumptions under which this probability-weighted average wage is a valid measure of the value of outside-occupation job options. In the appendix, we also show that the same probability-weighted average wage can be justified as a measure of outside-occupation job options in a simple matching model with heterogeneity in outside options and without search frictions. The true structure of the labor market is likely to be somewhere in between: the fact that both extreme case models rationalize our measure simply suggests that the exact structure of the search frictions in the labor market are less important for our measure than the assumption that there is *on average* some heterogeneity across workers in terms of the occupation that is their best outside job option.

3.1 Model setup

This model, based on the search-and-matching framework in labor markets (for a review, see Mortensen and Pissarides (1999); Rogerson et al. (2005)), has two core building blocks:

- Each employed worker bargains with her existing employer over the wage at the start of each period. The outcome of the bargain depends on the worker's outside option if she does not continue to work at the firm. She does not know her outside option with certainty: instead, the expected value of her outside option is her expected wage if she leaves her current job to search for other jobs.

- Job seekers apply for jobs to all employers they could feasibly work for, and receive offers from a subset of these employers. Each job seeker accepts the job offer which pays the highest wage.

These are the key items that will generate our formula for the average value of outside options to the worker as a probability-weighted average of the wages at different jobs.

The more detailed set-up is as follows:

Employed workers: Each employed worker Nash-bargains with her employer i at the start of each period. The outcome of wage bargaining is a wage w_i equal to the value of the worker's outside option oo_i, plus

a share β of the match surplus created by the worker in working for that firm:

$$w_i = \beta(MPL_i - oo_i) + oo_i$$

$$= \beta MPL_i + (1 - \beta)oo_i \tag{7}$$

The worker's outside option is to leave her current employer and search for a job in the rest of the labor market (as described below). We assume that, in expectation, all employed workers at the same firm have the same outside option.

Job seekers: Each job seeker working in occupation o and city k applies to all feasible employers j. Each employer offers the worker a job paying w_j with probability α_j[32]. Once she has received all her offers, the job seeker accepts the offer with the highest wage. If she does not receive an offer from any employers in her feasible set N, she moves to unemployment for the period and receives payoff b. She can then search again for a job in the next period.

Job displacement: Each period, fraction ξ of workers are exogenously displaced from their job. They become job seekers and search for a new job. While employed workers can choose to leave their job, in equilibrium they will not because their employer will always offer them a wage which is weakly greater than the expected value of their outside options.

3.1.1 The value of workers' outside options

The probability a worker moves to any one employer j given that she leaves her existing job is the product of the probability that she receives an offer from that employer, α_j, and the probability that the wage offered to her by that employer is the maximum of all the wages offered to her this period:

$$Pr(\text{move to employer } j) = \alpha_j \cdot Pr(w_j \text{ is best offer}) \tag{8}$$

The value of the worker's outside option oo_i is equal to her expected payoff if she leaves her current employer and applies for jobs at other firms. The expected value of this outside option is therefore:

$$oo_i = \sum_{j=1}^{N} Pr(\text{move to employer } j) \cdot w_j + \prod_{j=1}^{N_i}(1 - \alpha_j) \cdot b \tag{9}$$

Note that $\prod_{j=1}^{N}(1 - \alpha_j)$ is the probability that worker i receives no offers from any firms and is therefore

---

[32]This could encompass any employer-specific characteristic which influences the propensity to make a job offer, such as the employer's current labor demand as well as aggregate labor market tightness.

equivalent to the probability that worker i becomes unemployed if she leaves her current job.

Therefore, the expected value of the worker's outside option - leaving her employer and searching across all other feasible firms in the labor market - is a weighted average of the wages she would be paid at all those firms, where the weight on each firm's wage is the probability that she ends up moving to that firm if she leaves her current job, and of the unemployment benefit b[33], where the weight is her probability of becoming unemployed if she leaves her current job.

3.2 Within-occupation and outside-occupation options

Since we focus in this paper on occupational labor markets and outside-occupation job options, we segment the worker's set of feasible employers into two categories: the outside options represented by employers in the same occupation o, which we denote oo^{own}, and the outside options represented by employers in other occupations p, which we denote oo^{occs}. For simplicity, we do not consider job options outside the worker's own city or the outside option value of unemployment[34]. In the appendix, we show how our theory can be extended to consider job options outside workers' own city, or indeed to consider any other definition of the base labor market.

We can therefore segment equation (9) from above into the probability that the worker's best job offer is in occupation p, the probability that the best job offer *within* occupation p is in firm j, and the wage offered by firm j.

$$oo_{i,o} = Pr(\text{job in own occ } o \text{ is best offer}) \cdot \sum_{j=1}^{N_{\text{occ } o}} Pr(\text{move to employer } j) \cdot w_{j,o} \tag{10}$$

$$+ \sum_{p}^{N_{\text{other occs}}} Pr(\text{job in occ } p \text{ is best offer}) \cdot \sum_{j=1}^{N_{\text{occ } p}} Pr(\text{move to employer } j | \text{move to occ } p) \cdot w_{j,p}$$

$$\tag{11}$$

We then make two assumptions which will enable an empirical application. First, as an empirical analog for the probability that a job in occupation p is the worker's best job offer, we use the national average transition share between occupation o and occupation p as measured in our BGT resume data set, $\pi_{o \to p}$, multiplied by the relative employment share of occupation p in city k (compared to the national

---

[33]More precisely, this can be considered to be the value of any unemployment benefit the worker receives *plus* the monetary-equivalent of any utility the worker receives from being unemployed.

[34]Since unemployment rates are generally in the single digits, and unemployment benefits are low in the U.S., the outside option value of unemployment is likely to be small for most workers. Jaeger et al. (2018) find that the outside option value of non-employment is negligible for most workers in Austria, which has both higher unemployment and more generous unemployment benefits than the U.S.

average), $\frac{s_{p,k}}{s_p}$:

$$Pr(\text{job in occ } p \text{ is best offer}) = \frac{\text{workers moving from occ } o \text{ to occ } p}{\text{workers leaving job in occ } o} \cdot \frac{\text{emp. share in occ } p \text{ in city } k}{\text{national emp. share in occ } p}$$

$$= \pi_{o \rightarrow p} \cdot \frac{s_{p,k}}{s_p} \tag{12}$$

Second, we assume that the probability that a worker would move to a job in firm j in occupation p, conditional on moving to *some* job in occupation p, is proportional to firm j's employment share in that occupation, $\sigma_{j,p}$, following Burdett and Mortensen (1980)[35].

This implies the following expected value of outside options for workers in firm i in occupation o and city k:

$$oo_{i,o,k} = oo_{i,o,k}^{own} + oo_{i,o,k}^{occs}$$

$$= \pi_{o \rightarrow o} \cdot \underbrace{\sum_{\substack{j \neq i}}^{N_{i,o,k}} \sigma_{j,o,k} \cdot w_{j,o,k}}_{\text{within-occupation job options}} + \underbrace{\sum_{\substack{p \neq o}}^{N_{occs}} \pi_{o \rightarrow p} \cdot \frac{s_{p,k}}{s_p} \cdot \bar{w}_{p,k}}_{\text{outside-occupation job options}} \tag{13}$$

This expression states that the ex-ante value of the component of workers' outside options based on jobs in *other* occupations is the weighted average of wages in other occupations, weighted by the share of workers from the initial occupation transitioning to each of the other occupations p and the relative local availability of jobs in occupation p. In the next section, we take this expression for the outside-occupation options to the data to measure how outside options vary across different locales in the U.S.[36]

4 Outside-occupation options and wages

4.1 Empirical outside-occupation option indexes

Our theoretical framework in section 3 gave us an expression for the outside option value of jobs outside workers' own occupation for workers in occupation o and city k: the weighted average of local wages in other occupations $\bar{w}_{p,k}$, with the weights the product of national occupational transition probabilities,

---

[35]Burdett and Mortensen (1980) assume that the conditional probability that a job offer received by a searching worker is from firm i is equal to firm i's employment share, σ_i.

[36]Our model assumes rational expectations on the part of both workers and employers to arrive at this formula: on average, both employers and workers know what the expected probability is of each worker being able to find a job at a different employer. Greater uncertainty around these expectations should not affect the specification of our average outside option index, but will increase noise in the measure. Systematically biased expectations by either workers and/or employers will affect the level of the outside option index but not our regression results. The only case in which workers'/firms' uncertainty around the probabilities of occupational transitions might bias our regression results is if workers and firms in certain occupations are *both* more likely to systematically under-/over-estimate the ability of workers to leave the occupation, *and* more likely to have higher/lower wages than other occupations.

$\pi_{o \to p}$, and the local relative employment share of occupation p, $\frac{s_{p,k}}{s_p}$:

$$oo_{o,k,t}^{occs} = \underbrace{\sum_{p \neq o}^{N_{occs}} \pi_{o \to p} \cdot \frac{s_{p,k,t}}{s_{p,t}} \cdot \bar{w}_{p,k,t}}_{\text{outside-occupation job options}}$$

We construct this index at the annual level for as many SOC 6-digit occupations and US cities over the years 1999–2016 as our data allows[37]. We use our data on occupation-to-occupation transitions from the Burning Glass Technologies resume data to construct $\pi_{o \to p}$[38], and we use employment data and wage data from the BLS Occupational Employment Statistics (OES) to construct the 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. The BLS OES data does not exist for many of the occupation-city pairs: of the possible 786,335 occupation-city pairs, wage data in the BLS OES only exists for approximately 115,000 each year. The missing occupations and cities are primarily the smaller ones.

For the sake of simplicity, in this paper we only consider outside-occupation job options *within a worker's own city*, and do not consider workers' outside job options in other cities. Empirically, annual outward residential mobility from metropolitan areas is approximately 3%[39], suggesting that while migrating to other cities can be an important outside option, occupational mobility is substantially more important for most workers than geographic mobility[40]. Nonetheless we hope that extending our probabilistic measure of outside options to account for geographic mobility may be a fruitful avenue for future research.

4.2 Wages and outside options

To study the relationship of outside-occupation job options with wages, we regress the log of average wages by occupation and city on the log of our index of outside-occupation options and various combinations of fixed effects $\Gamma_{o,k,t}$ in the following specification, where coefficient β estimates the relationship

---

[37] As noted above, 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. We would rather use Commuting Zones than CBSAs and NECTAs, since they are better measures of local geographic labor markets. However, occupational wage data is not available for Commuting Zones.

[38] As in the descriptive section, we exclude all occupations for which we have fewer than 500 person-year observations in the BGT data, leaving us with 786 SOC 6-digit occupations out of a possible 840. Our regression results are robust to including these occupations.

[39] According to county-to-county mobility data constructed from IRS tax returns.

[40] Annual outward mobility from a SOC 6-digit occupation exceeds 10% for more than half of all occupations, as we showed earlier in Table2.

between outside-occupation job options and wages:

$$log(\bar{w}_{o,k,t}) = \alpha + \beta log(oo_{o,k,t}^{occs}) + \Gamma_{o,k,t} + \epsilon_{o,k,t} \tag{14}$$

As in the construction of our outside option index, we use BLS OES data for average occupational wages by city and year for the dependent variable $\bar{w}_{o,k,t}$. Our full data set for the regressions comprises 1.95 million occupation-city-year observations: 394 cities and 753 6-digit SOC occupations over 17 years[41].

Table 6 shows the results of this regression across all occupation-city labor markets at an annual frequency over 1999 to 2016 inclusive, with progressively more fixed effects. Column (1) shows that there is a strong positive correlation in the raw data between outside-occupation options and wages. Column (2) has occupation-year and city fixed effects and column (3) has city-year and occupation fixed effects: they show that in the cross-section, occupation-city-year cells which have higher oo^{occs} compared to the national average for their occupation have significantly higher wages. Column (4) has occupation-by-city and occupation-by-year fixed effects, and so identifies only off annual variation in outside-occupation options compared to their mean for each occupation-by-city and occupation-by-year unit. The coefficients are positive and significant at the 1% level in all specifications, with the magnitudes in columns (2) through (4) suggesting that a 10 log point higher value of outside options in other occupations is associated with 0.5-1.1 log points higher wages in the workers' own occupation[42]; or a 1 standard deviation[43] higher value of outside options in other occupations is associated with 1.7-3.7 log points higher wages in the workers' own occupation.

4.3 Instrumental variable regressions

Endogeneity issues may be expected to bias the coefficients on our outside-occupation option measure upwards in our simple regressions. Shocks to the demand or supply of a similar occupation in your own city in a given year may also be direct shocks to the demand or supply of your own occupation in your city in that year (driven, for example, by a common product market shock or a regulatory change). In addition, there is a reverse causality or reflection problem: if occupation p is an outside option for workers in occupation o, and occupation o is an outside option for workers in occupation p, then a wage increase in o will increase wages in p and vice versa.

Ideally therefore, we could identify exogenous shocks to the wages in workers' outside-occupation

---

[41]Although not all SOC occupations have data for all cities or all years

[42]The results are similar if we run an employment-weighted regression, with coefficients between 0.3 and 1.8.

[43]This represents the average standard deviation of the logged outside-occupation option index *within* each occupation and year across different cities.

options which do not affect and are not affected by the wages in their own occupation. At the micro level with individual occupations this may be possible, but it is more difficult when looking to identify aggregate relationships. We therefore 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 a worker's own city k, we use the leave-one-out national mean wage for occupation p, excluding the wage for occupation p in city k.

In addition, 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[44]. 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, $\bar{w}_{p,\cancel{k},t}$, where the weights are the product of the year 1999 relative employment share in each of those occupations in the worker's own 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, $\pi_{o \to p}$.

$$oo^{occs,inst}_{o,k,t} = \sum_{p}^{N_{occs}} \left(\pi_{o \to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\cancel{k},t} \right) \tag{15}$$

The key identifying assumption for the wage instrument is that the national leave-one-out mean wage in outside option occupation p is correlated with the local wage in occupation p, but is not correlated with the local wage in initial occupation o (after controlling for the fixed effects: occupation o-by-city and occupation o-by-year). Identification is achieved from two factors. Identifying variation within the same occupation across different cities comes from differences in each city's initial exposure to outside option occupations[45]. Identifying variation over time within the same occupation-city cell comes from national (leave-one-out) changes over time in wages of occupations that represent local outside options. 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 these workers were more exposed to the shock to their outside options). This instrumental variable strategy is closely related to that of Beaudry et al. (2012), who 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.

---

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

[45]This refers to the relative employment share of each occupation p in city k compared to the national average, in either 1999, or the first year in data if there is no data for that occupation and city in 1999 in the OES.

We show the reduced form results of our instrumented regressions in Table 7. The results for the instrumented oo^{occs} index remain positive and strongly significant, with magnitudes only slightly smaller than the non-instrumented regressions. Columns (2) and (3) show that workers in cities which have a relatively high proportion of their employment in their outside-option occupations have higher wages, compared to workers in the same occupation in cities with a lower proportion of employment in their outside-option occupations. Column (4) shows that for a given occupation-city unit, in years where the national wage in workers' outside-option occupations rises, the wage in workers' own occupation also rises. The coefficient magnitudes suggest that a 10 log point higher outside-occupation option index is associated with 0.4-1.0 log points higher wages in the workers' own occupation[46]; and so that a 1 standard deviation higher outside-occupation option index is associated with 1.4-3.4 log points higher wages in the workers' own occupation.

4.4 Alternate channels: wage bargaining and mobility

Our theoretical model focuses on the effect of improved outside-occupation job options on workers' wages through the bargaining channel: higher wages in an outside-occupation job option lead to higher wages in the worker's own occupation because of more bargaining leverage. We note, however, that there is another channel by which outside-occupation job options can affect wages: the mobility channel. As the wages in an outside option occupation p rise, some workers from initial occupation o will move to occupation p. The supply of workers in occupation o falls, and so the wage rises[47]. Note that *both* of these mechanisms imply that the alternative occupation p is a relevant outside option for workers in occupation o: the difference is simply that in the bargaining case, workers don't exercise that option, and in the mobility case, workers do exercise that option.

Our results in Table 8 demonstrate that the mobility channel is indeed present. In column (1) we regress the local employment share of initial occupation o in city k in year t on the the naive (non-instrumented) outside-occupation option index, with occupation-by-city and occupation-by-year fixed effects; in column (2) we repeat the specification, but instrument for the outside-occupation option index as above. As one would expect, in years where the wages of outside-option occupations rise in a given city, the employment share of the initial occupation o falls (as workers, presumably, move on net to the improving outside-option occupations).

Our results in Table 9 suggest that even taking into account the mobility channel, the bargaining

---

[46]As with the naive regressions, the results are very similar if we run an employment-weighted IV regression.

[47]Assuming a downward-sloping labor demand curve.

channel still matters. In columns (1) and (3) we repeat the naive and instrumented wage regressions from Tables 6 and 7 respsectively, controlling for occupation-by-city and occupation-by-year fixed effects. In columns (2) and (4), we re-run these regressions while *also* controlling for the employment share of the initial occupation o in city k in year t. The coefficients for the effect of the outside-occupation option index on the wage are still large, positive and significant: even controlling for the decrease in supply of workers to the initial occupation, an increase in the value of outside-occupation options increases the wage of the initial occupation.

Our results therefore suggest that nationwide demand shocks to relevant outside option occupations are associated with positive, significant and meaningful changes in local occupational wages. Since our instrument is plausibly exogenous, our results suggest that on average, workers' relevant outside options and therefore their relevant labor markets extend substantially beyond their own occupation – and that these outside options matter both as an option workers actually exercise (through mobility) and as an option in the wage bargaining process.

5 Labor market concentration, outside-occupation options, and wages

Our analysis so far suggests that jobs outside a worker's occupation form an important part of her labor market, and that the availability of job options outside a worker's own occupation matters for her wages. When trying to estimate the degree of local labor market concentration and monopsony power, therefore, job options outside workers' own occupation should be taken into account. We show in this section that failure to consider workers' options outside their occupation and city can lead to overestimates of the size of the relationship between labor market concentration and wages, and obscures substantial heterogeneity in this relationship.

5.1 Recent work on labor market concentration and monopsony power

In a perfectly competitive model of the labor market, workers move frictionlessly between jobs, while firms are price-takers. Models of imperfect competition relax these assumptions, introducing search frictions or switching costs for workers and firms, worker and firm heterogeneity, and differential firm size (Boal and Ransom, 1997; Ashenfelter et al., 2013; Manning, 2003). Common to all models of imperfect competition in labor markets is the feature that workers are limited in their ability to find better job opportunities elsewhere, giving firms some discretion over the wage. In the framework of a two-sided

matching market with heterogeneous workers and firms and search frictions[48], for example, the worker's outside option (her expected wage if she left her current job) gives a lower bound on the wage, and the firm's outside option (the expected cost of filling the job with an equally productive worker) gives an upper bound. Between these two bounds, the wage is determined by the relative bargaining power of the worker and the firm[49].

The outside options of workers are therefore an important dimension in understanding the relative market power of workers and employers. The outside options determining workers' bargaining power may be jobs in the worker's own occupation and city, or in other occupations and/or other cities, as discussed in section 3. Whether implicitly or explicitly, all analysis on workers' labor market power must take a stance on which jobs are included in the workers' outside option set.

Recent research on labor market concentration and monopsony power has adopted the "market definition approach" common in antitrust policy, which defines the relevant market of substitutable jobs and excludes all other jobs from the analysis. Commonly-used labor market definitions are occupation or industry by geographic area (Commuting Zone, county or metropolitan area). Azar et al. (2017) and Azar et al. (2018) find a large, negative and significant relationship between wages and employer concentration in online vacancy data within an SOC occupation group (6-digit or 4-digit), commuting zone and quarter. Benmelech et al. (2018) similarly find a large, negative and significant relationship between wages and employer concentration using employment HHIs at a 3- or 4-digit SIC code level for county-industry-year cells over three decades, using establishment-level data from the Census of Manufacturing. Using a broader set of industries, Rinz (2018) and Lipsius (2018) find similar results on the relationship between wages and employer concentration calculated as HHIs by industry and geography using Longitudinal Business Database data for the entire US[50]. Considering a broader set of affected outcomes, Hershbein et al. (2019) show that employment HHIs at the industry-CZ and vacancy HHIs at the occupation-CZ level are negatively related to wages, and further show that firms in concentrated labor markets demand higher skills in their job postings.

5.2 Wage and HHI regressions

If the boundaries of an occupation were impermeable, so that workers could rarely switch occupation, then the degree of local labor market concentration within an occupation may indeed be a good measure

---

[48] As in the search-and-matching literature on the labor market (Mortensen and Pissarides, 1999; Rogerson et al., 2005).

[49] In a Nash bargaining setup, for example, this split is determined by the bargaining coefficient.

[50] In particular, Rinz (2018) uses data for 1976-2015 at the commuting zone level, and Lipsius (2018) uses data for 1980-2012 at the MSA level.

of workers' outside options. However, our analysis in this paper suggests that jobs outside workers' occupation are relevant parts of their labor market and do impact their wages - so ignoring workers' ability to move outside their occupation and city may exclude jobs which are important outside options for worker bargaining.

This implies that HHIs measured at the level of a SOC 6-digit occupation by city on average overstate the degree of employer concentration faced by workers. It also has two further testable implications. First, the empirical relationship between wages and employer concentration (HHI) should be stronger for occupations which are better definitions of workers' true labor market. In cities where few workers have the option to get a job outside their own occupation, local employer concentration in that occupation would be expected to have a much greater effect on workers' wages than in cities where workers are easily able to get jobs outside their own occupation. Second, the empirical relationship between wages and HHI may be biased if employer concentration within an occupation is correlated with the availability of outside-occupation job options.

Before testing these implications, we first confirm that we can replicate the result in other studies of a negative correlation between wages and employer concentration. In Table 10, we regress the log average wage on the log vacancy HHI with city and occupation-by-year fixed effects[51]. We follow Azar et al. (2018) and Hershbein et al. (2019) in using vacancies from Burning Glass Technologies' database of online vacancy postings, calculating the vacancy HHI indexes at the level of the SOC 6-digit occupation by city by year. As in the other studies, in our data there is a negative and significant relationship between mean hourly wages and annual vacancy concentration for SOC 6-digit occupations by city over 2013-2016. The elasticity of mean wages to the annual vacancy HHI is -0.019 in a specification with occupation-by-year and city fixed effects, which is smaller than but of the same order of magnitude as the estimates in Azar et al. (2017) and Rinz (2018) (although we note that this is a correlation and cannot necessarily be interpreted as a causal relationship).

To test whether the empirical relationship between wages and HHI is stronger for occupations which are better definitions of workers' true labor market, we segment our data into four quartiles using the national average occupation leave share[52], and re-run the regression of log wage on log HHI at the occupation-city level on each of the the four quartiles separately. As Table 10 shows, the coefficient for the quartile of occupations with the lowest outward mobility (lowest occupation leave share) is more

---

[51]The timespan of our HHI data is too short to analyze changes over time in the HHI within a city-occupation unit.

[52]The occupation leave share measures the following: of the people observed in the BGT resume data in occupation o in year t who are observed in a different job in year $t + 1$, the leave share is the proportion who are no longer observed in their initial occupation o but remain in the data in year $t + 1$. It is defined and discussed in more detail in section 2.

than 50% higher than the average and the coefficient for the quartile of occupations with the highest outward mobility (highest occupation leave share) is 50% lower than the average[53]. These results are consistent with our hypothesis that occupations with very high outward mobility are substantially worse approximations of workers' true labor markets than occupations with low outward mobility, and that the relationship between concentration and wages will be over-estimated (will appear to be too negative) for these high mobility occupations.

To test whether the empirical wage-HHI relationship is biased by ignoring outside-occupation job options, we regress the log wage on the log vacancy HHI while controlling for our outside-occupation options index, estimating a specification of the form:

$$log(\bar{w}_{o,k,t}) = \alpha + \beta_1 log(HHI_{o,k,t}) + \beta_2 log(oo_{o,k,t}^{occs}) + \delta_k + \delta_{o,t} + \epsilon_{o,k,t}, \tag{16}$$

again including city as well as occupation-year fixed effects. We estimate this expression first using OLS with our simple outside-occupation index and then using two-stage least squares and our instrumented outside-occupation index. Since the effect of controlling for outside-occupation options should vary across differentially mobile occupations, we also run both versions of this regression separately by quartile of the occupation leave share.

The results using the simple outside-occupation option index are shown in Table 11 and using the instrumented outside-occupation option index are shown in Table 12. In both cases, when we control for outside options in the full sample (column 2), the coefficient on the vacancy HHI falls statistically significantly by a large amount: from -0.019 to -0.010 when controlling for the simple outside-occupation option index, or to -0.013 when controlling for the instrumented outside-occupation option index. This fall in the coefficient is consistent with omitted variable bias from not accounting for outside options: in our data on US occupation-city labor markets over 2013–2016, the vacancy HHI is strongly negatively correlated with workers' outside-occupation options - workers in occupation-city labor markets with worse options *within* their occupation also have worse options *outside* their occupation - so that the estimated coefficient on the HHI is biased upward in magnitude when outside options are not controlled for. Moreover, the regressions by quartile of the occupation leave share (columns 3-6 of Tables 11 and 12) show that the coefficient falls most strongly for the highest mobility quartile. As is intuitive, the bias in the HHI-only regression from omitting outside-occupation options is larger, the more mobile workers in an occupation are.

---

[53]The average (pooled) coefficient is not statistically significantly different from the coefficients for the 2nd and 3rd quartiles of the occupation leave share, but is strongly statistically significantly different from the coefficients for the 1st and 4th quartiles.

Overall these results suggest that coefficient estimates for the relationship between an HHI measure and local wages are likely to be biased upward in size (estimating a relationship that is too negative) if they do not control for local differences in outside options across occupations. Moreover, this bias is likely to be largest for highly mobile occupations, for which labor markets are least well represented by a single occupation.

5.3 Augmenting the HHI to incorporate outside-occupation job options

Marinescu and Hovenkamp (forthcoming) argue that antitrust analyses should use labor market HHIs and, specifically, that the HHI should be defined at the level of the 6-digit SOC occupation by commuting zone. Our analysis in this paper however suggests that the HHI defined at the level of a 6-digit SOC occupation is unlikely to reflect the true job options available to workers and, importantly, will reflect job options differentially well or poorly for different occupations.

To illustrate this point, compare the example of pharmacists in Charlotte-Concord-Gastonia (North and South Carolina) to the example of amusement and recreation attendants in Cape Coral-Fort Myers (Florida). Both occupation-city labor markets have an HHI of around 0.25, which by product market guidelines would be considered to be highly concentrated (Marinescu and Hovenkamp, forthcoming; Azar et al., 2017). Pharmacists, however, have an occupation leave share of only 9%, whereas amusement and recreation attendants have an occupation leave share of 27%. The within-occupation HHI suggests that both groups face similar labor market concentration, but in reality, the amusement and recreation attendants in in Cape Coral-Fort Myers are likely to face a much less concentrated labor market because they have more job options outside their occupation[54].

We therefore propose an augmented HHI index to take account of workers' differential occupational mobility and local availability of feasible outside job options. Given the analysis in this paper, we believe that any such augmented HHI should satisfy the following two properties:

1. Outward occupational mobility should matter: For two occupations with the same within-occupation HHI, the one with higher outward mobility should have a lower augmented HHI than the one with lower outward mobility.

2. Employer concentration in outside-option occupations should matter: For two occupations with the same within-occupation HHI and the same degree of aggregate outward mobility, the one with

---

[54]Our simple regressions of the HHI on the wage discussed above (and in Table 10) would suggest that the relationship between employer concentration and the wage is more than twice as strong for the pharmacists in Charlotte-Concord-Gastonia, who are in the bottom quartile of occupation leave share, compared the the amusement attendants in Cape Coral-Fort Myers, who are in the top quartile.

more employer concentration in their outside-option occupations should have a higher augmented HHI than the one with less employer concentration in their outside-option occupations.

From the theory, it is not a priori clear exactly what this index should be - the precise form of the index would depend on the purported mechanism through which labor market concentration affects wages in a particular setting.

However, for an illustration of the probabilistic labor market approach, we propose a simple "probabilistic labor market" HHI index (HHI^{PLM}) which is a weighted average of the HHIs in different occupations in a worker's labor market. Assume that a job-searching worker considers considers all jobs in her current occupation, as well as all jobs in one other occupation p, chosen with probability $\pi_{o \to p}$ (our measure of the occupation transition probability conditional on leaving a job). Then, the expected probabilistic labor market HHI of the occupations a worker currently in occupation o will search in is given by

$$HHI_o^{PLM} = \phi_{o,o} HHI_o + \sum_{p \neq o}^{N_{occs}} \phi_{o,p} HHI_p,$$

where $\phi_{o,p} = \frac{\pi_{o \to p}}{1 + \sum_{p \neq o}^{N_{occs}} \pi_{o \to p}}$ and $\phi_{o,o} = \frac{1}{1 + \sum_{p \neq o}^{N_{occs}} \pi_{o \to p}}$ (the weights are scaled such that they add to one to ensure that the resulting HHI_o^{PLM} values are not mechanically larger than the simple HHI_o).

If the probabilistic labor market-based HHI measure improves upon a single-occupation HHI, we would expect the former to have a statistically significant coefficient in a horserace with the latter in wage regressions of the form

$$log(\bar{w}_{o,k,t}) = \alpha + \beta_1 log(HHI_{o,k,t}) + \beta_2 log(HHI_{o,k,t}^{PLM}) + \delta_k + \delta_{o,t} + \epsilon_{o,k,t}. \tag{17}$$

The results of estimating this specification are shown in Table 13.

The first column of the table shows that the probabilistic HHI has a strong reduced form relationship with wages. The negative coefficient is about 50% larger than that for the single-occupation HHI (estimated in Table 10). When both HHI measures are included in a horse race (column 2), only the probabilistic measure is significant in predicting lower wages. Note that if the single-occupation specification were correct, we would expect the HHI^{PLM} measure to consist of the correct measure with added noise, and so most of the estimated relationship should load onto the single-occupation HHI. The fact that the single-occupation HHI coefficient is not statistically significant at all suggests that *if* this regression measures a true negative effect of labor market concentration on wages, then the probabilistic

33

labor market HHI better captures this relationship than the single-occupation HHI.

Columns 3-6 further explore the relevance of the different HHI measures for different quartiles of occupational mobility. Two aspects of the pattern of results support our assertion that the probabilistic HHI measure better takes into account worker mobility. First, the estimated coefficient on HHI^{PLM} is very similar across different mobility quartiles - in contrast to the much larger coefficients in low-mobility quartiles that we found for the single-occupation HHI (Table 10). This is consistent with the interpretation that coefficient sizes are heterogeneous when using the single-occupation HHI because employer concentration in other relevant occupations had been omitted – and since the probabilistic HHI measure better captures employer concentration in other relevant occupations, it allows us to estimate a more stable association of labor market concentration with lower wages.

Second, the coefficients on the single-occupation HHI in the horse race are much smaller in absolute size and are only significant (at a 5% level) for the lowest-mobility quartile of occupations. This pattern aligns well with our justification for using a probabilistic measure: for highly mobile occupations a single-occupation HHI substantially mismeasures the relevant labor market. However, for very immobile occupations, the single-occupation HHI comes close to capturing the correct labor market, which is likely why we still find a significant relationship with wages for the least mobile occupations.

To establish whether the larger estimated coefficients when using the probabilistic HHI are robust to controlling for outside options, Tables 19 and 20 in the Appendix repeat the analyses from Tables 11 and 12, using the probabilistic HHI instead of the single-occupation HHI. The estimated coefficients exhibit the same qualitative pattern as those for the simple HHI, but their absolute size is again 40% larger. We note however that even when using the probabilistic HHI, the measure of outside-occupation options still has a positive and significant effect on wages. This suggests that probabilistic labor market concentration and better outside-occupation job options may reflect different channels through which worker outside options affect wages. The former reflects the firm size distribution of a worker's options but does not contain information on wages, while the latter ignores firm size distribution but incorporates outside-occupation job options' wages. Our analysis shows that changes along both of these dimensions seems to be associated with changes in wages.

Overall, the analysis in this section suggests that labor market definition when measuring labor market concentration has a substantial effect on results, highlighting the importance of choosing a good approximation to workers' actual labor markets.

6 Conclusion

In this paper we argue that the binary approach to labor market definition is inappropriate for much labor market analysis. It inherently excludes jobs which are relevant outside options or includes jobs which are not true outside options for the workers under consideration. Using conventional proxies for labor markets, such as geographies, current industries, or current occupations, fails to take into account worker mobility.

We support this argument by documenting a number of new facts on occupational mobility in the U.S., using a large new data set of U.S. worker resumes. Our data shows that workers are highly mobile across occupations, that there is a very large difference across occupations in the degree to which the SOC 6-digit occupation is an appropriate definition of workers' true labor market, that aggregating to a higher level of SOC code for occupations is not an appropriate way to fix this issue, and that the directed nature of the occupation-to-occupation transition matrix suggests that outside-occupation job options should not be considered symmetric across occupations. Furthermore, the sparse nature of the occupation-to-occupation transition matrix suggests that for many occupations, workers' true labor markets can be constructed out of relatively small clusters of similar occupations (as we do in this paper).

Since the binary approach to labor market definition ignores important outside options, we argue instead that workers' labor markets should be defined probabilistically to approximate the actual realm of jobs that are available to them. We suggest one feasible approach: using empirical occupational mobility patterns to identify job options outside workers' own occupation. We apply this probabilistic definition of outside options to U.S. data, constructing an index of worker outside-occupation job options for over one hundred thousand SOC 6-digit occupations and metropolitan areas in the U.S. over 1999–2016.

Our index shows that workers differ substantially by occupation and location in the size of their local labor market, and that this notion of an expanded labor market that includes other local occupations contributes to differences in wages, in line with the predictions of standard bargaining models. Specifically, using plausibly exogenous Bartik-style shocks to the wages of workers' outside option occupations, we show that workers in cities with more outside-occupation job options see bigger wage gains when those outside-occupation job options are in higher demand.

Finally, we show the importance of appropriate labor market definition with an application of our approach to the recent literature on labor market concentration and wages. The empirical correlation between local labor market concentration (within an occupation) and wages is much stronger for occupations with low outward occupational mobility, which could be thought of as better definitions of

workers' true labor markets, and much weaker for occupations with high outward occupational mobility. This suggests that a measure of local labor market concentration within an occupation, which ignores the availability of job options outside occupations, can be a misleading indicator of the true availability of outside options for workers. Further, when controlling for the availability of outside-occupation job options, we show that the magnitude of the estimated relationship between local labor market concentration and wages is reduced, and this reduction in bias is greater for more mobile occupations. This leads us to propose an alternative, probabilistic HHI measure which can account for the availability of job options outside workers' own occupation.

Overall, our results suggest that labor markets for workers are complicated objects that vary across geographies and depend on links between different occupations - and that we can improve upon simplistic binary definitions by inferring probabilistic connections from actual labor market behavior. 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

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

ASHENFELTER, O., H. FARBER, AND M. RANSOM (2013): "Modern Models of Monopsony in Labor Markets: A Brief Survey," *IZA Discussion Paper*.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2006): "The polarization of the US labor market," *American economic review*, 96, 189–194.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 118, 1279–1333.

AZAR, J., I. MARINESCU, AND M. I. STEINBAUM (2017): "Labor Market Concentration," Tech. rep., National Bureau of Economic Research.

AZAR, J. A., I. MARINESCU, M. I. STEINBAUM, AND B. TASKA (2018): "Concentration in US Labor Markets: Evidence From Online Vacancy Data," Tech. rep., National Bureau of Economic Research.

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75, 1331–1370.

BEAUDRY, P., D. A. GREEN, AND B. SAND (2012): "Does industrial composition matter for wages? A test of search and bargaining theory," *Econometrica*, 80, 1063–1104.

BENMELECH, E., N. BERGMAN, AND H. KIM (2018): "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" Tech. rep., National Bureau of Economic Research.

BIDNER, C. AND B. SAND (2018): "Job Prospects and Pay Gaps: Theory and Evidence on the Gender Gap from U.S. Cities," Working paper.

BLAU, F. D. AND L. M. KAHN (2017): "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 55, 789–865.

BOAL, W. M. AND M. R. RANSOM (1997): "Monopsony in the labor market," *Journal of Economic Literature*, 35, 86–112.

BURDETT, K. AND D. T. MORTENSEN (1980): "Search, Layoffs, and Labor Market Equilibrium," *Journal of Political Economy*, 88, 652–672.

BUREAU OF LABOR STATISTICS, U.S. DEPARTMENT OF LABOR (2018): "BLS Website," https://www.bls.gov/cps/demographics.htm#women.

CALDWELL, S. AND O. DANIELI (2018): "Outside Options in the Labor Market," Working paper.

CALDWELL, S. AND N. HARMON (2018): "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," Working paper.

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

GATHMANN, C. AND U. SCHÖNBERG (2010): "How general is human capital? A task-based approach," *Journal of Labor Economics*, 28, 1–49.

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

HERSHBEIN, B., C. MACALUSO, AND C. YEH (2019): "Concentration in U.S. local labor markets: evidence from vacancy and employment data," *Working Paper*.

JAEGER, S., B. SCHOEFER, S. YOUNG, AND J. ZWEIMUELLER (2018): "Wages and the Value of Nonemployment," *Working Paper*.

KAMBOUROV, G. AND I. MANOVSKII (2009): "Occupational specificity of human capital," *International Economic Review*, 50, 63–115.

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

LIPSIUS, B. (2018): "Labor Market Concentration does not Explain the Falling Labor Share," Tech. rep., working paper.

MACALUSO, C. (2019): "Skill remoteness and post-layoff labor market outcomes," Working paper.

MANNING, A. (2003): *Monopsony in motion: Imperfect competition in labor markets*, Princeton University Press.

MANNING, A. AND B. PETRONGOLO (2017): "How local are labor markets? Evidence from a spatial job search model," *American Economic Review*, 107, 2877–2907.

MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60, 531–542.

MARINESCU, I. AND H. HOVENKAMP (forthcoming): "Anticompetitive Mergers in Labor Markets," *Indiana Law Journal*.

McFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, 105–142.

MOLLOY, R., C. L. SMITH, AND A. WOZNIAK (2011): "Internal migration in the United States," *Journal of Economic Perspectives*, 25, 173–196.

MORTENSEN, D. T. AND C. A. PISSARIDES (1999): "New Developments in Models of Search in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, 2567–2627.

NEFFKE, F., A. OTTO, AND A. WEYH (2017): "Inter-industry labor flows," *Journal of Economic Behavior & Organization*, 142, 275–292.

NIMCZIK, J. S. (2018): "Job Mobility Networks and Endogenous Labor Markets," .

RINZ, K. (2018): "Labor Market Concentration, Earnings Inequality, and Earnings Mobility," *U.S. Census Bureau Carra Working Paper Series*, 2018-10.

ROGERSON, R., R. SHIMER, AND R. WRIGHT (2005): "Search-theoretic models of the labor market: A survey," *Journal of economic literature*, 43, 959–988.

SHAW, K. L. (1987): "Occupational change, employer change, and the transferability of skills," *Southern Economic Journal*, 702–719.

TALAMÀS, E. (2018): "Nash Bargaining with Endogenous Outside Options," Working paper.

XU, M. (2018): "Understanding the Decline in Occupational Mobility," *Working Paper*.

7   Figures

Figure 1: Distribution of the "occupation leave share": the probability that a worker will leave their occupation conditional on leaving their job, calculated from Burning Glass Technology resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample mean.
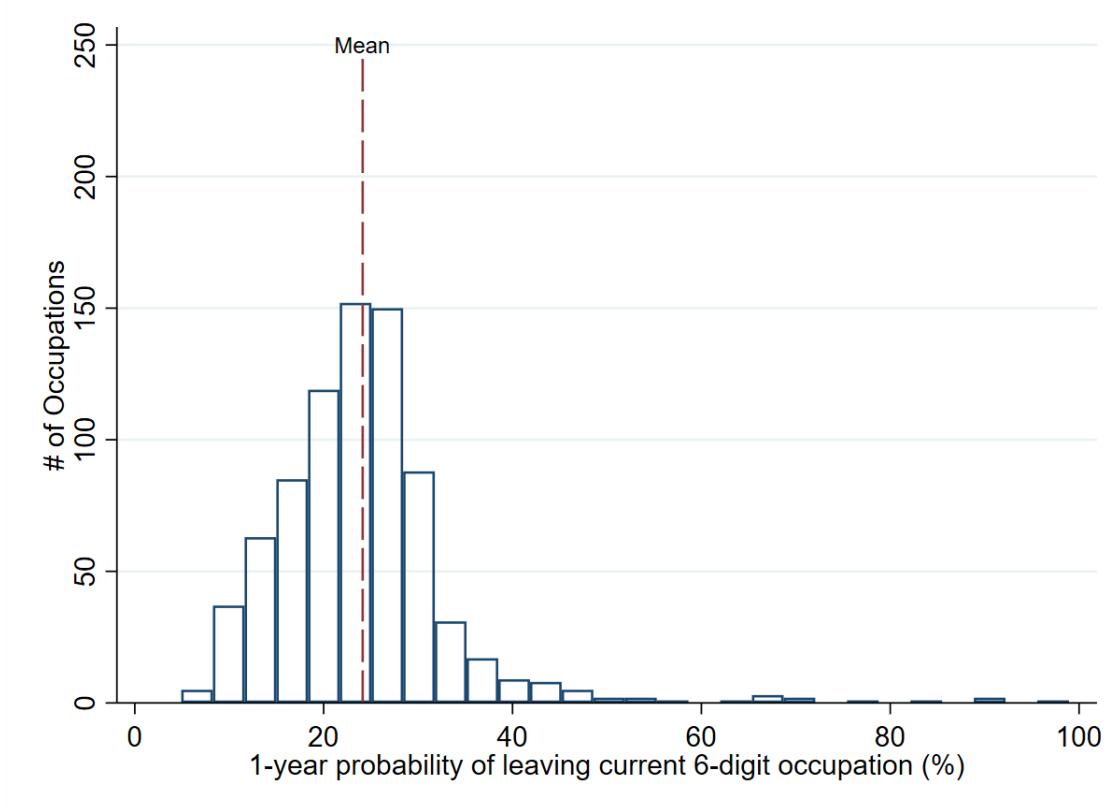
Figure 2: Distribution of the proportion of workers moving 6-digit SOC occupation who *also* move 2-digit SOC occupation, by occupation, calculated from Burning Glass Technology resume data for 2002-2015 period. Histogram shows 786 occupations.
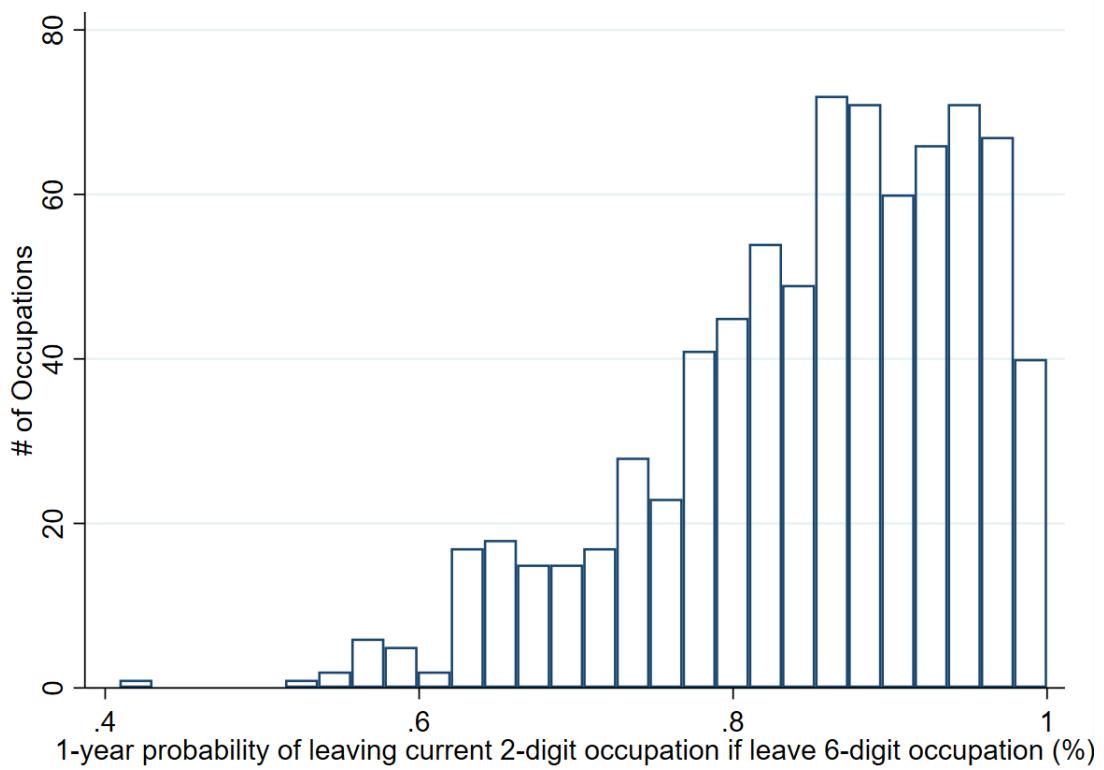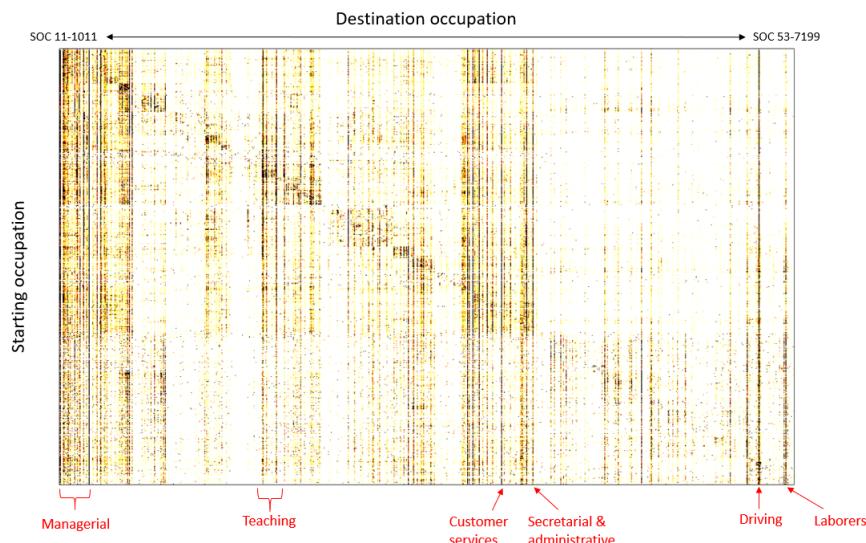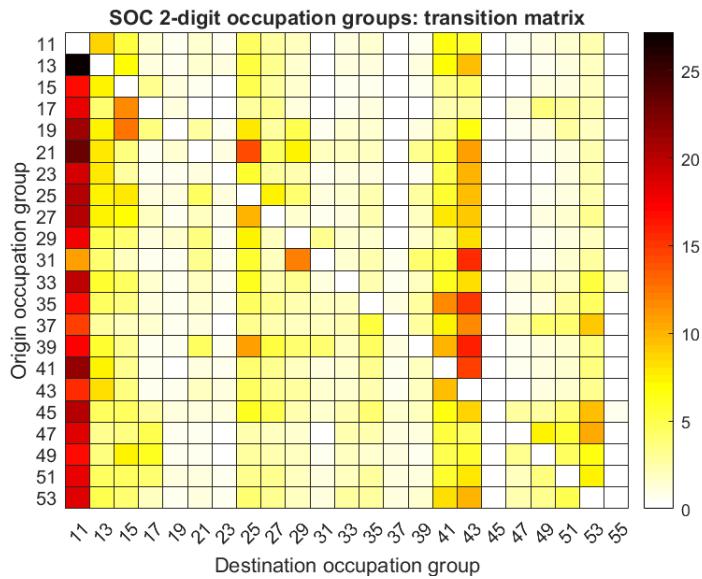
Figure 3: Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial job. Occupations are sorted in SOC numerical order. Cells colored black have a transition probability of 1% or greater conditional on leaving the initial job. Transitions to own occupation are excluded. Data computed from Burning Glass Technology 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 4: 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 Burning Glass Technology resume data set for 2002-2015.

Figure 5: 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 6: 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.

Which occupations do registered nurses go to?
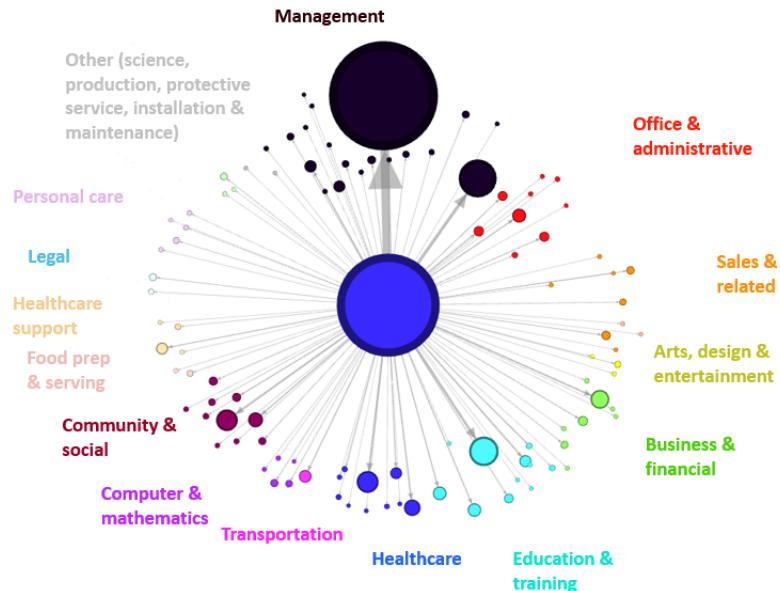
Figure 7: Coefficients and 95% confidence intervals from regression of 2002-2015 average probability of moving into another occupation (conditional on any job move) on absolute difference in occupational characteristics. All regressions also include a constant, absolute avg. hourly wage differences, and origin occupation fixed effects - except for the amenities regressions, where wage differences are omitted. Standard errors are clustered at the origin occupation level.
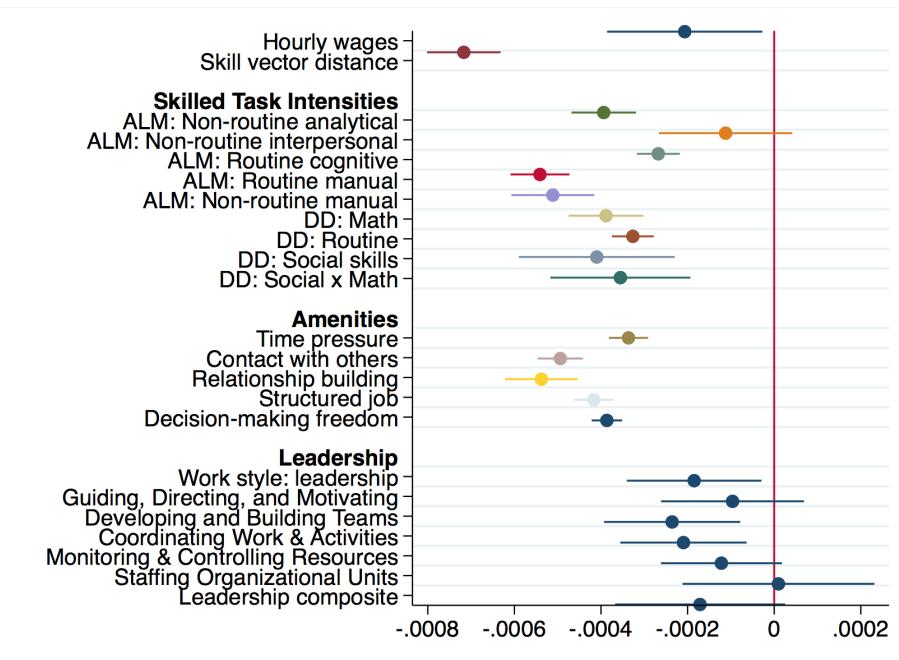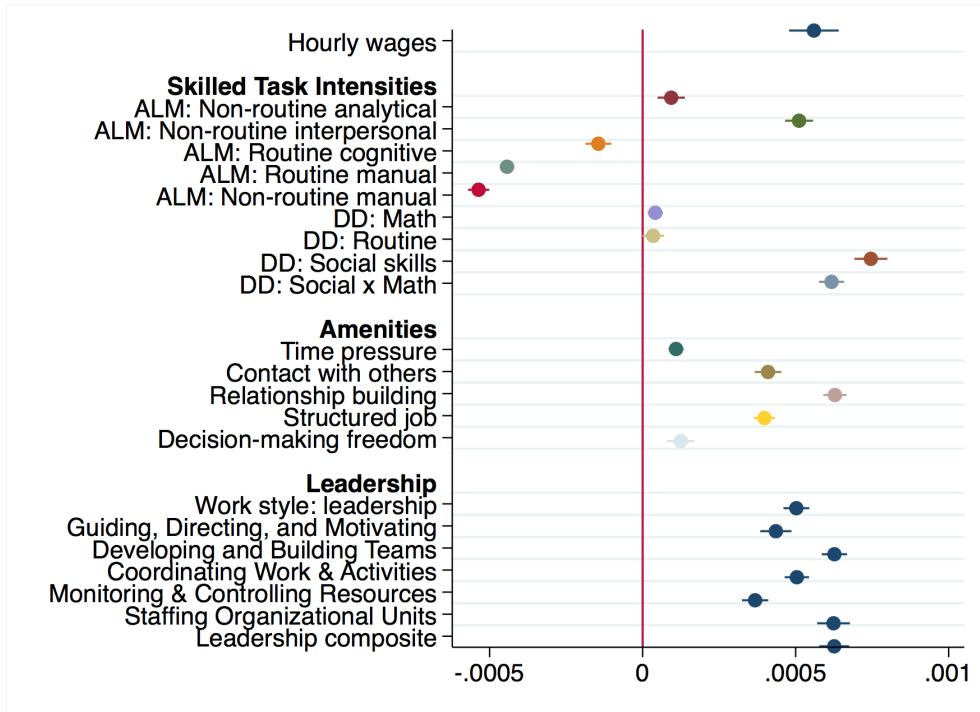
Figure 8: Coefficients and 95% confidence intervals from regression of 2002-2015 average probability of moving between two occupations (conditional on any job move) on *relative* difference (target minus origin) in stated characteristic between target and origin occupation. All regressions also include a constant, relative avg. hourly wage differences, and origin occupation fixed effects - except for the amenities regressions, where wage differences are omitted. Standard errors are clustered at the origin occupation level.

8 Tables

Table 1: Number of observations in the BGT occupational mobility data, by occupation (2002-2015)

Percentile	1	5	10	25	50	75	90	95	99
Obs.	595	1054	1,649	4,940	20,774	112,258	466,814	853,891	3,471,904

This table shows summary statistics of number of observations by occupation in our occupational mobility data set, calculated from the Burning Glass Technology resume data. An observation in our data is a person-year unit, as long as that person is also observed in the data in the following year (so that we can calculate annual occupational mobility). We exclude occupations with fewer than 500 observations in the BGT data.

Table 2: Share leaving job and occupation, by occupation (2002-2015)

	Share in different job	Share leaving occupation (6d)	"Occupation leave share" Share leaving occupation conditional on leaving job
Avg. (emp.weight)	0.46	0.11	0.23
Average (simple)	0.47	0.11	0.24
P1	0.30	0.047	0.09
P5	0.35	0.062	0.11
P10	0.37	0.074	0.14
P25	0.40	0.90	0.19
Median	0.45	0.10	0.24
P75	0.52	0.12	0.28
P90	0.61	0.14	0.33
P95	0.66	0.18	0.38
P99	0.74	0.29	0.69

This table shows summary statistics of the share of workers leaving their job and occupation, by SOC 6-digit occupation. The statistics cover workers observed in the BGT resume data over 2002-2015, for all 6-digit SOCs with at least 500 data points. The employment-weighted average takes the average across SOC 6-digit occupations, weighting them by their total U.S. employment from 2017 OES data; the simple average takes the average across SOC 6-digit occupations.

Table 3: Share of outward occupational moves which cross SOC 2d boundary, by occupation (2002-2015)

Percentile	1	5	10	25	50	75	90	95	99
	0.55	0.65	0.70	0.79	0.87	0.93	0.97	0.98	1.00

This table shows summary statistics of the share of all occupational coincidences in subsequent years that are moves which cross SOC 2-digit boundaries, by origin occupation (for origin 6-digit SOCs with at least 500 data points). This implies that for the median occupation, 87% of all occupational moves are to a different SOC 2-digit occupation.

Table 4: Twenty large occupations with lowest leave shares and highest leave shares

Initial occupation	Leave share	Employment (2017)	Obs. (in BGT data)	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 representatives, wholesale and manufacturing, technical and scientific products	.31	327,190	198,337	Sales representatives, wholesale and manufacturing, except technical and scientific 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
Molding, coremaking, and casting machine setters, operators, and tenders, metal and plastic	.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).

Table 5: 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 representatives, wholesale and manufacturing, technical and scientific products	Sales representatives, wholesale and manufacturing, except technical and scientific 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
Wholesale and retail buyers, except farm products	Purchasing agents, except wholesale, retail, and farm products	.16		31,254
Physical therapists	Medical and health services managers	.16	225,420	44,314
Architectural and engineering managers	Managers, all other	.15	179,990	749,670
Biological scientists	Operations research analysts	.15		9,005
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
Electromechanical equipment assemblers	Aircraft mechanics and service technicians	.14		1,803
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 representatives, wholesale and manufacturing, except technical and scientific 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 representatives, wholesale and manufacturing, technical and scientific 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 representatives, wholesale and manufacturing, except technical and scientific 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 administrative assistants, except legal, medical, and executive	.12	185,870	132,543
Mechanical engineers	Architectural and engineering managers	.12	291,290	408,178

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 are shown.

Table 6: Regression of wage on outside-occupation option index

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
oo^{occs}	0.286***	0.106***	0.111***	0.051***
	(0.009)	(0.003)	(0.005)	(0.005)
Fixed effects	Year	Occ-Year	City-Year	Occ-Year
		City	Occ	Occ-City
Observations	1,944,370	1,944,370	1,944,370	1,944,370

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

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. As noted in the paper, 'cities' refers to CBSAs (metropolitan and micropolitan statistical areas) or NECTAs (New England city and town areas).

Table 7: Two-stage least squares regression of wage on instrumented outside-occupation option index

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
oo^{occs}, instrumented	0.292***	0.099***	0.099***	0.042***
	(0.011)	(0.004)	(0.006)	(0.005)
Fixed effects	Year	Occ-Year	City-Year	Occ-Year
		City	Occ	Occ-City
Observations	1,944,370	1,944,370	1,944,370	1,944,370

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

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. 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.

Table 8: Regression of employment share on outside-occupation options, simple and instrumented

Dependent variable:	Employment share	
	(1)	(2)
oo^{occs}	-0.056***	
	(0.021)	
oo^{occs} , instrumented		-0.203***
		(0.022)
Fixed effects	Occ-Year	Occ-Year
	Occ-City	Occ-City
Observations	1,944,370	1,944,370

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

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. 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.

Table 9: Regression of wages on outside-occupation options, controlling for employment share

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
oo^{occs}	0.051***	0.050***		
	(0.005)	(0.004)		
oo^{occs} , instrumented			0.042***	0.039***
			(0.005)	(0.005)
Employment share		-0.015***		-0.015***
		(0.001)		(0.001)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	Occ-City	Occ-City	Occ-City	Occ-City
Observations	1,944,370	1,944,370	1,944,370	1,944,370

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

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. The instrumented outside-occupation option index uses the national leave-one-out mean wage in outside option occupations to instrument for the local (CBSA-level) wage, and the initial local employment share in outside option occupations to instrument for the current local employment share.

Table 10: Regression of wage on single-occupation Vacancy HHI, by quartile of occupation leave share

Dependent variable:	Log wage				
	Full Sample	By quartile of occupation leave share			
	Baseline	Q1	Q2	Q3	Q4
Vacancy HHI	-0.019***	-0.026***	-0.018***	-0.017***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City
Observations	420,292	112,451	107,101	106,937	93,765

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

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. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2015).

Table 11: Regression of wage on single-occupation vacancy HHI and simple outside occupation options

Dependent variable:	Log wage					
	Full sample		By quartile of leave share			
	Baseline	Incl. oo^{occs}	Q1	Q2	Q3	Q4
Vacancy HHI	-0.019***	-0.010***	-0.017***	-0.009***	-0.006***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
oo^{occs}		0.108***	0.112***	0.091***	0.109***	0.118***
		(0.004)	(0.006)	(0.006)	(0.005)	(0.006)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City	City
Observations	420,254	420,254	112,451	107,101	106,937	93,765

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

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. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2015).

Table 12: Regression of wage on single-occupation vacancy HHI and instrumented outside occupation options

Dependent variable:	Log wage					
	Full sample		By quartile of leave share			
	Baseline	Incl. oo^{occs}	Q1	Q2	Q3	Q4
Vacancy HHI	-0.019***	-0.013***	-0.019***	-0.011***	-0.009***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
oo^{occs}, instrumented		0.089***	0.097***	0.079***	0.090***	0.092***
		(0.005)	(0.007)	(0.006)	(0.006)	(0.007)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City	City
Observations	420,254	420,254	112,451	107,101	106,937	93,765

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

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. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2015). 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.

Table 13: Regression of wage on single-occupation and probabilistic HHI, by quartile of occupation leave share

Dependent variable:	Log wage					
			By quartile of leave share			
	Baseline		Q1	Q2	Q3	Q4
Probabilistic HHI	-0.028***	-0.025***	-0.027***	-0.031***	-0.031***	-0.026***
	(0.001)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)
Single-occ. HHI		-0.003	-0.009***	0.002	0.004	0.006*
		(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City	City
Observations	415,252	415,252	111,134	105,734	105,647	92,737

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

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. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2015). Probabilistic HHI is calculated as a weighted average over all relevant occupations in the labor market of an occupation's workers - details are noted in the text.

9 Data Appendix: Burning Glass Technologies Resume Data

This Data Appendix contains further information about our resume data set from Burning Glass Technologies ("BGT"). This is a new proprietary data set of 23 million unique resumes, covering over a hundred million jobs over 2002–2018.

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.

9.1 Data cleaning and transition data construction

We apply a number of different filters to the Burning Glass resume data before calculating our occupational mobility matrices: First, we retain only resumes that are from the U.S. Next, we keep only jobs on these resumes that last for longer than 6 months to ensure that we are only capturing actual jobs rather than short-term internships, workshops etc. We also apply a number of filters to minimize the potential for mis-parsed jobs, by eliminating all jobs that started before 1901 or lasted longer than 70 years. Moreover, we impute the ages of workers based on their first job start date and education and limit our sample to resumes submitted by workers between the ages of 16 and 100. As we are interested in occupational transitions during the last two decades, we then restrict the data set to jobs held after 2001. The final number of resumes that contain at least two years of job data under these restrictions is 15.8 million. The main job information retained for each resume are the occupation and duration of each job held.

For each of these resumes, we start by extracting separate observations for each occupation that the worker was observed in, in each year. These observations are then matched to all other occupation-year observations on the same resume. We retain all matches that are in sequential years - either in the same occupation or in different occupations. For instance, if a worker was a Purchasing Manager in the period 2003-2005, and a Compliance Officer in 2005-2007, we would record 1-year horizon sequential occupation patterns of the form shown in Table 14.

Table 14: Illustrative example of sequential job holding data.

Year:	2003	2004	2005	2006
Current Occ.		*1-Year Horizon Occ.*		
Purchasing Mgr. (11-3061)	11-3061	11-3061 13-1040		
Compliance Off. (13-1040)			13-1040	13-1040

In our data we have 80.2 million job-year observations. This results in 178.5 million observations of year-to-year occupation coincidences (including year-to-year pairs where workers are observed in the same occupation in both years). Below, we describe the characteristics of this data and how it compares to other data sets - with all statistics referring to this final set of filtered sequence observations, or the 15.8 million resumes, unless otherwise noted.

We use these occupation 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 year-to-year occupation 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, eliminating any potential bias from the skewed age distribution of our sample.

9.2   Summary statistics

Below, we describe the characteristics of this data and how it compares to other data sets. All statistics referring to the final set of 15.8 million filtered resumes, or 178.5 million observations of year-to-year occupation coincidences ('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 the resumes list 10 or fewer jobs (note that a change of job under our definition could include a change of 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 15 shows more information on the distribution of job incidences and job durations on our resumes.

Table 15: Distribution of number of jobs on resume and duration of jobs in BGT data set.

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

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 come from males and 50% are 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 (Bureau of Labor Statistics, U.S. Department of Labor, 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 9). 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 - however, 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 10 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 observations[55], the 10th percentile is 1,226 observations, 25th percentile is 4,173, the median

---

[55] As defined above, for our purposes, an observation is a person-occupation-year observation for which we also observe

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. [56]

Figure 11 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 12, 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. 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 the 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.

---

another occupation in the following year: i.e. the start of a year-to-year occupation coincidence sequence.

[56]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.

9.3 Advantages over other datasets

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

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 23 million resumes covering over 100 million jobs, 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. 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 [57]. 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. 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.

9.4 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

---

[57]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.

who apply for the types of jobs which are listed through online applications.
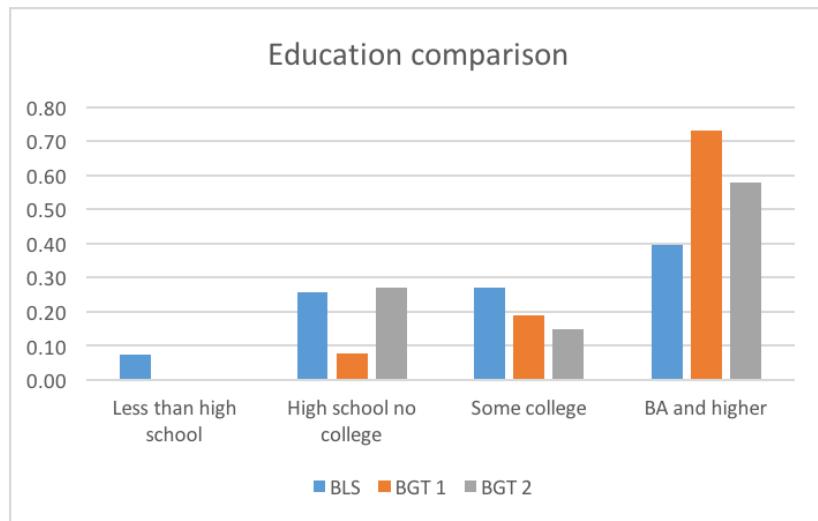
2/ Individuals choose what to put on their resume: We only observe whatever individuals have chosen to put on their resume. To the extent that people try to present the best possible picture of their education and employment history, and even sometimes lie, we may not observe certain jobs or education 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.

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.

9.5 Data Appendix Figures

Figure 9: Comparison of distribution of highest educational attainment in the labor force, according to BLS data, to distribution in BGT 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 10: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT job sequence data.
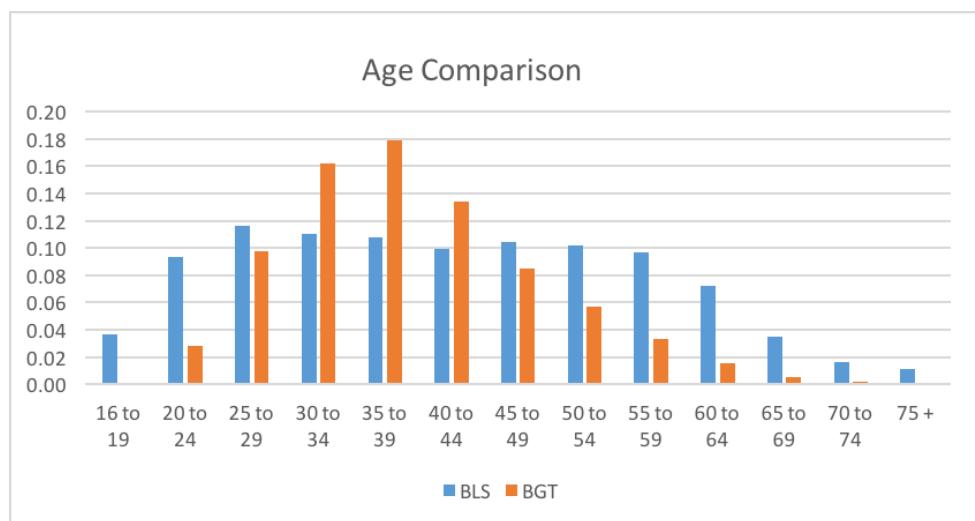
Figure 11: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT job sequence data.
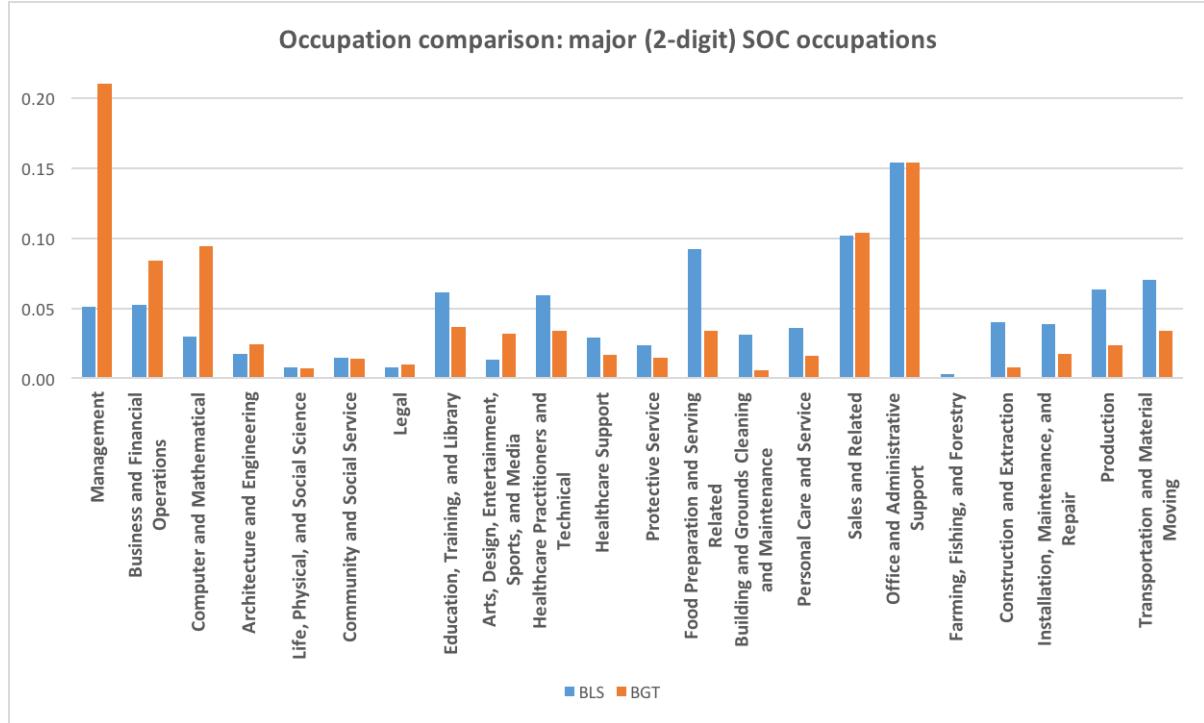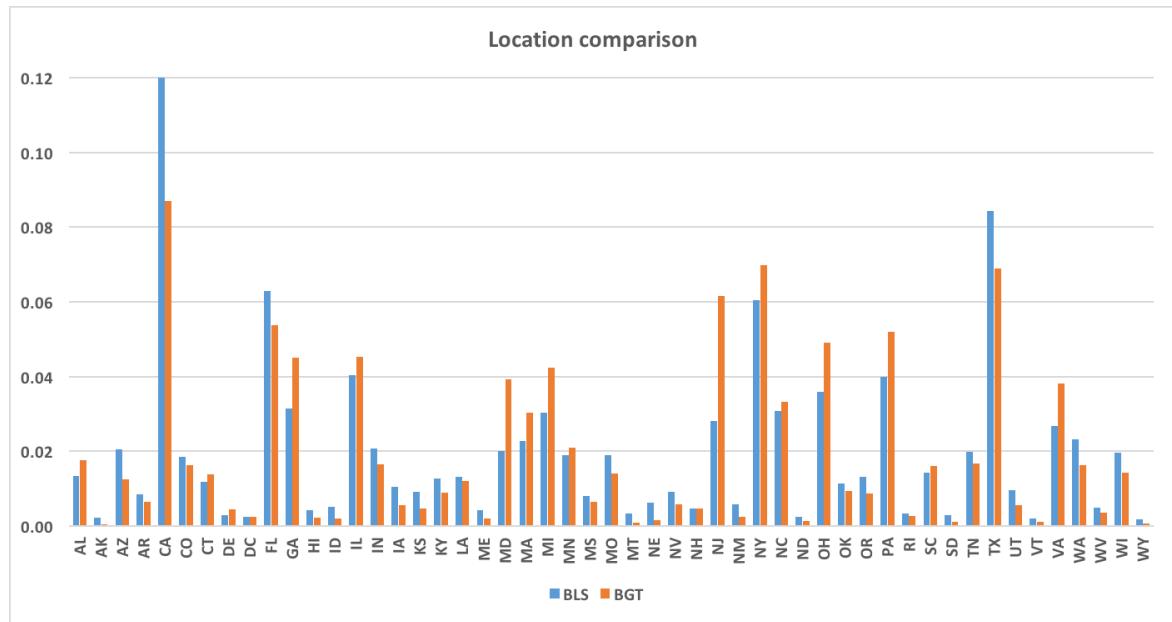

Figure 12: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT job sequence data. Graph shows share of total in each state.

10   Theory Appendix: Simple matching model

As discussed in section 3, the transition-weighted average wage can also be justified as a measure of outside options using a simple matching model. In this model, there are no search frictions. Instead, heterogeneous workers all work at the firm at which they are most productive. Workers also have job offers from other firms, so they know the value of their next-best outside option. This next-best outside option outside their current firm could be a job in the worker's own occupation and city, or in a different occupation or city. As in the search model, the worker and firm Nash bargain, so that the worker's wage is a weighted average of her marginal product in the job and her outside option:

$$w_i = \beta(MPL_i - oo_i) + oo_i = \beta MPL_i + (1 - \beta)oo_i \tag{18}$$

Each worker's wage is different, since each worker has a different best outside job option, and since she and the firm both know the value of that best outside option. To construct the average wage in a given occupation and city, we can segment the workers within that occupation and city into five groups: those whose best outside job option is in their own occupation and city, those whose best outside job option is their own occupation but outside their city, in their own city but outside their occupation, outside their city and occupation, or unemployment. Within each of these labor markets, assume that workers are offered a wage equal to the average wage in that labor market. This gives us an expression for the average value of the outside option in occupation o and city k:

$$\bar{oo}_{o,k} = \varsigma_{o,k}\bar{w}_{o,k} + \sum_{p \neq o}^{N_{occs}} \varsigma_{p,k}\bar{w}_{p,k} + \sum_{l \neq k}^{N_{cities}} \varsigma_{o,l}\bar{w}_{o,l} + \sum_{p \neq o}^{N_{occs}} \sum_{l \neq k}^{N_{cities}} \varsigma_{p,l}\bar{w}_{p,l} + \left(1 - \sum_{p}^{N_{occs}} \sum_{l}^{N_{cities}} \varsigma_{p,l}\right) b$$

where $\varsigma_{p,l}$ is the share of workers in occupation o & city k with best outside option in occupation p & city l.

We can assume that workers' actual occupational moves reflect moves to their best outside job option - either because they were involuntarily displaced and had to find their next-best job, or because they left their job by choice after a preference shock. Then, if we assume that the distribution of best outside options for workers who *remain* in their jobs in occupation o and city k is equal to the distribution of best outside offers for workers who used to be in jobs in occupation o and city k, we can use occupational and geographic transitions to approximate for $\varsigma_{p,l}$ in the outside options expression. Then, the average value of the best outside-labor-market option for workers in occupation o and city k is the weighted average of wages in other occupations and cities, weighted by the proportion of workers who moved from occupation

o and city k to each of the other relevant labor markets.

11 Appendix: Additional Tables

Table 16: Regression of wage on outside-occupation option index, employment-weighted

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
oo^{occs}	0.489***	0.135***	0.177***	0.034***
	(0.055)	(0.009)	(0.008)	(0.010)
Fixed effects	Year	Occ-Year	City-Year	Occ-Year
		City	Occ	Occ-City
Observations	1,944,370	1,944,370	1,944,370	1,944,370

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

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. As noted in the paper, 'cities' refers to CBSAs (metropolitan and micropolitan statistical areas) or NECTAs (New England city and town areas). Observations are weighted by the average employment of their occupation-city over the sample period.

Table 17: Two-stage least squares regression of wage on instrumented outside-occupation option index, employment-weighted

Dependent variable:	Log wage			
	(1)	(2)	(3)	(4)
oo^{occs}, instrumented	0.544***	0.143***	0.166***	0.043***
	(0.057)	(0.012)	(0.014)	(0.009)
Fixed effects	Year	Occ-Year	City-Year	Occ-Year
		City	Occ	Occ-City
Observations	1,944,370	1,944,370	1,944,370	1,944,370

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

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. As noted in the paper, 'cities' refers to CBSAs (metropolitan and micropolitan statistical areas) or NECTAs (New England city and town areas). The instrumented outside-occupation option index uses the national leave-one-out mean wage in outside option occupations to instrument for the local (CBSA-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.

Table 18: Adj. R-squared from regressions of occupational relevance on characteristics

Dependent variable:	$\pi_{o \rightarrow p}$	
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

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

Here, $\pi_{o \rightarrow p}$ 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.

Table 19: Regression of wage on probabilistic vacancy HHI and simple outside occupation options

Dependent variable:	Log wage					
	Full sample		By quartile of leave share			
	Baseline	Incl. oo^{occs}	Q1	Q2	Q3	Q4
Probabilistic HHI	-0.028***	-0.014***	-0.023***	-0.015***	-0.009***	-0.006***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
oo^{occs}		0.106***	0.109***	0.089***	0.108***	0.117***
		(0.004)	(0.006)	(0.006)	(0.005)	(0.006)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City	City
Observations	415,489	415,489	111,190	105,779	105,685	92,835

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

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 2013–2016 inclusive. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2018). Probabilistic HHI is calculated as a weighted average over all relevant occupations in the labor market of an occupation's workers - details are noted in the text.

Table 20: Regression of wage on probabilistic vacancy HHI and instrumented outside occupation options

Dependent variable:	Log wage					
	Full sample		By quartile of leave share			
	Baseline	Incl. oo^{occs}	Q1	Q2	Q3	Q4
Probabilistic HHI	-0.028***	-0.018***	-0.028***	-0.018***	-0.014***	-0.010***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
oo^{occs}, instrumented		0.086***	0.092***	0.076***	0.087***	0.090***
		(0.005)	(0.007)	(0.006)	(0.006)	(0.007)
Fixed effects	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year	Occ-Year
	City	City	City	City	City	City
Observations	415,489	415,489	111,190	105,779	105,685	92,835

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

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 2013–2016 inclusive. Occupations are split into quartiles by the average occupation leave share in the Burning Glass Technologies resume data (averaged over 2002–2018). Probabilistic HHI is calculated as a weighted average over all relevant occupations in the labor market of an occupation's workers - details are noted in the text. 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.