Getting labor markets right: outside options and occupational mobility *

Gregor Schubert, Anna Stansbury and Bledi Taska†

PRELIMINARY DRAFT. COMMENTS WELCOME.

Abstract

Many analyses of important questions in labor economics use occupations as proxies for workers' labor markets, yet high occupational mobility suggests 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. In an IV design, we show that plausibly exogenous wage shocks to local outside-option occupations have a large, positive, and significant relationship with wages. We then show that failing to consider outside-occupation options has important implications for labor market research, with two applications. First, we 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. Second, we analyze the role of occupational linkages in propagating the effect of the China shock on local labor markets. Throughout, we demonstrate substantial heterogeneity in the effect of outside-occupation options by occupational task intensity. Overall, our work suggests that outside-occupation options are important for labor market outcomes, and provides a tractable framework to incorporate them easily into labor market analyses.

*The authors thank Justin Bloesch, Gabriel Chodorow-Reich, Oren Danieli, David Deming, Karen Dynan, Martin Feldstein, Ed Glaeser, Claudia Goldin, Emma Harrington, Simon Jaeger, Max Kasy, Larry Katz, Bill Kerr, Robin Lee, Jeff Miron, Nancy Rose, Maya Sen, Isaac Sorkin, Betsey Stevenson, Larry Summers and participants of the briq Workshop on Firms, Jobs and Inequality and the Harvard Labor lunch, Labor breakfast, Industrial Organization lunch and Multidisciplinary Seminar on Inequality and Social Policy for comments and suggestions. Anna Stansbury gratefully acknowledges financial support from the James M. and Kathleen D. Stone PhD Scholarship in Inequality and Wealth Concentration.

†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. Some frequently used labor market definitions include occupations, industries, workers' educational qualifications, and/or workers' geographic location (state, metropolitan area, commuting zone or county). This approach is inherently binary: all jobs within the chosen market definition are considered perfectly substitutable for each other, and all other jobs are excluded.

In this paper we propose an alternative method for defining labor markets: a probabilistic occupational mobility network derived from empirically observed worker transitions. This methodology enables us to analyze differences in worker outside options across occupations and geographies and yields new insights into the determination of wages, the effects of labor market concentration, and the the impact of labor demand shocks.

If the boundaries of an occupation, industry or geographic area were impermeable, so that workers could never switch, then a binary approach to labor market definition would be correct. Even if not, if it is sufficiently rare or difficult for workers to switch occupation or metropolitan area, then the binary approach may still be a good-enough approximation.

However, empirically, workers are extremely mobile across occupations, industries and geographies - and their mobility is highly heterogeneous. Kambourov and Manovskii (2009) find that annual occupational mobility is 13%-18% and annual industry mobility is 10%-12%[1] Similarly, in our new resume data set from Burning Glass Technologies (described later in more detail), we find that 11% of workers who are in any one SOC 6-digit occupation in one year are no longer in that occupation in the following year (Figure 1 shows the distribution across 786 SOC 6-digit occupations), that the probability of changing SOC 6-digit occupation given that a worker changes her job is around 20%, and that most moves between SOC 6-digit occupations are also moves to different SOC 2-digit occupational groups[2].

[1] They use the PSID, and define occupational mobility to have occurred when the occupation reported is different from the most recently reported previous occupation. Since the PSID is annual, this is effectively annual occupational mobility, but it also incorporates people who report being non-employed in some years. They find 13% occupational mobility at the one-digit level using Census occupation codes, and 18% at the three-digit level. They find 10% industry mobility at the one-digit level using Census industry codes, and 12% at the three-digit level. For Austria, Nimczik (2018) shows that about three quarters of job movers leave their 2-digit industry annually. In addition, Molloy et al. (2011) find that 13% of US workers move to a different commuting zone within 5 years. Occupational and geographic mobility also varies across place and across time: geographic mobility for US workers has declined substantially over recent decades (Molloy et al., 2011) whereas occupational mobility rose until the 2000s and then started to fall gently (Kambourov and Manovskii, 2009; Xu, 2018).

[2] 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.

Ignoring workers' ability to switch occupation or move location when defining labor markets will therefore likely 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. Inevitably therefore, the binary approach to labor market definition will either exclude jobs which are somewhat substitutable, or include jobs which are not good substitutes for each other[3]. In either case, the choice of labor market definition is likely to affect results on a range of important labor market questions.

We argue that this problem can be avoided by moving beyond the binary approach to labor market definition, where all jobs of a particular type are considered to be either in the relevant labor market or not. Instead, researchers can use a probabilistic approach to labor market definition, where workers' job options outside their own occupation are identified using measures of occupational mobility.

In this paper, our empirical focus is on defining occupational labor markets and analyzing the implications of an empirically grounded labor market definition for labor market research. To identify workers' options outside their occupation, we use observed worker transitions between pairs of occupations. Using observed 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. These transitions implicitly reflect a number of factors that may not be visible in the most common alternative approach to measuring occupational similarity - using skill or task data - such as differences in amenities between occupations, or explicit labor market barriers like occupational licensing requirements.

We obtain occupational transition data from a new and unique dataset of resumes, collected by Burning Glass Technologies. 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 probabilities 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, providing new high-dimensional estimates of 'revealed' pairwise occupational similarity. We demonstrate that these transitions

[3] As noted by Kaplow (2015) in the case of antitrust in product markets.

capture underlying similarity in occupational characteristics, showing that occupational mobility is greater between occupations that are more similar in their types of tasks, skills and amenities, and is partially motivated by moves up the career ladder.[4]

Using occupational transitions may help identify relevant outside options, but does so at an extremely high-dimensional level. For feasible use in labor market analysis, a more tractable measure is valuable. We use our occupational transition probabilities to create an index of the average value of workers' outside-occupation job options within their local area, calculated as the weighted average of local wages in all occupations except the worker's own. The weights used are the occupational transitions, which proxy for the likelihood that the average worker's best job option outside her own occupation is in each of these other occupations. While not the focus of our paper, it is also possible to calculate these indexes in the same manner for options outside a worker's local area.

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 equal to 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 worker transitions.

This framework gives structure to the way in which options outside the narrow occupational labor market can be expected to affect workers' 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 labor market, and/or the higher the relative wage in that labor market, the more valuable outside options are in that labor market 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-CBSA-year level over 1999-2016, we find that our indexes of outside-occupation options are significantly and positively related to wages. This relationship is both economically and statistically significant, and exists both cross-sectionally within

[4] Nimczik (2018) also finds that most job moves in Austria involve moving up the career ladder.

occupations and within cities, and over time within the same occupation and city. There is a concern, however, that the relationship could be driven by common shocks to similar occupations in the same city. We therefore generate quasi-exogenous shocks to workers' outside options, instrumenting for local demand shocks to workers' outside-option occupations using the national leave-one-out mean wage in those occupations (using 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 j on the local average wage of occupation i. The positive and significant results persist, with a 1 standard deviation increase in the value of value of outside-occupation options associated with 0.9%-1.4% higher wages. The effect of outside-occupation options on wages appears to be stronger for workers in higher-wage occupations, and for workers in occupations more intensive in cognitive tasks and less intensive in manual tasks.

These results show that workers' wages respond to the value of their outside options *outside* their own occupation. We are able to demonstrate this fact in data that comprises almost the entire set of U.S. occupations and metropolitan areas over 17 years. Our results suggest that the commonly-used labor market definitions of occupation-by-CZ or occupation-by-CBSA are too narrow to reflect workers' true labor markets, and that revealed occupational mobility patterns can be used to infer workers' relevant job options outside their occupations.

In addition to showing the fundamental importance of defining labor markets correctly and suggesting a method to do so, we also apply our framework to two empirical applications from the recent literature on local labor markets. In the context of these applications, we show that failure to consider outside-occupation job options matters quantitatively and qualitatively for the results of analyses of labor markets.

First, we revisit recent analyses of local labor market concentration and wages. We show that the coefficients on HHI in wage regressions are biased upward when workers' options outside their occupation are not considered, and that the coefficient on HHI in wage regressions is substantially higher for occupations with little outward mobility, consistent with the hypothesis that a lack of job options outside the occupation compounds the effects of labor market concentration within the occupation.

Second, we explore the effect of outside options in mitigating or exacerbating the impact of labor demand shocks on workers. We build on the extensive literature documenting employment effects of Chinese import shocks by contributing a number of new insights: On the one hand, we use the fact that we have data on local occupational-level wages to analyze how the effect of the "China Shock" differs across occupations *within* a geographic area. On the other hand, we show that the differential impact on occupations is in part driven by the difference in the quality of local outside options between occupations. In particular, we document that negative shocks to workers' outside-option occupations have an indirect effect on worker outcomes that arises in addition to any direct impact of the "China Shock" on the worker's own occupation. Moreover, we show that the degree to which these indirect effects of labor demand shocks matter for workers depends on the task composition of their occupation.

While our application of a transition-based probabilistic labor market definition focuses on occupations and metropolitan areas (CBSAs), we note that our method is agnostic to the exact definition of the labor market used and could be applied to other settings or geographies.

The results in our paper build on, and relate to, a substantial literature on labor market definition, occupational similarity, and worker outside options. A small body of work uses worker mobility to estimate the geographic extent of local labor markets. Manning and Petrongolo (2017) use unemployment and vacancy flows among UK census wards to infer that workers search across spatially proximate areas, which leads to interdependent effects in response to local shocks. Nimczik (2018) shows that the job mobility network among Austrian firms reveals connected clusters of firms that are time-varying, don't align well with traditional geographic units, and 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.

Our work also relates to the literature estimating the similarity between pairs of occupations, using skill requirements (Macaluso, 2017), task similarity (Gathmann and Schönberg, 2010), worker demographic similarity (Caldwell and Danieli, 2018), or mobility 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' pairwise occupational similarity in the form of sequential job incidence in resumes, and demonstrates that occupational mobility reflects many dimensions of occupational task, skill and amenity similarity.

Another literature that is closely related to our contribution consists of papers exploring the effect of outside options in a bargaining setting on wage outcomes. 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.[5]

A role for outside options in wage determination emerges from imperfect competition models of the labor market, where a degree of firm monopsony power arises either from search frictions or from firm size (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 in particular microeconomic settings. 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. The first empirical paper that we know of to try to construct an outside options index at the aggregate level is by Caldwell and Danieli (2018). In their analysis, they find that the degree to which workers face a more diverse set of outside options in their local labor market in Germany is associated with higher wages. Our index differs methodologically from that in Caldwell and Danieli (2018), as the dynamic nature of our occupational mobility data allows us to incorporate the directed nature and asymmetry of job moves. 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.

In showing the effect of shifts in occupational wages on the wages in other occupations, our paper is also similar in spirit, if not in methodology, to Beaudry et al. (2012), who show that local changes in the availability of high-wage jobs in some industries have spillover effects on wages in all other industries, as would be expected if those jobs represented relevant worker outside options in a Nash bargaining setting.[6] Our paper differs from Beaudry et al. (2012) in estimating the scope of workers' labor markets based on empirical estimates of oc-

[5]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.

[6]Their empirical estimation uses an IV approach based on national industry wages and employment dynamics to identify fundamental shifts in these variables that are unrelated to local unobservable trends - this allows them to estimate the general equilibrium effects and spillovers of exogenous wage changes. We use a similar IV strategy in our wage regressions.

cupational mobility, rather than assuming that *all* industries in a city matter equally to all workers.

Using occupational mobility to identify better- and worse-defined labor markets, and using our outside-occupation option index, we re-analyze recent work on the link between 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 and Macaluso, 2018). 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.

Our results on the way in which outside options mediate the effect of labor demand shocks builds on various papers exploring the effect of the "China Shock" of exposure to import competition from China during the 2000s. For example, Autor et al. (2013) find that commuting zones that are more exposed to import-competing manufacturing experience lower wages and higher unemployment during the 1990-2007 period. Similar effects of labor demand shocks in employment and wages have been found in a number of different settings (see, e.g. Hummels et al. (2014); Garin and Silvério (2018); Yagan (forthcoming)).[7] Acemoglu et al. (2016) find that the impact of the import shock extended to an "employment sag" in other sectors of the U.S. economy through input-output linkages across sectors, but that the effect was concentrated in exposed tradables sectors.

Our paper contributes novel insights to this analysis of labor demand shocks propagating through economic links by being the first - to our knowledge - in formally defining the network of occupational links and analyzing the propagation of shocks through this network at the occupational level and within U.S. labor markets. Moreover, our approach of defining a local measure of the quality of outside options at an occupational level allows us to connect the aforementioned literatures: If outside options affect a worker's ability to exploit labor market opportunities, then the impact of labor demand shocks should depend on the quality of outside options to which affected workers have access in their area.

Overall, our theory and empirical results suggest that a broader concept of local labor

[7]Specifically, researchers have found that local employment shocks in the US during the Great Recession persisted over long periods (Yagan, forthcoming); that labor demand shocks due to offshoring affect low-skilled workers in Denmark negatively, and high-skilled ones positively, while greater exports raise wages for both group (Hummels et al., 2014); and that negative export demand shocks to Portuguese firms reduce employment and wages (Garin and Silvério, 2018).

markets and worker outside options - taking into account occupational and geographic mobility - is important to labor market analysis. 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' options outside their narrowly-defined labor markets, as revealed through occupational and geographic mobility, into labor market analysis. This allows economists and policymakers to easily adjust existing administrative definitions of labor markets to capture the effects of differences in local labor markets.

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 simple search framework that motivates our empirical measure of worker outside options. Section 4 estimates the effect of outside options on wages. Section 5 shows that accounting for differences in outside options changes our interpretation of the effect of labor market concentration on wages. In section 6, we study the indirect spillovers of labor demand shocks on occupational wages and employment arising from shocks other occupations in the same labor market. The last section concludes.

2 Using occupational mobility to identify workers' outside-occupation job 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 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. On one dimension is feasibility: the likelihood that the worker could easily become a typically productive worker in the new occupation (the new occupation's distance from the worker's current skillset). On the other dimension is desirability: the degree to which that worker would like to do a job in the new occupation relative 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

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 previous authors create measures of occupational similarity in this way. Macaluso (2017) 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 the task vectors for each occupation. 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 will not 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, and the kind of people that work in one occupation may well not be likely to want to work in the other. In addition, skill- or task-based similarity measures require substantial information on the skill and task content of different occupations (much of which is now available from O*NET) as well as substantial assumptions as to how these data can 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 captures occupational similarity in terms of skills and tasks required, based on 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. 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 than the old occupation, incorporating the value of amenities such as work-life balance, status, or career concerns. Unlike the other two approaches, a transition-based approach also does not require the imposition of symmetry on occupational feasibility and desirability: 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.

We adopt the 'revealed' approach to identify outside-occupation options, based on occupational transitions, in this paper. 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, since this explicitly captures workers' decisions between jobs in their own occupation and in other occupations (as defined formally in equation 1). The higher is the proportion of workers of occupation o who transition 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.

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

[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)

2.2 Data: resume data from Burning Glass Technologies

Our data on job moves and transitions between SOC 6-digit occupations is from a new proprietary data set of 23 million unique resumes covering 100 million jobs over 2002–2018, provided by 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 construct the share of workers moving from occupation o to occupation p as 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 $t+1$. Similarly, we construct the share of workers leaving their job in occupation o as the share of all workers observed in a given job in occupation o at any point in year t who are not observed in that same job at any point in year $t+1$. This leads to the empirical construction of our occupational transition probability as defined in equation 2.

$$\begin{aligned}\pi_{o\to p} &= \frac{\text{Share moving from occ } o \text{ to occ } p}{\text{Share leaving job}} \\ &= \frac{\frac{\text{number working in occ } o \text{ in year } t \text{ and in occ } p \text{ in year } t+1}{\text{number working in occ } o \text{ in year } t \text{ observed in any other occ in year } t+1}}{\frac{\text{number working in occ } o \text{ in year } t \text{ who are not observed in that job in year } t+1}{\text{number working in occ } o \text{ in year } t \text{ who remain in the data in year } t+1}} \qquad (2)\end{aligned}$$

We estimate these occupation transition probabilities at the national level between almost all 840 x 840 pairs of SOC 6-digit occupations, averaging over all observations in years 2002–2016[9], as well as the shares of people leaving their job in each SOC 6-digit occupation in each of these years.

The BGT resume data set is largely representative of the U.S. labor force by age, gender and location. It over-represents younger workers and white-collar occupations. Since

[9]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 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). .

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 transitions by the U.S. population age shares by occupation, provided by the BLS for 2012-2017. (Further discussion on the data representativeness, including on sample selection concerns, is in the Data Appendix).

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 occupations, then job options outside of a worker's occupation are likely to matter in theory but not in practice. This does not appear to be the case. In this section we present descriptive statistics on occupational mobility over 2002-2016 in the BGT resume database,[10] showing that occupational transitions are 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. Together these facts imply that SOC occupations are not appropriate definitions of workers' true labor markets for a large number of occupations, suggesting that job options outside workers' current occupations must be considered for analysis seeking to capture workers' true labor markets.

The average share of workers leaving their occupation in our data - which we define as the probability of *no longer being observed* in your initial SOC 6-digit occupation from one year to the next, weighted by employment in that SOC 6-digit occupation - is 0.11.[11] The average share of workers leaving their 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 0.5, consistent with the average length of a job in our data being 2 years.[12] Combining these statistics, the average

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

[11]Our measure of outward occupational mobility 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 dataset 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 because of the nature of our resume data a different concept than the concept of strict annual occupational mobility: 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.

[12]Note that leaving your job does not necessarily entail leaving your firm. The CPS reports that median employee

probability of a worker leaving her 6-digit occupation given that she leaves her job in our data is over 20%. This suggests that failing to consider jobs outside a worker's current occupation substantially understates the true job options available to her.

There is fairly large variation in the average share of workers leaving their occupation when they leave their job. Of the 781 6-digit occupations for which we have more than 500 observations, ranking occupations by their outward occupational mobility, the median occupation has a proportion leaving of 0.22, with the 25th percentile 0.17 and the 75th percentile 0.26. The 5th percentile is 0.11 and the 95th percentile is 0.42 (see Table 3).

Almost all of the 'stickiest' occupations (those with the lowest shares leaving the occupation conditional on leaving their job) are highly specialized, such as various medical, legal and educational occupations, people with specific skills such as firefighters or graphic designers, and people in (presumably) desirable unionized occupations like truck drivers (see Table 5). In contrast, many of the least 'sticky' occupations require few specialized skills, such as restaurant hosts and hostesses, cashiers, tellers, counter attendants, and food preparation workers. The difference between these large occupations can be substantial: over 30% of hosts and hostesses at restaurants, lounges and coffee shops, and of telemarketers leave their occupation when they leave their job, which is around three times greater than the 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 for others.

The SOC hierarchy structure groups occupations with other 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, 82% of moves out of the SOC 6-digit occupation are also out of the SOC 2-digit occupation, but this is only 65% at the 10th percentile and is 92% at the 90th percentile (see Table 4). For example, 92% of flight attendants leaving their 6-digit SOC occupation also leave their 2-digit SOC occupation (which includes other transportation occupations like truck drivers, taxi drivers and air traffic controllers), and most flight attendants who leave their occupation move into sales or other white collar jobs. In contrast only 38% of systems software developers who leave their 6-digit occupation also leave their 2-digit occupation: most move to other computer-related occupations like applications software developers, computer programmers, and computer

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.

systems analysts, which are all within the same 2-digit SOC occupation group. 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 2). While many people are observed leaving their occupation each year, there are only a few 'thick' occupational transition paths - i.e., occupational pairs where the transition probability is greater than negligible (as listed in Table 6). For example, conditional on leaving the initial occupation, there are only 382 pairs of 6-digit occupations which have a transition probability of 10% or greater (out of 284,797 observed pairs with greater than 500 observations in the BGT data). Many of these transitions are within very close occupational families. Of computer programmers who leave their occupation, 35% become either web developers, software developers, computer systems analysts or "computer occupations, all other"; 30% of licensed practical and vocational nurses who leave their occupation become registered nurses and a further 11% of them become health services managers; 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: 26% of human resources specialists become human resources managers, 18% of legal secretaries who leave their occupation become paralegals the following year, 18% of accountants and auditors become financial managers, 12% of mechanical engineers who leave their occupation become engineering managers. Finally, some thick transition cells demonstrate occupational similarity across conventionally defined occupational boundaries: 15% of biological scientists who leave their occupation become operations research analysts; 13% of meat, poultry and fish cutters who leave their occupation become truck drivers.

The occupational transition matrix is also highly asymmetric. Many occupational transition paths 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, and the correlation between the absolute size of the flows is only 0.05. 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 is a better proxy on average for the 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 occupation-by-city narrow labor market from occupation-to-occupation transition probabilities. This probabilistic method of labor market definition can also be applied to other questions requiring labor market definition using aggregate data.

2.4 Occupational mobility, task similarity and amenities

In interpreting worker transitions as describing the network of worker outside options, we assume that two occupations with more frequent transitions between them are more similar to each other in the worker's ability to do the required work, and/or in the worker's desire to work in the occupation. There may be a concern however that occupational transition shares are reflective of something idiosyncratic to our data rather than latent similarities between occupations.[13] In addition, observed worker flows may mostly represent short-run contractions or expansions of different occupations, such that flows represent no particular pattern of preference but rather represent which occupations were expanding when others contracted. While our occupational mobility matrix is deliberately estimated by averaging data over a long time horizon, from 2002 to 2015, as with any finite sample, short-run fluctuations may lead us to pick up spurious variation in occupational flows that does not represent structural preferences. To allay these concerns, we therefore explore the degree to which the

[13]Though the size (23 million unique U.S. resumes) and relative representativeness of our data should do something to assuage this concern.

occupational mobility shares measured using our resume database reflect latent similarities between different occupations.

2.4.1 Task distance and occupational mobility

First, we ask whether greater similarity in task requirements between jobs predicts a greater likelihood of observing moves between them.

In order to quantify the similarity between occupations along the task dimensions, we will use a number of different approaches. One approach to measure occupational similarity, proposed by Macaluso (2017) is to use the average difference in characteristics across the full set of "Skill" task content items - there are 35 in total - provided by O*Net. We scale all of these to lie between zero and ten and aggregate them into an average task distance D_{op} between occupations o and p,[14] defined as

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

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

In addition, we source task composites from the literature on occupational task categories that have been linked to important economic outcomes. In particular, we consider six task composites first introduced in Autor et al. (2003) - denoted "ALM", 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. Moreover, we also consider a categorization by Deming (2017) - denoted "DD" - which recasts the occupational task composites and introduces the additional dimension of a composite capturing social skill-related task intensity. 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.

For each of these task composites, we run the following regression:

$$\pi_{o \to p} = \alpha + \beta |TC_{occ\ p} - TC_{occ\ o}| + \gamma (w_{occ\ p} - w_{occ\ o}) + \epsilon_{op}, \tag{3}$$

where $\pi_{o \to p}$ again represents the probability that an occupation o worker will start work-

[14] For a similar notion of task distance, see (Gathmann and Schönberg, 2010). Macaluso (2017) applies a similar formula using task differences to measure local worker-job mismatch.

ing in occupation p in the following year, conditional on working in a different job.[15] In all regressions, we are controlling for wage differences between the occupations, and include origin occupation fixed effects to control for differences in the levels of mobility between occupations - but the results are qualitatively similar without these controls. For the analysis using the task distance measure, we replace $|TC_{occ\ o} - TC_{occ\ p}|$ with D_{op}, as defined above.

If our occupational mobility measures capture a notion of feasible job transitions in the sense of jobs requiring similar tasks, we would expect the coefficient on the task distance measures to be negative. That is, the greater the difference in tasks required for a job relative to the worker's current occupation, the less likely we should be to observe the worker moving into that job. Figure 3 shows the coefficients obtained from estimating equation 3 for the different task distance measures.

We can see that all the coefficients are negative and statistically significant, with the exception of the "non-routine interactive" and "social skill" composites, which capture the extent to which the occupation involves social activities and interpersonal tasks, respectively. The latter have coefficients that are not significantly different from zero, indicating that workers are no less likely to move into occupations that require different levels of social skills than their current job. This suggests that skills to execute social tasks are either more widespread or more easily acquired on-the-job, as differences in this task dimension are not associated with lower occupational transitions.

In contrast, requiring a different amount of cognitive / analytical or manual tasks along different dimensions seem to be associated with lower observed transitions between occupations. This suggests that "harder" tasks that require education, or innate abilities, may represent stronger barriers to taking up jobs if workers are currently in jobs that do not require those tasks.

2.4.2 Job amenities and occupational mobility

Another potential factor in determining moves between jobs - and a source of non-monetary benefits - may be job amenities in the form of "temporal flexibility" of jobs. These amenities are of importance, because, as Goldin (2014) notes, "certain occupations impose heavy penalties on employees who want fewer hours and more flexible employment" (p. 1106),

[15] More precisely, it measures for each year the share of the workers in occupation o in one year that are observed working in occupation p in the next year, divided by the share of workers in that year who are observed having left their job by the next year. Each of the variables used in this calculation represents an average of 2002-2015 observations.

which in turn may contribute to gender gaps in earnings.

We first ask the question of whether workers in our data are more likely to move to jobs that have *similar* time flexibility amenities to their current occupation. 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.[16]

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.

In order to answer the question of whether workers are more likely to move to jobs that are more similar in terms of time flexibility, we estimate the following regression for the absolute difference between each of the aforementioned occupational rigidity characteristics Rig_o, analogous to the analysis in the previous section:

$$\pi_{o \to p} = \alpha + \beta |Rig_{occ\ o} - Rig_{occ\ p}| + \epsilon_{op}. \tag{4}$$

The results of this variation are shown in figure 5. It shows that all the coefficients on absolute differences in time flexibility amenities are negative. That is, workers are less likely to move into occupations that have very different amenities from their current occupation - suggesting that there are latent preferences or a need for flexibility that restrict workers in their mobility.

This specification imposes symmetry on occupational transitions. In fact, it might be the case that workers who switch occupations are more likely to move to occupations with less or more temporal flexiblity. To estimate this, we run the regression laid out in 4, but with the relative difference between the occupational rigidity characteristics $Rig_{occ\ o} - Rig_{occ\ p}$ rather than the absolute difference.

The individual coefficient estimates for the effect of the characteristic differences between the target and origin occupations on transition probabilities are shown in figure 4.[17] Note

[16] These 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.

[17] Note that we do not control for differences in wages in this analysis, in contrast to the task distance regressions above. The reason is that job amenities should be priced in the wage, a fact that Goldin (2014) highlights, so controlling for wages will absorb the effect of amenities. However, in unreported regressions we have found that the results are very similar when we control for wage differences, with the exception that, when controlling for wages,

that this analysis involves *directed* relationships between occupations, so if the same share of moves in each direction would be observed for an occupation pair, then the estimated effect of differences between them would be zero. This analysis therefore relies on the special feature of our data that we can observe asymmetric transition dynamics to determine the effect of differences in characteristics.

From the figure, we can see that occupational transitions have on average been *towards* occupations that require contact and working relationships with others. That is, when workers move into another occupation, they are on average less likely to have time flexibility in that occupation.

2.4.3 Career advancement and occupational mobility

Another reason for observing *directed* career moves may be career advancement by workers. Over the lifecycle of a career, workers might move into positions of increasing responsibility or seniority, which is reflected in changes of occupation. As our measure of transitions is sequential - that is, we measure whether an occupation is observed *following* another - it should reflect the evolution of jobs held during a typical career towards greater rather than less responsibility.

To study whether workers do indeed tend to move *up* rather than *down* the organizational hierarchy, we first identify occupational characteristics that measure managerial responsibilities. In particular, we used the following algorithm to determine the applicable characteristics: 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. In addition, O*Net has a work style category that explicitly measures "leadership" in a job - so we added that measure to the list as well.[18]

workers have been less likely to move to environments that give them discretion in the form of in decision-making without supervision.

[18] 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.

The final list obtained from this selection algorithm comprises 7 different occupational characteristics. 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. Finally, we also create a "leadership" composite, which represents the mean score across these 7 characteristics.[19]

For each of these measures, we estimate the following regression equation, analogous to the previous sections' analysis:

$$\pi_{o\rightarrow p} = \alpha + \beta(Lead_{occ\ p} - Lead_{occ\ o}) + \gamma(w_{occ\ p} - w_{occ\ o}) + \epsilon_{op}, \tag{5}$$

where we are now interested in the β coefficient on the difference in leadership tasks $Lead_{occ}$, between the target and origin occupation.

Figure 6 plots the estimates for the leadership coefficients. For all 7 leadership characteristics, the coefficient is positive and significant. This means that, on average, workers transition *towards* jobs that require more leadership tasks and for which managerial tasks are more important - as would be expected from moves up the career ladder over the course of a typical work history. Similarly, the leadership composite positively predicts occupational moves. As a result, our transition probabilities do not just capture lateral moves, but also the the outside option of moving up to a job with greater responsibility. This supports our claim that our occupational mobility measure captures directed moves revealing worker preferences - for instance for career advancement.

In summary, this section has explored various dimensions - task distance, amenities, and career advancement - that represent intuitive motivations for deliberate occupational moves and we have found that our measure of occupational mobility relates in an intuitive way to worker preferences in that occupational mobility is higher for jobs that require similar tasks and offer similar amenities; at the same time, workers tend to move into jobs requiring greater responsibility and leadership.

3 Theoretical framework

In section 2, we argued that the relative frequency of occupational transitions reflects the relative relevance of destination occupations as outside options for workers in any given initial

[19] 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.

occupation. This implies that we can use occupational transition data to create a ranking of occupations in terms of their relevance as outside options to any given initial occupation. The greater the number of transitions there are from one occupation to another, the more relevant that occupation is assumed to be as an outside option for the initial occupation.

Not only can we define *which* jobs are outside options to a given occupation on average - we can also estimate the average *value* of workers' outside-occupation job options. Specifically, we can use occupational transition data to estimate the dollar value to workers in a given occupation of their average or expected best job option outside their occupation, by weighting the average wages in each alternative occupation by the probability that workers from the initial occupation move to that occupation, given that they leave their job. Under the assumption that the *observed* transitions we see are representative of the *potential* transitions of workers still in the initial occupation - that is, that the relative occupational composition of the best outside options of the workers who have left the occupation is the same as the best outside option of the workers who are still in the occupation - the observed transitions out of the initial occupation represent the best outside options for workers still in the initial occupation.

$$\text{value of outside-occupation option}_o = \sum_{p}^{occs} \frac{\text{workers moving from occ } o \text{ to occ } p}{\text{workers leaving job in occ } o} \cdot \text{wage}_p \quad (6)$$

Consider a simple example, shown in Table 1: 100 workers start in initial occupation o. They all leave their job at the end of the year and find new jobs either in their own occupation or one of three other occupations p, q and r:

Table 1: Outside option example.

Occupation x	Number moving from occ o to occ x	Wage in occ x
o	80	\$12
p	10	\$10
q	5	\$12
r	5	\$8

Here, 20% of workers in occupation o leave their occupation when they switch job. They are twice as likely to move to occupation p than occupation q or r, implying that jobs in occupation p are twice as relevant outside options for occupation o workers, on average. The expected wage outside workers' own occupation o would be $0.5 \times 10 + 0.25 \times 12 + 0.25 \times 8 = \$10$, and the average expected value of their outside-occupation job option is therefore $\$10 \times 0.2 = \$2$.

This measure of the value of the average outside-occupation job option is intuitively plausible: the value of workers' average outside option outside their occupation is a weighted average of all the jobs outside their occupation, with the weights being the observed likelihood of workers from their own occupation to transition to each new occupation, given that they leave their job.

In the next section, we outline a simple search model which lays out more formally the assumptions under which this transition-weighted average wage is a valid measure of the value of outside-occupation job options. In the appendix, we also show that the same transition-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.

In our search framework in section 3.1, each worker receives only a subset of all feasible offers each period and takes the best of those options. Occupational transition shares are used to proxy for a worker's ex ante expected likelihood of a job in a given occupation being her best outside option, and the overall value of her outside-occupation options is the expected value over each of these possible outcomes.

3.1 Search framework

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. They accept the offer which pays the highest wage.

3.2 Model setup

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.

$$\begin{aligned} w_i &= \beta(MPL_i - oo_i) + oo_i \\ &= \beta MPL_i + (1-\beta)oo_i \end{aligned} \tag{7}$$

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[20]. 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.

3.2.1 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

[20]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.

her current employer and applies for jobs at other firms. The expected value of this outside option is therefore:

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

Note that $\prod_{j=1}^{N_i}(1-\alpha_j)$ is the probability that worker i receives no offers from any firms and is therefore 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, where the weight is her probability of becoming unemployed if she leaves her current job.

3.3 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 also do not consider the outside option value of unemployment[21]. 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.

Analogous to before (8), the probability that the worker moves to *any* job in occupation p is the sum of the probabilities of getting job offers from each of the firms j in that occupation, and the probability that that job offer is the best job offer the worker receives. We make three assumptions which will enable an empirical application. First, we assume that observed occupational transitions (at the national level) are representative of incumbent workers' likelihood of moving to a given occupation. Second, we assume that the probability of moving to firm j in occupation p, given that the worker moves to *some* job that occupation-by-city

[21]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.

labor market, is proportional to firm j's employment share in that labor market, $s_{j,p}$, following Burdett and Mortensen (1980)[22]. The worker's probability of moving from firm i in occupation o to firm j in occupation p is therefore:

$$\begin{aligned} & Pr(\text{move from firm } i \text{ in occ } o \rightarrow \text{firm } j \text{ in occ } p) \\ &= \sum_{j}^{N_{occs}} Pr(\text{get offer from job } j \text{ in occ } p) \cdot Pr(w_j \text{is best offer}) \\ &= \pi_{o\rightarrow p} s_{j,p} \end{aligned} \tag{10}$$

We also assume that all workers in occupation o have the same probability of moving between occupations, and would all be offered the same wage by each firm j in any occupation p (in expectation).

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

$$\begin{aligned} oo_{i,o} &= oo_{i,o}^{own} + oo_{i,o}^{occs} \\ &= \underbrace{\pi_{o\rightarrow o} \cdot \sum_{j\neq i}^{N_{i,o}} s_j \cdot w_{j,o}}_{\text{jobs in own occ}} + \sum_{p\neq o}^{N_{occs}} \underbrace{\pi_{o\rightarrow p} \cdot \bar{w}_p}_{\text{jobs in other occs}} \end{aligned} \tag{11}$$

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.

4 Outside-occupation options and wages

In section 2, we argued that high annual occupational mobility implies that occupations are poor reflections of workers' true labor markets, and that highly heterogeneous mobility by occupation implies that for some occupations, the occupation is a better approximation of their true labor market than for others. If these arguments are correct, then the quality of workers' job options outside their current occupation will positively affect workers' wages in their current occupation, and the strength of this relationship will differ across occupations.

[22]Burdett and Mortensen (1980) assume that the conditional probability that an offer received by a searching worker is that of firm i is equal to $\frac{n_i}{n}$.

In section 3, we argued from a simple intuitive standpoint and using search and matching models that a transition-weighted average of the wages in other occupations can be used as an index for workers' job options outside their current occupation. In this section, we construct this outside-occupation option index for most occupations and metropolitan areas in the U.S. over 1999–2016, and evaluate its relationship with 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 o as the weighted average of wages in other occupations $\bar{w}_p$, with the weights the occupational transition probabilities $\pi_{o \to p}$:

$$oo_o^{occs} = \sum_{p \neq o}^{N_{occs}} \pi_{o \to p} \cdot \bar{w}_p \tag{12}$$

We apply this at the level of individual metropolitan areas[23] ("cities"), considering outside-occupation job options only *within workers' own city*. We do not consider workers' outside job options in other cities, either in their own or other occupations. Annual outward residential mobility from metropolitan areas is approximately 3%[24]. While migrating to other cities can be an important outside option, occupational mobility appears to be substantially more important for most workers: annual outward mobility from a SOC 6-digit occupation is over 10%.

We want to construct our outside-occupation option index at the level of individual cities, but our occupational transition probabilities $\pi_{o \to p}$ are calculated at the national level[25], and so implicitly reflect the national average availability of jobs in outside occupations. The abundance of job options in some occupations in some local labor markets may differ substantially from the national average, however, so we re-weight to reflect this using the local relative employment share of jobs in p: the employment share of jobs in p in city k in year t, $s_{p,k,t}$, relative to the national average employment share of jobs in p in year t, $s_{p,t}$.

Our empirical outside-occupation option index for workers in occupation o in city k in year t (expression 13) is therefore a weighted average of the average wages in other occupations p in city k in year t, where the weights are the product of two components: the national

[23] We would prefer to use Commuting Zones, since they are better measures of local geographic labor markets. However, occupational wage data is only available at the level of the metropolitan area and not Commuting Zone.

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

[25] They are calculated from the Burning Glass Technologies resume data, as described in section 2.

average occupation o to p transition probability, $\pi_{o \to p}$, and the local relative employment share of jobs in occupation p,$\frac{s_{p,k,t}}{s_{p,t}}$. We use employment data and wage data from the BLS Occupational Employment Statistics (OES) for the relative employment shares and average wages by SOC 6-digit occupation and metropolitan area.

$$oo^{occs}_{o,k,t} = \sum_{p}^{N_{occs}} \left(\pi_{o \to p} \cdot \frac{s_{p,k,t}}{s_{p,t}} \cdot \bar{w}_{p,k,t} \right) \tag{13}$$

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.

$$log(\bar{w}_{o,k,t}) = \alpha + \beta_1 log(oo^{occs}_{o,k,t}) + \Gamma_{o,k,t} + \epsilon_{o,k,t} \tag{14}$$

We use data from the BLS OES (Occupational Employment Statistics) on employment and the average hourly wage by SOC 6-digit occupation and CBSA for each year of 1999-2016. This data does not exist for many of the occupation-CBSA pairs. Of the possible 786,335 occupation-CBSA pairs, wage data in the BLS OES only exists for approximately 115,000 each year. The missing occupations and CBSAs are primarily the smaller ones.

Table 7 shows the results of this regression across all occupation-CBSA 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 CBSA fixed effects and column (3) has CBSA-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-CBSA and occupation-by-year fixed effects, and so identifies only off annual variation in outside options from their mean for each occupation-by-CBSA 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 (5) suggesting that a 10 log point higher value of outside options in other occupations is associated with 0.7-0.8 log points higher wages in the workers' own occupation; or a 1 standard deviation[26] higher value of outside options in other occupations is associated with 3.2-3.6 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 options which do not affect 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 occupation with plausibly exogenous national demand shocks to the 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 (or the first year the occupation-CBSA cell is in the data if it is not present in 1999). Our instrument for the oo^{occs} index, $oo^{occs,inst}$, therefore becomes the weighted average of national leave-one out mean wages in occupation p, $\bar{w}_{p,\not{k},t}$, with the weights 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, $\pi_{o \to p}$.

$$oo_{o,k,t}^{occs,inst} = \sum_{p}^{N_{occs}} \left(\pi_{o \to p} \cdot \frac{s_{p,k,1999}}{s_{p,1999}} \cdot \bar{w}_{p,\not{k},t} \right) \tag{15}$$

The key identifying assumption for the wage instrument is that the national leave-one-

[26] This represents the average standard deviation of the outside-occupation option index *within* each occupation.

out mean wage in occupation p is correlated with the local wage in occupation p, but is not correlated with the local wage in occupation o. Identification is achieved from two factors. Identification comes from differences in initial exposure to related occupations[27] across different cities: taking occupation o in city k and in city l, if city k had a higher initial proportion of its employment in outside-option occupations than city l did, the instrumented oo^{occs} index is higher and therefore we would expect the wage in occupation o in city k to be higher than in city l. This instrumental variable strategy is closely related to that of Beaudry et al. (2012), who avoid the reflection problem in their index of cities' industrial composition by using national industry wage premia to substitute for city-level industry wages.

We show the reduced form results of our instrumented regressions in Table 8. The results for the instrumented oo^{occs} index remain positive and strongly significant, with magnitudes about one third of the size of 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 similar occupations rises, the wage in the initial occupation also rises. The coefficient magnitudes suggest that a 10 log point higher outside-occupation option index is associated with 0.2-0.3 log points higher wages in the workers' own occupation; and a 1 standard deviation higher outside-occupation option index is associated with 0.9-1.4 log points higher wages in the workers' own 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.

4.4 Heterogeneity by occupations: wages and task intensities

We explore heterogeneity in the relationship between outside-occupation options and wages for occupations at different points in the wage distribution. Splitting the occupations into terciles based on their national average wage in 2016 (with cutoffs shown in Table 9), we

[27] 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.

run our preferred specification on each of the occupational wage terciles for the simple outside options measure in Table 10 and the instrumented outside options measure in Table 11. In both sets of regressions, the coefficients are smaller for low-wage occupations than for medium- or high-wage occupations; in the instrumented regressions, we see no relationship at all between the value of outside-occupation options and the value of the wage for low-wage occupations.

This stark difference between low-wage occupations and middle- and high-wage occupations could be explained by four factors. First, it is possible that outside-occupation job options are not very feasible or desirable outside options for low-wage occupations. This strikes us as unlikely, since the group of low-wage occupations has high outward mobility: the average occupation leave share conditional on leaving the initial job is 27% for the low-wage tercile of occupations, compared to 21% for the high-wage tercile of occupations. Second, it is possible that (in a Nash bargaining framework) workers in low-wage occupations have a substantially higher bargaining power or rent-sharing elasticity than workers in high-wage occupations. This also strikes us as unlikely: if anything, one would expect workers in medium- and high-wage occupations to have higher bargaining power. Third, it is possible that low-wage labor markets are better approximated by a competitive model of wage determination than a model of imperfect competition and/or frictions. In this case the bargaining channel whereby outside options affect wages would not be relevant. Finally, it is possible that the leave-one-out mean wage instrument does not effectively identify demand shocks to local occupations for low-wage workers.

We also explore heterogeneity by a number of dimensions of task intensity. Specifically, we run our baseline and instrumented regression of wages on outside-occupation options by tercile of various measures of task intensity: non-routine cognitive analytical tasks, non-routine cognitive interpersonal tasks, routine cognitive tasks, routine manual tasks, and non-routine manual tasks as measured by Autor et al. (2003), and our composite task measure of leadership (described in section 2). The regression results are shown in Tables 12 and 13. The instrumented results suggest that for occupations which are intensive in cognitive tasks, or not intensive in manual tasks, shocks to outside-option occupations matter for wages. On the other hand for occupations which are intensive in manual tasks, or not intensive in cognitive tasks, shocks to outside-option occupations appear to have little effect on wages. These results are compatible with the results by wage tercile: in particular, non-routine cognitive analytical task intensity and leadership task intensity are highly predictive of occupational wages[28].

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

We demonstrated in section 4 that our index of outside-occupation options and, to a lesser degree, our index of outside-city options are strongly and significantly related to wages. Here, we apply this logic to recent work on local labor market concentration and monopsony power. We demonstrate that failure to consider workers' options outside their occupation and city can lead to biased inference on the relationship between wages and local labor market concentration.

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[29], 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 (in a Nash bargaining setup, for example, this split is determined by the bargaining coefficient).

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 a variety of occupations and locations: in the worker's

[28] A cross-sectional regression of the average national occupational wage in 2016 on the occupation's non-routine cognitive analytical task intensity and leadership task intensity has an R-squared of 38%.

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

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 (or products), and excludes all other jobs (or products) from this analysis. 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[30]. Considering a broader set of affected outcomes, Hershbein and Macaluso (2018) 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.

If the boundaries of an occupation or city were impermeable, so that workers could never switch, then labor market concentration within an occupation and city may indeed be a good proxy for workers' outside options. However, our results in section 4 suggest that jobs outside workers' occupation 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.

5.2 Wage and HHI regressions

In Table 14, we regress the log average wage on the log vacancy HHI with occupation, CBSA and year fixed effects. We follow Azar et al. (2018) and Hershbein and Macaluso (2018) 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 CBSA

[30] 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.

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 CBSA over 2013-2016. The elasticity of mean wages to the annual vacancy HHI is -0.019 in a specification with occupation-by-year and CBSA fixed effects, which is smaller than but of the same order of magnitude as the estimates in Azar et al. (2017) and Rinz (2018).

Our results above show, however, that options outside a worker's occupation matter significantly for their wages. Therefore, if workers in different occupations and cities have different availability of these outside-occupation options - if they are differently mobile across occupations - local employer concentration should have very different effects on different narrow occupation-city labor markets. In occupation-city labor markets where few workers have the option to get a job outside that labor market, local employer concentration would be expected to have a much greater effect on worker outside options - and therefore on the wage - than in occupation-city labor markets where workers are easily able to get jobs outside that labor market. To the extent that the regressions of wages on local employer HHIs capture effects of employer concentration on wages, rather than omitted variables, therefore, one would expect the coefficient to be lower, the more outside-occupation options are present.

We segment our data into four quartiles by the national average occupation leave share (annual outward mobility from the occupation) over 2002–2016[31]. We re-run the regression of log wage on HHI at the occupation-CBSA level on each of the the four quartiles separately (14). The coefficient on the vacancy HHI is about the same as the overall baseline coefficient for the 2nd and 3rd quartiles of the leave share, and the coefficients are not statistically significantly different from each other; but the coefficient for the quartile of occupations with the lowest outward mobility (lowest occupation leave share) is more 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. These results are consistent with the interpretation that occupations with very low outward mobility (low leave shares) are substantially better approximations of workers' true labor markets than occupations with higher leave shares.

In addition, we find that in our data, the vacancy HHI in occupation-by-city labor markets is strongly negatively correlated with workers' outside-occupation options and weakly negatively correlated with our instrument for workers' outside-occupation options. This suggests

[31]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 $T+1$.

that coefficient estimates for the relationship between the HHI and the wage, from regressions which do not control for outside-occupation job options, are likely to be biased upward. We show this by regressing the log wage on the vacancy HHI as above, controlling for our simple outside-occupation index in column (2) and our instrumented outside-occupation index in column (3) (Table 15). In the regression with the simple outside-occupation option index, the coefficient on the vacancy HHI falls statistically significantly by roughly half, from a semi-elasticity of -0.019 to a semi-elasticity of -0.009, while in the regression with the instrumented outside-occupation option index the coefficient is roughly the same as in the original regression. The fall in the coefficient when controlling for the simple outside-occupation option index is consistent with the omitted variable bias hypothesis above: that estimates of the relationship of the HHI with wages in recent papers which do not control for differences in workers' availability of options outside their occupation-city labor market have an upward bias (in terms of the magnitude of the coefficient).

6 Outside options and labor demand shocks

There has been an active and growing literature on the effect of labor demand shocks on local labor market outcomes. One part of this literature consists of various papers exploring the effect of the "China Shock" of exposure to import competition from China during the 2000s in manufacturing industries on the labor markets in the US that are most dependent on employment in the affected industries. For example, Autor et al. (2013) find that commuting zones that are more exposed to import-competing manufacturing experience lower wages and higher unemployment during the 1990-2007 period. Acemoglu et al. (2016) finds that this impact extended to an "employment sag" in other sectors of the US economy through input-output linkages across sectors, but that the effect was concentrated in exposed tradables sectors, in line with the import shock explanation.

At the same time, there has been an increasing awareness that worker outside options matter for local labor market outcomes - whether they are measured in the form of other local industries Beaudry et al. (2012); a greater diversity of employment choices for individual workers with particular characteristics Caldwell and Danieli (2018); or as a network of former coworkers Caldwell and Harmon (2018).

Our approach of defining a local measure of the quality of outside options at an occupational level connects these two literatures: If outside options affect a worker's ability to exploit labor market opportunities, then the impact of labor demand shocks should depend on the quality of outside options to which affected workers have access in their area.

We explore this prediction by extending the Chinese import shock analysis by Acemoglu et al. (2016) and other authors in several ways.

It has been well established that labor demand shocks from Chinese import competition may adversely affect local wages and employment in areas that are more exposed to these shocks. However, there has so far not been any attempt to explore the heterogeneity in local impact on specific occupations. We use the fact that we have data on local occupational-level wages to analyze how the effect of the "China Shock" differs across occupations *within* a geography.

As we found in section 4 above, labor demand shocks to workers' outside-option occupations affect their own wages, and this pattern differs by type of occupation, with stronger effects for workers in occupations with high cognitive task intensities and low manual task intensities. These results would suggest that labor demand shocks, such as those from Chinese import competition, will not only have direct effects on occupations in import-competing industries, but will also transmit through outside-option linkages to other occupations.

To test these hypotheses in the data, we estimate the degree to which local labor demand shocks spill over between occupations that form part of a probabilistic labor market under our definition and how these spillovers depend on the task components of occupations. We show that the differential impact on occupations is in part driven by the difference in the quality of local outside options between occupations - an *indirect* effect of labor demand shocks that operates in addition to any direct effects.

We focus on labor demand shocks in the form of Chinese import competition shocks as defined by Autor et al. (2013). In that paper, the change in import competition IP in industry m at time τ is originally defined as

$$\Delta IP_{m\tau} = \frac{\Delta M^{US}_{m,\tau}}{Y_{m,q_1} + M_{m,91} - E_{m,91}}$$

which uses the change in imports from China $\Delta M^{US}_{m,\tau}$ divided by a measure of initial period trade absorption, defined as the sum of industry shipments and net imports. We obtain the data for this measure from Acemoglu et al. (2016) and adapt it to our occupational-level analysis as described below.

6.1 MSA-level impact of labor demand shocks

We start by documenting the effects of the China labor demand shock at the MSA level, in order to be able to contrast the results with the disaggregated effects at the occupation level.

To do this, we aggregate the expected effect of the China Shock at an MSA level during various time periods using local industry employment shares from County Business Patterns data. We estimate an equation of the form

$$\Delta Y_{\tau}^{MSA} = \alpha_{\tau} + \beta \Delta IP_{\tau}^{CZ} + \gamma X_{\tau}^{MSA} + e_{MSA,\tau},$$

where the dependent variable can be the change in total MSA employment ΔE_{τ}^{MSA} or the change in the log of avg. MSA wages Δw_{τ}^{MSA}.

The IV estimation results for the MSA-level effect on employment are shown in table 16. All columns of the table correspond to the same specification, but for subsets or aggregations of the data that correspond to different time periods or geographies. Our baseline results in columns (1)-(3) suggest that, over the 2000-2011 period, a 1 percentage point increase in import penetration is associated with a reduction in the local MSA employment rate by 2.21 percentage points, which replicates similar results found in other research, including in Acemoglu et al. (2016)[32].

The estimated effects of import shocks on MSA-level wages are shown in Table 17. They are - somewhat counterintuitively - *positive*, suggesting that a 1 percentage point increase in import penetration is associated with a increase in the growth rate of nominal wages in the MSA by 2.85 percentage points over the 2000-2011 period. Obtaining a qualitatively similar result, Acemoglu et al. (2016) argued that this positive effect may be due to a change in the composition of employment. This would be the case, for instance, if the negative effect on employment of the labor demand shock predominantly affects lower-paid workers, such that those remaining in employment are high-wage workers, which increases the average wage in the MSA *among the employed*. We will explore this hypothesis further in our occupation-by-geography-level analysis.[34] The results are shown in table 17 - with the specifications in each column corresponding to the first four columns in table 16.

[32]The remaining columns of the table show that this result is robust with regard to using different time periods (columns (4)-(6))[33] and using commuting zones as the geographic units (columns (7)-(10)). CZs include some rural and small urban areas that are not part of the MSA sample. The results show that the significant negative effects persist for CZs and are of similar size. These findings are similar to those in other China Shock papers. In fact, columns 9 and 10 of table 16 correspond exactly to results in Acemoglu et al. (2016).

[34]The OES wage data is not provided at a CZ level for this time period, which is why we only do the wage analysis at the MSA level.

6.2 Estimating occupation-level effects of China shocks

To translate the national industry-level shock into local occupation-level impact measures, we make use of the disaggregated occupation-by-MSA OES wage data, and U.S. occupation-by-industry matrices, both available from the BLS. Using the industry-level import shock measure $\Delta IP_{j\tau}$ as an input, we define the measure of exposure to demand shocks from import competition for an occupation o in geography k as

$$\Delta IP^k_{o\tau} = \sum_m \frac{\tilde{L}^k_{om\tau}}{\tilde{L}^k_{o\tau}} \Delta IP_{m\tau}, \tag{16}$$

Here, the term $\tilde{L}^k_{om\tau}$ measures the expected number of local workers in industry m that have occupation o, and $\tilde{L}^k_{o\tau} = \sum_k \tilde{L}^k_{om\tau}$ is the expected total number of location k employees in occupation o.[35] Intuitively, this expression applies weights to the national industry-level import competition shocks – based on the expected local prevalence of an occupation's workers in those industries – to translate them into local occupation-level shocks.

To compute the expected number of local industry m workers $\tilde{L}^k_{om\tau}$ in occupation o, we combine data on local industrial employment, from the County Business Patterns, and the occupational content of industries, measured at the national level. We use data from the three years 1999-2001 for which the BLS provides occupation-by-industry matrices that map between SIC industry codes and the SOC codes used for later occupational data. Conveniently, these years also coincide with the beginning of our occupational wage sample. We average the occupation-by-industry shares over these three years to make our estimate of the breakdown more robust to sampling error. The result is an SIC 3-digit industry-level vector of shares ϕ_{om}, which represents the national share of occupation o workers in industry m

We assume that the baseline period occupational structure within industries does not vary systematically across geographies, such that we can proxy for local occupational structure using the national industry average. Thus, we compute expected local occupation structure as

$$\tilde{L}_{o\tau k} = E^k_{m\tau} \phi_{om\tau},$$

where $E^k_{m\tau}$ is the area k employment in industry m. While the primary motivation for using a national industry-occupation mapping to construct local occupational structures is data

[35]Note that this expression is quite similar to equation (8) in Acemoglu et al. (2016), but with the difference that we ultimately aggregate to the occupation-by-MSA level instead of the commuting zone level. Thus, while our approach is comparable, we are interested in a more disaggregated level of analysis.

availability, an additional motivation is that it avoids some potential endogeneity issues. That is, local occupational structures in particular industries may adapt to the availability of labor markets for talent due to the presence of other local industries with similar skill demands. In that case, the extent to which occupations are locally exposed to different industries would be a function of local outside options, which we will later also use to measure indirect exposure to the import shocks. Consequently, differential import shock exposure might be mechanically correlated with occupational outcomes through local occupational structure. In contrast, the method of using national industry-to-occupation mappings to construct *expected* local occupational breakdowns avoids this endogeneity arising from local adaptation in occupational structures.

However, the industry-level import competition shocks raise endogeneity issues, as the evolution of local US industries may be causing import dynamics. To focus on the exogenous effect of import competition, we therefore follow Acemoglu et al. (2016) and instrument for $\Delta IP_{m\tau}$ using a proxy for the supply-driven component of import competition based on the analogue for the US measure of Chinese import competition, but calculated for eight other high-income countries:

$$\Delta IPO_{m\tau} = \frac{\Delta M^{OC}_{m,\tau}}{Y_{m,88} + M_{m,88} - X_{m,88}}$$

In all the estimations below, the instrumented versions replace the actual industry-level import shock with $\Delta IPO_{m\tau}$ in constructing the respective measures of exposure at the geography- or occupation-level.

6.3 Occupation-level *direct* impact of labor demand shocks

At the occupation-by-geography level, we estimate an equation of the form

$$\Delta Y_{o\tau k} = \alpha_\tau + \beta \Delta IP^{Dir}_{o\tau k} + \gamma X^{MSA}_{o\tau k} + e_{o\tau k},$$

where the dependent variable and import shocks are now also indexed by occupation o. The dependent variables Y are, again, employment and wages. First, we focus on the *direct* effect of China shocks on occupational labor market outcomes without considering indirect channels - these will be explored in the next section.

As mentioned before, the expected effect of direct import shocks on wages can be ambiguous, as the negative employment effects may disproportionately fall on lower-income workers in an occupation. In that case, average measured wage effects for an occupation may be insignificant or positive even though the effect on individual workers is negative. However, the occupational employment effects of a negative labor demand shock should be unambiguously negative.

Occupational wage effects of direct shocks. We begin by estimating the direct effect of China shock exposure on wages by regressing the occupation-by-MSA change in wages over 2000-2011 on the estimated import shock exposure, computed as in equation (16). The results of the OLS estimation of this regression are shown in columns 1-3 of table 18 with the most basic specification in column 1, and controlling for MSA fixed effects in column 2, and for major occupational group fixed effects in column 3. The shock impact on wages is positive and significant in the latter specification. The finding of positive effects on average wages is robust to instrumenting for the import shock using the change in Chinese exposure of other countries, as we do in columns 7-9.[36]

This means that, within an MSA, occupations that have higher import shocks relative to their city and occupational group mean on average experience higher wage increases. However, it is important to stress that this does not mean that import shocks have a salutary effect on impacted occupations: more likely, this increase in average wages among those employed in the occupations represents selection in the workers who are dismissed. This interpretation is also consistent with the negative employment effects and the heterogeneity of indirect effects by task intensity presented below.

Occupational employment effects of direct shocks. For employment, the effects of greater import shocks affecting each occupation-by-MSA unit are unambiguously negative, across all specifications, in line with the effect estimated at the MSA level. The effect sizes are similar to, or somewhat bigger in magnitude, than those found at the MSA level. The full IV specification in column 9 suggests that a 1 percentage point annual increase in import penetration, relative to the average local and occupation group import shock exposure, is associated with a reduction in local occupation employment share by 2.85 percent annually. The MSA average effect therefore obscures the fact that the shock impact on more exposed occupations far exceeds that on less exposed occupations. This means that defining the entire MSA as

[36] In order to ensure that the initial difference in results between the MSA-level and the occupation-by-MSA level is not due to an issue with our method for imputing local occupational import shock exposure, we re-aggregate our measure of the occupation-level import shocks to the MSA level, weighting occupations by actual employment in the year 2000. Using this measure in MSA-level regressions that are not reported here, the estimated MSA-level effects of the import shocks using this average of the predicted occupation-level shocks is again positive and significant, with a coefficient of 1.85 when all controls are included, so the coefficient sign reversal when occupational group fixed effects are not controlled for reflects an aggregation effect, not an issue in our occupational shock measure.

the relevant labor market for evaluating the impact of shocks to a worker's labor market is likely to lead to misleading results. Moreover, these results do not yet take into account the possibility for spillovers between occupations that are closely connected, as we do in the next section.[37]

6.4 Spillovers of labor demand shocks to other occupations

While the analysis above shows that labor demand shocks affect local occupational wages and employment directly, we are particularly interested in what role labor market linkages in the form of job options outside workers' own occupations play in transmitting shocks.

Ex ante, if job options outside workers' own occupation matter, we would expect wages of workers in occupation o to be negatively affected by import shocks to their labor market. That is, worker bargaining power will diminish if employment opportunities in other parts of their occupation-specific labor market decline. Estimating the effect of *indirect* labor demand shocks also avoids the composition bias for wage effects noted above - that lower-income workers may be let go in greater numbers. While the expected sign of the estimated direct effects of labor demand shocks was ambiguous due to that bias, indirect labor demand shocks should have a clear negative effect on wages if outside options matter for wage bargaining.

At the same time, the effect of shocks in related labor markets on employment in occupation o would be expected to operate through flows of laid-off workers from those other occupations. That is, those workers who are most likely to consider occupation o as part of their labor market should look for work in occupation o when their original occupation is hit by a negative labor demand shock - driving up labor supply and employment in occupation o. The first channel therefore captures the role of *unexercised* outside-occupation job options in the wage bargaining process, while the second channel captures the role of *exercised* outside-occupation job options.

Measuring indirect labor demand shocks. For both of the channels through which shocks on other occupations may affect workers - indirect wage effects and indirect employment effects - we can use our measure of occupational mobility to define an empirical measure of the relevant labor market connections.

[37] Note that spillovers between occupations that are closely connected through the labor market drive a wedge between the partial equilibrium estimates at the occupation level and the MSA-level effects that incorporate these externalities, but the sign of this difference is not clear ex ante.

For the analysis of indirect effects on wages, we define the measure of import shock effects on the outside options of occupation o in city k as

$$\Delta IP^{OO}_{o\tau k} = \sum_{occ \neq o} \pi_{o\rightarrow p} \rho_{p\tau k} \Delta IP_{p\tau k},$$

where $\pi_{o\rightarrow p}$ captures the probability of workers in occupation o working in occupation p within a year, conditional on working anywhere else, taken directly from the average occupational connection matrix for 2002-2015. Here, $\rho_{p\tau k}$ is the ratio of the local share of employment in occupation p relative to the national share, capturing the relative local prevalence of outside option p:

$$\rho_{p\tau k} = \frac{\tilde{L}_{p\tau k}}{\tilde{L}_{\tau k}} \left(\frac{\tilde{L}_{p\tau}}{\tilde{L}_{\tau}} \right)^{-1}$$

Thus, $\Delta IP^{OO}_{o\tau k}$ captures the average expected impact of import shocks on other local occupations, but weighting them to account for their relative prevalence in city k and their relevance for the labor market of workers in occupation o. Similarly, the measure of the local exposure to indirect import shocks through occupations likely to *send* worker flows (*WF*) to occupation o is constructed as

$$\Delta IP^{WF}_{o\tau k} = \sum_{occ \neq o} \pi_{p\rightarrow o} \rho_{p\tau k} \Delta IP_{p\tau k}.$$

Wage effects of indirect labor demand shocks. Using our measure of indirect import shocks, we can estimate a wage equation of the form

$$\Delta w_{o\tau k} = \alpha_\tau + \beta_{Dir} \Delta IP^{Dir}_{o\tau k} + \beta_{OO} \Delta IP^{OO}_{o\tau k} + \gamma X^{MSA}_{o\tau k} + e_{o\tau k},$$

Similarly, we can analyze the differential effects of direct and indirect trade shock exposure on employment by replacing the dependent variable with $\Delta E_{i\tau k}$, and measuring the indirect shock exposure by $\Delta IP^{WF}_{o\tau k}$.

The results from using this equation to estimate the effect of indirect import shock exposure on wages are shown in columns (4)-(6) (OLS) and (10)-(12) (IV results) of table 18. Columns (4)-(5) and (10)-(11) show that the effect of import shocks on workers' wages is negative, when we control for MSA and occupation group fixed effects - no matter whether we use OLS, or instrument for shock variables. Moreover, the negative effect of indirect import shocks on wages is statistically significant in all the IV specifications. This means that the indirect shock effects matter for wages - independently of the direct effect of labor demand shocks on wages.

One concern that might arise from the way our measures of indirect shocks are constructed is that they might capture an effect arising from the difference in overall local availability of *outside options* for different occupations[38]. To address this concern, we control for the local *presence* of outside options for each occupation, which is simply determined as

$$OO_{o,2000,k} = \sum_{occ \neq o} \pi_{o \to p} \rho_{p \tau k},$$

so it corresponds to the indirect import shock through outside options, if all the import shocks were equal to one.

Columns (6) and (12) of table 18 show the results when we include this control for the level of outside options in the baseline period - here, the year 2000. The estimated negative effect of instrumented indirect shocks on wages in column 12 becomes larger and is now significant at a 1% level, in line with the theoretical prediction from a bargaining model where worse outside options lower wages for workers. This strengthening of the result when controlling for the amount of local outside options also suggests that the negative effect of indirect import shocks is not solely due to *more* outside options, but rather due to the *shock exposure* of the outside options.

We think of the specification in column (12) as best estimating the effect of interest for a bargaining setting - as it most closely captures the effect of changes in the quality rather than the size of the labor market outside one's own occupation.

It is important to note that the indirect effect regressions control for the *direct* impact of import shocks on each local occupation as well - such that the indirect shock effect arises from an outside option channel that is not simply explained by similar occupations experiencing correlated shocks.

Employment effects of indirect labor demand shocks. The results from applying a similar empirical approach to estimating the effect of indirect import shock exposure on employment are shown in columns (4)-(6) and columns (10)-(12) of table 19. The estimated negative effect from *direct* import shocks on employment is highly robust to controlling for indirect labor supply effects as a result of workers leaving affected occupations that are related. The indirect labor supply effect has the expected positive and significant effect when

[38]Unlike in our national wage shock regressions in section 4, we are unable to have occupation-by-MSA fixed effects since we are estimating cross-sectionally, using the changes over the period 2000-2011.

we control for MSA and occupation group fixed effects, but in our preferred specification in columns (6) and (12) only the OLS estimate - not the IV one - is statistically significant. However, as we will see in the task heterogeneity section below, when we allow for variation in the effect by task content of the occupation, we obtain the expected positive and significant effect on employment from labor supply spillovers.

These results are exactly in line with ex ante expectations of how spillovers between related labor markets should operate: a given occupation will receive an increase in labor supply when *other* occupations that are part of its labor market experience negative shocks through imports, even though a *direct* import shock has an unambiguously negative effect on employment.

This means that, while individual occupations that are impacted see reductions in employment that are large, *some* of the workers who lose their job may find new employment in other occupations within their relevant labor market to the extent that those outside options exist locally. Thus, workers in different geographies may differ in their experience of negative labor demand shocks to the degree that their local labor market provides more opportunities in the form of alternative jobs to move to, in the case they are laid off.

This analysis suggest that estimating the relevant size of a worker's labor market is important from a policy perspective. The size of the public effort in assisting affected workers should be scaled with the size of the actual impact on the *affected* worker groups, rather than the average effect on *all* local workers in the same geography. That is, defining the entire MSA as the relevant labor market for evaluating the impact of shocks to a worker's labor market is likely to lead to misleading results if the impact is more narrow and concentrated in particular occupations, and if this differs across geographies. This supports our earlier contention that trying to approximate local labor markets more accurately - for instance using the methodology we propose in this paper - enables researchers to produce more policy-relevant descriptions of labor market adjustments in response to shocks.

6.5 Task determinants of indirect import shock spillovers

In order to test whether there is heterogeneity in terms of responsiveness to indirect shocks to related occupations, we focus on a dimension of occupational wage determination that has been the subject of much debate in the recent literature on the effect of technological change on wages: the intensity of task requirements in that occupation which fall along different skill dimensions.[39] In particular, we focus on the task composites defined by Acemoglu and Autor (2011),[40] which distinguish between routine and non-routine manual and cognitive tasks, and further subdivide non-routine cognitive tasks into analytical and interpersonal dimensions. Moreover, we also consider the leadership task composite introduced earlier to capture differences in effects between occupations that have more or less management responsibility.

Task heterogeneity of indirect wage effects. Table 20 shows the results of adding the level of these task intensities as controls to the occupation-by-MSA wage IV regressions, and interacting them with the indirect import shock exposure through local outside options. The coefficients on the latter should indicate whether the indirect shock effect on wages is greater or smaller for occupations with particular task characteristics.

All specifications include MSA and occupation group fixed effects, the level of year 2000 outside options, and control for the level of the interacted task composite.

The results show that the effect of indirect shocks on wages varies substantially with the task requirements of different occupations, with both more routine manual and more routine cognitive occupations seeing a greater negative effect of outside options on their wages, albeit the effect is only statistically significant for the routine cognitive dimension. This makes intuitive sense, as the demand for routine tasks was falling over this time period, leading to the well-documented "hollowing-out" of the middle of the wage distribution(Autor et al., 2003; Acemoglu and Autor, 2011).

A different way of framing this result is to note that in Nash bargaining models, workers with less bargaining power will have wages that are *more* sensitive to changes in their outside options. As routine task jobs were declining over this time period, routine task workers should have had low bargaining power, which explains the finding of a greater sensitivity to shocks to their outside options.

Column (6) shows that jobs with greater leadership responsibilities were associated with a less negative effect of import shocks to outside options, albeit this effect is not statistically significant.

Task heterogeneity of indirect employment effects. When we consider the heterogeneity of indirect effects by occupational task requirement on employment, we obtain the results shown in table 21. The first thing to note is that once we allow for variation along different

[39]See, e.g. Autor et al. (2013); Acemoglu et al. (2016); Deming (2017).

[40]These represent an update and expansion of similar concepts introduced by Autor et al. (2003).

skilled task dimensions, the indirect labor supply effect is always positive and almost always statistically significant -in line with the theoretical prediction of greater labor supply as a result of spillovers from other affected occupations.

Considering the interaction terms, we find that workers from routine occupations - in columns (3) and (4) - are significantly more likely to spill over into related occupations and find employment there - while non-routine task workers - columns (1), (2), and (5) - and those in leadership roles - column (6) - are less likely to do so.

This makes intuitive sense, as workers in occupations that have high task intensity in these dimensions are more likely to be higher-skilled workers who are more likely to be retained in response to negative shocks and are also more likely to have built up firm- and occupation-specific capital that is not transferable.

In summary, we find significant variation in the effects of indirect shocks on wages and employment along different occupational task dimensions that align closely with overall trends towards lower bargaining power for workers in routine jobs during this time period. Moreover, these results highlight that a probabilistic labor market definition enables us to explore important channels of bargaining power and labor supply spillovers that drive the transmission of shocks between occupations.

7 Conclusion

In this paper we have tried to show that labor markets for workers should be defined to approximate the actual realm of jobs that are available to them. Using conventional proxies for labor markets, such as geographies, current industries, or current occupations, fails to take into account worker mobility.

We suggest one feasible empirical approach to taking worker mobility into account by constructing a probabilistic occupational mobility matrix and implement it for the U.S. by making use of a large new data set of U.S. worker resumes.

Applying this probabilistic definition of outside options to U.S. data, we show that workers differ substantially 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.

Moreover, we show the relevance of this approach to defining labor markets by applying it to two recent debates in labor economics. First, we show that our definition of outside options enables a more nuanced view of the degree to which labor market concentration is associated with lower wages: we find that controlling for outside-occupation job options reduces the magnitude of the estimated relationship between local labor market concentration and wages, and that the relationship is much stronger for occupations with low outward occupational mobility (which could be thought of as better definitions of workers' true labor markets).

Second, we explore the effect of labor demand shocks in the form of Chinese import shocks on occupational wages and employment. We find that labor demand shocks have heterogeneous effects on occupations within a given geography. At the same time, labor demand shocks not only have a direct effect on workers through the local industries in which their local occupation is employed, but also affect workers indirectly through shocks to other occupations that are part of their probabilistic labor market. Moreover, workers are more or less exposed to such shocks to their outside options depending on their skill level.

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

On the other hand, worker exposure to shocks can vary substantially depending on their local circumstances and as researchers we should try hard to figure out in which ways these differences exacerbate or mitigate the effects of other cleavages and disadvantages among workers. 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.

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): "Import competition and the great US employment sag of the 2000s," *Journal of Labor Economics*, 34, S141–S198.

Ashenfelter, O., H. Farber, and M. Ransom (2013): "Modern Models of Monopsony in Labor Markets: A Brief Survey," *IZA Discussion Paper*.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 103, 2121–68.

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.

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.

Garin, A. and F. Silvério (2018): "How Responsive are Wages to Demand within the Firm? Evidence from Idiosyncratic Export Demand Shocks," Working paper.

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. and C. Macaluso (2018): "Labor market concentration, skills, and wages across U.S. local labor markets," *Working Paper*.

Hummels, D., R. Jørgensen, J. Munch, and C. Xiang (2014): "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 104, 1597–1629.

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.

Kaplow, L. (2015): "Market Definition, Market Power," *International Journal of Industrial Orgniazation*, 10.

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

Macaluso, C. (2017): "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.

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*.

Yagan, D. (forthcoming): "Employment Hysteresis from the Great Recession," *Journal of Political Economy*.

8 Figures

Figure 1: Distribution of the 1-year horizon probability that a worker will be working, but no longer in their current occupation, calculated from Burning Glass Technology resume data for 2002-2015 period. Histogram shows 786 occupations, with dashed line indicating the sample mean.

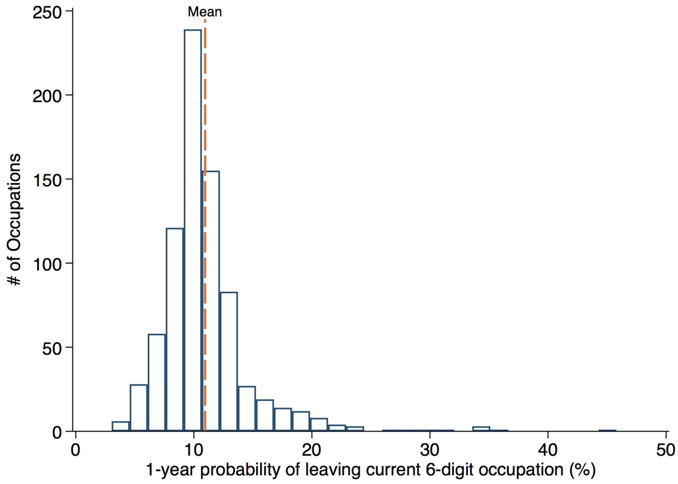

Figure 2: [**NOTE: GRAPH IS NOT UP-TO-DATE - DATA ONLY INDICATIVE**] Occupational transition matrix showing transition probability between 6-digit SOC occupations conditional on leaving the initial occupation. Data computed from Burning Glass Technology resume data set for 2002-2015.

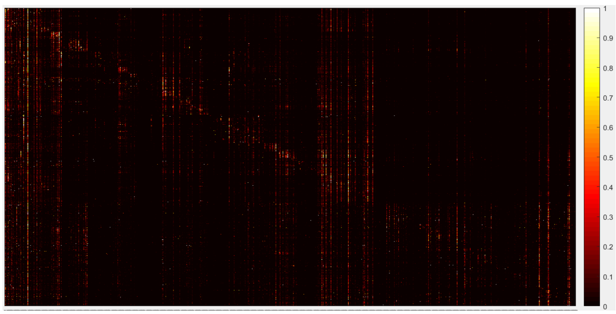

Figure 3: Coefficients from regression of 2002-2015 average probability of moving into another occupation (conditional on any job move) on absolute difference in stated skills. All regressions also include a constant, avg. hourly wage differences, and origin occupation fixed effects. Standard errors are clustered at the origin occupation level.

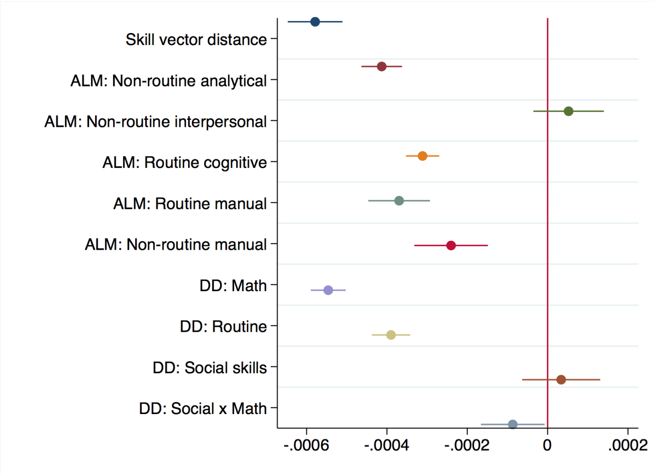

Figure 4: Coefficients from regression of 2002-2015 average probability of moving between two occupations (conditional on any job move) on stated characteristic differences between target and origin occupation. All regressions also include a constant, and origin occupation fixed effects. Standard errors are clustered at the origin occupation level. Variable labels are explained in footnote 16.

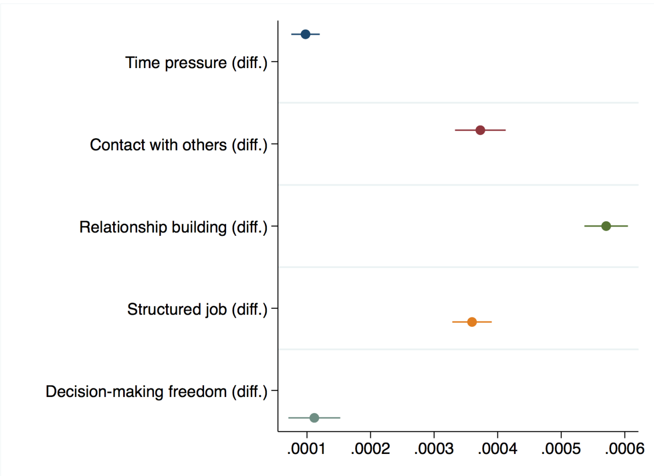

Figure 5: Coefficients from regression of 2002-2015 average probability of moving between two occupations (conditional on a move) on stated characteristic *absolute* differences between target and origin occupation. All regressions also include a constant, and origin occupation fixed effects. Standard errors are clustered at the origin occupation level. Variable labels are explained in footnote 16.

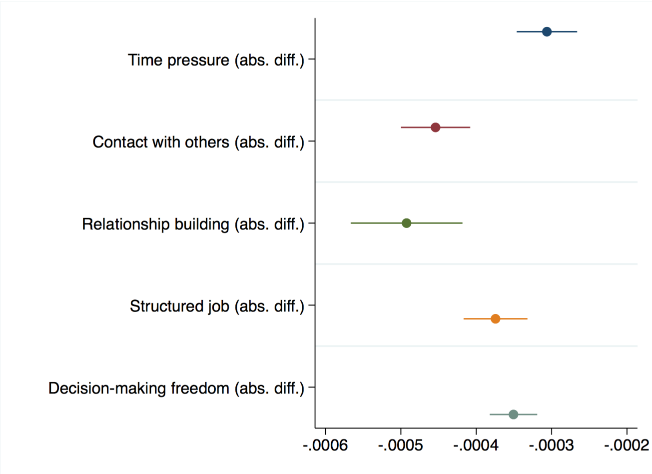

Figure 6: Coefficients from regression of 2002-2015 average probability of moving between two occupations (conditional on a move) on *relative* differences between target and origin occupation in each leadership characteristic or composite. All regressions also include a constant, avg. hourly wage differences, and origin occupation fixed effects. Standard errors are clustered at the origin occupation level. Variable labels are explained in footnote 18.

9 Tables

Table 2: Number of observations in the BGT occupational mobility data, by occupation (2002-2015)

Percentile	1	5	10	25	50	75	90	95	99
Obs.	109	742	1,215	4,302	19,952	110,574	456,745	825,879	2,842,297

This table shows summary statistics of number of observations by occupation in our occupational mobility data set, which we calculate from the Burning Glass Technology resume data. An observation in our occupational mobility 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).

Table 3: Share leaving job and occupation, by occupation (2002-2015)

	Share leaving job	Share leaving occupation (6d) conditional on leaving job	Share leaving occupation
Avg. (emp.weight)	0.49	0.11	0.22
Average (simple)	0.51	0.11	0.24
P1	0.34	0.047	0.082
P5	0.38	0.062	0.11
P10	0.41	0.074	0.13
P25	0.44	0.90	0.17
Median	0.49	0.10	0.22
P75	0.56	0.12	0.26
P90	0.63	0.15	0.34
P95	0.68	0.18	0.42
P99	0.76	0.29	0.80

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

Table 4: Share of outward occupational moves which cross SOC 2d boundary, by occupation (2002-2015)

Percentile	1	5	10	25	50	75	90	95	99
	0.52	0.60	0.65	0.74	0.82	0.88	0.92	0.93	0.96

This table shows summary statistics of the share of all outward occupational moves which are across SOC 2-digit boundaries, by occupation. This implies that for the median occupation, 82% of all outward occupational moves are to a different SOC 2-digit occupation.

Table 5: Twenty large occupations with least mobility ('stickiest') and most mobility ('least sticky')

Initial occupation	Leave share	Employment (2017)	Obs. (in BGT data)	Modal new occupation
Dental hygienists	.068	211,600	17,458	Dental assistants
Nurse practitioners	.076	166,280	57,830	Registered nurses
Pharmacists	.082	309,330	121,887	Medical and health services managers
Physical therapists	.09	225,420	44,314	Medical and health services managers
Firefighters	.093	319,860	60,039	Emergency medical technicians and paramedics
Graphic designers	.098	217,170	439,953	Art directors
Self-enrichment education teachers	.098	238,710	169,369	Teachers and instructors, all other
Postsecondary teachers, all other	.1	189,270	825,879	Managers, all other
Lawyers	.11	628,370	667,960	General and operations managers
Licensed practical and licensed vocational nurses	.12	702,700	254,787	Registered nurses
Emergency medical technicians and paramedics	.12	251,860	111,180	Managers, all other
Fitness trainers and aerobics instructors	.12	280,080	281,903	Managers, all other
Registered nurses	.13	2,906,840	1,427,102	Medical and health services managers
Heavy and tractor-trailer truck drivers	.13	1,748,140	2,174,486	Managers, all other
Chief executives	.13	210,160	1,425,400	General and operations managers
Radiologic technologists	.13	201,200	80,347	Magnetic resonance imaging technologists
Hairdressers, hairstylists, and cosmetologists	.14	351,910	107,167	Managers, all other
Health specialties teachers, postsecondary	.14	194,610	41,963	Medical and health services managers
Education administrators, elementary and secondary school	.14	250,280	394,459	General and operations managers
Teachers and instructors, all other	.14	611,310	1,009,894	Postsecondary teachers, all other
...				
Stock clerks and order fillers	.28	2,046,040	597,137	Laborers and freight, stock, and material movers, hand
Hotel, motel, and resort desk clerks	.28	253,540	663,574	Customer service representatives
Combined food preparation and serving workers, including fast food	.28	3,576,220	661,252	Retail salespersons
Helpers–production workers	.28	402,140	112,759	Production workers, all other
Packaging and filling machine operators and tenders	.28	392,910	36,793	Laborers and freight, stock, and material movers, hand
Dishwashers	.28	503,540	72,610	Cooks, restaurant
Bill and account collectors	.28	271,700	310,951	Customer service representatives
Food batchmakers	.28	151,950	12,729	Industrial production managers
Order clerks	.28	169,120	46,880	Customer service representatives
Cooks, institution and cafeteria	.29	404,120	5,174	Cooks, restaurant
Loan interviewers and clerks	.29	227,430	234,933	Loan officers
Cement masons and concrete finishers	.29	178,640	9,555	Managers, all other
Cooks, short order	.29	174,230	39,906	Cooks, restaurant
Counter and rental clerks	.29	445,530	41,340	Customer service representatives
Tellers	.3	491,150	468,829	Customer service representatives
Hosts and hostesses, restaurant, lounge, and coffee shop	.3	414,540	159,098	Waiters and waitresses
Counter attendants, cafeteria, food concession, and coffee shop	.31	476,940	118,131	Retail salespersons
Telemarketers	.31	189,670	47,409	Customer service representatives
Meat, poultry, and fish cutters and trimmers	.41	153,280	2,383	Heavy and tractor-trailer truck drivers
Food servers, nonrestaurant	.43	264,630	13,199	Waiters and waitresses

This table shows the twenty large occupations with the lowest and the highest 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 (the 'stickiest' occupations), 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 6: 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 is the share of those in the initial occupation who leave their occupation who move to the new occupation.

Table 7: Regression of wage on outside-occupation option index

	(1)	(2)	(3)	(4)
oo^{occs}	0.404***	0.072***	0.077***	0.079***
	(0.008)	(0.002)	(0.002)	(0.003)
Fixed effects	Year	Occ-year CBSA	CBSA-year Occ	Occ-year Occ-CBSA
Observations	1,944,477	1,944,477	1,944,477	1,944,477

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 1999–2016 inclusive.

Table 8: Two-stage least squares regression of wage on instrumented outside-occupation option index

	(1)	(2)	(3)	(4)
oo^{occs}, instrumented	0.456***	0.020***	0.033***	0.019***
	(0.009)	(0.003)	(0.002)	(0.006)
Fixed effects	Year	Occ-year CBSA	CBSA-year Occ	Occ-year Occ-CBSA
Observations	1,944,477	1,944,477	1,944,477	1,944,477

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$. Units of observation are 6 digit SOC by CBSA 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 9: Occupational wage terciles, by 2016 national hourly wage

	Low wage	Medium wage	High wage
Minimum	9.84	18.11	27.85
Maximum	18.09	27.84	129.62

This table splits occupations into terciles by the 2016 average national hourly wage, and shows the cut-offs for those terciles. These are used in the regressions of wages on outside-occupation options by wage tercile in Tables 10 and 11.

Table 10: Regression of wage on outside-occupation options, by wage tercile

	Low wage	Medium wage	High wage
oo^{occs}	0.062***	0.076***	0.095***
	(0.002)	(0.003)	(0.003)
Fixed effects	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year
Observations	646,750	654,432	643,164

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 1999–2016 inclusive. Occupations are split into low-wage, medium-wage and high-wage terciles by their 2016 national average hourly wage, with cut-offs shown in Table 9.

Table 11: Regression of wage on instrumented outside-occupation options, by wage tercile

	Low wage	Medium wage	High wage
oo^{occs}, instrumented	0.001	0.025***	0.028***
	(0.008)	(0.006)	(0.006)
Fixed effects	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year
Observations	646,750	654,432	643,162

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 1999–2016 inclusive. Occupations are split into low-wage, medium-wage and high-wage terciles by their 2016 national average hourly wage, with cut-offs shown in Table 9. 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 12: Regression of wage on outside-occupation options, by tercile of task intensity

	Panel A: non-routine cognitive analytical		
	Low	Medium	High
oo^{occs}	0.063***	0.081***	0.090***
	(0.002)	(0.003)	(0.003)
Observations	654,278	642,173	648,028
	Panel B: non-routine cognitive interpersonal		
	Low	Medium	High
oo^{occs}	0.071***	0.083***	0.083***
	(0.003)	(0.003)	(0.003)
Observations	653987	640349	650143
	Panel C: routine cognitive		
	Low	Medium	High
oo^{occs}	0.086***	0.080***	0.072***
	(0.003)	(0.003)	(0.003)
Observations	650278	647696	646505
	Panel D: routine manual		
	Low	Medium	High
oo^{occs}	0.087***	0.082***	0.068***
	(0.003)	(0.003)	(0.003)
Observations	654595	645887	643997
	Panel E: non-routine manual		
	Low	Medium	High
oo^{occs}	0.084***	0.084***	0.071***
	(0.003)	(0.003)	(0.003)
Observations	654564	646869	643046
	Panel F: leadership		
	Low	Medium	High
oo^{occs}	0.076***	0.079***	0.082***
	(0.003)	(0.003)	(0.003)
Observations	634772	635120	635467
Fixed effects	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,*** $p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 1999–2016 inclusive. Occupations are split into terciles by their intensity in various tasks. All task measures except Leadership are taken from Autor et al. (2003); leadership is constructed from O*NET data as detailed in section 2.

Table 13: Regression of wage on instrumented outside-occupation options, by tercile of task intensity

	Low	Medium	High
Panel A: non-routine cognitive analytical			
oo^{occs}	0.008	0.023***	0.025***
	(0.007)	(0.007)	(0.006)
Observations	654278	642173	648026
Panel B: non-routine cognitive interpersonal			
oo^{occs}	0.019***	0.021***	0.018***
	(0.007)	(0.006)	(0.006)
Observations	653987	640349	650141
Panel C: routine cognitive			
oo^{occs}	0.009	0.019***	0.027***
	(0.006)	(0.007)	(0.006)
Observations	650276	647696	646505
Panel D: routine manual			
oo^{occs}	0.022***	0.030***	0.007
	(0.006)	(0.006)	(0.007)
Observations	654595	645885	643997
Panel E: non-routine manual			
oo^{occs}	0.028***	0.027***	0.004
	(0.006)	(0.006)	(0.007)
Observations	654564	646867	643046
Panel F: leadership			
oo^{occs}	0.011*	0.029***	0.019***
	(0.007)	(0.006)	(0.006)
Observations	634772	635120	635465
Fixed effects	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year	Occ-CBSA, Occ-year

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $*p < .1$,$**p < .05$,$***\ p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 1999–2016 inclusive. Occupations are split into terciles by their intensity in various tasks. All task measures except Leadership are taken from Autor et al. (2003); leadership is constructed from O*NET data as detailed in section 2. 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 14: Regression of wage on Vacancy HHI, by quartile of occupation leave share

	Dependent variable: Log wage				
		By quartile of leave share			
	Baseline	Q1	Q2	Q3	Q4
Vacancy HHI	-0.019***	-0.025***	-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
	CBSA	CBSA	CBSA	CBSA	CBSA
Observations	420,546	111,737	108,425	107,528	92,577

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $*p < .1, ^{**}p < .05, ^{***} p < .01$. Units of observation are 6 digit SOC by CBSA 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).

Table 15: Regression of wage on Vacancy HHI and outside occupation and city options

	(1)	(2)	(3)
Log Vacancy HHI	-0.019***	-0.009***	-0.018***
	(0.001)	(0.001)	(0.001)
oo^{occs}	0.073***		
		(0.002)	
oo^{occs}, instrumented	0.012***		
			(0.002)
Fixed effects	Occ-Year	Occ-Year	Occ-Year
	CBSA	CBSA	CBSA
Observations	420,443	420,223	420,223

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

Heteroskedasticity-robust standard errors clustered at the CBSA level shown in parentheses: $*p < .1, ^{**}p < .05, ^{***} p < .01$. Units of observation are 6 digit SOC by CBSA by year, for all observations with available data over 2013–2016 inclusive.

Table 16: Employment Effect of China Shock at the Geography Level

Geography:	MSA						CZ			
Time period:	2000-2011			2000-2007	1991-2011	1991-2007	2000-2011	2000-2007	1991-2011	1991-2007
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Import shock exposure	-1.26**	-1.52**	-2.21***	-2.54***	-1.51**	-1.79***	-2.17***	-2.32***	-1.70**	-1.89***
	(0.49)	(0.78)	(0.69)	(0.67)	(0.59)	(0.49)	(0.70)	(0.63)	(0.78)	(0.65)
Period FEs					Yes	Yes			Yes	Yes
Baseline Mfg Emp Share	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Census Div. FEs	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	545	545	545	545	1090	1090	722	722	1444	1444

Heteroskedasticity-robust standard errors clustered at the geography (CBSA or CZ) level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$. The dependent variable is 100 times the annualized change in the ratio of total employment to working-age population over the time period. Period FEs refer to fixed effects for the 1991-1999 and 1999-end periods for any periods beginning in 1991. US import shocks are instrumented using the evolution of Chinese imports in other developed countries.

Table 17: Wage Effect of China Shock at the Geography Level

Dependent var.:	MSA			
Time period:	2000-2011			2000-2007
	(1)	(2)	(3)	(4)
Import shock exposure	0.42	4.84***	2.85***	3.26***
	(1.22)	(1.23)	(1.10)	(1.11)
Baseline Mfg Emp Share	No	Yes	Yes	Yes
Census Div. FEs	No	No	Yes	Yes
N	319	319	319	318

Heteroskedasticity-robust standard errors clustered at the geography level shown in parentheses: $^*p < .1$,$^{**}p < .05$,*** $p < .01$. The dependent variable is the annualized growth rate, in percent, in the average hourly wage at the MSA level over the time period, measured across all OES-reported occupations. Period FEs refer to fixed effects for the 1991-1999 and 1999-end periods for any periods beginning in 1991, while the periods 2000-2007 and 2000-2011 only contain a constant. US import shocks are instrumented using the evolution of Chinese imports in other developed countries.

Table 18: Wage Effect of China Shock at the MSA-by-Occupation Level

Dependent var.:	Δ MSA$\times$Occ. Hourly Wage 2000-2011											
Estimation:	OLS						GMM-IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Direct import shock	-0.13***	-0.07*	0.15***	0.16***	0.20***	0.15***	-0.14***	-0.08**	0.26***	0.26***	0.30***	-0.04
	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.06)	(0.05)	(0.05)
Indirect outside option effect				-2.25***	-0.51***	-0.15				-3.58***	-0.62**	-1.99***
				(0.63)	(0.17)	(0.13)				(0.66)	(0.28)	(0.52)
MSA fixed effect		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
Occ. group fixed effect			✓		✓	✓			✓		✓	✓
Outside options in 2000						✓						✓
Observations	73904	73904	73904	73885	73885	73885	73904	73904	73904	73885	73885	73885

Heteroskedasticity-robust standard errors clustered at the geography level shown in parentheses: $^*p < .1$,$^{**}p < .05$,*** $p < .01$. The dependent variable is the annualized growth rate, in percent, in the average hourly wage at the occupation-by-MSA level over the time period. For the IV estimations, US import shocks are instrumented using the evolution of Chinese imports in other developed countries. GMM-IV estimation is implemented using feasible two-stage GMM. Direct import shocks for occupations are estimated as the weighted average of the industry-level import shocks from Acemoglu et al. (2016), with weights based on the local share of occupation workers in different industries based on national occupation shares by industry and local industry composition of employment. The indirect outside options effect for a local occupation x is calculated as the weighted average direct shock to *other* local occupations, with the weights being the product of the average share of workers from x taking a job in that other occupation within a year (conditional on working in any other job, and computed as the average over 2002-2015) and the predicted prevalence of those other occupations, relative to their national employment share. Baseline period outside options are computed as the indirect outside options effect, if all the direct shocks in the MSA were equal to 1 - so it captures the availability of other occupations to move to from occupation x, without accounting for differences in import competition.

Table 19: Employment Effect of China Shock at the Occupation Level

Dependent var.:	Δ MSA$\times$Occ. Employment 2000-2011											
Estimation:	OLS						GMM-IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Direct import shock	-3.32***	-3.58***	-1.85***	-3.56***	-1.90***	-1.86***	-4.86***	-4.93***	-2.85***	-4.91***	-2.91***	-4.99***
	(0.56)	(0.53)	(0.34)	(0.53)	(0.34)	(0.33)	(0.34)	(0.32)	(0.29)	(0.32)	(0.28)	(0.32)
Indirect labor supply effect				-0.11*	0.17***	0.19***				-0.14	0.23**	-0.17
				(0.07)	(0.05)	(0.05)				(0.10)	(0.11)	(0.10)
MSA fixed effect		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
Occ. group fixed effect			✓		✓	✓			✓		✓	✓
Outside options in 2000						✓						✓
Observations	73904	73904	73904	73904	73904	73885	73904	73904	73904	73904	73904	73885

Heteroskedasticity-robust standard errors clustered at the geography level shown in parentheses: *$p < .1$,**$p < .05$,*** $p < .01$. The dependent variable is the annualized growth rate, in percent, in each occupation's local MSA employment share over the time period. For the IV estimations, US import shocks are instrumented using the evolution of Chinese imports in other developed countries. GMM-IV estimation is implemented using feasible two-stage GMM. Direct import shocks for occupations are estimated as the weighted average of the industry-level import shocks from Acemoglu et al. (2016), with weights based on the local share of occupation workers in different industries based on national occupation shares by industry and local industry composition of employment. The indirect outside options effect for a local occupation x is calculated as the weighted average direct shock to *other* local occupations, with the weights being the product of the average share of workers from x taking a job in that other occupation within a year (conditional on working in any other job, and computed as the average over 2002-2015) and the predicted prevalence of those other occupations, relative to their national employment share. Baseline period outside options are computed as the indirect outside options effect, if all the direct shocks in the MSA were equal to 1 - so it captures the availability of other occupations to move to from occupation x, without accounting for differences in import competition.

Table 20: Occupational Wage Effect of China Shock by Task Intensity

Dependent var.:	Δ MSA×Occ. Hourly Wage 2000-2011					
Estimation:	GMM-IV					
	(1)	(2)	(3)	(4)	(5)	(6)
Direct import shock	0.24***	0.24***	0.22***	0.22***	0.19***	0.25***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Indirect outside option effect	−0.07	0.06	0.18	0.22	0.25	−0.03
	(0.28)	(0.28)	(0.28)	(0.27)	(0.26)	(0.28)
Indirect × NR Cogn. Analyt.	−0.20					
	(0.22)					
Indirect × NR Cogn. Interpers.		−0.03				
		(0.15)				
Indirect × Routine Cognitive			−0.25**			
			(0.12)			
Indirect × Routine Manual				−0.13		
				(0.17)		
Indirect × NR Manual					−0.06	
					(0.22)	
Indirect × Leadership						0.16
						(0.15)
Skill intensity (level)	✓	✓	✓	✓	✓	✓
MSA fixed effect	✓	✓	✓	✓	✓	✓
Occ. group fixed effect	✓	✓	✓	✓	✓	✓
Outside options in 2000	✓	✓	✓	✓	✓	✓
Observations	73,785	73,785	73,785	73,785	73,785	73,785

Heteroskedasticity-robust standard errors clustered at the geography level shown in parentheses: $^{*}p < .1$,$^{**}p < .05$,*** $p < .01$. The dependent variable is the annualized growth rate, in percent, in the average hourly wage at the occupation-by-MSA level over the time period. For the IV estimations, US import shocks are instrumented using the evolution of Chinese imports in other developed countries. GMM-IV estimation is implemented using feasible two-stage GMM. Direct import shocks for occupations are estimated as the weighted average of the industry-level import shocks from Acemoglu et al. (2016), with weights based on the local share of occupation workers in different industries based on national occupation shares by industry and local industry composition of employment. The indirect outside options effect for a local occupation x is calculated as the weighted average direct shock to *other* local occupations, with the weights being the product of the average share of workers from x taking a job in that other occupation within a year (conditional on working in any other job, and computed as the average over 2002-2015) and the predicted prevalence of those other occupations, relative to their national employment share. Baseline period outside options are computed as the indirect outside options effect, if all the direct shocks in the MSA were equal to 1 - so it captures the availability of other occupations to move to from occupation x, without accounting for differences in import competition. Task composites for different skills are computed from O*Net task intensity scores for tasks related to the composite theme, and are standardized Z-scores with higher values corresponding to higher relative task intensity - see the Data Appendix for details.

Table 21: Occupational Employment Effect of China Shock by Task Intensity

Dependent var.:	Δ MSA×Occ. Employment 2000-2011					
Estimation:	GMM-IV					
	(1)	(2)	(3)	(4)	(5)	(6)
Direct import shock	−2.78***	−2.75***	−2.85***	−2.73***	−2.87***	−2.74***
	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)
Indirect labor supply effect	0.20	0.22**	0.21*	0.25**	0.33**	0.22*
	(0.12)	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)
Indirect × NR Cogn. Analyt.	−0.20**					
	(0.09)					
Indirect × NR Cogn. Interpers.		−0.11**				
		(0.05)				
Indirect × Routine Cognitive			0.10**			
			(0.05)			
Indirect × Routine Manual				0.12**		
				(0.06)		
Indirect × NR Manual					−0.16***	
					(0.05)	
Indirect × Leadership						−0.09**
						(0.04)
Skill intensity (level)	✓	✓	✓	✓	✓	✓
MSA fixed effect	✓	✓	✓	✓	✓	✓
Occ. group fixed effect	✓	✓	✓	✓	✓	✓
Outside options in 2000	✓	✓	✓	✓	✓	✓
Observations	73,785	73,785	73,785	73,785	73,785	73,785

Heteroskedasticity-robust standard errors clustered at the geography level shown in parentheses: $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. The dependent variable is the annualized growth rate, in percent, in each occupation's local MSA employment share over the time period. For the IV estimations, US import shocks are instrumented using the evolution of Chinese imports in other developed countries. GMM-IV estimation is implemented using feasible two-stage GMM. Direct import shocks for occupations are estimated as the weighted average of the industry-level import shocks from Acemoglu et al. (2016), with weights based on the local share of occupation workers in different industries based on national occupation shares by industry and local industry composition of employment. The indirect outside options effect for a local occupation x is calculated as the weighted average direct shock to *other* local occupations, with the weights being the product of the average share of workers from x taking a job in that other occupation within a year (conditional on working in any other job, and computed as the average over 2002-2015) and the predicted prevalence of those other occupations, relative to their national employment share. Baseline period outside options are computed as the indirect outside options effect, if all the direct shocks in the MSA were equal to 1 - so it captures the availability of other occupations to move to from occupation x, without accounting for differences in import competition. Skill composites are computed from O*Net task intensity scores for tasks related to the composite theme, and are standardized Z-scores with higher values corresponding to higher relative task intensity - see the Data Appendix for details.

10 Data Appendix

10.1 Burning Glass Technologies Resume Data

This section of the 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 several 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 dataset.

10.1.1 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 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 first convert each job (i.e. occupation worked in) into separate observations for each year that the job was held, which are then matched to all other job-year observations on the same resume. We retain all matches that are in sequential years - either in the same job or in different jobs. 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 job holdings of the form shown in table 22.

Table 22: 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 observe 80.2 million jobs, and 178.5 million instances of sequential occupation occurrences at a 1-year horizon, which we use to construct our measures of occupational mobility. 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.

When computing the occupational mobility matrix, we convert the counts of occurrences of 1-year horizon occupations for current workers in occupation o into mobility probabilities by dividing them by the total count of workers in an occupation o that are still in the sample in the next year. We compute those probabilities separately for different age categories and aggregate them while reweighting based on the relative prevalence of those ages in the labor force, according to the BLS, relative to the prevalence in our sample. Thus, the aggregate occupational mobility matrix has been reweighted to correspond to the age distribution in the labor force, eliminating any potential bias from the issue of a skewed age distribution of our sample, which we discuss below.

10.1.2 Summary statistics

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 23 shows more information on the distribution of job incidences and job durations on our resumes.

Table 23: 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 sequential job observations, 88% are on resumes were 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 the sequential job 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 the sequential job data points are on resumes containing some information about education. The breakdown of education in our data for these data points is that 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 7). 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, howerver, 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 8 for the full distribution. One would expect a sample drawn from online resume submissions and consisting of 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 job sequences in the data set, 169.6 million were able to be coded into non-military 6-digit SOC occupations by the BGT parser. 833 of the 840 6-digit SOC occupations are present, some with few observations and some with very many. Ranking occupations by the number of job sequences we observe starting in each one, the 10th percentile is 1,226 observations, 25th percentile is 4,173, the median is 20,526, 75th percentile is 117,538, and the 90th percentile is 495,699. We observe 216 occupations with more than 100,000 job sequences, 83 occupations with more than 500,000 job observations, and 19 occupations with more than 2 million job sequence observations. [41]

Figure 9 compares the prevalence of occupations at the 2-digit SOC level in our BGT job sequence data to the share of employment in that occupation group in the labor force according to the BLS in 2017. As the figure shows, at a 2-digit SOC level, management occupations, business and finance, and computer-related occupations are substantially overweight in the BGT data relative to the labor force overall, while manual occupations, healthcare and education are substantially underrepresented

Location: Since not all workers list the location where they work at their current job, we assign workers a location based on the address they list at the top of their resume. 115.4 million of our job sequences 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 10, 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

[41] 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.

workers from the West region.

10.1.3 Advantages over other datasets

As a large, nationally-representative sample with information about labor market history over the past year, the CPS March Supplement 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 [42]. 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.

[42] 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.

10.1.4 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 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.

10.2 Occupational shares by industry

For our analysis of the effects of local labor demand shocks on wages in Section 6, we need to translate the demand shocks caused by exposure to Chinese import competition from the industry level to the occupation level. The allocation of occupations to industries trades off the loss of precision from smaller coverage in the occupational classifications in the years before 2004 and the need to crosswalk any OES files at the industry level from NAICS to the SIC classification after 2001.

As a result, we use all the OES files under the SOC 2000 classification that still use the SIC industry classification, which is true for the years 1999-2001, which also conveniently coincide with the beginning of the period when the China shock is expected to take effect, which starts in 2000.

For each of these files, we extract the percent of each 3-digit SIC's workers that work in certain 6-digit SOC 2000 occupations. Then, we average these percentages over the 1999-2001, where available, and keep only the average.

Lastly, we use a crosswalk to allocate data under the SOC 2000 occupation codes to the SOC 2010 classification.

11 Appendix Figures

Figure 7: 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

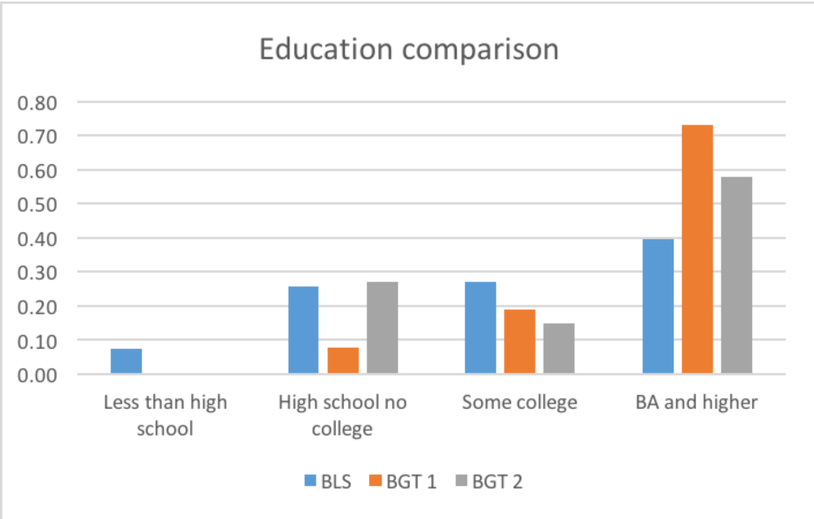

Figure 8: Comparison of distribution of age in the labor force, according to 2018 BLS data, to distribution of imputed worker ages in BGT job sequence data.

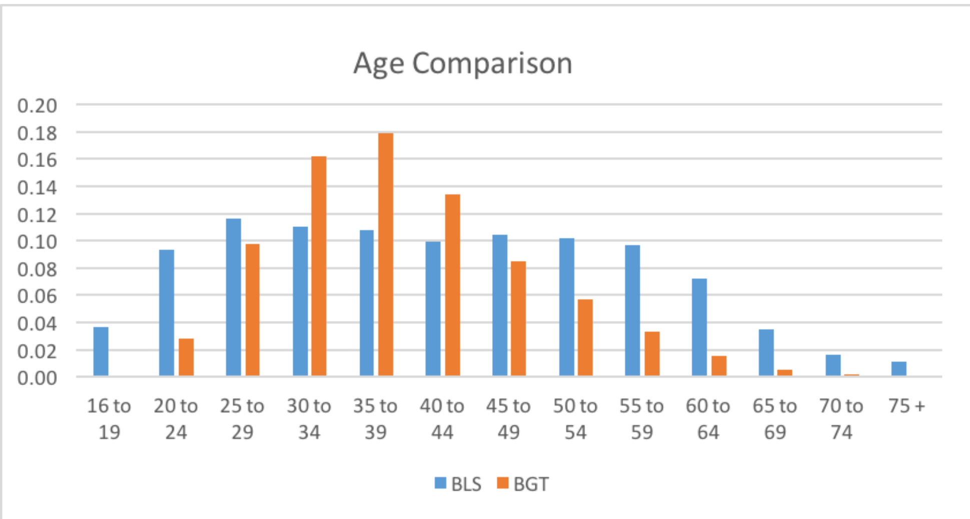

Figure 9: Comparison of distribution of 2-digit SOC occupations in the labor force, according to 2017 BLS data, to distribution of occupations in BGT job sequence data.

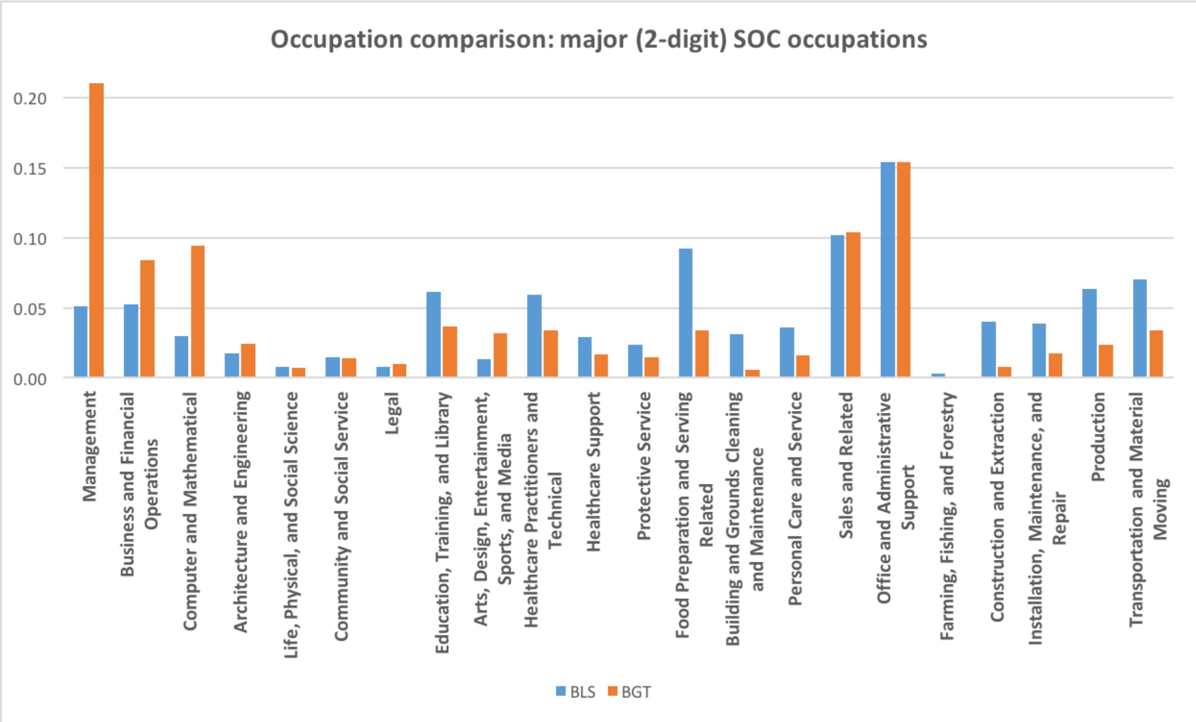

Figure 10: Comparison of distribution of employment by U.S. state, according to 2017 BLS data, to distribution of resume addresses in BGT job sequence data. Graph shows share of total in each state.