How Local Are U.S. Labor Markets?: Using an Assignment Model to Forecast the Geographic and Skill Incidence of Local Labor Demand Shocks

Richard K. Mansfield*
University of Colorado-Boulder

August 19, 2021

Abstract

This paper examines how spatial frictions that differ among heterogeneous workers and establishments shape the geographic and skill incidence of alternative local labor demand shocks, with implications for the appropriate level of government at which to fund local economic initiatives. LEHD data featuring millions of job transitions facilitate estimation of a rich two-sided labor market assignment model. The model generates simulated forecasts of many alternative local demand shocks featuring different establishment compositions and local areas. Workers from the targeted areas (generally containing around 100,000 workers) receive only 9.9% (5.6%) of nationwide welfare (employment) short-run gains, with at least 35.7% (60.5%) accruing to out-of-state workers, despite much larger per-worker impacts for the closest workers. Local incidence by prior earnings category is very sensitive to shock composition, but different shocks produce similar skill incidence farther from the shock. The results suggest that reduced-form approaches using distant locations as controls can produce accurate estimates of local shock impacts on local workers, but that the distribution of local impacts badly approximates shocks' statewide or national incidence.

*This paper has benefitted greatly from an extraordinary effort by Henry Hyatt, as well as from considerable early assistance from Evan Buntrock, John Abowd, and Erika McEntarfer and valuable comments from Jeronimo Carballo, Terra McKinnish, Brian Cadena, Francisca Antman, Miles Kimball, Martin Boileau, Alessandro Peri, Murat Iyigun, Stephan Heblich, Yanos Zylberberg, and participants at the NBER Trade and Geography Conference, the SOLE/EALE Joint Conference, the IZA World of Labor Conference, the American and European Urban Economics Association Conferences, and seminars at the University of Illinois, Princeton University, Colorado State University, the University of Bristol, and the University of Colorado-Boulder. This research uses data from the US Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by NSF grants SES-9978093, SES-0339191, and ITR-0427889; National Institute on Aging grant AG018854; and grants from the Alfred P. Sloan Foundation. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

Billions of dollars in local aid are spent each year by state, federal, and local agencies to support city-level or county-level economic development initiatives that seek to enhance labor market opportunities for workers who live or work within the local jurisdiction (Bartik (2004)). These often take the form of local infrastructure spending, discounted loans or subsidies aimed at startup companies, or tax breaks to lure firms to relocate. In order to determine which types of firms or projects to support, federal, state, and local policymakers must predict not only which types of workers from which locations would be directly hired by the tax-supported firms, but also how the resulting ripple effects that operate through vacancy chains and pressure on local wages would indirectly benefit both local and more distant workers. In particular, whether to fund such initiatives at the city, county, or state level depends critically on the shares of the initiative's employment and welfare incidence expected to redound to workers within the city, county, and state borders, respectively. Officials need to be able to predict which types of targeted firms and locations will yield a geographically concentrated impact in which the labor demand shock primarily trickles down to lower skill levels rather than out toward more distant locations.

While a large literature in economics seeks to evaluate the incidence of place-based labor demand policies and shocks, most reduced-form methods focus on quite local impacts. More distant towns, counties or states are either excluded from the sample or used as control groups, thereby ignoring the possibility that these more distant areas might collectively account for a sizeable share of shock incidence, even if no single area is strongly affected. Furthermore, by virtue of their focus on particular policies or shocks occurring in one or a small number of locations, these studies are usually ill-equipped to compare the incidence of shocks featuring different labor demand compositions on locations featuring different local labor supply compositions, or to examine differential skill incidence among local and less local areas (due to small samples of workers within a small radius around the shock and/or a lack of detailed data on distant locations).

The primary difficulty is that either evaluating or predicting worker-level welfare incidence across a variety of alternative local labor market shocks requires a spatial equilibrium model that accommodates ripple effects by incorporating the network of spatial linkages among overlapping local labor markets while simultaneously featuring heterogeneity in worker and firm preferences, search costs, and match productivities along a variety of observable dimensions.

Motivated by this challenge, this paper makes two central contributions. First, we adapt the marriage market assignment model of Choo and Siow (2006) to develop a theoretically-motivated empirical framework for assessing and forecasting welfare incidence across location-by-demographic group categories from labor demand shocks featuring alternative geographic and establishment type compositions. Second, after estimating the parameters of the model, we analyze a large set of model simulations that illustrate several general properties of U.S. local labor markets. These simulations create a useful national prior about which types of workers and locations are most sensitive to which types of local labor demand shocks.

We estimate the model using matched employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) database on a subset of 19 U.S. states that approved the use of their employment records. The data display three key properties that make it suitable for a rich assignment model. Namely, 1) they capture the (near) universe of job matches from the participating states, mitigating selection problems; 2) they include hundreds of millions of job matches, allowing precise estimates of the large number of parameters necessary to capture complex two-sided multidimensional sorting; and 3) workers' establishments are geocoded to the census tract level.

These properties, when combined, make it feasible to study incidence across worker types at the very local level necessary to make the estimates useful to local policymakers, while still allowing for complex spatial ties between the local area and the surrounding towns, counties, and states that make the estimates useful to state and federal policymakers. In particular, when combined with the assignment model, these data provide the necessary inputs to compute the shares of employment and welfare gains or losses from alternative local labor demand shocks that accrue to workers of particular skill levels located within particular jurisdictions both near and far from the shock.

Several key features of Choo and Siow (2006)'s (hereafter CS) version of the assignment game also facilitate these goals. First, it can accommodate multidimensional heterogeneity based on unordered categorical characteristics for agents on both sides of the matching market. This allows the model to feature arbitrary spatial links between workers and establishments in different geographic units, including geographic units of both very small and large sizes. It also permits analysis of incidence across groups defined by past income categories, age groups, or industries (or combinations thereof). Second, the assignment game requires market clearing, optimizing behavior by all agents, and explicit payoffs to each agent from each possible job match, making it well-suited for forecasting welfare effects from exogenous shocks.

Third, the key model parameters, mean relative joint surpluses among matched pairs of workers and firms belonging to observable types, can be identified from a single labor market matching of workers (with associated origin jobs) to destination positions, and act as sufficient statistics for counterfactuals that yield the resulting allocation and payoff changes for both workers and firms from any arbitrary change in the composition of labor supply and/or demand (or the surplus parameters themselves). Importantly, this sufficient statistics approach does not require the specification of a more fundamental structural model of utility, firm production, and moving costs, ensuring that none of the heterogeneity present in the matching patterns is lost in paring down to a small number of interpretable structural parameters.

While we model heterogeneity on both sides of the labor market much more richly than other structural models, we do not explicitly model the housing and product markets (though their impact may nonetheless be captured by the estimated surplus parameters through the way they affect job-to-job flows). Thus, the model predictions capture "labor-related" welfare changes induced by these shocks, and can act as complementary inputs to local policy decisions along with estimates of house and product price elasticities. For example, policymakers concerned that local initiatives creating new high-skilled positions could increase rent for low-income renters might wish to know whether their downstream earnings and employment opportunities will increase enough to compensate.

The counterfactual simulations we consider involve plant or store openings and closings that create or destroy 500 job positions in particular U.S. locations (census tracts) featuring alternative combinations of establishment size, average pay, and industry supersector.

On one hand, the simulations suggest that per-worker effects of local labor demand shocks decline rapidly with distance from the focal tract. A randomly chosen worker in the targeted tract is about 4, 23 and over 13,000 times more likely to fill one of the new vacancies than counterparts in an adjacent tract, an adjacent PUMA[1], and a non-adjacent state, respectively. Since workers who join the incoming firms are likely to already be employed, so that their transitions generate further openings for others, the predicted utility and employment gains decline more slowly with distance. Nonetheless, utility and employment gains (in parentheses) for initially local workers are 3.1 (3.1), 18.4 (11), and 2,952 (436) times as large as for workers in an adjacent tract, an adjacent PUMA, and a non-adjacent state, respectively, with expected utility gains (scaled in $ of annual earnings) of $1,181 for focal tract workers and just $0.40 for the most distant workers. Such rapid declines and tiny mean impacts for far away workers confirm that local labor markets are sufficiently isolated to allow accurate reduced-form estimates of treatment effects of local demand shocks on local workers when distant locations serve as control groups.

On the other hand, the local workforce naturally makes up a very small share of the national labor market: 73,057 tracts and 2,378 PUMAs were defined in the 2010 U.S. census, so that a single census tract generally only contains a few thousand workers, and a single PUMA contains around 100,000. Consequently, even quite disproportionate welfare gains for the most local workers cannot account for more than a tiny share of the aggregate welfare gains. Thus, reduced-form models miss a different sense in which job-related welfare gains or losses from very local shocks are widely distributed: the simulations suggest that as little as 9.9% of the utility gains and 5.7% of the net employment gains from such local stimuli accrue to workers already working in (or seeking a job from) the surrounding PUMA at the beginning of the year, while 36.6% and 60.7% of the job-related utility and employment gains accrue to workers beginning the year outside the state.

The simulations also reveal a rich tapestry of impact heterogeneity among the most local workers. The chief beneficiaries of local shocks vary widely with the establishment composition of the newly created jobs, suggesting opportunities for local officials to craft local development initiatives that target particular local subpopulations. For example, younger nonemployed workers seem to reap the largest welfare gains ($1958) from the creation of positions at large, low paying retail/wholesale firms and the smallest gains ($696) from positions at small, high paying construction firms. By contrast, older unemployed workers benefit the most ($2081) from positions at small, low paying firms in the other services sector and the least ($666) from large, high paying retail/wholesale firms, and initially employed but low-paid workers initially employed benefit the least ($670) from

[1]PUMAs or "public-use microdata areas" are mutually exclusive and exhaustive collections of contiguous counties and census tracts encompassing at least 100,000 residents. They are used in this paper as a geographic unit that captures a small city-sized population regardless of nearby population density.

the creation of positions at large, high paying manufacturing firms and the most ($1665) from positions at large, low-paying leisure/hospitality firms. Positions at small, high paying firms in the information sector generate very large gains for the already highest-paid local workers ($2707) and much smaller gains for unemployed and initially lower-paid workers, and incorporating existing estimates of job multipliers only slightly alters these findings.

We also find that the share of employment gains from stimuli that accrue to workers initially working or seeking a job within the chosen PUMA is twice as high in rural as in urban areas (7.8% vs. 3.5%), and that requiring the newly created jobs to be filled exclusively by local PUMA workers (or jobseekers) increases their share of net employment gains from under 6% to around 17%.

However, as with mean impacts, the substantial heterogeneity in local impacts presents a misleading guide to heterogeneity in impacts at the county, state or national level. Regardless of establishment composition, as the simulated shocks ripple outward, they become less skill-biased: predicted differences in welfare (or employment) gains among past earnings categories converge quickly as one considers workers at initial locations further from the site of the shock. Furthermore, focal tract characteristics that predict relatively greater benefits for local low-paid and unemployed workers from local job creation fail to predict any such redistribution at the national level. Thus, from the wider perspective of state-level funders, the local heterogeneity in incidence is of second-order importance, allowing them to focus more narrowly on job creation per dollar of funding.

To capture such multidimensional heterogeneity, the surplus parameters that determine equilibrium elasticities of substitution (and thus govern the impact of simulated shocks) exploit for identification the full sample of worker flows generated by an implicit mix of local and national shocks to both supply and demand composition, rather than exclusively using flows involving locations experiencing local labor demand shocks. To show that the model is nonetheless capable of generating accurate forecasts for moderately sized local shocks, we perform a model validation exercise in which parameters estimated using flows from prior years are used to predict the realized reallocation around 180 census tracts that experienced gains or losses of more than 100 jobs within one year between 2003 and 2012. The model predicts these out-of-sample reallocations quite well and considerably better than even relatively rich one-sided parametric models that fit firm or worker conditional choice probabilities with over 100 parameters. This validation exercise illustrates that the very large set of estimated parameters is not causing overfitting, but is instead necessary to capture the highly nonlinear and multidimensional nature of the U.S. job matching technology.

This paper builds primarily on three literatures. The first consists of evaluations of particular place-based policies or local economic shocks. Most papers in this branch use average wages or employment rates in the targeted location as the outcome of interest, seek to define a control group of alternative locations, and evaluate the policy or shock's impact using a treatment effect framework. This literature is vast, and is thoroughly discussed by survey articles such as Glaeser et al. (2008), Moretti (2010), Kline and Moretti (2013), and Neumark and Simpson (2014).[2]

[2] A particularly prominent paper in this branch of the literature is Greenstone et al. (2010), who compare employment gains in counties making winning bids for "million-dollar" plants to control counties who made losing bids. More recent

Busso et al. (2013)'s evaluation of the U.S. empowerment zone system stands out as one of the few quasi-experimental papers to explicitly evaluate social welfare impact, which they accomplish by deriving a set of sufficient elasticity parameters that can be cleanly identified. Interestingly, they find that while empowerment zones significantly increase wages and employment of zone residents, they do not meaningfully affect rent prices. This suggests that commuting adjustments may largely facilitate the response to very local shocks, limiting the impact on rent.[3]

A related sub-literature seeks to estimate local job multipliers due to the increased product demand and agglomeration and congestion externalities created by an initial infusion of new positions (e.g. Moretti (2010) or Bartik and Sotherland (2019)). Our approach complements this sub-literature: those papers are generally silent about which types of workers from which initial locations benefit most from the estimated net change in local job opportunities, while the assignment model here takes the new spatial distribution of positions (possibly including jobs generated by multipliers) as an exogenous input and evaluates the resulting skill and spatial incidence. We demonstrate this point by also evaluating a composite shock of 500 new manufacturing positions combined with 342 service jobs spread throughout the PUMA in accordance with the relevant job multiplier estimate from Bartik and Sotherland (2019).

The paper also contributes to a fast-growing literature on structural spatial equilibrium models designed to forecast the incidence of economic shocks across spatially-linked geographic areas. Schmutz and Sidibe (2016) estimate a search-and-matching model with data on worker flows among French metropolitan areas. They show that search frictions play a greater role than moving costs in limiting worker mobility and determining the incidence of local shocks, suggesting the potential promise of efforts to disseminate information about distant jobs. Monte et al. (2015) and Caliendo et al. (2015) (hereafter CDP) each estimate trade-theoretic models with labor, housing, and product market clearing and arbitrary spatial frictions in both labor and product markets. The former features joint choices of residential and work locations, and highlights the role of commuting in determining local shock incidence.[4] The latter shows that counterfactual dynamic equilibrium paths can be evaluated for alternative structural shocks (changes in trade costs, mobility costs, productivities, etc.) without estimating all the model primitives. The present paper relies on a very similar sufficient statistics approach, in that it evaluates the distribution of welfare impacts from demand shocks of alternative compositions without identifying any of the fundamental utility, production function, and moving cost parameters. The model below imposes even less structure on the form of production and utility than CDP, but focuses on a one-year (static) horizon for reallocation and is more limited in the set of counterfactuals it can evaluate.

Each of these papers aggregates locations to at least the county level. Manning and Petrongolo (2017), by contrast, represents the most notable attempt to determine the equilibrium incidence across nearby areas of small scale shocks. Like Schmutz and Sidibe (2016), they propose a search

contributions include Gregory (2013), Freedman (2013), and LeGower and Walsh (2017).

[3]The authors point out that the limited rent price impact might be due to the particularly depressed nature of the targeted locations, which could make them undesirable residential locations (or subject to rent control).

[4]Due to a lack of residential microdata, we do not consider whether new job matches involve residential mobility.

and matching model and fit the model-predicted geographic distribution of vacancy outflows to data on changes in vacancy stocks from local job search centers in Britain. Like this paper, they simulate the impact on the geographic distribution of unemployment of an exogenous increase in vacancies (new jobs) within particular census wards (similar in size to the census tracts used here). They also find evidence that labor markets are quite local, in the sense that moderate distance to vacancies substantially decreases the probability of an application. Nonetheless, they also find that ripple effects from overlapping markets cause the unemployment incidence to spread widely, with very little of the employment gain accruing to the ward receiving the shock.

While the present paper lacks the dynamics and explicit housing and product markets modeled in CDP, the commuting links modeled in Monte et al. (2015), and the distinction between search and moving costs highlighted by Schmutz and Sidibe (2016) and Manning and Petrongolo (2017), it features a much richer labor market. Among the papers just described, none feature any worker heterogeneity beyond initial location, and only CDP (industry differences) features any observable firm heterogeneity besides firm location. Similarly, several spatial equilibrium models in the labor literature (e.g. Piyapromdee (2017) or Diamond (2016)) feature imperfect substitution among observable worker types, but only differentiate firms by location. Because none of these models feature multidimensional two-sided sorting, the model featured in this paper is the only one equipped to evaluate differential incidence both across space and across skill/demographic groups from local labor demand shocks with alternative firm compositions. While Lindenlaub (2017) and Bonhomme et al. (2019) each estimate multidimensional labor market sorting models, they consider assignment of workers to occupations and jobs without incorporating geography or spatial frictions.

Indeed, Nimczik (2017), who characterizes labor markets as networks of firms disproportionately sharing worker flows, shows that the geographic and industrial scope of labor markets varies substantially across occupation and education categories. However, the stochastic block model he employs separately defines labor markets for each skill category, so that the various skill categories cannot be incorporated into a single integrated labor market. Thus, unlike an equilibrium model with two-sided optimal choice, it is not designed to analyze the tradeoffs firms and workers make following local demand shocks between settling for skill mismatch and paying the moving and search costs needed to overcome spatial mismatch.

Finally, this paper also draws heavily from the theoretical literature on the identification and estimation of two-sided assignment games. To my knowledge this is the first large-scale empirical application of a two-sided assignment model to the national labor market.[5] The theoretical properties of assignment games have been well-established for at least a generation.[6] However, the empirical content of assignment games and hedonic models for contexts in which the universe (or a large random sample) of all market entrants on both sides and their matches can be observed has only recently attracted interest, with pioneering work by Ekeland et al. (2004), Heckman et al. (2010) and CS leading to contributions by Chiappori and Salanié (2016), Menzel (2015), and

[5] See Tervio (2008) and Chen (2017) for applications of the assignment game to the narrower market for CEOs.

[6] See Koopmans and Beckmann (1957), Shapley and Shubik (1972), Roth and Sotomayor (1992), and Sattinger (1993).

Galichon and Salanié (2015), among others. We make three contributions to this literature.

First, we consider implementation in a context featuring millions of match observations and thousands of types on both the supply and demand side. We address the challenge of a somewhat sparse matching matrix by developing a smoothing procedure to aggregate matching patterns across "nearby" match types without smoothing away the heterogeneity the model is designed to highlight. Second, we allow separate surplus values for job stayers relative to within-job-type movers, and show that this reveals the welfare losses and gains from negative and positive demand shocks to be asymmetric. Third, we consider the limits to identification when the number of unmatched partners of each type is either unobserved or only observed on one side of the market: while unemployment by type may be inferred with reasonable accuracy by combining LEHD and public-use ACS data, unfilled vacancy counts by detailed type are not available. The existing identification results of CS and Menzel (2015), among others, rely on observing the number of singles on both sides. We discuss conditions under which ignoring unmatched partners would not affect the incidence of local shocks, and show robustness of results to endogenizing the set of positions to be filled via a fixed point algorithm that incorporates updates to the level and composition of demand in response to changes in the required transfers implied by the previous iteration's equilibrium assignment.

The rest of the paper proceeds as follows. Section 2 describes the two-sided assignment game that forms the theoretical basis for the empirical analysis. Section 3 applies the insights of CS to the labor market context to identify a set of joint surplus parameters that are sufficient to perform counterfactual simulations of labor demand shock incidence. Section 4 describes the LEHD database and presents summary statistics that motivate the subsequent analysis. Section 5 describes sample selection and the smoothing procedure. In addition, Section 5 also introduces the various labor demand shocks and the methods used to aggregate the resulting counterfactual allocations of workers to positions into interpretable statistics that effectively characterize variation in shock incidence. Section 6 presents the main findings and the model validation results, and Section 7 concludes.

2 The Two-Sided Assignment Model

In this and the following sections we model the labor market as a static assignment game played by workers and establishments. The model introduces several features and extensions necessary to adapt CS's marriage market model to a labor market setting. The exposition closely mirrors Galichon and Salanié (2015) (hereafter GS), which generalizes CS. Section 2.1 defines the matching game. Section 2.2 describes how the workers and positions and the job matches that determine the game's payoffs are aggregated to types and groups, respectively. Section 3.1 imposes additional structure that facilitates the identification and estimation of the underlying group-level match surpluses that determine the frequencies of particular kinds of job matches. Section 3.2 shows how to use the estimated surpluses to construct counterfactual simulations capturing the incidence of local labor supply and demand shocks of varying worker and position compositions.

2.1 Defining the Assignment Game

Suppose that in a given year there are I potential workers comprising the set $\mathcal{I}$ who participate in the labor market. Each worker i enters the market with an existing job match with a position $j(i)$ at establishment $m(j(i))$ taken from the set of possible positions $\mathcal{J}$. Let $m(j) = 0$ represent unemployment so that positing an initial "job" for each worker is without loss of generality.

The payoff to worker i, initially at position j, of accepting a position k in the current year is denoted $U(i, j(i), k)$, or more parsimoniously as $U(i, k)$. The worker's potential earnings in the chosen year from accepting position k, denoted w_{ik}, is assumed to be additively separable from all other determinants of the worker's payoff, so that $U(i, k)$ has a quasi-linear money-metric form:[7]

$$U(i, k) = \pi^i_{ik} + w_{ik} \tag{1}$$

π^i_{ik} captures the combined value to worker i of a variety of payoff components. We show below that the researcher need not specify any of the fundamental components or the functions governing their links to payoffs to construct the counterfactual simulations that form the primary contribution of the paper. Any payoff function specification in which current worker earnings are additively separable will suffice. That said, careful thought about which determinants of the payoff are likely to be large and differential across alternative workers, positions, and job matches is necessary to guide the choice of characteristics used to assign workers and positions to types in section 2.2 below, as well as to evaluate the plausibility of assumptions laid out in section 3.2 that underlie the simulations.

Such components might include worker i's valuation of various non-pecuniary amenities offered by position k (including the appeal of its geographic location), as well as any search, moving, or training costs paid by worker i to find, move to, or settle into position k from initial position j.[8] They might also include the continuation value associated with beginning the next year as an incumbent, trained worker at position k, which might depend on the productivity gains from firm-specific experience and the availability of other opportunities in position k's local labor market.

On the other side of the market there are K potential positions comprising set $\mathcal{K}$ at establishments that seek workers in the chosen year. The intersection of $\mathcal{K}$ and $\mathcal{J}$ may be quite large, so that many positions in $\mathcal{K}$ can potentially be "filled" by retaining an existing worker. We assume that each establishment makes independent hiring decisions for each position so as to model positions' preferences over individual workers rather than establishments' preferences over collections of workers.[9]

[7] Since we have data on annual earnings but not wages or hours, for simplicity we assume that the hours associated with a job match are fixed by contract and common across positions for a given worker.

[8] The traditional assignment game precludes stochastic search frictions, so that each agent may potentially match with any agent on the opposite side. However, Menzel (2015) shows that one can introduce a probability that i and k meet that is independent of other payoff determinants, assign their joint surplus to $-\infty$ if the pair does not meet, and use these alternative payoffs to determine the stable matching. Alternatively, search costs might be modeled as a deterministic cost that must be paid to an intermediary (e.g. a headhunter or a matching website) to reveal certain opposite side agents.

[9] One justification for treating positions as independent is that there are nontrivial costs of coordinating multiple independent hires/retentions that outweigh the gains from better exploiting production complementarities. Roth and Sotomayor (1992) highlight the complications that arise when establishments have preferences over collections of workers.

Let $V(i,k)$ denote the value to position k in establishment $m(k)$ of hiring (or retaining) worker i. The potential earnings paid by k to worker i in the chosen year is assumed to be additively separable from all other determinants of the position's payoff, so that $V(i,k)$ can be written as:

$$V(i,k) = \pi_{ik}^{k} - w_{ik} \tag{2}$$

Akin to π_{ik}^{i}, π_{ik}^{k} combines several payoff components that need not be fully specified by a particular functional form. These components might include worker i's contribution to establishment $m(k)$'s annual revenue, any recruiting, moving, and training costs borne by $m(k)$ in hiring worker i, as well as any continuation value from starting next year's market with i already installed in position k, including the fact that retaining i next year avoids further recruiting/training costs.

One can then define the joint surplus from the match between worker i and position k as the sum of the worker and position valuations of the match:

$$\pi_{ik} \equiv U(i,k) + V(i,k) = \pi_{ik}^{i} + \pi_{ik}^{k} \tag{3}$$

Since worker annual earnings in the current year are additively separable in both the worker's and position's payoffs, this assignment model exhibits transferable utility. Written in this form, the game's structure mimics the classic assignment game analyzed by Shapley and Shubik (1972).

A matching or market-wide allocation in this labor market is an $I \times K$ matrix μ such that $\mu_{i,k}$ = 1 if worker i matches with position k, and 0 otherwise. As in Galichon and Salanié (2015), we focus on stable matchings, which require a division of joint surplus in each proposed job match such that no currently unmatched worker-position pair can find any division of the joint surplus from their potential match that makes both the worker and position strictly better off than under the proposed matching. Shapley and Shubik (1972) show that the set of stable matchings coincides with both the core of the game and with the set of competitive equilibria from a decentralized market. Furthermore, they show that with transferable utility there will exist a unique assignment (or, equivalently, competitive equilibrium allocation) of workers to positions as long as preferences are strict on both sides of the market. This equilibrium allocation/stable assignment maximizes the aggregate surplus and is the solution to a linear programming problem.[10]

Equivalently, the unique stable assignment can also be found by solving the dual problem: identifying a set of worker utility values $\{r_i\}$ and position profit values $\{q_k\}$ that minimize the total "cost" of all workers and positions, $\sum_{i \in \mathcal{I}} r_i + \sum_{k \in \mathcal{K}} q_k$, subject to the constraint that these values cannot violate the underlying joint surplus values: $r_i + q_k \geq \pi_{ik} \; \forall \, (i,k)$. Crucially, inspection of the problem reveals that the stable assignment is fully determined by the joint surplus values $\{\pi_{ik}\}$; no separate information on the worker and firm components π_{ik}^{i} and π_{ik}^{k} is needed. This dual problem

[10]The joint surplus is given by $\sum_{(i,k) \in \mathcal{I} \times \mathcal{K}} \mu_{i,k} \pi_{ik} + \sum_{i \in \mathcal{I}: \mu_{i,k}=0 \; \forall \; k} \mu_{i,0} \pi_{i0} + \sum_{k \in \mathcal{K}: \mu_{i,k}=0 \; \forall \; i} \mu_{0,k} \pi_{0k}$, where π_{i0} and π_{0k} denote i's payoff from unemployment and k's payoff from remaining vacant. Constraints must also hold requiring that each position and worker match with at most one counterpart: $\sum_i \mu_{i,k} \leq 1 \; \forall \; k \in \mathcal{K}$ and $\sum_k \mu_{i,k} \leq 1 \; \forall \; i \in \mathcal{I}$.

yields the following conditions that define the optimal assignment (GS):

$$\mu_{ik} = 1 \text{ iff } k \in \arg\max_{k \in \mathcal{K} \cup 0} \pi_{ik} - q_k \text{ and } i \in \arg\max_{i \in \mathcal{I} \cup 0} \pi_{ik} - r_i \tag{4}$$

We aggregate these conditions in the next section to deliver identification of "group"-level surpluses. Given optimal worker and position payoffs $\{r_i\}$ and $\{q_k\}$ from the dual solution, Shapley and Shubik (1972) show how to decentralize this optimal assignment via a set of earnings transfers w_{ik}:

$$w_{ik} = \pi^k_{ik} - q_k \tag{5}$$

Because $r_i + q_k = \pi_{ik} \equiv \pi^i_{ik} + \pi^k_{ik}$ for any matched pair (i, k) in the stable match, this implies:

$$w_{ik} = r_i - \pi^i_{ik} \tag{6}$$

Using (5) and (6), the conditions (4) can be rewritten as the standard requirements that worker and establishment choices must be utility- and profit-maximizing, respectively:

$$\mu_{ik} = 1 \text{ iff } k \in \arg\max_{k \in \mathcal{K} \cup 0} \pi^i_{ik} + w_{ik} \text{ and } i \in \arg\max_{i \in \mathcal{I} \cup 0} \pi^k_{ik} - w_{ik} \tag{7}$$

This shows that the market-clearing earnings amounts will in general be specific to worker-position pairs (i, k). By contrast, the market-clearing utilities r_i and profit contributions q_k will be worker-specific and position-specific, respectively, which will be exploited below. Importantly, while the stable assignment μ is generally unique, the equilibrium payoffs and transfers are not: all $\{r_i\}$ values can generally be shifted slightly up or down (with offsetting $\{q_k\}$ shifts) without violating stability. The exact equilibrium payoffs/wages depend on the particular market clearing mechanism.

While the model does not require a particular earnings determination process, one candidate is a simultaneous ascending auction in which all positions bid on all workers. Workers set reservation utilities based on their values of remaining unemployed for a year. Each position bids utility values of a one year commitment U_{ik} (which include the value of beginning the next period as an incumbent), and may only win the bidding for a single worker (or it may choose to remain vacant). The position k that bids the highest utility r_i retains worker i and pays an annual earnings amount w_{ik} that, combined with the non-pecuniary component π^i_{ik}, equals the worker's promised valuation $U_{ik} = r_i$. The auction ends when no position wishes to change its bid for any worker. Some workers may remain unemployed, and some positions may remain unfilled. Importantly, though positions start at different π^i_{ik} baselines, with transferable utility bid changes can always take the form of annual earnings increases. Thus, one may scale the changes in equilibrium utilities r_i following labor demand shocks in terms of annual earnings gains (though sometimes workers achieve utility gains by taking an earnings cut to join an establishment offering superior non-pecuniary values).

2.2 Modeling the Match Surpluses

Part of the joint surplus π_{ik} from match (i, k) will be common to all matches that share certain salient characteristics. For example, positions at larger firms may face smaller per-position costs of recruiting distant workers due to economies of scale; highly skilled workers may generate larger surplus at positions whose output is particularly sensitive to worker skill. Thus, we assign each potential match (i, k) to one of a set of mutually exclusive groups $g \in \mathcal{G}$ (with $G \equiv |\mathcal{G}|$), and use $g(i, k)$ to denote the group assignment of match (i, k). Importantly, these groups are always defined by a combination of *observable* characteristics of the worker, the destination position, or the match (i, k). These characteristics should be chosen to capture as comprehensively but parsimoniously as possible the underlying (structural) preferences, productivities, moving costs, and search costs that determine the joint surpluses $\{\pi_{ik}\}$.

Some observed characteristics may only relate to worker i or worker i's previous position $j(i)$ or match $(i, j(i))$, and thus are common to all current positions. We use such characteristics to assign each worker to a worker type $l \in \mathcal{L}$, with $l(i)$ denoting this assignment. In the empirical work, worker types (detailed further below) are defined by combinations of a prior earnings/ unemployment category (proxying for worker skill) and the location of workers' previous establishments.[11]

Analogously, a second subset of the characteristics defining match groups may only describe the destination position, and may be used to assign each position $k \in \mathcal{K}$ to a position type $f \in \mathcal{F}$, denoted $f(k)$. In the empirical work, position types consist of unique combinations of the following establishment characteristics: $m(k)$'s geographic location (detailed further below), its prior year quartiles in the national establishment-level employment and average earnings distributions (the latter used to proxy for average skill requirements), and its industry supersector.[12]

Finally, let $z(i, j(i), k) \equiv z(i, k)$ denote any characteristics defining the match group that depend on both $(i, j(i))$ and k. In the empirical work below, the only z characteristic is an indicator for continued employment at the same establishment, $1(m(k) = m(j(i)))$, so that job stayers and movers are placed into different groups within each type pair with feasible stayers, reflecting the fact that search, recruiting, and training costs need not be repaid for existing workers. This in turn allows establishments to retain existing employees more frequently than they hire other local workers (important for predicting which workers ultimately accept newly created jobs).[13]

Thus, one can rewrite the mapping $g(i, k)$ as $g(l(i), f(k), z(i, k)) \equiv g(l, f, z)$. While knowledge of g is sufficient to recover l and f, knowledge of l and f need not uniquely identify g due to the presence of z. In a slight abuse of notation, we will sometimes use $l(g) = l(g(i, k)) = l(i)$ to refer to group g's worker type and $f(g) = f(g(i, k)) = f(k)$ to refer to its position type.

[11]Ideally, residential location would define the worker type and establishment location would define the position type. In the absence of data on workers' residential locations, initial establishment locations are used as proxies.

[12]Note that i's earnings in the previous year is used to proxy for worker skill, but $m(k)$'s average pay in the previous year is used to proxy for k's skill requirements. This can be rationalized by assuming that a newly hired worker develops the required skills by the end of the year (perhaps incurring training costs affecting π_{ik} that are paid by either side).

[13]Mourifié and Siow (2017) use the same approach to distinguish marriage from cohabitation.

Given these definitions, one can decompose the surplus π_{ik} into a part common to all matches classified as group $g(i,k)$, denoted θ_g, and an idiosyncratic component ϵ_{ik} specific to (i,k):

$$\pi_{ik} = \theta_{g(i,k)} + \sigma\epsilon_{ik} \tag{8}$$

ϵ_{ik} might reflect, for example, the low psychic costs of a particular worker who is moving back to a familiar location, or perhaps particular skill requirements of position k that worker i uniquely possesses. Following Decker et al. (2013), σ is a scaling parameter that captures the relative importance of idiosyncratic surplus components compared to components that are common to all group g matches in determining the variation in match surpluses across potential pairs $(i,k) \in \mathcal{I} \times \mathcal{K}$. We show below that counterfactual unique stable job assignments will not depend on σ, but σ plays a key role in determining the scale of changes in offered utilities r_i for particular workers in particular locations necessary to facilitate the reallocation that yields the stable assignment.

The goal is to use the observed matching μ to recover the set of group mean surplus values $\{\theta_g\}$. As GS emphasize, one could impose further structure on the production, utility, search cost, and recruiting cost functions that comprise the joint surpluses and estimate the model via maximum likelihood. Driven by computational considerations and an interest in being agnostic about the various structural functions that underlie $\{\theta_g\}$, we follow CS and leave $\{\theta_g\}$ unrestricted, achieving identification instead by assuming that ϵ_{ik} draws are i.i.d across all alternative matches (i,k') and $(i',k) \in \mathcal{I} \times \mathcal{K}$ and follow a Type 1 extreme value distribution.[14] Unlike in CS and GS, equation (8) allows the idiosyncratic component to be truly pair-specific: the combined surplus from two matches changes if the workers swap positions, even if they share a worker type and the positions share a position type (Section 6.6 compares simulation results from the current model with the CS alternative). Given coarsely-defined worker and position types, pair-level heterogeneity in job match quality is likely to be substantial. As discussed in Section 3.2 and Appendix A5, allowing such heterogeneity prevents the use of observed transfers to recover group-level worker and position subcomponents θ_g^l and θ_g^f analogous to the match-level components π_{ik}^i and π_{ik}^k defined above. Fortunately, this decomposition is not necessary to generate key measures of worker-level incidence.

3 Identification

3.1 Identification of the Set of Group-Level Match Surpluses $\{\theta_g\}$

Recall from section 2.1 that a necessary condition for a matching μ to be stable (and thus sustainable as a competitive equilibrium) is that there exists a set of worker payoffs $\{r_i\}$ such that $\mu_{ik} = 1$ implies that $i \in \arg\max_{i \in \mathcal{I}} \pi_{ik} - r_i$ for any potential match $(i,k) \in \mathcal{I} \times \mathcal{K}$. Given candidate

[14]Menzel (2015) shows that imposing that ϵ_{ik} draws are i.i.d is the key assumption (the choice of the Type 1 extreme value distribution is not critical), since it causes each sides' conditional choice probabilities to satisfy the independence of irrelevant alternatives property. GS and Chiappori et al. (2009) discuss the possibility of allowing certain forms of correlation in the idiosyncratic component across matches featuring shared characteristics. However, we maintain the standard i.i.d. assumption in this paper in order to minimize an already substantial computational burden.

equilibrium payoffs $\{r_i\}$ combined with the i.i.d. Type 1 EV assumption for ϵ_{ik}, Decker et al. (2013) show that the probability that hiring (or retaining) i maximizes k's payoff is given by:

$$P(i|k) = \frac{e^{\frac{\theta_g - r_i}{\sigma}}}{\sum_{i' \in \mathcal{I}} e^{\frac{\theta_{g'} - r_{i'}}{\sigma}}} \tag{9}$$

Next, define $n(l)$ as the share of workers assigned to type l, define C_l as the mean of $e^{-\frac{r_i}{\sigma}}$ among type l workers, and define $\overline{S}_{g|l,f}$ as the mean among type f positions of the share of type l workers whose hire/retention would be assigned to group g (i.e. the incumbent share if $z(g) = 1$, the mover share if $z(g) = 0$). With two additional assumptions, one can aggregate equation (9) to derive a tractable expression for the conditional probability $P(g|f)$ that a randomly chosen position of type f wishes to hire a worker whose job match would be assigned to group g:

$$P(g|f) = \frac{e^{\frac{\theta_g}{\sigma}} \overline{S}_{g|l,f} n(l) C_l}{\sum_{l' \in \mathcal{L}} \sum_{g' \in (l,f)} e^{\frac{\theta_{g'}}{\sigma}} \overline{S}_{g'|l',f} n(l') C_{l'}} \tag{10}$$

In particular, this expression depends only on the group g and the types l and f rather than the individual workers i and positions k.[15] Appendix A1 presents and proves this result formally as Proposition A1. Intuitively, the first assumption imposes that the utility payoffs required in equilibrium by workers from the same skill class and local area must not differ systematically across initial establishments. This becomes a better approximation as more characteristics and finer geography are used to define worker types, so that workers of the same type become close substitutes for one another. The second assumption imposes that establishments of the same position type feature roughly the same number and worker type distribution of incumbent workers. This approximation improves as narrower worker location and skill categories and establishment location, size and average pay categories are used to define worker and position types.[16]

Next, let $\hat{\mu}$ denote an observed matching. Since each job match can be assigned to a unique group g, one can easily aggregate the individual-level matching into an empirical group-level distribution. Specifically, let $\hat{P}_g$ denote the fraction of observed matches that are assigned to group g: $\hat{P}_g \equiv \frac{1}{|I|} \sum_{(i,k) \in \mathcal{I} \times \mathcal{K}} \hat{\mu}_{ik} 1(g(i,k) = g)$. Similarly, $\hat{n}(l) = \frac{1}{|I|} \sum_{i \in \mathcal{I}} 1(l(i) = l)$ denotes the fraction of matches featuring type l workers and $\hat{h}(f) = \frac{1}{|K|} \sum_{k \in \mathcal{K}} 1(f(k) = f)$ denotes the fraction featuring type f destination positions.[17] Given these definitions, one can estimate the conditional choice

[15] Note that in contrast to CS, the probability that a worker of worker type l is chosen depends on the share of workers of type l in the population, $\eta(l)$. This difference stems from allowing the idiosyncratic surplus component to be pair-specific. Menzel (2015) derives a similar formula in his nontransferable utility assignment model.

[16] These assumptions are necessary because the probability of filling a position with an existing employee depends on how many employees one already has, so that the group average of $e^{\frac{-r_i}{\sigma}}$, a non-linear function of the random variable r_i, depends on the entire establishment size distribution among firms who are at risk of creating a match that could be classified into g. The assumptions essentially impose that Jensen's inequality is close to an equality for $e^{\frac{-r_i}{\sigma}}$.

[17] Note that in the empirical work each match will have both a worker and a position because we do not observe unfilled vacancies and we augment $\mathcal{K}$ to include a sufficient number of nonemployment "positions". As a result, the number of workers seeking jobs will equal the number of "positions".

probability $P(g|f)$ by calculating the observed fraction of type f positions that were filled via group g matches: $\hat{P}(g|f) = \frac{\hat{P}_g}{\hat{h}(f)}$. As the number of observed matches gets large, each member of the set of empirical CCPs $\{\hat{P}(g|f)\}$ should converge to the corresponding expression in (10). Note also that the average shares $\{\overline{S}_{g|l,f}\}$ can be estimated using the average of the incumbent indicator $1(m(j(i)) = m(k))$ across all possible matches (i, k) sharing type pairs $(l(i), f(k))$.

One may now assess the amount of information contained in the observed empirical choice probabilities $\{\hat{P}(g|f)\}$ about the mean match surplus values $\{\theta_g\}$. First, using (10), we can derive an expression for the log odds between two CCPs involving an (arbitrarily chosen) position type f_1 and two (arbitrarily chosen) match groups g_1 and g_2 for which $f(g_1) = f(g_2) = f_1$:

$$\ln(\frac{\hat{P}_{g_1|f_1}}{\hat{P}_{g_2|f_1}}) = (\frac{\theta_{g_1} - \theta_{g_2}}{\sigma}) + \ln(\frac{\overline{S}_{g_1|l(g_1),f_1}}{\overline{S}_{g_2|l(g_2),f_1}}) + \ln(\frac{n(l(g_1))}{n(l(g_2))}) + \ln(\frac{C_{l(g_1)}}{C_{l(g_2)}}) \tag{11}$$

Since the initial worker type shares $n(l(g_1))$ and $n(l(g_2))$ and shares of potential stayers $\overline{S}_{g_1|l(g_1),f_1}$ and $\overline{S}_{g_2|l(g_2),f_1}$ are either directly estimable or observed (depending on whether a sample or the full population is available), to establish identification one can treat their terms as known and bring them to the left hand side. Even these adjusted log odds still conflate the relative mean (re-scaled) surplus values from match groups g_1 and g_2, $(\frac{\theta_{g_1} - \theta_{g_2}}{\sigma})$, with the log ratio of mean exponentiated worker re-scaled utilities between the two worker types, $\ln(\frac{C_{l(g_1)}}{C_{l(g_2)}})$.

However, consider two additional groups g_3 and g_4 for which $f(g_3) = f(g_4) = f_2$ and $l(g_3) = l(g_1)$ and $l(g_4) = l(g_2)$.[18] The four groups g_1 to g_4 can be chosen to represent the two ways to match a given pair of positions to a given pair of workers. Taking the ratio of equation (11) to its analogue using g_3 and g_4 conditional on f_2 and rearranging, one obtains:

$$\ln(\frac{\hat{P}_{g_3|f_2}/(\overline{S}_{g_3|l(g_3),f_2} n(l(g_3)))}{\hat{P}_{g_4|f_2}/(\overline{S}_{g_4|l(g_4),f_2} n(l(g_4)))} / \frac{\hat{P}_{g_1|f_1}/(\overline{S}_{g_1|l(g_1),f_1} n(l(g_1)))}{\hat{P}_{g_2|f_1}/(\overline{S}_{g_2|l(g_2),f_1} n(l(g_2)))}) = \frac{(\theta_{g_3} - \theta_{g_4}) - (\theta_{g_1} - \theta_{g_2})}{\sigma} \tag{12}$$

Thus, the adjusted log odds ratio identifies the expected gain in scaled joint surplus from swapping partners in any two job matches. Note that differencing and conditioning, respectively, necessarily remove any information about the mean payoffs or welfare of worker types and position types. However, the identified set of surplus difference-in-differences $\Theta^{D-in-D} \equiv \{\frac{(\theta_g - \theta_{g'}) - (\theta_{g''} - \theta_{g'''})}{\sigma} \forall (g, g', g'', g''') : l(g) = l(g''), l(g') = l(g'''), f(g) = f(g'), f(g'') = f(g''')\}$ preserves the critical information about the relative efficiency of alternative matchings present in the observed group frequencies.

3.2 Counterfactual Simulations

In this subsection we show that identification of Θ^{D-in-D} is sufficient to generate the unique counterfactual aggregated assignment $P^{CF}(g)$ following arbitrary changes in the marginal distributions of worker types $n(l)$ and position types $h(f)$. If more than one matching is observed, then σ can

[18] f_2 could be (but need not be) the same position type as f_1.

be (roughly) estimated as well, permitting a proper welfare analysis of the approximate mean utility and profit gain for each worker type and position type.

One can characterize the set of workers to be reallocated via their worker type distribution, $n^{CF}(l)$. The "CF" superscript indicates that this distribution could be counterfactual (e.g. capturing a proposed influx of refugees). Similarly, the set of counterfactual positions to be filled can be represented by $h^{CF}(f)$, and the prevailing matching technology can be denoted $\{\theta_g^{CF}\}$. $n^{CF}(l)$, $h^{CF}(f)$, and $\{\theta_g^{CF}\}$ are all inputs that are either observed or chosen by the researcher, who wishes to predict the equilibrium aggregate distribution of matches, $P^{CF}(g)$.

To take a concrete example, suppose that a local development board has already forecasted the number and location of new manufacturing positions (and perhaps spillover-driven retail positions) that a plant opening would generate, and suppose the existing group mean surpluses $\{\theta_g^{CF}\}$ have been estimated. The board may wish to predict the extent to which the plant opening will increase job-related utility and the employment rate among existing local workers/job seekers in the chosen and surrounding neighborhoods (and perhaps the profits of local and less local firms).

We assume that the counterfactual assignment also satisfies the assumptions of Proposition A1 above. We also assume that the set of position type averages of the shares of workers of each worker type who would be incumbents at the establishment, $\{\overline{S}_{g|l,f}^{CF}\}$, is known, and treat it as an input. In particular, when $n^{CF}(l) = n^{y'}(l)$ and $h^{CF}(f) = h^{y'}(f)$ for some observed year y', then the appropriate existing employee fractions can be obtained via $\overline{S}_{g|l(g),f(g)}^{CF} = \overline{S}_{g|l(g),f(g)}^{y'}$ $\forall$ g, which is observed. Then the counterfactual conditional choice probability $P^{CF}(g|f)$ can be expressed as (10) with $(\theta_g^{CF}, n^{CF}(l), h^{CF}(f), \overline{S}_{g|l(g),f(g)}^{CF}, C_l^{CF})$ replacing $(\theta_g, n(l), h(f), \overline{S}_{g|l(g),f(g)}, C_l)$. The worker type-specific mean exponentiated (and rescaled) utility values $\mathbf{C}^{CF} \equiv \{C_1^{CF} \dots C_L^{CF}\}$ are ex ante unknown equilibrium objects affected by the counterfactual changes reflected in $(\theta_g^{CF}, n^{CF}(l), h^{CF}(f))$. Thus, each counterfactual CCP must initially be treated as a function of the set $\mathbf{C}^{CF}$.

GS and Decker et al. (2013) each show that a unique probability distribution over match groups $P^{CF}(g)$ satisfies the aggregate analogues to the stability and feasibility conditions. However, these papers as well as CS assume when proving identification that one observes the numbers of unmatched partners (and thus the total number of agents) of each type on both sides of the market. While counts of unemployed workers by type can be accurately constructed, the LEHD database contains no information about unfilled vacancies.[19] Because each submatching of a stable matching must also be stable, observing only the subset of positions that match with workers does not threaten identification of the remaining elements of Θ^{D-in-D}; the estimated surplus relationships would not be reversed if data were augmented with unmatched agents.

In principal, though, unfilled positions may put upward pressure on wages that affect the division of surplus between workers and positions, even if they do not affect the final job assignment. However, many unfilled vacancies may not be the second-best option for any worker, or may only be slightly more appealing than a third-best position that eventually settles for another worker rather

[19] Furthermore, constructing counts of vacancies for the position types used in this paper from publicly available vacancy data is also not straightforward.

than remaining vacant for the year, so that a large share of unfilled vacancies ignored by $h^{CF}(f)$ may negligibly affect the division of surplus. A related concern is that firms might wish to alter the number of positions they choose to fill when wages rise or fall in the wake of labor demand shocks. However, for relatively small and localized shocks, firms may display an extremely inelastic extensive margin response if the costs of adjusting the number of positions at establishments (and thus changing the assignment of production tasks to workers) are large relative to the small shock-induced changes in the minimized cost of an efficiency unit of labor. In this scenario establishments will prefer to only adjust the composition of workers they choose to fill a fixed set of positions.

While our baseline estimates maintain an assumption of perfect inelasticity on the extensive margin, for larger shocks that make this untenable, one can use existing wage elasticity and multiplier estimates to incorporate the endogenous response into $h^{CF}(f)$ and re-interpret $h^{CF}(f)$ as a post-adjustment distribution. As a robustness check in Section 6.6, we explicitly endogenize the extensive margin in this way by iterating between assignment model equilibria and calibrated extensive margin responses to changes in a position's expected profitability until a fixed point is found.

Treating the set of positions that will be filled as exogenous (at least within an iteration) simplifies the choice of variation used to identify relative surplus values. One need not isolate labor supply shocks in order to identify extensive margin labor demand elasticities by type. Instead, surplus diff-in-diffs Θ^{D-in-D} (along with σ) are essentially determining equilibrium elasticities of substitution for each position type among different worker types. Elasticities of substitution are fully determined by *relative* prices, so they should be insensitive to the source of relative cost changes for different worker types: upward (downward) shifts in the number of local (distant) workers seeking positions and downward (upward) shifts in the number of local (distant) positions tending to prefer local (distant) workers are all valid sources of variation in relative prices of workers from different initial locations. So there is no inconsistency in using the full set of year-to-year job flows, implicitly driven by a mix of many small and large local supply and demand shocks, to recover Θ^{D-in-D}.

Requiring all positions in h_f^{CF} to fill also dramatically simplifies the otherwise demanding computation of counterfactual equilibria. With an unknown number of unmatched partners on each side, GS show that one must solve $L + F$ non-linear equations that combine the feasibility and stability conditions for the mean equilibrium payoffs of all worker and firm types ($\{C_l^{CF}\}$ and $\{C_f^{CF}\}$). By contrast, when the "supply" of positions by position type is assumed known, these values can be set equal in equilibrium to worker "demand" for such positions to create a set of F type-level market clearing conditions that determine $\{C_f^{CF}\}$.[20] Equivalently, if a dummy "position" type is added with mass equal to the share of workers who will end up unmatched, then the augmented demand (including "demand" from nonemployment) for each worker type l will equal the supply $n^{CF}(l)$, facilitating market clearing on the worker side.[21] Since relative payoffs among worker types fully

[20] Koopmans and Beckmann (1957) point out that when unmatched agents only exist on one side of the market, the dual problem payoffs need only be recovered on one side of the market in order to construct the stable assignment.

[21] These dummy nonemployment positions represent a computational mechanism for incorporating workers' payoffs from nonemployment, $\{\pi_{i0}^i\}$, akin to "balancing" an unbalanced assignment problem (Hillier and Lieberman (2010)). A formal proof of equivalence is proposed as Proposition A2 and proved in Appendix A3.

determine the equilibrium assignment (so one can normalize $C_1^{CF} = 0$), and the worker type distribution $n^{CF}(*)$ must sum to one, we obtain the following $L-1$ market clearing conditions:

$$\begin{aligned}
&\sum_{f\in\mathcal{F}} h^{CF}(f)\Big(\sum_{g:l(g)=2} P^{CF}(g|f,\mathbf{C^{CF}})\Big) = n^{CF}(2)\\
&\vdots\\
&\sum_{f\in\mathcal{F}} h^{CF}(f)\Big(\sum_{g:l(g)=L} P^{CF}(g|f,\mathbf{C^{CF}})\Big) = n^{CF}(L)
\end{aligned} \tag{13}$$

Given a solution to (13), one can then construct the counterfactual probability for any match group via $P^{CF}(g) = \sum_f h^{CF}(f) P^{CF}(g|f, \mathbf{C^{CF}})$. Since the solution also satisfies the stability and feasibility conditions, it must be the unique aggregate counterfactual stable assignment.

Because only $\min\{L, F\}$ equations must be solved, this approach provides considerable computational savings when the number of types is much larger on one side of the market. Below we present results that average over 500 counterfactual allocations featuring around 1,000 worker and 10,000 position types that would be prohibitive to compute with unmatched agents on both sides.

3.3 Interpreting the Counterfactual Simulations

In the simulations below, we generally use data from the 2012-2013 set of job transitions (including stayers) to construct the simulation inputs, so that $\Theta^{CF} = \Theta^{2012}$, $n^{CF}(*) = n^{2012}(*)$, and $h^{CF}(*)$ will equal $h^{2012}(*)$ plus a shock consisting of additional positions added to or subtracted from a chosen type f (or relocated among types). We wish to interpret the difference between the resulting counterfactual reallocation and the observed 2012-2013 reallocation as the one-year impact that such a stimulus or plant closing would have caused in that economy. However, a few additional assumptions and clarifications are necessary to justify and elaborate on this interpretation.

First, constructing the market-clearing conditions (13) requires a full set of group joint surpluses $\Theta^{2012} \equiv \{\theta_g^{2012}\ \forall\, g \in \mathcal{G}\}$, but the identification argument in section 3 suggests that only the set of diff-in-diffs $\Theta^{D-in-D,2012}$ is identified. In Appendix A2, we prove the following proposition:

Proposition 1:

Define the set $\Theta^{D-in-D} \equiv \{\frac{(\theta_g-\theta_{g'})-(\theta_{g''}-\theta_{g'''})}{\sigma}\ \forall\,(g,g',g'',g''') : l(g) = l(g''), l(g') = l(g'''), f(g) = f(g'), f(g'') = f(g'''))\}$*. Given knowledge of* Θ^{D-in-D}*, a set* $\tilde{\Theta} = \{\tilde{\theta}_g\ \forall\, g \in \mathcal{G}\}$ *can be constructed such that the unique group level assignment* $P^{CF}(g)$ *that satisfies the market-clearing conditions (13) using* $\theta_g^{CF} = \tilde{\theta}_g\ \forall\, g$ *and arbitrary marginal PMFs for worker and position types* $n^{CF}(*)$ *and* $g^{CF}(*)$ *will also satisfy the corresponding market-clearing conditions using* $\theta_g^{CF} = \theta_g\ \forall\, g \in \mathcal{G}$ *and the same PMFs* $n^{CF}(*)$ *and* $g^{CF}(*)$*. Furthermore, denote by* $\mathbf{\tilde{C}^{CF}} \equiv \{\tilde{C}_1^{CF}, \ldots, \tilde{C}_L^{CF}\}$ *and* $\mathbf{C^{CF}} \equiv \{C_1^{CF}, \ldots, C_L^{CF}\}$ *the utility vectors that clear the market using* $\theta_g^{CF} = \tilde{\theta}_g$ *and using* $\theta_g^{CF} = \theta_g$*, respectively. Then* $\mathbf{\tilde{C}^{CF}}$ *will satisfy* $\tilde{C}_l^{CF} = C_l^{CF} e^{\frac{-\Delta_l}{\sigma}}\ \forall\, l \in \mathcal{L}$ *for some set of worker* *type-specific constants* $\{\Delta_l : l \in [1, L]\}$ *that is invariant to the choices of* $n^{CF}(*)$ *and* $g^{CF}(*)$.

Essentially, the proposition states that the identified set of surplus difference-in-differences Θ^{D-in-D} contains sufficient information to generate the unique counterfactual group-level assignment $P^{CF}(g)$ associated with the complete set of surpluses Θ. Furthermore, the utility premia $\tilde{\mathbf{C}}^{\mathbf{CF}}$ that clear the market using the artificially completed surpluses $\tilde{\Theta}$ will always differ from the "true" premia $\mathbf{C}^{\mathbf{CF}}$ that clear the counterfactual market under Θ by the same l-type-specific constants regardless of the compositions of supply $n^{CF}(l)$ and demand $h^{CF}(f)$ that define the counterfactual.

The "bias" terms $\{\Delta_l\}$ in Proposition 1 imply that even relative levels of baseline utility among worker types in counterfactual allocations (or observed allocations) are not identified. However, because the Δ_l values are constant across counterfactuals featuring different $n^{CF}(l)$ and $h^{CF}(f)$ distributions, the relative *changes* $[(\ln(C_l^{CF1})-\ln(C_l^{CF2}))-(\ln(C_{l'}^{CF1})-\ln(C_{l'}^{CF2}))] \approx (\frac{(\bar{r}_l^{CF1}-\bar{r}_l^{CF2})-(\bar{r}_{l'}^{CF1}-\bar{r}_{l'}^{CF2})}{\sigma})$ in mean rescaled utilities across worker types among two counterfactuals can be identified.[22] Below, pairs of counterfactuals include one that features a targeted local demand shock and an otherwise identical counterfactual that does not. In some cases one may reasonably impose a priori values for utility changes for one or more worker types, so that utility changes $\frac{\bar{r}_l^{CF1}-\bar{r}_l^{CF2}}{\sigma}$ for other types can be identified. Below we assume that the small, very local stimuli and plant closings we consider do not alter utility for the least affected (usually quite distant) worker type. Such restrictions allow the recovery of each worker type's share of total welfare gains or losses from a labor demand (and/or supply) shock. The model's symmetry between workers and positions implies that mean changes in profits and shares of profit gains or losses by position type also are identified. Thus, given data on even a single matching, the model can produce a reasonably complete analysis of job-related welfare incidence from labor supply and demand shocks.

Second, besides these normalizations, in order for the predicted allocation and welfare gains to accurately reflect what would have happened had the simulated shocks occurred, one must also assume that the joint surpluses diff-in-diffs $\Theta^{D-in-D,CF}$ and marginal type distributions $n^{CF}(*)$ and $h^{CF}(*)$ that act as simulation inputs are exogenous to (i.e. unaffected by) the shock itself. Any reallocation and welfare changes are assumed to be driven exclusively by the changes in transfers across worker types required to eliminate shock-induced imbalances between supply and demand.

As discussed above, exogeneity of $h^{CF}(*)$ imposes that the shock does not cause further changes in firms' location and size decisions. To highlight heterogeneity in the scope of labor markets by firm size, average pay, and industry, below we generally consider simple "apples-to-apples" comparisons where each shock involves adding or subtracting a common number of jobs to a single position type. However, in addition to endogenizing firm responses to shock-induced wage changes, we also consider a second robustness check that incorporates product market spillovers into the simulation by adding additional service positions in locations near the original "exogenous" shock in accordance

[22] This approximation requires that the variation in utility values among workers assigned to the same worker type is limited, so that $\ln(C_l) \equiv \ln(\frac{1}{|l|}\sum_{i:l(i)=l} e^{\frac{r_i}{\sigma}}) \approx \ln(\frac{1}{|l|}\sum_{i:l(i)=l} e^{\frac{\bar{r}_l}{\sigma}}) = \frac{\bar{r}_l}{\sigma}$.

with existing job multiplier estimates from Bartik and Sotherland (2019). Other agglomeration and congestion forces could be similarly built into the stimulated shock.

There are also plausible mechanisms by which the joint surpluses Θ^{2012} might respond to the shock.[23] However, for reasonably small local shocks, the most obvious endogenous surplus changes are likely to be minuscule relative to the size of existing surplus variation in the worker types' relative productivities across firm types, relative valuations of firm and location amenities, and moving costs from alternative transitions, so that such exogeneity violations generate minimal bias. Note also that only changes in surplus diff-in-diffs Θ^{D-in-D} affect the counterfactual assignment, so that the components of endogenous changes to productivities, amenities, or continuation values among position types that are common to all workers do not affect the shock's incidence among workers.[24]

Another caveat relates to a shock's duration. We focus on forecasting reallocations and welfare changes that occur within one year of the shocks, and we assume that job matches with relocated and shock-generated positions create the same surplus as those to existing positions of the targeted position type. Implicitly, this requires that the new positions have the same expected duration as other positions of their type. To evaluate a particularly temporary construction stimulus, for example, one might wish to estimate separate surplus parameters for temporary and permanent construction jobs.[25] As is, the model is designed to show that the incidence of even very local shocks may spread quite widely across space (and skill levels) even over a short period, despite strong tendencies of movers to take nearby jobs consistent with large short-run mobility frictions.

A final, important caveat relates to the absence of a housing market in the model (and residential choices in the data). Standard models of spatial equilibrium in urban economics (e.g. Roback (1982) or Kline and Moretti (2013)) emphasize that if housing supply is quite inelastic and workers are very mobile, a large share of the incidence of a positive place-based shock accrues to landholders via higher rents (which offset the utility gains to workers from any wage increases). However, this same argument suggests that for local workers who are also nearby renters, the simulated employment-related welfare gains are likely to place an upper bound on the full welfare gains these workers would experience.[26] Thus, local low-skill workers would be well justified in resisting local initia-

[23]For example, the existence of a new establishment nearby might increase the demand for intermediate products produced by other local firms, raising the productivity of workers for such firms. Alternatively, if the new positions are permanent and search/recruiting/moving costs increase with distance, then jobs at nearby establishments might now have greater continuation value because future job searches will begin in a local area featuring greater labor demand.

[24]From the proof of Proposition 1 in Appendix A2, we see that such surplus changes only change Δ_f^2, which does not enter into equilibrium mean payoffs for worker types $\{C_l^{CF}\}$. Instead, any increase in joint surplus for a position type that is common to all workers will be fully reflected in its profit payoff, through higher revenue for the same costs or lower salaries that offset the change in worker continuation values. The profit gains among nearby position types will be understated, however. This partly motivates the focus on incidence among workers, for whom differential agglomeration effects among firms across shock compositions may be less important.

[25]Fully elucidating the differences in welfare effects between shocks of different expected durations requires a fully dynamic assignment model along the lines of Choo (2015) that specifies both worker expectations and the serial correlation in (now time dependent) idiosyncratic surplus components ϵ_{ikt}. Similarly, the Markov-style model used here cannot accommodate optimal location sequences and return migration choices highlighted by Kennan and Walker (2011). See Weinstein (2018) for an evaluation of the dynamic consequences of a particular local shock.

[26]Exceptions to this claim might occur, for example, if house price increases yielded property tax revenue that was disproportionately spent on services these workers/residents valued.

tives focused on bringing "good" jobs to town if they are likely to generate an employment-related incidence that is either geographically dispersed or concentrated among higher skilled workers.

Furthermore, house prices may not change much in places where housing supply is likely to be relatively elastic (such as rural areas or areas with weak zoning laws) or where commuting costs are low (so that adjustment to shocks occurs primarily via changing commuting patterns). In these cases, abstracting from the housing market may produce minimal bias in incidence forecasts. Indeed, since commuting costs from job transitions involving location changes are implicitly captured in the model as a component of the joint surplus θ_g, they will be reflected in incidence estimates.[27] While a complete welfare analysis requires incorporating housing and product markets, the goal of this paper is to highlight the heretofore underappreciated roles of differential geographic scopes of local labor markets for different types of workers and firms and the skill vs. spatial mismatch tradeoff in determining the incidence of alternative local labor demand interventions.

3.4 Identifying σ

While the share of welfare gains or losses for workers (or firms) can be identified without knowledge of σ, it is nonetheless a parameter of interest. Since payoffs are additive in worker earnings, knowledge of σ allows the estimated utility gains $\frac{\overline{r}_l^{CF1}-\overline{r}_l^{CF2}}{\sigma}$ and profit gains $\frac{\overline{q}_f^{CF1}-\overline{q}_f^{CF2}}{\sigma}$ to be re-scaled in dollar terms, making it easy to gauge whether the welfare changes from various labor market shocks are economically meaningful. Conditional on Θ, σ governs the elasticity of matching choices with respect to relative wages or required utility bids, so that it determines that magnitude of utility reallocations across skill types and locations following changes in labor demand composition.[28]

As Galichon et al. (2017) have noted, identification of σ requires combining information from multiple matchings, so we estimate σ using observed matchings between 2003-2004 and 2012-2013.[29] Because the procedure (described fully in Appendix A4) requires additional strong assumptions, the estimate of σ is likely to be quite rough.[30] In practice, the estimates we obtain for σ are surprisingly consistent across years. We use the mean estimate of σ^y across all sample years,

[27] Differential willingness to pay for locational amenities will be reflected in the relative propensities for different worker types to move to positions at particular locations, which are captured by the odds ratios used to identify Θ^{D-in-D}.

[28] Intuitively, when destination position type C disproportionately chooses type A workers over type B workers compared to position type D, it could be because $\theta_{AC} - \theta_{AD} \gg \theta_{BC} - \theta_{BD}$ and σ is substantial, or because $\theta_{AC} - \theta_{AD}$ is marginally larger than $\theta_{BC} - \theta_{BD}$ but σ is tiny. When the former is true, large changes in required utility bids are necessary to engender sufficient substitution across worker types to overcome strong comparative advantages from matching certain types of workers and positions. If the latter is true, small utility changes suffice to clear the market after a shock.

[29] The existence of a single σ parameter governing relative wage elasticities stems from assuming additive separability of θ_g and ϵ_{ik} and an i.i.d type 1 extreme value distribution for the vector of ϵ_{ik} values.

[30] Essentially, differences among worker types in their observed mean annual earnings changes between years $y-1$ and y are regressed on model-generated log differences in predicted scaled utility values $\ln(C_l^{CF,y}) - \ln(C_{l'}^{CF,y}) \approx (\overline{r}_l^{CF,y} - \overline{r}_{l'}^{CF,y})/\sigma^y$. These predicted values are constructed by computing counterfactual allocations in which worker and position type distributions evolve as they actually did but joint surplus values are held fixed at their estimated 2003-2004 values. Under the assumptions that a) the evolution in the utility premia enjoyed by workers in particular locations and skill categories was due primarily to changes in supply and demand composition rather than changes in the moving costs, recruiting costs, tastes, and relative productivities that compose the joint surplus values Θ, and b) mean utility gains for each worker type generally took the form of increases in annual earnings in the chosen year rather than increases in amenities or continuation values, the coefficient on $(\overline{r}_l^{CF,y} - \overline{r}_{l'}^{CF,y})/\sigma^y$ will approximately equal σ^y.

$\overline{\sigma} = 18,400$, to assign dollar values to all utility gains presented below.

As noted by GS, in the CS assignment model observed earnings in job matches also can be used to decompose each group-level joint surplus value θ_g into the sum of worker and position subcomponents θ_g^l and θ_g^f (group-level analogues of π_{ik}^i and π_{ik}^k). However, in Appendix A5 we show that clean identification of θ_g^l and θ_g^f breaks down without the particular structure CS place on the unobserved match quality component ϵ_{ik} unless further strong assumptions are imposed. We do not pursue this approach because we have shown that this decomposition is not needed to determine the incidence across worker and position types of alternative local labor demand shocks.

4 Data

We construct a dataset of year-to-year worker job transitions (pairs of primary jobs in consecutive years) using the Longitudinal Employer-Household Dynamics (LEHD) database. The core of the LEHD consists of state-level records collected for unemployment insurance purposes containing quarterly job earnings and unique worker and establishment IDs for a near universe of jobs in the state.[31] The worker and establishment IDs are then linked across states, and the data are augmented with firm and establishment characteristics (notably establishment locations and industry codes) from an extract of the ES-202/QCEW report and worker demographics from the Social Security Administration (including age, race and sex but not occupation nor education for most workers).[32]

4.1 Sample Selection

Our sample consists of all LEHD records from the 19 U.S. states that opted to provide data to our FSRDC project.[33] An individual is included in the initial sample if he/she is ever observed as employed in one of the sample states in any of the years for which data was provided, and for disclosure avoidance reasons a 50% random sample of all initial sample members is taken to form the final sample. Unlike previous versions, the 2014 LEHD snapshot (still the most recent available to external researchers) includes a file that indicates whether a worker was employed in some U.S. state in each quarter, even among states not providing records to a particular project, as long as the state has provided UI data to the Census Bureau. This allows job-to-job transitions into and out of the 19 observed states to be distinguished from transitions to and from nonemployment.

While the estimation of σ and the model validation exercise described in Section 6.7 use all the data from 2003 forward (when the last sample state begins reporting data), the simulations of counterfactual shocks are based on surplus parameters estimated from 2012-2013 data. Preliminary work suggested that the shock incidence forecasts were quite insensitive to the pair of years chosen.[34]

[31] The database does not include farm jobs, self-employed workers, or federal employees.

[32] For further details about the contents and creation of the LEHD, see Abowd et al. (2009) and Vilhuber et al. (2018).

[33] By agreement with the Census Disclosure Avoidance Review staff, the identities of the states cannot be revealed, but they include large, medium, and small states, and are spread throughout the U.S., albeit unevenly.

[34] This was true despite the decreasing job-to-job mobility over this time period documented by Hyatt et al. (2016).

To create year-to-year transition/retention observations, we select each worker's highest earnings job in each year among those lasting at least one full quarter, sum earnings from this primary job across all four quarters, and then append the analogous primary job from the next year.[35] Workers are considered nonemployed in a given year if they did not earn above $2,000 at any job in any full quarter in any observed state and are not reported as employed in some out-of-sample state.

We limit each worker's presence in the sample to begin and end with his/her first and last years of observed employment. We exclude nonemployment spells before and after any observed employment in an attempt to remove workers who are out of the labor force. We also drop observations featuring workers with ages below 20 or over 70. These choices limit the influence of "nonemployment" spells consisting of full-time education or retirement followed by part-time work, so that parameters related to unemployment are identified primarily from prime-aged workers who were unemployed or temporarily out of the labor force.

Since most of the results presented below rely on parameters estimated using 2012-2013 matches and sample coverage ends in 2015Q1, excluding nonemployment spells without an observed resumption of employment may cause a slight undercount of E-to-U and U-to-U transitions, since a small share of unemployed workers in 2013 likely remained in the labor force but did not find jobs by 2015Q1. We address this by using data from the American Community Survey, which distinguishes unemployment from labor force departure, to construct estimated counts of E-to-U and U-to-U transitions for each combination of initial U.S. state, destination U.S. state, 5-year age bin, and initial earnings category (for E-to-U transitions only). Since match groups defined by these characteristics are coarser than those in the model, we use the E-to-U and U-to-U transitions in the LEHD only to distribute the ACS group counts across the finer groups in the model. We supplement the ACS data with BLS national unemployment counts by age group to align the scale of the labor force with standard measures. Appendix A6 further describes these imputation procedures.

A drawback of the LEHD sample is that the establishment characteristics and pay of employed workers are only observed among the 19 states providing data. Rather than exclude job-to-job transitions into or out of the 31 out-of-sample states, which would cause the counterfactual simulations to overstate the geographic concentration of demand shock incidence by removing effects on out-of-sample workers, we aggregate all employment in the remaining 31 states into a single out-of-sample "state" and "tract". As with flows to unemployment, we use aggregated counts from the ACS to set the scale of flows between in-sample and out-of-sample states, and combine the ACS counts with the LEHD data to impute the joint distribution of worker and firm characteristics among flows into and out of each in-sample census tract.[36] Because forecasts of shock incidence may be particularly sensitive to observing flows of workers to and from states adjacent to the focal state, the simulations below generally sample target census tracts only from 10 states in the west/southwest/great plains area where coverage is nearly complete, so that almost all adjacent states are observed.[37]

[35] A job is observed in a full quarter if it features positive earnings in the preceding and following quarter as well.

[36] This procedure is also described further in Appendix A6.

[37] Using ACS 1-year residential mobility data and weighting states by their census tract count, we estimate that for

4.2 Assigning Workers and Positions to Types and Job Matches to Groups

For each pair of years $(y-1, y)$ we assign each observation to a worker type $l^{y-1}(i)$, a position type $f^y(k)$, and a match group $g^{y-1,y}(i,k)$ (year superscripts will henceforth be dropped except where necessary). Workers are assigned to worker types based on the combination of their year $y-1$ primary establishments' locations (discussed in Section 5.1) and the earnings quartile associated with their earnings from this establishment.[38] For workers who were not employed in $y-1$, the location of their most recent establishment is used (or, for new entrants, the location is imputed using ACS/LEHD data) and the earnings quartile is replaced by one of two unemployed categories, differentiated by age (< 25 or ≥ 25) to distinguish new entrants/recent graduates from more experienced workers, since employers might treat the two as quite imperfect substitutes. Workers' year y primary positions are assigned to position types based on the combination of their establishment $m(k)$'s location, industry supersector, and size and average worker earnings quartiles based on establishment-level employment and average pay distributions. These characteristics were chosen because they are consistently observable and likely to be important determinants of productivity complementarities, recruiting, search and moving costs, and the other components of the match surpluses that heterogeneous positions create with heterogeneous workers. The match (i,k) is assigned to a group $g(i,k) \equiv g(l(i), f(k), z(i,k))$ based on the worker type $l(i)$, the position type $f(k)$, and indicator $z(i,k) \equiv 1(m(j(i)) = m(k))$ for whether the match is a retention or a job change.

4.3 Summary Statistics

Column 1 of Table 3 (Figure A1a) presents the distribution of distance between the locations of origin and destination establishments for workers who changed primary jobs $(m(j) \neq m(k))$ between 2012 and 2013. 3.4% of job switchers took new jobs within the same census tract, while another 5.6%, 6.3%, and 12.6% percent moved to jobs one, two, or 3+ tracts away within the same PUMA. 57.3% found jobs in another PUMA within the same state, while 14.8% changed states. The sizable share of workers accepting new jobs very near their previous jobs is prima facie evidence that either search/moving costs are large or preferences for particular locations are strong, so that conditions in workers' local labor markets may still hold outsized importance for their job-related welfare.

Row 1 of Table 2 Panel A shows that 15.6% of sample observations involve job-to-job transitions. A full 69.4% of workers keep the same primary job from one year to the next, while 9.3%, 2.8% and 2.9% of observations involve U-to-E, E-to-U, and U-to-U transitions, respectively. Collec-

the 10 states supplying target tracts, about 47% of year-to-year worker inflows from other states and about 92% of total job-to-job changes ending in one of these 10 states (including within-state flows) originate in one of 19 in-sample states.

[38] Earnings quartile cutoffs are defined relative to the distribution of primary job annual earnings among all sample workers in year $y-1$, and are based on prorating earnings from full quarters only to ensure that the quartile captures a worker's skill rather than the share of the year he/she worked. A worker's location must be imputed for multi-establishment firms. However, the Census Bureau's unit-to-worker imputation procedure assigns establishments with probabilities that depend on the distance between that establishment and the worker's residence, so any mistakes will likely misattribute the worker's job to another nearby establishment, limiting scope for significant measurement error. We exploit the provided Successor-Predecessor file to reclassify as retentions any spurious job-to-job transitions stemming from changes to a firm's structure that do not alter a worker's location.

tively, the 2012-2013 estimation sample for the set $\{\Theta^{D-in-D}\}$ features 24.2 million observations.

Examining other rows of Panel A, we see that about 90% of workers younger than 25 who were unemployed in 2012 found jobs in 2013, while only 66.2% of older unemployed workers transitioned to a job in 2013, highlighting the need to consider these two groups of unemployed workers separately. We also see that employed workers in the lowest earnings quartile in 2012 were far less likely than higher paid workers to stay at their job (65.5%) and far more likely to transition to unemployment (6.5%) or another job (28.0%); 87.0% of highest quartile workers in 2012 were retained, while only 11.7% transitioned to new jobs. However, conditional on changing jobs, the highest earners were most likely to leave their original PUMA and to change states, suggesting that the geographic scope of labor markets may differ across skill categories. These differences motivate the use of nonemployment status/earnings quartiles as characteristics defining worker types.

Panel B of Table 2 shows that highest paying quartile of establishments are considerably more likely to retain their workers, but are slightly more likely to hire distant workers when filling a vacancy: 9.3% of their new hires were previously working out of state and 25.8% were working in the same PUMA, compared to 6.2% and 32.7% for the lowest paying quartile. The largest quartile of establishments (based on employment) are more likely than their smaller counterparts to retain workers, but are the least likely to hire from within the same PUMA (20.8% of new hires vs. 37.9% for the smallest quartile). Interestingly, an establishment's size poorly predicts its tendency to hire from out of state (7.3% vs. 6.9% for the top vs. bottom quartiles). While these statistics motivate the choices of types and the need for labor demand shocks featuring different establishment compositions, they do not condition on any other firm, location, or worker characteristics. Comparing incidence across counterfactual shocks that hold all but one characteristic fixed will be more informative about how the scope of labor markets differs across types of workers and establishments.

5 Estimation

5.1 Collapsing the Type Space for Distant Geographic Areas

Since match groups g are defined by several other worker and establishment characteristics in addition to initial and destination locations, treating all 28,000 census tracts in the 19 state sample as separate locations would generate trillions of match groups. Given the particular interest in the incidence of alternative demand shocks across locations relatively close to the shock, we combine initial types (and thus groups) that share the same worker and establishment characteristics and are geographically proximate to each other but far from the shock. Specifically, beyond a five tract radius surrounding the targeted tract, a type's location is defined by a PUMA rather than a tract. Beyond the targeted state, a type's location is defined by a state rather than a PUMA.

Coarsening the type space for distant geographic locations dramatically decreases the total number of groups and the sparsity of the empirical group distribution $\hat{P}(g)$. While many job-to-job transitions are between nearby tracts, very few transitions occur between most distant pairs of tracts, so relative surplus parameters for match groups featuring tracts in different states would be weakly identified without such coarsening. Note that this approach still incorporates all available job matches and all locations in the 19 state sample as well as the out-of-sample "state" into each simulation, so that each local labor market is still nested within a single national labor market.

Even after combining types, there are relatively few observed matches per group g, particularly for groups local to the shock. Thus, following Hotz and Miller (1993) and Arcidiacono and Miller (2011), we smooth $\hat{P}(g)$ prior to estimation by replacing each element's value with a kernel-density weighted average of $\hat{P}(g)$ among groups featuring "similar" worker and position characteristics.

Because excessive smoothing across other groups erodes the signal in the data about the degree of heterogeneity in the joint surpluses from job matches featuring different worker and position characteristics and locations, we create a customized smoothing procedure, detailed in Appendix A7. It is based on the intuition that the destination establishment's location will be critical in determining the origin locations from which worker job matches create the most surplus (i.e. least moving/search cost), while non-location characteristics (size, average pay, and industry) matter more than location for determining the worker initial earnings/unemployment category that generates the most surplus.

Note that the type aggregation and smoothing procedures imply that type and group spaces change whenever a new target tract is selected. Furthermore, FSRDC disclosure restrictions prevent the release of results that are specific to a particular substate location. Thus, we only report averages of incidence measures across 500 simulations for each shock type, where each simulation targets a different randomly chosen tract from the 10 state southwest/west/great plains subsample.[39]

After the simulations have been run, match groups are redefined in order to average simulation results across alternative targeted tracts. Worker and position type locations are replaced with bins capturing distance to the targeted tract, and we report estimates of incidence for various distance rings surrounding the shock.[40] During the simulations the spatial links between adjacent and nearby tracts are left entirely unrestricted. Thus, no a priori assumption about the role of distance is imposed by the model beyond the initial aggregation of distant tracts to PUMAs and states.

5.2 Defining the Local Labor Demand Shocks

Baseline simulated shocks either add or remove 500 jobs from the stock of positions to be filled in a chosen census tract and remove or add 500 national unemployment "positions".[41] This represents about a 20% increase in labor demand for an average tract with around 2,500 jobs. For each chosen tract, we first simulate 38 "stimulus packages" featuring different kinds of new establishments

[39] A census tract is only eligible to be a target tract in the simulations if it features at least 250 jobs, so that the parameters governing local firms' and workers' choices are well-identified. The same set of 500 randomly chosen target tracts is used for each shock specification to facilitate fair comparisons among alternative specifications.

[40] When defining types and presenting results we mostly focus on distance bins defined by tract, PUMA, and state pathlengths, since the number of potential workers contained within circles defined by the same pathlength is likely to be more consistent across urban and rural areas featuring very different densities than circles with a miles-based radius.

[41] We experimented with "plant relocations" that move jobs to a new location from a distant state. These shocks had nearly identical employment and welfare incidence to their stimulus analogues among workers within the receiving state.

represented by combinations of the non-location establishment attributes that define a position type: quartiles of establishment size and average pay along with industry supersector. Table 1 details the composition of each shock. The first 32 shocks examine the heterogeneity in incidence across industry/size/avg. pay cells. The next three packages add a requirement that the new positions may only be filled by workers from the surrounding PUMA, reflecting stipulations included in some economic development contracts between cities and incoming firms.[42] Comparing these "restricted" specifications to their unrestricted counterparts illustrates the value of these provisions to cities or states. The final three stimulus packages are robustness checks that evaluate sensitivity of results to key model assumptions. Finally, to examine asymmetry between positive and negative shocks as well as sensitivity to shock scale, we consider several pairs of analogous positive and negative shocks of various magnitudes involving either large high-paying manufacturing firms ("plant openings" and "plant closings") or large low-paying retail firms ("store openings" and "store closings").

5.3 Inference

Given that we observe the universe and not a sample of job matches within the available states, it is unclear how to define the relevant population for the purposes of inference. Furthermore, since we estimate nearly a million surplus parameters $\theta_g \in \Theta$, and each counterfactual incidence statistic depends on the full set Θ, any confidence intervals should provide information about the precision of incidence forecasts as opposed to specific parameters. Rather than characterizing sampling error in isolation, we rely on the model validation results presented in section 6.7 to assess the combined contribution of sampling error and misspecification to out-of-sample forecast accuracy.[43]

6 Results

6.1 How Local Are Labor Markets? Aggregated Incidence by Distance to Focal Tract

Before comparing stimulus packages of different establishment compositions, we focus first on characterizing the geographic scope of labor markets for a "typical" local stimulus by averaging the predicted changes in assignments across the 32 baseline stimuli. This effectively integrates over the joint distribution of establishment industries, sizes, and average pay levels. While we focus on tabled results, many tables have accompanying Appendix figures (listed in parentheses).

Column 3 of Table 3 (Figure A2a) displays the mean probability of receiving one of the 500 new stimulus jobs among individuals initially working or seeking employment at different distances from the focal tract. The table highlights a sense in which U.S. labor markets are quite local: the probability of obtaining a stimulus job is four times higher for a worker initially within the target

[42] For example, Empowerment Zones only subsidize wages for employees that are local residents (Busso et al. (2013)).

[43] The one exception is that in the first few results tables, we provide standard errors that only reflect the sampling error stemming from averaging over one possible draw of 500 census tracts, while the population parameter of interest is the average among all tracts in the 19 state sample (around 28,000). These standard errors are tiny, suggesting little value to running more than 500 simulations per specification. As a result, subsequent tables do not report standard errors.

tract (.013) than for one in an adjacent tract (.003), over 6 times higher than for a worker 2 tracts away (.002), and 14 times higher than for one initially 3 or more tracts away within the same PUMA (.001). Furthermore, additional distance from the focal tract continues to matter at greater distances: the probability of obtaining a stimulus job is 23 times higher for a local (target tract) worker than for one from an adjacent PUMA, 46 and 273 times higher than for a worker two PUMAs away or 3 or more PUMAs away within the same state, respectively, and 2,075 and 13,330 times higher than for a random worker one state or two or more states away, respectively.

However, the vast differences in $P(\text{new job }|\text{distance from target})$ present a misleading guide to the overall geographic incidence of new jobs. This is because the target tract initially contains only 0.0021% of the workforce at risk of obtaining these jobs, while other tracts within the same PUMA contain 0.146%, other PUMAs within the state contain 5.95%, and other states contain 93.8% (Table 3, col. 2, Figure A1b). Consequently, one obtains a very different impression of incidence by swapping the terms in the conditional probability and calculating the share of stimulus jobs obtained by workers initially working or nonemployed in each distance bin, $P(\text{distance from target }|\text{new job})$.

Column 4 of Table 3 (Figure A2b) displays the mean share of new jobs by distance bin across the 32 simulated stimulus packages. 3.3% of new jobs are obtained by workers from the target tract, another 22.9% by other workers in the PUMA, 55.9% by workers in different PUMAs within the state, and 17.8% by out-of-state workers. So a very large share of the new jobs are likely to be taken by workers far from the local jurisdiction that receives the stimulus.

One might obtain similar forecasts of the shares of workers by distance bin who would obtain jobs at a new establishment simply by looking at the distance composition of workers who obtained jobs from actual past plant openings. As emphasized in the introduction, the probabilities of obtaining the particular new jobs created by the stimulus package may not be very informative about the true incidence of the shock. This is because many workers who take the new jobs would have obtained other similar jobs in the absence of the stimulus, and other workers may now obtain these jobs, and so on, creating ripple effects through vacancy chains that determine the true employment and welfare incidence. This is where a flexible equilibrium model provides additional insight.

Column 5 of Table 3 (Figure A3a) is analogous to column 3 (Figure A2a), except that instead of the probability of obtaining a stimulus job, it captures the change in the probability of any employment due to the stimulus, relative to a no-stimulus counterfactual, by workers' distance from the target tract. The change in employment rate is still quite locally concentrated, but less so than the probability of obtaining a stimulus job. Workers from the target tract are 0.18% more likely to be employed at the end of the year than in the absence of the stimulus. This is 3.1, 5.0, and 8.2 times greater than the employment rate changes for workers 1, 2, or 3+ tracts away within the same PUMA, 11, 18, and 51 times greater than for workers 1, 2, or 3+ PUMAs away within the same state, and 168 and 436 times greater than for workers one state and 2+ states away, respectively. In particular, the odds of net employment gains for workers 2+ states away relative to focal tract workers are 30 times higher than they were for the probability of obtaining a stimulus job.

Column 6 of Table 3 (Figure A3b), the employment rate analogue to column 4 (Figure A2b), displays the share of the 500 job increase in national employment that accrues to workers initially in each distance bin. Only 0.5% of the net employment change redounds to workers from the target tract, with 5.2% of employment gains going to workers in other tracts within the PUMA, 33.8% to workers in other PUMAs within the target state, and 60.6% to workers from out of state.

The simulation procedure also generates counterfactual changes in mean job-related utility for each worker type following the various stimuli.[44] Since only relative changes are identified, the estimated utility impact is normalized to 0 for the worker type estimated to experience the smallest impact, which varies by the stimulus composition, but is generally young, initially unemployed workers in some distant state. Thus, all presented utility changes are relative to this worker type.

Column 7 of Table 3 (Figure A4a) provides the average utility impact, scaled in $ of 2012 annual earnings, by distance bin from the target tract for the "typical" stimulus package. Workers from the focal tract receive an estimated $1,181 increase in money metric utility (relative to the least affected worker type), while workers initially 1, 2, and 3 or more tracts away receive expected utility gains of $377, $210, and $99 respectively. Workers initially 1, 2, and 3+ PUMAs away within the state receive the utility equivalent of $64, $41, and $22 in annual earnings gains, while workers one state away, 2+ states away, and out-of-sample receive average gains of $3, $0.4, and $0.4, respectively. Column 8 (Figure A4b) plots the share of total utility gains (relative to the normalized type) that accrue to workers in each distance bin. Only 0.9% of total worker welfare gains accrue to workers from the focal tract, with another 9.0% accruing to other workers originally within the PUMA of the focal tract. 54.5% of the gains accrue to workers from other PUMAs within the same state, while 35.6% go to workers from out of state. Thus, examining welfare incidence rather than employment incidence suggests a slightly less geographically integrated labor market.

Table 4 provides the expected employment and welfare gains and the shares of total employment and welfare gains among distance bins constructed based on miles from the focal tract rather than tract, PUMA, or state pathlength. The story is the same: only 6.3% of employment gains and only 10.7% of welfare gains accrue to workers within 10 miles of the target tract even though 27.9% of stimulus jobs are filled by such workers. 73.8% of employment gains and 54.6% of welfare gains accrue to workers initially more than 250 miles away or in an out-of-sample state.

Table 5 (Figure A5) illustrates the impact on stimulus incidence of requiring incoming establishments to fill their positions using only workers from the same PUMA as the tract receiving the stimulus. The unemployment rate for target tract workers falls by 1.0% instead of 0.2%, while the decrease in unemployment rate is 2-3 times as large in the restricted vs. unrestricted version of the stimulus for workers from other tracts within the PUMA. Overall, the within-PUMA share of net employment gains increases from 5.9% to 17.2% when hiring is restricted to local PUMA workers.

Restricting hiring also increases the expected money metric utility gains by over eight-fold

[44] Recall that scaling utility premia in dollars of annual earnings requires estimating σ. As discussed in section 3.4 and Appendix A4, the assumptions underlying the estimate of σ are stronger than for the relative joint surplus values (they are more like approximations), so the estimated dollar values of predicted welfare gains should be treated cautiously.

($1077 to $9142) for workers from the focal tract, with 2-4 fold increases in utility gains for other workers initially in the PUMA, depending on initial distance from the target tract. The share of utility gains accruing to the targeted PUMA increases from 9.7% to 27.6%. Thus, local development initiatives such as empowerment zones that add stipulations restricting hiring or wage subsidies to only the local workers seem likely to generate a much more locally concentrated labor market incidence. This is despite the fact that any additional downstream hiring caused by initially employed workers vacating jobs to take the 500 new positions remains unrestricted in these counterfactuals.

6.2 Heterogeneity in Incidence by Worker Initial Earnings Category

Table 6 (Figure A6a) provides separate distance profiles of employment rate incidence among workers whose initial jobs (or unemployment) place them in different earnings quartile/distance bin combinations. Employment gains decline with distance in a fairly similar pattern for all initial earnings quartiles and both younger and older initially unemployed workers. However, the magnitude of employment gains varies considerably by initial employment/earnings status. Older unemployed workers from the focal tract enjoy a sizable 0.97% decrease in their unemployment rate, while the decrease is around 0.25% for both the younger (age <25) unemployed local workers and the lowest paid quartile among employees, and falls monotonically to only 0.05% for the 4th (highest) earnings quartile. Column 1 of Table 7 (Figure A7a) captures the share of the 500 job net employment gain enjoyed by workers in each initial earnings quartile/unemployed status (aggregated across distance bins). Age 25+ (<25) unemployed workers account for 33.7% (10.0%) of total employment gains, despite constituting 6.4% (3.2%) of the workforce, while earnings quartiles 1-4 account for 24.6%, 14.4%, 9.1%, and 8.3%, respectively. The smaller values for initially high earning workers reflect the fact that they were much less likely to become unemployed in the absence of the shock.

Table 8 (Figure A6b) displays the welfare analogue to Table 6 (Figure A6a), while column 6 of Table 7 (Figure A7b) shows the corresponding national shares of welfare incidence by initial earnings status. In contrast to the employment results, the high paid workers seem to receive a disproportionate share of the welfare gains from a typical shock. Among focal tract workers, the highest paid quartile enjoy expected gains of $1443 compared to $1129 for the lowest paid and $1119 and $1319 for younger/older unemployed workers. Here we see evidence of more geographically integrated labor markets for higher paid workers, as their welfare gains decline faster with distance, converging quickly to the profiles of lower paid and unemployed workers. Overall, the top earnings quartile accounts for 26.8% of welfare gains, with quartiles 1-3 accounting for 18.1%, 20.8%, and 22.0% and younger and older unemployed workers accounting for 5.9% and 6.4%.

6.3 Heterogeneity in Incidence by Establishment Characteristics

Columns 2-9 of Table 9 (Figure A8a) shows the utility gains only for workers initially in the target tract by initial employment status/earnings quartile of the worker and industry supersector of the stimulus. The rankings of utility gains for local workers across industries differ strikingly across unemployment/initial earnings categories. For example, stimuli featuring positions in the construction and other services industries yield utility gains for older unemployed workers equivalent to \$1,549 and \$1,421 in annual earnings, compared to \$1102 for stimuli involving retail/wholesale trade.[45] By contrast, construction and other services offer the lowest welfare gains to younger unemployed workers likely to be new entrants (\$905 and \$986), with leisure/hospitality and retail/wholesale trade stimuli delivering the greatest payoffs (\$1479 and \$1261). Among employed workers, leisure/hospitality, government, and information positions produce the highest payoffs for quartiles 1, 2 and 3, and 4, respectively.

Figure A8b (Table 10) shows the expected utility gains for workers in the focal tract by establishment size and pay quartile combinations. As expected, the highest paying quartile of firms (regardless of size) generate much larger gains for initially high paid workers and smaller gains for both low-paid and unemployed workers. Stimuli featuring positions at large, high paying firms create much smaller payoffs for initially unemployed workers (\$934 and \$987 for young and older unemployed workers, respectively), while generating a substantial \$1793 for 4th quartile workers. Small, low paying (2nd quartile) firms show the opposite pattern, with payoffs of \$1367, \$1731, and \$976 for the same three groups. Firm size seems to be most important for experienced unemployed local workers; they reap considerably larger utility gains from jobs created at the smallest (1st quartile) firms relative to the largest (4th quartile), suggesting that local development projects may help such workers more by encouraging startups than luring one large establishment to open or relocate.

In addition, the substantial heterogeneity in local skill incidence across industries and size/pay categories misses further heterogeneity at the three-dimensional supersector/size/pay cell level. Figure A9 plots welfare gains by skill proxy among workers from the focal tract for all 32 stimulus compositions. The range of predicted gains is huge. Welfare gains for young unemployed workers range from \$696 (small, high paying construction positions) to \$1958 (large, low paying wholesale/retail), while their older counterparts exhibit a range from \$666 (large, high paying wholesale/retail) to \$2081 (small, low paying other services). For 1st earnings quartile workers, they range from \$670 (large, high paying manufacturing) to \$1665 (small, low paying leisure/hospitality). For 4th quartile workers, they range from \$619 (large, low paying wholesale/retail) to \$2707 (small, high paying information). For small precinct councilors concerned with very local incidence, these results represent massive differences in the scale and skill intensity of utility incidence that demonstrate the need to tailor the design of economic development packages to the targeted local subpopulation, and that would be obscured by an analysis that ignored skill heterogeneity or used coarser geography.

On the other hand, city-level and particularly state- and federal-level policymakers may safely ignore incidence heterogeneity, since shocks featuring different establishment industry, size, and average pay composition exhibit increasingly similar spatial and skill incidence at distances greater than two or three tracts from the stimulus site. Aggregating across initial earnings/employment status and size/average pay combinations, Table 11, Col. 2-9 shows that the change in employment rate among workers at different distances varies shockingly little across stimuli featuring differ-

[45] "Other Services" includes repair and maintenance, personal and laundry services, and religious/civic organizations.

ent supersectors, with all industries displaying within-PUMA shares of employment gains between 5.0% (retail/wholesale trade) and 6.0% (education/health) and within-state shares between 39.0% and 40.0%. Analogous results for utility (Table 12, Col. 2-9) display a similar uniformity among industries in spatial incidence away from the shock. Similarly, when focusing on skill incidence and averaging over locations, all supersectors feature shares of employment and utility gains accruing to each skill category within 1% of the overall average across all stimuli. Essentially, the set of positions vacated by workers taking stimulus jobs better approximate the U.S. establishment distribution than the original shock, and each successive ripple of shock-induced reallocation yields an increasingly generic composition of vacated positions.

Changing the establishment size/pay composition of stimuli also produces limited heterogeneity in geographic, and more surprisingly, skill incidence beyond tracts close to the site of the shock (Tables 7, 13, and 14). Stimuli featuring low-paying firms only generate 1-2% higher shares of employment and welfare gains for initially low-paid or unemployed workers relative to those featuring high paying firms (Table 7). By contrast, initially unemployed workers are predicted to take 28% of stimulus jobs at high paying firms versus 38% for low paying firms, and the initially highest paid workers take 29% vs. 5% of jobs at high vs. low paying firms, suggesting that the skill incidence of the particular stimulus jobs understates the degree to which employment gains from labor demand shocks biased toward high skilled workers "trickle down" to unemployed workers.

6.4 Heterogeneity in Incidence by Focal Tract Characteristics

Heterogeneity in geographic incidence also stems from the choice of focal tract. Tables 15 and 16 (Figure A10) provide the mean employment and welfare incidence among the least dense (most rural) and most dense (most urban) 100 tracts of the 500 receiving shocks, as well as among the 100 tracts with smallest/largest # of jobs within 5 miles, average two-bedroom rent and poverty rates.

Both welfare and employment gains are more geographically concentrated for tracts with lower population density. The expected utility gain for workers within the focal tract, 1 tract, 2 tracts, and 3+ tracts away within the PUMA are all several times larger for the most rural relative to the most urban focal tracts (\$2765 vs. \$818, \$933 vs. \$79, \$408 vs. \$62, and \$137 vs. \$45, respectively). The differences in expected welfare gains among nearby workers are even larger for tracts featuring few vs. many jobs within 5 miles (e.g. \$3029 vs. \$558 for focal tract workers). These large differences partly stem from the fact that 500 additional jobs represents a larger per-worker shock to low density areas, since they tend to have fewer total workers in the focal and surrounding tracts. However, substantial density-based differences also exist in shares of welfare and employment gains enjoyed by within-PUMA workers, so that larger per-worker gains in low density areas more than offset smaller shares of the national labor force: the average share of welfare (employment) gains enjoyed by workers from the targeted PUMA is 14.4% (7.8%) among the 100 most rural tracts versus 5.9% (3.5%) for the 100 most urban tracts (and 9.9% (5.7%) among all 500 tracts selected).

Comparisons for tracts with low vs. high average rent for a two bedroom apartment closely mirror the rural/urban results. Since low rent may indicate a high elasticity of housing supply, the job-related welfare gains may better approximate total welfare gains for such tracts. Tracts with higher poverty rates exhibit larger mean welfare gains for local workers as well as larger shares of gains that stay within PUMA, suggesting that targeting local development projects at high poverty areas may yield greater local labor market benefits than for the representative U.S. tract.

Since residential sorting leads to high correlations among many tract characteristics, Table 17 reports the results of a set of regressions that relate various measures of shock incidence to a broader set of focal tract characteristics, where each has been standardized to have zero mean and unit standard deviation to ease coefficient comparability. To improve power, the sample here consists of 2,500 plant opening simulations featuring 500 new positions at large, high paying manufacturing firms but different randomly selected focal tracts.

Columns 2-5 confirm that the unconditional correlations from Tables 15 and 16 also survive as partial correlations: one s.d. increases in two-bedroom rent and population density still strongly predict lower average welfare gains for target PUMA residents ($36 and $237, respectively) and target PUMA shares of total welfare gains (1.5% and 2.7%, respectively) even conditional on each other and other tract characteristics. Similarly, both lower median household income and higher poverty rates predict larger welfare gains for target PUMA residents, and a greater number of jobs w/in 5 miles continues to predict considerably smaller average welfare gain within the target PUMA, and is in fact a stronger predictor of smaller local incidence than job density within the target tract.

Columns 6-9 focus more narrowly on employment and welfare gains for initially low-paid workers in the target PUMA, and show similar patterns, but with larger coefficient magnitudes for employment gains and smaller for welfare gains, consistent with the initial earnings incidence results from above. However, columns 7-10, which present mean incidence among all U.S. low-paid workers, reveal an interesting result: tract characteristics that predict greater welfare gains for local low-paid workers tend to predict smaller gains for low-paid workers nationwide. One possible explanation is that these characteristics may predict higher search costs that make it appealing to hire initially low-paid local workers rather than more distant low-paid or even high-paid workers from farther away (since jobs vacated by high paid workers may be taken by their lower-paid neighbors).

This finding highlights the limitations of a reduced-form difference-in-difference approach that compares focal areas with "untreated" distant areas before and after a set of local shocks. Specifically, regressions capturing treatment effect heterogeneity on low-paid workers might lead authors to make incorrect inferences about which choice of focal areas would best alleviate poverty, since larger gains for the local poor in certain local areas captured (and slightly overstated!) by such regressions would be outweighed by smaller expected gains among many less proximate workers.

Table 18 mimics Table 17 but replaces "plant opening" stimuli with "store openings" featuring positions at large low-paying retail firms so as to test the importance of mismatch between the skills of local workers and those required by the newly created jobs. Evidence of a role for mismatch is weak: focusing on the columns 6-9 that capture incidence for initially low-paid local workers, the coefficients on poverty rate barely change from Table 17, while lower median income predicts incidence among local low paid workers only slightly more strongly than before. Changing the shock's firm composition also does not affect the relationship between tract characteristics and low-paid workers' national share of total employment and welfare gains.

Finally, focusing on contrasts among observed tract characteristics masks additional unexplained heterogeneity in incidence among alternative focal tracts. For each shock specification, the within-PUMA shares of employment gains range from below 2% to around 10% and the within-state shares (partly driven by state size) range from below 25% to above 55%, though these ranges may partly reflect sampling error. Shares of welfare gains display even greater variation: within-PUMA shares range from below 2% to near 20% and within-state shares range from below 40% to above 80%.

6.5 Heterogeneity in Incidence by Sign and Magnitude of the Shock

Columns 3, 6, 9, and 12 of Table 19 (Figure A11) display the change in employment rate and expected welfare by distance from the focal tract for both "plant openings" and "plant closings" consisting of the creation or destruction of 500 positions at large, high paying manufacturing firms. These estimates average across 200 focal tracts randomly selected from the smaller set of tracts that initially contained at least 500 positions of this position type, so that such openings and closings are realistic scenarios for tracts in this set. Because these tracts have high baseline job counts, the plant opening represents a smaller percentage change, and so only raises the employment rate and expected welfare gain among workers initially in the focal tract by 0.03% and $310, respectively. However, a dramatic asymmetry is instantly apparent in the figure: the equivalently-sized plant closing lowers the employment rate for such workers by 0.6%, and they lose a whopping $12,300 in annual earnings-scaled welfare. The share of total employment and welfare gains/losses among focal tract workers are 0.4% and 1.0% for the plant openings and 7.1% and 27.4% for the plant closings. By contrast, the gains and losses from these plant openings and closings exhibit similar magnitudes and patterns of decay with distance among workers outside the focal tract.

What is the source of this asymmetry? Plant openings or expansions require new workers to be hired, and even hiring local workers generates substantial search and training costs, so that the surplus from hiring them is only moderately larger than for more distant workers, and relative labor demand for more local workers increases modestly. To capture this fact, the plant opening simulations and the stimulus packages analyzed above impose that newly created positions cannot be filled by "job stayers" (match groups featuring $z(g) = 1$). On the other hand, plant closings eliminate what had been a large source of joint surplus from retaining existing workers, since recruiting and moving costs had already been paid and workers had developed considerable firm-specific skill. The very high retention rate in all industries chronicled in Table 2 is a manifestation of the generally large surpluses from maintaining existing matches. Thus, as the mass layoff literature has also documented, with considerably inferior outside options, laid off workers suffer large welfare losses.

In a companion paper featuring an analogous labor market assignment model (Carballo and

Mansfield (2021)), we show that the asymmetry completely disappears when we pool the previously distinct joint surplus values associated with retention and with replacement by a worker of the same type (operationalized as setting $z(i, k) = 0$ for all job matches). This asymmetry illustrates the importance of both distinguishing retention from replacement by a similar worker and using job-level microdata rather than aggregate counts of job matches by worker type-firm type combination.

Table 19 (Figure A12) shows how the geographic employment and welfare incidence evolves as the shock size is scaled from 125 to 250 to 500 positions. For both openings and closings, the changes in employment rate and expected welfare scale nearly linearly with increased shock magnitude. Closings do exhibit a slight convexity in local employment rate changes with scale, as the share of employment losses accruing to focal tract workers rises from 5.9% to 6.4% to 7.1% for the three shocks. For smaller shocks, local workers are less likely to be the marginal workers; they disproportionately retain the remaining jobs at the expense of distant workers who would have been hired in the absence of the closing but whose transitions to the focal tract generate less surplus. As shock size grows, the local workers become the marginal workers. By contrast, the local share of welfare gains is slightly concave in shock size, since larger shocks cause enough of an exodus to meaningfully affect labor supply to more distant areas. Combining the nearly linear relationship between shock size and average impact with the urban/rural differences in geographic concentration of incidence above, the results suggest that targeting several rural areas with small development initiatives might generate larger per-worker local employment and welfare gains than a single larger plant opening in a densely populated area (barring large differences in job multipliers).

Table 20 (Figure A13) provides separate estimates of geographic employment and welfare incidence by initial earnings/unemployment category for 500 position plant closings of large high-paying manufacturing firms. Increases in unemployment rates are quite small for unemployed workers (since they were less likely to become employed in the absence of the shock). They are slightly higher for high-paid than low-paid workers locally, but this relationship reverses away from the target tract. High-paid workers are more likely to lose their jobs when positions at high-paying firms are removed, but in general initially high-paid workers are able to outcompete initially low-paid workers for now scarcer positions, so that the employment incidence passes down the skill ladder. Indeed, among out-of-state workers, the lowest-paid quartile are ten times more likely to endure shock-induced employment loss than the highest-paid quartile. Welfare losses display a similar pattern, except that high-paid local workers experience much larger losses than low-paid local workers (their higher baseline retention rates indicate greater relative surplus from retention vs. job change), and the gap disappears at larger distances as the shock becomes more skill-neutral.

Table 21 (Figures A14 and A15) reinforces this intuition by comparing plant closings featuring large high-paying manufacturing firms with store closings featuring large low-paying retail firms. Among workers from the focal tract, the store closing creates much larger employment rate increases among low-paid than high-paid workers (16% vs. 1% for the lowest and highest quartiles), since low-paid workers are both disproportionately targeted and less able to compete for other jobs, while the plant closing generates nearly equal employment losses for low- and high-paid local workers.

Similarly, welfare losses from store closings are only slightly larger for local low-paid workers than for high-paid workers, since low-paid workers' greater exposure is partially offset by smaller differences between their retention and outside options than for their managers, perhaps due to less firm-specific skill. Both shock types become more generic with distance, though, so the national shares of employment and welfare losses accruing to the various earnings/unemployment categories differ far less across shock compositions than the local incidence would suggest.

While quantifying the employment and utility incidence of negative labor demand shocks is important for allocating relief funds, policymakers and local communities also care about flows of workers away from sites experiencing negative shocks. Thus, Table 22 (Figures A16a and A16b) displays the change in the probability of ending up employed in each distance bin for workers initially in the target tract. Overall, the probability of remaining employed in this tract only decreases by 4.0% and 3.4% for plant and store closings, respectively, even though these closings reduce employment by closer to 10% on average for these tracts. This is both because local workers retain a disproportionate share of remaining jobs relative to would-be job movers from afar, but also because a large minority of workers would have taken jobs elsewhere even without the shock. An extra 0.6% of local workers become unemployed due to the closings, while an extra 0.8% (0.5%) move to other tracts in the PUMA following plant (store) closings, an extra 1.8% move to other PUMAs within the state for both plant and store closings, and an extra 0.9% (0.5%) find jobs out of state.

Columns 2-7 of Table 22 (Figure A17) paint a richer picture by presenting destination distributions for each initial earnings/unemployment category. Despite similar rates of extra unemployment after plant closings, high-paid workers are much more likely to find distant jobs than low-paid workers, with an additional 1.6% of local top quartile earners switching states and extra 2.7% switching PUMA, compared to 0.4% and 1.3% for the lowest earners (Panel A). For the store closing (Panel B), which targets low paid workers, we see large flows of low paid workers to unemployment and to nearby tracts and other PUMAs, but small flows out of state, reflecting their less integrated labor markets. We also see a small additional outflow by local unemployed workers who would have found local jobs but instead are forced elsewhere, again demonstrating the value of examining equilibrium reallocation rather than merely the destinations of the particular displaced workers.

6.6 Robustness Checks

Table 23 compares the baseline simulation results capturing the pattern of geographic incidence from Section 6.1 to those from three sets of alternative simulations that test the sensitivity of results to model assumptions. In each case, we consider the standard 500 job "plant opening" from above. First, columns 2 and 6 display results from a model that incorporates job multipliers generated by the stimulus. Specifically, we adopt Bartik and Sotherland (2019)'s estimate that the creation of each high-tech manufacturing job (presumed to be associated with large, high paying firms) generates an additional 0.71 jobs after one year. While this is a reduced-form estimate that captures the net effect of all spillover sources, we assume that increased product demand for local services is the dominant source. Thus, we add 500*.71 = 342 additional retail/wholesale sector jobs, distributed across tracts within the local PUMA in proportion to their workers' shares of expected within-PUMA earnings gains from the baseline results that exclude spillovers. The augmented shock increases employment and welfare gains by roughly the same 71 percent as the spillover multiplier, with only slight changes in the distribution of employment and welfare gains toward surrounding tracts and away from the target tract and more distant locations. These results indicate that explicitly introducing job multipliers rather than treating the simulated shocks as implicitly inclusive of job multipliers (as we have done throughout) would not alter the paper's key findings.

Next, columns 3 and 7 display results from a specification that allows firms to endogenously update the number of positions they wish to fill in response to shock-induced changes in labor costs. We assume a constant elasticity of desired positions with respect to the per-position change in each position type's expected payoff ($\overline{q}_f$), and assign this parameter a value of -0.197 based on the mean short-run employment elasticity estimate from Lichter et al. (2015)'s meta-analysis of the minimum wage literature. We then iterate between 1) computing equilibrium assignments and payoffs given a vector $h^{CF}(f)$ of type-specific position counts and 2) updating $h^{CF}(f)$ for each position type by applying the elasticity to $\%\Delta\overline{q}_f$. We include a fixed cost of adjusting the stock of positions equal to 1% of average salaries to prevent fractional worker adjustments by a large share of firms. This process converges to a fixed point in which the final vector $h^{CF}(f)$ aligns with firms' optimal number of positions to fill given the expected firm payoff from filling a position. Across 500 simulations featuring different focal tracts, the mean adjustment is a net decrease in shock size from 500 to 486 positions, with a standard deviation of 19. Changing the size of the shock very slight decreases the magnitude of expected employment and welfare gains, but, as with job multipliers, it barely changes the shares of gains accruing to workers in different distance bins. Thus, failing to explicitly incorporate endogenous responses in desired employment is unlikely to invalidate the results presented, particularly given the one-year horizon and modestly-sized shocks.

Finally, columns 4 and 8 display results from a specification that adopts the Choo and Siow (2006) structure of unobserved idiosyncratic surplus components, which features interactions between an individual worker and a position type and a worker type and an individual position ($\epsilon^1_{if(k)} + \epsilon^2_{l(i),k}$) rather than between a worker and a position (ϵ_{ik}). This approach assumes perfect correlation rather than zero correlation in individuals' preferences for positions within firm types, and thus tests the model's sensitivity to the assumed structure of the unobserved components.

We were unable to make the model fully converge with the Choo-Siow error components. Nonetheless, the geographic distribution of employment rate changes and shares of total employment gains among workers are surprisingly similar to those of the baseline specification, reflecting very similar re-assignments of workers following shocks. This suggests that the predictions about worker reallocation and unemployment are quite robust to specification. However, the distribution of welfare gains is notably different. The Choo-Siow specification generates smaller gains for local workers and a much slower rate of decay with distance, with workers two states away still experiencing an expected gain of $111 relative to the normalized (least affected) type, generally involving workers from a different distant state. These welfare results seem to suggest an unrealistic degree of geographic integration, with 68.8% of welfare gains accruing to out-of-state workers. Since these results seem at odds with the same model's estimated employment incidence, we worry that the lack of convergence may be undermining a valid comparison of welfare effects.[46]

6.7 Model Validation

The estimated surplus parameters $\hat{\Theta}^{D-in-D}$ that underlie the simulations are identified from millions of quotidian job transitions driven by small firm expansions/contractions, labor force turnover, and preference or skill changes over the life cycle that generate considerable offsetting churn in the U.S. labor market. Thus, one might wonder whether parameters governing ordinary worker flows are capable of capturing the response to sizable, locally focused positive or negative shocks. To address this concern, we perform a model validation exercise in which surplus parameters estimated on pre-shock ordinary worker flows were used to forecast the reallocation of workers after actual local economic shocks observed in the sample. We evaluate model fit using the index of dissimilarity between the predicted and actual match group distributions $P(g)$, and average this index across 180 shocks defined by tract-year combinations featuring large positive or negative changes in employment (at least 100 workers and 10% of the pre-shock tract employment level), without offsetting contemporaneous shocks to other tracts in the PUMA or shocks to the same tract in other years. Appendix A8 describes the exercise in detail, while Table 24 reports the results.

To summarize, on average only 6.2% of the resulting job matches nationally would need to be reallocated across match groups in order for the two-sided assignment model to perfectly match the actual allocation that occurred following the shocks. Because match groups feature much more narrowly defined locations within the target PUMA (census tracts rather than PUMAs or states), the model would need to reallocate 34.1% of job matches of workers originating in the target PUMA to other groups g to perfectly match the true within-PUMA distribution. However, a large share of "incorrect" predictions involve either slight differences in destination tract within the same distance bin or slightly mismatched combinations of establishment size/avg. pay/supersector categories.

When the group space is collapsed post-simulation so that origin and destination locations are defined by 42 distance bins from the target tract and non-location establishment characteristics are excluded (though worker skill categories remain), the share of job matches that must be reallocated across collapsed groups to match the data falls to 1.0% nationally and 4.2% among transitions originating in the target PUMA. This is despite the fact that $P(g)$ still contains 10,752 match groups with only 294 restrictions imposed by $n(l)$ and $h(f)$. The model also fits well the worker and position type distributions among workers who either exit or enter unemployment after the shock, particularly when distance bins are aggregated to those in earlier figures: on average only 1.3% of job matches originating within the target PUMA require reassignment to match the actual allocation. This suggests that the counterfactual estimates of the share of employment incidence accruing to dif-

[46]Preliminary versions of the baseline model that failed to converge exhibited similar volatility of welfare estimates.

ferent skill/distance bin combinations are likely to be accurate. The model fit is negligibly affected by further restricting the LEHD shock sample to tract-years in which the shock is particularly likely to be driven by demand rather than supply, as judged by the existence of a single establishment that adds or lays off at least 100 workers in the chosen year.

Furthermore, the assignment model vastly outperforms a one-sided parametric conditional logit model fit to the same pre-shock CCPs $P(g|f)$. Thus, with many million observed job matches, it appears that the risk of overfitting from using a highly saturated, just identified model is far outweighed by the inability of a more parsimonious parametric model (still featuring over a hundred parameters!) to fully capture the rich multidimensional matching patterns contained in the data. The two-sided model also outperforms (though by a much smaller margin) other one-sided non-parametric forecasts that hold fixed the full set of either raw or smoothed CCPs (so $P(g)^{y,CF} = h^y(f)P^{y-1}(g|f)$), suggesting that requiring market clearing does have additional predictive value, even for the reasonably small shocks considered. A comparison with the Choo and Siow (2006) model, which only alters the correlation structure among the idiosyncratic surplus components, reveals very similar model fit, with the Choo-Siow model better fitting the full disaggregated allocation and the current model better fitting flows into and out of unemployment. Taken together, the model does quite a good job of predicting the reallocation of workers across job types and particularly employment statuses that follows substantial local labor market shocks.

7 Conclusion

This paper models the U.S. labor market as a large-scale assignment game with transferable utility, and uses the model estimates to simulate the employment and welfare incidence across locations and worker skill categories of a variety of local labor demand shocks representing different local development initiatives and establishment openings or closings.

We find that U.S. labor markets are quite local, in that the per-worker employment and welfare gains from a locally targeted labor demand shock are substantially larger for workers in the focal and adjacent census tracts than even workers several tracts away. Nonetheless, because these very local workers are a tiny share of the U.S. labor force competing for positions, we also find that, regardless of establishment composition, around 60% (35%) of the employment (welfare) gain from a large establishment opening redounds to workers initially working out of state, with only around 6% (10%) going to existing workers in the PUMA containing the focal tract.

We also document a high degree of heterogeneity in incidence by initial income among very local workers across demand shocks featuring different establishment composition and/or different focal tract characteristics, suggesting that the type of establishment and community targeted by a local development policy has major implications for the groups of workers most likely to benefit. That said, as these alternative shocks ripple across space through a chain of job transitions, their incidence across initial income categories becomes increasing similar, so that the overall skill and spatial composition of welfare gains across workers slightly farther from the site is extremely similar across different types of demand shocks and target areas. Thus, state-level funders of local projects who internalize these ripple effects can safely devolve the selection of local projects to local leaders.

These findings demonstrate both the value and the limitations of reduced-form research analyzing local economic development policy. On one hand, the simulation results suggest that per-person employment and welfare impacts of local labor demand shocks become quite small at greater distances, so that research designs treating distant but similar locations as control groups may be valid for estimating treatment effects on local populations. On the other hand, the results also indicate that the distribution of local impacts need not resemble the distribution of state-level or national impacts. In fact, some characteristics of the target tract that predict relatively larger treatment effects for disadvantaged workers locally also predict smaller effects for disadvantaged workers nationally.

We also find that negative shocks display a much greater concentration of employment and welfare losses than the corresponding gains from positive shocks of equivalent magnitude. This is because many local workers have jobs at risk from negative local shocks, but would have been working anyway without a positive shock, and removing the option to stay at one's job generates large welfare losses, presumably due to both job switching costs and the loss of firm-specific experience.

Methodologically, we show that one can still produce forecasts of welfare incidence on both sides of the market from changes in agent type composition on either side of the market even when singles are either not observed or observed on only one side. By basing simulations on millions of composite joint surplus parameters rather than a much smaller set of fundamental utility or production function parameters, the sufficient statistics approach used here can fully exploit the massive scale of the administrative LEHD database to capture multidimensional heterogeneity on both sides of a two-sided market without placing unjustified structure on the job matching technology.

The method can be customized to forecast the incidence of any particular shock composition or magnitude in any location, and incidence can be assessed among groups of agents on either side of the market defined by any arbitrary combination of observed characteristics, including categorical characteristics without a natural ordering such as race, industry or location. Given appropriate administrative matching data, the approach here could also be easily adapted to the student-college or patient-doctor contexts, among other applications.

References

Abowd, John M, Bryce E Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L McKinney, Marc Roemer, and Simon Woodcock, "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in "Producer Dynamics: New Evidence from Micro Data," University of Chicago Press, 2009, pp. 149–230.

Arcidiacono, Peter and Robert A Miller, "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 2011, *79* (6), 1823–1867.

Bartik, Timothy J., "Economic Development," in J. Richard Aronson and Eli Schwartz, eds., *Management Policies in Local Government Finance*, 2004.

Bartik, Timothy J and Nathan Sotherland, "Local Job Multipliers in the United States: Variation with Local Characteristics and with High-Tech Shocks," 2019.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A distributional framework for matched employer employee data," *Econometrica*, 2019, *87* (3), 699–739.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *The American Economic Review*, 2013, *103* (2), 897–947.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "The Impact of Trade on Labor Market Dynamics," Technical Report, National Bureau of Economic Research 2015.

Carballo, Jeronimo and Richard Mansfield, "The Skill and Spatial Incidence of Changing Patterns of U.S. International Trade among American Workers," 2021. Working Paper.

Chen, Liang, "Compensation, Moral Hazard, and Talent Misallocation in the Market for CEOs," 2017.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

Chiappori, Pierre-André and Bernard Salanié, "The Econometrics of Matching Models," *Journal of Economic Literature*, September 2016, *54* (3), 832–61.

Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss, "Investment in Schooling and the Marriage Market," *The American Economic Review*, 2009, *99* (5), 1689–1713.

Choo, Eugene, "Dynamic Marriage Matching: An Empirical Framework," *Econometrica*, 2015, *83* (4), 1373–1423.

_ **and Aloysius Siow**, "Who Marries Whom and Why," *Journal of Political Economy*, 2006, *114* (1), 175–201.

Decker, Colin, Elliott H Lieb, Robert J McCann, and Benjamin K Stephens, "Unique Equilibria and Substitution Effects in a Stochastic Model of the Marriage Market," *Journal of Economic Theory*, 2013, *148* (2), 778–792.

Diamond, Rebecca, "The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Ekeland, Ivar, James J Heckman, and Lars Nesheim, "Identification and estimation of hedonic models," *Journal of political economy*, 2004, *112* (S1), S60–S109.

Freedman, Matthew, "Targeted Business Incentives and Local Labor Markets," *Journal of Human Resources*, 2013, *48* (2), 311–344.

Galichon, Alfred and Bernard Salanié, "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," 2015.

_ **,** _ **et al.**, "The Econometrics and Some Properties of Separable Matching Models," *American Economic Review*, 2017, *107* (5), 251–255.

Glaeser, Edward L, Joshua D Gottlieb et al., "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, *39* (1 (Spring)), 155–253.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Gregory, Jesse, "The Impact of Post-Katrina Rebuilding Grants on the Resettlement Choices of New Orleans Homeowners," *Work. Pap., Univ. Wisconsin–Madison*, 2013.

Heckman, James J, Rosa L Matzkin, and Lars Nesheim, "Nonparametric identification and estimation of nonadditive hedonic models," *Econometrica*, 2010, *78* (5), 1569–1591.

Hillier, Frederick S and Gerald J Lieberman, "Introduction to Operations Research. Ninth," 2010.

Hotz, V Joseph and Robert A Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Hyatt, Henry R, Erika McEntarfer, Ken Ueda, and Alexandria Zhang, "Interstate Migration and Employer-to-Employer Transitions in the US: New Evidence from Administrative Records Data," *US Census Bureau Center for Economic Studies Paper No. CES-WP-16-44*, 2016.

Kennan, John and James R Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kline, Patrick and Enrico Moretti, "Place Based Policies with Unemployment," Technical Report, National Bureau of Economic Research 2013.

Koopmans, Tjalling C and Martin Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica: Journal of the Econometric Society*, 1957, pp. 53–76.

LeGower, Michael and Randall Walsh, "Promise Scholarship Programs as Place-Making Policy: Evidence from School Enrollment and Housing Prices," *Journal of Urban Economics*, 2017.

Lichter, Andreas, Andreas Peichl, and Sebastian Siegloch, "The own-wage elasticity of labor demand: A meta-regression analysis," *European Economic Review*, 2015, *80*, 94–119.

Lindenlaub, Ilse, "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 2017, *84* (2), 718–789.

Manning, Alan and Barbara Petrongolo, "How Local Are Labor Markets? Evidence from a Spatial Job Search Model," *American Economic Review*, 2017.

Menzel, Konrad, "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 2015, *83* (3), 897–941.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," Technical Report, National Bureau of Economic Research 2015.

Moretti, Enrico, "Local multipliers," *The American Economic Review*, 2010, pp. 373–377.

Mourifié, Ismael and Aloysius Siow, "The Cobb Douglas Marriage Matching Function: Marriage Matching with Peer and Scale Effects," Technical Report 2017.

Neumark, David and Helen Simpson, "Place-Based Policies," Technical Report, National Bureau of Economic Research 2014.

Nimczik, Jan Sebastian, "Job Mobility Networks and Endogenous Labor Markets," 2017.

Piyapromdee, Suphanit, "The Impact of Immigration on Wages, Internal Migration and Welfare," Technical Report, Puey Ungphakorn Institute for Economic Research 2017.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Roth, Alvin E and Marilda Sotomayor, "Two-Sided Matching," *Handbook of Game Theory with Economic Applications*, 1992, *1*, 485–541.

Sattinger, Michael, "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 1993, *31* (2), 831–880.

Schmutz, Benoît and Modibo Sidibe, "Frictional Labor Mobility," 2016.

Shapley, LS and Martin Shubik, *Game Theory in Economics: Chapter 3: the" Rules of the Games"*, Rand, 1972.

Tervio, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, *98* (3), 642–68.

Vilhuber, Lars et al., "LEHD Infrastructure S2014 files in the FSRDC," *US Census Bureau, Center for Economic Studies Discussion Papers, CES*, 2018.

Weinstein, Russell, "Dynamic Responses to Labor Demand Shocks: Evidence from the Financial Industry in Delaware," *Journal of Urban Economics*, 2018.

Table 1: Specifications for Alternative Counterfactual Labor Demand Shocks

Spec. No.	Number of Jobs	Firm Avg. Earn. Quartile	Firm Size Quartile	Industry Supersector	Shock Type
1	500	2	1	Information	Stimulus
2	500	2	4	Information	Stimulus
3	500	4	1	Information	Stimulus
4	500	4	4	Information	Stimulus
5	500	2	1	Manufacturing	Stimulus
6	500	2	4	Manufacturing	Stimulus
7	500	4	1	Manufacturing	Stimulus
8	500	4	4	Manufacturing	Stimulus
9	500	2	1	Trade/Trans./Utilities	Stimulus
10	500	2	4	Trade/Trans./Utilities	Stimulus
11	500	4	1	Trade/Trans./Utilities	Stimulus
12	500	4	4	Trade/Trans./Utilities	Stimulus
13	500	2	1	Other Services	Stimulus
14	500	2	4	Other Services	Stimulus
15	500	4	1	Other Services	Stimulus
16	500	4	4	Other Services	Stimulus
17	500	2	1	Education & Health	Stimulus
18	500	2	4	Education & Health	Stimulus
19	500	4	1	Education & Health	Stimulus
20	500	4	4	Education & Health	Stimulus
21	500	2	1	Leisure & Hospitality	Stimulus
22	500	2	4	Leisure & Hospitality	Stimulus
23	500	4	1	Leisure & Hospitality	Stimulus
24	500	4	4	Leisure & Hospitality	Stimulus
25	500	2	1	Government	Stimulus
26	500	2	4	Government	Stimulus
27	500	4	1	Government	Stimulus
28	500	4	4	Government	Stimulus
29	500	2	1	Construction	Stimulus
30	500	2	4	Construction	Stimulus
31	500	4	1	Construction	Stimulus
32	500	4	4	Construction	Stimulus
33	500	4	1	Information	Restr. Stim.
34	500	2	4	Manufacturing	Restr. Stim.
35	500	2	1	Trade/Trans./Utilities	Restr. Stim.
36	500	4	4	Manufacturing	Stimulus w/ Endog. Vacancies
37	500	4	4	Manufacturing	Stimulus w/ Job Multiplier
38	500	4	4	Manufacturing	Stimulus w/ Choo-Siow Model
39	500	4	4	Manufacturing	Stimulus
40	500	4	4	Manufacturing	Plant Closing
41	125	4	4	Manufacturing	Stimulus
42	125	4	4	Manufacturing	Plant Closing
43	250	4	4	Manufacturing	Stimulus
44	250	4	4	Manufacturing	Plant Closing
45	500	2	4	Wholesale/Retail	Stimulus
46	500	2	4	Wholesale/Retail	Store Closing
47	125	2	4	Wholesale/Retail	Stimulus
48	125	2	4	Wholesale/Retail	Store Closing
49	250	2	4	Wholesale/Retail	Stimulus
50	250	2	4	Wholesale/Retail	Store Closing

Table 2: Summary Statistics Describing Heterogeneity in the Spacial Scope of Labor Markets by Worker and Establishment Characteristics

Panel A: By Origin Worker Unemployment or Earnings Category

		Share of All Transitions					Share of Job to Job Transitions		
Worker Subpop.	# of Obs.	Unemp. to Unemp.	Unemp. to Emp.	Emp. to Unemp.	Stay at Same Job	Job to Job	Same PUMA	New PUMA, Same State	New State
All Workers	24200000	0.029	0.093	0.028	0.694	0.156	0.270	0.549	0.180
Young (<25) Unemp.	1242000	0.101	0.899	0.000	0.000	0.000	0.291*	0.633*	0.076*
Older ($\geq$ 25) Unemp.	1703000	0.338	0.662	0.000	0.000	0.000	0.294*	0.597*	0.109*
1st Earn. Quart.	4935000	0.000	0.000	0.065	0.655	0.280	0.291	0.553	0.155
2nd Earn. Quart.	5327000	0.000	0.000	0.033	0.775	0.192	0.289	0.547	0.165
3rd Earn. Quart.	5494000	0.000	0.000	0.019	0.846	0.135	0.257	0.551	0.192
4th Earn. Quart.	5482000	0.000	0.000	0.012	0.870	0.117	0.211	0.543	0.245

Panel B: By Destination Establishment Pay Quartile and Size Quartile

		Share of All Transitions					Share of Job to Job Transitions		
Estab. Subpop.	# of Obs.	Unemp. to Unemp.	Unemp. to Emp.	Emp. to Unemp.	Stay at Same Job	Job to Job	Same PUMA	New PUMA, Same State	New State
1st Q. Avg. Earn.	5950000	0.000	0.188	0.000	0.634	0.178	0.327	0.611	0.062
2nd Q. Avg. Earn.	5635000	0.000	0.092	0.000	0.761	0.147	0.325	0.607	0.068
3rd Q. Avg. Earn.	5410000	0.000	0.058	0.000	0.806	0.136	0.302	0.627	0.071
4th Q. Avg. Earn.	5331000	0.000	0.044	0.000	0.820	0.136	0.258	0.649	0.093
1st Q. Size	5844000	0.000	0.133	0.000	0.698	0.168	0.379	0.548	0.073
2nd Q. Size	5589000	0.000	0.099	0.000	0.737	0.164	0.321	0.603	0.075
3rd Q. Size	5558000	0.000	0.085	0.000	0.765	0.150	0.275	0.658	0.067
4th Q. Size	5335000	0.000	0.072	0.000	0.813	0.115	0.208	0.723	0.069

Panel C: By Destination Establishment Industry

		Share of All Transitions					Share of Job to Job Transitions		
Estab. Industry	# of Obs.	Unemp. to Unemp.	Unemp. to Emp.	Emp. to Unemp.	Stay at Same Job	Job to Job	Same PUMA	New PUMA, Same State	New State
Nat. Resources	390000	0.000	0.133	0.000	0.705	0.162	0.444	0.460	0.094
Construction	1082000	0.000	0.112	0.000	0.711	0.177	0.281	0.630	0.08
Manufacturing	1983000	0.000	0.054	0.000	0.837	0.109	0.384	0.551	0.069
Whole/Retail/Trans.	4554000	0.000	0.107	0.000	0.749	0.144	0.268	0.652	0.079
Information	520000	0.000	0.069	0.000	0.769	0.162	0.262	0.655	0.077
Financial Activities	1319000	0.000	0.061	0.000	0.773	0.165	0.266	0.670	0.064
Prof. Bus. Services	3171000	0.000	0.119	0.000	0.681	0.200	0.261	0.664	0.076
Ed. Health	5033000	0.000	0.069	0.000	0.805	0.126	0.343	0.595	0.061
Leis. & Hosp.	2511000	0.000	0.182	0.000	0.633	0.184	0.335	0.594	0.071
Oth. Serv.	697000	0.000	0.121	0.000	0.730	0.149	0.337	0.596	0.066
Government	1076000	0.000	0.035	0.000	0.886	0.079	0.376	0.588	0.044

Notes: "1st Q. Avg. Earn": 1st quartile of the establishment-level (worker-weighted) per-worker annual earnings distribution. "1st Q. Size": 1st quartile of the establishment-level (worker-weighted) employment distribution. *: For initially unemployed workers, the share of unemployment-to-employment transitions by distance category is reported in place of share of job-to-job transitions. The locations of initially unemployed workers are assumed to be the location of their most recent employer if previously observed working, otherwise they are imputed from the conditional distribution among job-to-job transitions of origin locations given the destination employer location.

Table 3: Assessing the Impact of 500 Job Stimulus Packages at Different Distances from Focal Tract Across Several Outcomes (Averages Across All Stimulus Compositions)

Distance from Target Tract	Share of JtJ Dest.	Initial Locations	Prob. of Stim. Job	Share of Stim Jobs	Change in P(Employed)	Share of Emp. Gains	Avg. Welfare Change ($)	Share of Wel. Gains
Target Tract	0.034	2.1E-05	0.013	0.033	0.002	0.005	1181	0.009
			(6.4E-05)	(7.6E-05)	(8.0E-06)	(1.2E-05)	(18)	(2.2E-05)
1 Tct Away	0.056	1.1E-04	0.003	0.047	0.001	0.009	377	0.015
			(2.3E-05)	(1.1E-04)	(3.8E-06)	(2.0E-05)	(7)	(3.3E-05)
2 Tcts Away	0.063	2.5E-04	0.002	0.058	3.7E-04	0.012	210	0.021
			(1.0E-05)	(1.1E-04)	(1.6E-06)	(2.1E-05)	(3)	(3.6E-05)
3+ Tcts w/in PUMA	0.126	1.1E-03	0.001	0.124	2.3E-04	0.031	99	0.054
			(2.2E-06)	(1.4E-04)	(4.7E-07)	(3.6E-05)	(0.9)	(6.6E-05)
1 PUMA Away	0.085	1.4E-03	0.001	0.096	1.7E-04	0.029	64	0.051
			(1.9E-06)	(1.4E-04)	(3.0E-07)	(4.1E-05)	(0.6)	(7.8E-05)
2 PUMAs Away	0.143	4.1E-03	2.7E-04	0.144	1.0E-04	0.056	41	0.097
			(4.1E-07)	(1.7E-04)	(1.4E-07)	(5.7E-05)	(0.3)	(1.2E-04)
3+ PUMAs w/in State	0.345	0.054	4.6E-05	0.319	3.6E-05	0.253	22	0.397
			(1.9E-07)	(3.7E-04)	(8.3E-08)	(3.3E-04)	(0.2)	(4.7E-04)
1 State Away	0.028	0.052	6.0E-06	0.040	1.1E-05	0.074	3	0.099
			(1.4E-08)	(1.0E-04)	(9.1E-09)	(1.1E-04)	(0.0)	(1.6E-04)
2+ States Away	0.038	0.264	9.4E-07	0.032	4.2E-06	0.144	0.4	0.078
			(1.3E-09)	(4.7E-05)	(1.3E-09)	(8.1E-05)	(0.0)	(5.7E-05)
Out of Sample	0.082	0.622	1.3E-06	0.106	4.9E-06	0.389	0.4	0.179
			(1.5E-09)	(1.2E-04)	(1.6E-09)	(1.3E-04)	(0.0)	(1.4E-04)

Notes: The column labeled "Share of JtJ Dest." displays the share of all job-to-job transitions among 2012 and 2013 dominant jobs whose origin-destination distance fell into the distance bins given by the row labels. The column labeled "Initial Locations" captures the share of workers for whom the distance between their origin position and the targeted census tract fell into the chosen bin (averaged over 500 simulations featuring different target census tracts). The column labeled "Prob. of Stim. Job" indicates the probability that a randomly chosen worker in the row subgroup will receive one of the 500 new positions generated by the simulated stimulus package. The column labeled "Change in P(Employed)" indicates the change in the probability that a randomly chosen worker in the row subgroup will be employed in the destination year as a consequence of the simulated stimulus package. The column labeled "Avg. Welfare Change" indicates the change in job-related welfare (scaled to be equivalent to $ of 2012 annual earnings) that a randomly chosen worker in the subgroup indicated by the row label will experience as a consequence of the simulated stimulus package. The columns labeled "Share of Stim. Jobs", "Share of Emp. Gains" and "Share of Wel. Gains" indicate the share of all stimulus jobs and total employment and welfare gains, respectively, generated by the simulated stimulus package that accrue to workers in the subgroup indicated by the row label.
"Target Tract" indicates that the worker's origin establishment was in the tract receiving the stimulus package. 1/2/3+ Tct(s) Away" indicates that the origin establishment was one, two, or 3 or more tracts away (by tract pathlength) within the same PUMA. "1/2/3+ PUMAs Away" and "1/2+ States Away" indicate the PUMA pathlength (if within the same state) and state pathlength (if in different states), respectively. "Out of Sample" indicates that the worker's origin establishment was not among the 19 states providing data in the sample.
Standard errors are provided in parentheses, and are based on the sampling distribution among the sample of 500 target tracts simulated for each stimulus package specification.

Table 4: Average of Each Incidence Measure by Distance from Target Tract Across All Stimulus Packages, Measured in Miles - Each Column Averages Results 500 Simulations Featuring 500 Different Target Census Tracts

Distance from Centroid of Target Tract	Share of JtoJ Dest.	Prob. of Stim. Job	Share of Stim. Jobs	Change in P(Employed)	Share of Emp. Gains	Avg. Welfare Gain ($)	Share of Wel. Gains
Within 1 Mile	0.032	0.004	0.036	0.001	0.006	482	0.011
1-2 Miles Away	0.053	0.002	0.036	2.9E-04	0.007	261	0.011
3-5 Miles Away	0.092	0.001	0.096	2.0E-04	0.021	233	0.036
6-11 Miles Away	0.120	0.001	0.112	1.4E-04	0.029	258	0.049
11-26 Miles Away	0.160	3.9E-04	0.161	1.2E-04	0.048	190	0.083
26-50 Miles Away	0.069	2.7E-04	0.075	1.0E-04	0.028	79	0.049
51-100 Miles Away	0.063	1.7E-04	0.065	9.1E-05	0.034	39	0.061
101-250 Miles Away	0.201	3.3E-05	0.098	3.0E-05	0.090	12	0.153
>250 Miles Away	0.092	4.9E-06	0.216	7.9E-06	0.349	2	0.367
Out of Sample	0.117	1.3E-06	0.106	4.9E-06	0.389	0.4	0.179

Notes: See Table 3 for expanded definitions of the outcomes in the column labels. The row labels define subpopulations of workers for whom the distance between the establishment associated with their origin dominant jobs and the census tract receiving the simulated stimulus package fell in the listed distance bin.

Table 5: Assessing the Value of Restricting Stimulus Jobs to Workers Within the Target PUMA: Spatial Employment and Welfare Incidence for Restricted and Unrestricted Stimulus Packages (Each Featuring 500 Positions at a Large Low-Paying Manufacturing Firm)

Distance from Target Tract	Change in P(Employed)		Share of Emp. Gains		Avg. Welfare Change ($)		Share of Wel. Gains	
	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.
Target Tract	0.002	0.010	0.005	0.026	1077	9142	0.009	0.048
1 Tct Away	0.001	0.002	0.009	0.035	343	1468	0.014	0.055
2 Tcts Away	3.8E-04	0.001	0.012	0.041	198	701	0.020	0.063
3+ Tcts w/in PUMA	2.4E-04	0.001	0.033	0.070	95	221	0.054	0.110
1 PUMA Away	1.7E-04	1.4E-04	0.030	0.024	62	58	0.051	0.042
2 PUMAs Away	1.1E-04	8.9E-05	0.057	0.047	41	39	0.098	0.081
3+ PUMAs w/in State	3.7E-05	3.2E-05	0.258	0.222	22	21	0.407	0.330
1 State Away	1.1E-05	1.0E-05	0.073	0.067	3	4	0.101	0.084
2+ States Away	4.2E-06	3.8E-06	0.142	0.129	0.4	1	0.077	0.062
Out of Sample	4.8E-06	4.2E-06	0.381	0.338	0.4	1	0.169	0.126

Notes: See Table 3 for expanded definitions of the row labels and the outcomes in the column labels. Table entries consist of various measures of incidence by worker initial distance from the target census tract from a stimulus package consisting of 500 new jobs at large (top quartile of employment), low-paying (2nd quartile of avg. worker pay) manufacturing establishments. Columns labeled "Res." report results from specifications in which the new positions are constrained to be filled by workers initially working (or most recently working) in the same PUMA as the targeted tract, while columns labeled "Unres." report results from specifications in which the new positions may be filled by any worker in the nation.

Table 6: Change in Probability of Employment due to Stimulus for a Randomly Chosen Individual at Different Combinations of Initial Earnings Quartile (or Nonemployment) and Distance from Focal Tract: Average Across All Stimulus Specifications Featuring 500 New Jobs

Distance from	Employment Status/Earnings Quartile					
Focal Tract	UE $\leq$ 25	UE $>$ 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Target Tract	2.5E-03	9.7E-03	2.4E-03	1.2E-03	6.6E-04	4.8E-04
1 Tct Away	1.1E-03	3.1E-03	7.8E-04	3.9E-04	2.2E-04	1.4E-04
2 Tcts Away	6.9E-04	1.9E-03	4.7E-04	2.4E-04	1.4E-04	9.0E-05
3+ Tcts w/in PUMA	4.1E-04	1.2E-03	2.9E-04	1.5E-04	8.2E-05	5.8E-05
1 PUMA Away	2.9E-04	8.6E-04	2.0E-04	1.0E-04	6.0E-05	4.6E-05
2 PUMAs Away	1.9E-04	5.3E-04	1.3E-04	6.6E-05	3.9E-05	2.9E-05
3+ PUMAs w/in State	7.4E-05	1.7E-04	4.3E-05	2.3E-05	1.4E-05	9.5E-06
1 State Away	1.9E-05	5.3E-05	1.4E-05	7.4E-06	4.3E-06	3.5E-06
2+ States Away	8.3E-06	2.1E-05	5.3E-06	2.7E-06	1.6E-06	1.4E-06
Out of Sample	2.3E-05	2.4E-05	7.3E-06	3.2E-06	2.0E-06	2.0E-06

Notes: See Table 3 for expanded definitions of the row labels. Each cell contains the average change in the probability of employment in the destination year generated by a 500 job stimulus for workers whose distance between their origin establishment and the census tract receiving the stimulus package placed them in the distance bin indicated in the row label and whose employment status or earnings in the origin year placed them in the earnings/employment category listed by the column label. "UE $\leq$ Age 25": Workers who were unemployed in the origin year (defined as no full quarter of work with $>\$2,000$ in earnings at any establishment) and who were 25 years old or younger.
"UE $>$ Age 25": Workers who were unemployed in the origin year and who were more than 25 years old. "1st/2nd/3rd/4th Quartile": Workers whose average earnings among full quarters worked at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th quartile of the 2012 annual earnings distribution for the sample states. Each cell averages results across 32 stimulus packages featuring new jobs with establishments with different combinations of industry supersector, firm size quartile, and firm average pay quartile. Results are further averaged across 500 simulations for each of the 32 stimulus package specifications featuring different target census tracts.

Table 7: Shares of Additional Employment and Utility Produced by Stimulus among Workers Initially Employed (or Unemployed) at Different Initial Earnings Quartiles (or Unemployment): Stimuli Consist of 500 New Jobs at Firms in Different Firm Size/Firm Average Earnings Quartiles (Averaged across Different Industries)

Earnings	Share of Employment Gains					Share of Welfare Gains				
Category	Avg.	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi	Avg.	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi
UE $\leq$ Age 25	0.100	0.101	0.104	0.095	0.098	0.056	0.060	0.061	0.050	0.052
UE $>$ Age 25	0.337	0.348	0.332	0.340	0.328	0.053	0.059	0.052	0.052	0.047
1st Quartile	0.246	0.244	0.253	0.242	0.247	0.187	0.193	0.201	0.175	0.179
2nd Quartile	0.144	0.142	0.144	0.144	0.146	0.223	0.228	0.228	0.217	0.217
3rd Quartile	0.091	0.087	0.089	0.093	0.094	0.234	0.230	0.229	0.238	0.239
4th Quartile	0.083	0.077	0.079	0.087	0.088	0.249	0.230	0.230	0.269	0.266

Notes: See Table 6 for expanded definitions of the row labels. The first five columns capture the share of employment gains (scaled to be equivalent to $ of 2012 annual earnings) in the destination year attributable to a 500 job stimulus package accruing to workers whose employment status or earnings in the origin year places them in the earnings/employment category listed by the row label. The last five columns capture the share of all stimulus-driven welfare gains accruing to workers in each earnings/employment category. Columns 1 and 6 average across all 32 stimulus package specifications. Each of columns 2-5 and 7-10 averages results across 8 stimulus packages featuring jobs with establishments in the same firm size quartile/firm average pay quartile combination but in different industry supersectors (as well as simulated 500 simulations for each stimulus package specification featuring different target census tracts) "Sm./Low": The 500 stimulus jobs are generated by establishments whose employment levels place them in the smallest quartile of firms and whose average worker pay levels place them in the 2nd smallest quartile of firms. "Lg./Low": The 500 stimulus jobs are generated by establishments whose employment levels place them in the largest quartile of firms and whose average worker pay levels place them in the 2nd smallest quartile of firms. "Sm./Hi": The smallest quartile of firms and whose average worker pay levels place them in the highest quartile of firms. "Lg./Hi": The 500 stimulus jobs are generated by establishments whose employment levels place them in the largest quartile of firms and whose average worker pay levels place them in the highest quartile of firms.

Table 8: Expected Welfare Gain From New Stimulus Positions Among Workers Initially Employed at Different Combinations of Initial Earnings Quartile (or Nonemployed) and Distance from Focal Tract: Averaged Across All Stimulus Specifications Featuring 500 New Jobs)

Distance from	Employment Status/Earnings Quartile					
Focal Tract	UE ≤ 25	UE > 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Target Tract	1119	1319	1129	1173	1237	1443
1 Tct Away	447	493	349	361	376	403
2 Tcts Away	253	237	199	205	211	221
3+ Tcts w/in PUMA	116	110	94	98	99	101
1 PUMA Away	75	67	60	63	64	67
2 PUMAs Away	48	40	40	41	42	42
3+ PUMAs w/in State	25	20	21	22	22	22
1 State Away	3	2	3	3	3	3
2+ States Away	0.4	0.2	0.4	0.4	0.5	0.5
Out of Sample	0.7	0.1	0.4	0.4	0.4	0.5

Notes: See Table 3 for expanded definitions of the row labels. See Table 6 for expanded definitions of the column labels. Each cell contains the average job-related welfare gain (scaled to be equivalent to $ of 2012 annual earnings) generated by a 500 job stimulus for workers whose distance between their origin establishment and the census tract receiving the stimulus package placed them in the distance bin indicated in the row label and whose employment status or earnings in the origin year placed them in the earnings/employment category listed by the column label. Each cell averages results across 32 stimulus packages featuring new jobs with establishments with different combinations of industry supersector, firm size quartile, and firm average pay quartile. Results are further averaged across 500 simulations for each of the 32 stimulus package specifications featuring different target census tracts.

Table 9: Expected Job-Related Welfare Gain From New Stimulus Positions Among Workers Initially Employed in the Focal Tract at Different Earnings Quintiles (or Unemployed) by Industry Supersector (Averaged Across Firm Size/Firm Average Earnings Combinations)

Earnings		Industry							
Quintile	Avg.	Info.	Manu.	R/W Trd.	Oth. Serv.	Ed./Hlth	Lei/Hosp.	Gov.	Const.
UE $\leq$ Age 25	1119	1095	1143	1261	986	1033	1479	1048	905
UE $>$ Age 25	1319	1226	1218	1102	1421	1213	1406	1413	1549
1st Quartile	1129	1098	1019	929	1121	1233	1403	1214	1012
2nd Quartile	1173	1119	1164	934	1296	1214	1200	1338	1120
3rd Quartile	1237	1180	1192	984	1389	1234	1238	1412	1265
4th Quartile	1443	1726	1513	1264	1628	1294	1337	1448	1332

Notes: See Table 7 for expanded definitions of distance bins captured by the row labels. Each cell contains the average job-related welfare gain (scaled to be equivalent to $ of 2012 annual earnings) generated by a 500 job stimulus for workers initially employed (or most recently employed) in the focal tract whose employment status or earnings in the origin year placed them in the earnings/employment category listed by the row label. Each column averages results across four stimulus packages featuring jobs with establishments in the same industry supersector but in different quartiles of the establishment-level employment and average worker earnings distributions. Results are further averaged across 500 simulations featuring different target census tracts for each of the stimulus package specifications.
"Avg.": Average employment change across all 32 stimulus packages considered (and all 500 target tracts for each stimulus package specification. "Info": Information. "Manu.": Manufacturing. "Trd./Tns.": Trade/Transportation/Utilities. "Oth. Serv.": Other Services (includes repair, laundry, security, personal services). "Ed./Hlth": Education and Healthcare. "Lei/Hosp": Leisure and Hospitality. "Gov.": Government. "Const.": Construction.

Table 10: Expected Change in Utility From New Stimulus Positions Among Workers Initially Employed in the Focal Tract at Different Earnings Quintiles (or Nonemployed) by Firm Size Quartile/Firm Average Pay Quartile Combination (Averaged Across Industry Supersectors)

Earnings	Firm Size/Pay Level Combination			
Quintile	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi
NE $\leq$ Age 25	1367	1290	882	934
NE $>$ Age 25	1731	1127	1429	987
1st Quartile	1278	1371	935	930
2nd Quartile	1318	1237	1081	1056
3rd Quartile	1197	1141	1319	1291
4th Quartile	976	965	2038	1793

Notes: See Table 7 for expanded definitions of employment status/earnings quartile categories captured by the row labels. See Table 13 for expanded definitions of the establishment size/avg. pay combinations captured by the column labels. Each cell contains the average job-related welfare gain (scaled to be equivalent to $ of 2012 annual earnings) generated by a 500 job stimulus for workers initially employed (or most recently employed) in the focal tract whose employment status or earnings in the origin year placed them in the earnings/employment category listed by the row label. Each column averages results from eight stimuli that feature jobs with establishments from different industry supersectors but the same quartiles of the establishment-level employment and average worker earnings distributions (indicated by the column label). Results are further averaged across 500 simulations featuring different target census tracts for each of the stimulus package specifications.

Table 11: Change in Probability of Employment due to Stimulus for a Randomly Chosen Individual at Different Distances from Focal Tract: Stimuli Consist of 500 New Jobs at Firms in Alternative Industries (Averaged Across Firm Size/Firm Average Earnings Combinations)

Distance from		Industry							
Focal Tract	Avg.	Info.	Manu.	Trd./Tns.	Oth. Serv.	Ed./Hlth	Lei/Hosp.	Gov.	Const.
Target Tract	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(8.0E-06)	(8.4E-06)	(8.5E-06)	(8.7E-06)	(9.9E-06)	(9.6E-06)	(1.1E-05)	(9.6E-06)	(9.5E-06)
1 Tct Away	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(3.8E-06)	(3.8E-06)	(3.6E-06)	(3.5E-06)	(3.9E-06)	(4.4E-06)	(4.9E-06)	(4.5E-06)	(4.4E-06)
2 Tcts Away	3.7E-04	3.6E-04	3.7E-04	3.3E-04	3.8E-04	4.0E-04	3.7E-04	3.8E-04	3.7E-04
	(1.6E-06)	(1.7E-06)	(1.7E-06)	(1.4E-06)	(1.8E-06)	(1.8E-06)	(1.4E-06)	(1.7E-06)	(2.4E-06)
3+ Tcts w/in PUMA	2.3E-04	2.2E-04	2.2E-04	2.0E-04	2.2E-04	2.4E-04	2.2E-04	2.4E-04	2.3E-04
	(4.7E-07)	(5.0E-07)	(5.0E-07)	(4.6E-07)	(4.8E-07)	(5.6E-07)	(4.7E-07)	(5.4E-07)	(5.5E-07)
1 PUMA Away	1.7E-04	1.6E-04	1.6E-04	1.6E-04	1.7E-04	1.8E-04	1.7E-04	1.7E-04	1.7E-04
	(3.0E-07)	(3.1E-07)	(3.3E-07)	(2.8E-07)	(3.5E-07)	(5.4E-07)	(2.7E-07)	(3.4E-07)	(3.1E-07)
2 PUMAs Away	1.0E-04	1.0E-04	1.0E-04	1.0E-04	1.0E-04	1.1E-04	1.0E-04	1.1E-04	1.1E-04
	(1.4E-07)	(1.5E-07)	(1.5E-07)	(1.6E-07)	(1.5E-07)	(1.5E-07)	(1.6E-07)	(1.5E-07)	(1.6E-07)
3+ PUMAs w/in State	3.6E-05	3.6E-05	3.6E-05	3.7E-05	3.6E-05	3.6E-05	3.6E-05	3.7E-05	3.6E-05
	(8.3E-08)	(8.3E-08)	(8.5E-08)	(9.5E-08)	(8.5E-08)	(8.4E-08)	(9.2E-08)	(8.9E-08)	(9.2E-08)
1 State Away	1.1E-05	1.1E-05	1.1E-05	1.1E-05	1.1E-05	1.1E-05	1.1E-05	1.1E-05	1.1E-05
	(9.1E-09)	(9.6E-09)	(9.5E-09)	(9.9E-09)	(1.1E-08)	(1.0E-08)	(9.5E-09)	(9.6E-09)	(1.1E-08)
2+ States Away	4.2E-06	4.3E-06	4.2E-06	4.2E-06	4.2E-06	4.2E-06	4.3E-06	4.2E-06	4.2E-06
	(1.3E-09)	(1.5E-09)	(1.4E-09)	(1.5E-09)	(1.5E-09)	(1.5E-09)	(1.5E-09)	(1.4E-09)	(1.6E-09)
Out of Sample	4.9E-06	4.9E-06	4.9E-06	4.9E-06	4.9E-06	4.8E-06	4.9E-06	4.8E-06	4.8E-06
	(1.6E-09)	(1.7E-09)	(1.7E-09)	(1.7E-09)	(1.8E-09)	(1.6E-09)	(1.6E-09)	(1.6E-09)	(1.8E-09)

Notes: See Table 3 for expanded definitions of the row labels. See Table 9 for expanded definitions of the industry supersectors captured by the column labels. Each entry provides the average change in the probability of being employed in the destination year attributable to a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package placed them in the distance bin indicated in the row label. Different columns consider average employment impacts from stimuli featuring jobs with establishments representing different industry supersectors. Each column averages results across four stimulus packages featuring jobs with establishments in the same industry supersector but in different quartiles of the establishment-level employment and average worker earnings distributions. Standard errors are provided in parentheses, and are based on the sampling distribution among the sample of 500 target tracts simulated for each stimulus package specification.

Table 12: Expected Change in Utility from New Stimulus Positions for a Randomly Chosen Individual at Different Distances from Focal Tract: Stimuli Consist of 500 New Jobs in Different Industries (Averaged Across Firm Size/Firm Average Earnings Combinations)

Distance from		Industry							
Focal Tract	Avg.	Info.	Manu.	R/W Trd.	Oth. Serv.	Ed./Hlth	Lei/Hosp.	Gov.	Const.
Target Tract	1181	1170	1132	989	1268	1181	1277	1296	1131
	(18)	(21)	(19)	(23)	(26)	(22)	(25)	(23)	(23)
1 Tct Away	377	367	344	318	385	425	414	399	365
	(7)	(7)	(6)	(7)	(8)	(8)	(10)	(9)	(7)
2 Tcts Away	210	198	215	177	223	234	192	202	237
	(3)	(3)	(4)	(3)	(5)	(5)	(3)	(3)	(7)
3+ Tcts w/in PUMA	99	96	97	89	100	109	96	104	98
	(0.90)	(0.92)	(0.95)	(0.88)	(1)	(1)	(0.92)	(1)	(1)
1 PUMA Away	64	63	63	60	65	73	61	64	62
	(0.56)	(0.62)	(0.59)	(0.52)	(0.66)	(1)	(0.51)	(0.62)	(0.58)
2 PUMAs Away	41	41	41	42	41	42	41	42	41
	(0.28)	(0.29)	(0.29)	(0.31)	(0.30)	(0.30)	(0.30)	(0.29)	(0.30)
3+ PUMAs w/in State	22	22	22	23	22	21	22	22	21
	(0.17)	(0.17)	(0.18)	(0.20)	(0.19)	(0.17)	(0.19)	(0.18)	(0.17)
1 State Away	3	3	3	3	3	3	3	3	3
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
2+ States Away	0.42	0.43	0.42	0.43	0.42	0.43	0.43	0.41	0.42
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Out of Sample	0.42	0.43	0.43	0.43	0.42	0.40	0.42	0.40	0.42
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: See Table 3 for expanded definitions of the row labels. Each entry provides the expected change in job-related utility attributable to a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package placed them in the distance bin indicated in the row label. Different columns consider average welfare impacts from stimuli featuring jobs with establishments representing different industry supersectors. Each column averages results across four stimulus packages featuring jobs with establishments in the same industry supersector but in different quartiles of the establishment-level employment and average worker earnings distributions. Standard errors are provided in parentheses, and are based on the sampling distribution among the sample of 500 target tracts simulated for each stimulus package specification. "Avg.": Average employment change across all 32 stimulus packages considered (and all 500 target tracts for each stimulus package specification. "Info": Information. "Manu.": Manufacturing. "Trd./Tns.": Trade/Transportation/Utilities. "Oth. Serv.": Other Services (includes repair, laundry, security, personal services). "Ed./Hlth": Education and Healthcare. "Lei/Hosp": Leisure and Hospitality. "Gov.": Government. "Const.": Construction.

Table 13: Change in Probability of Employment and Share of Nationwide Employment Gains From New Stimulus Positions by Distances from Focal Tract: Stimuli Consist of 500 New Positions in Alternative Combinations of Firm Size Quartile/Firm Average Pay Quartile (Averaged Across Industry Supersectors)

Distance from	Change in P(Employed)				Share of Emp. Gains			
Focal Tract	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi
Target Tract	0.002	0.002	0.002	0.002	0.006	0.005	0.005	0.004
1 Tct Away	0.001	0.001	0.001	0.001	0.010	0.009	0.008	0.008
2 Tcts Away	4.1E-04	3.7E-04	3.7E-04	3.3E-04	0.013	0.012	0.012	0.011
3+ Tcts w/in PUMA	2.4E-04	2.3E-04	2.2E-04	2.1E-04	0.034	0.032	0.030	0.029
1 PUMA Away	1.8E-04	1.7E-04	1.6E-04	1.6E-04	0.031	0.030	0.028	0.027
2 PUMAs Away	1.1E-04	1.1E-04	1.0E-04	1.0E-04	0.058	0.057	0.054	0.053
3+ PUMAs w/in State	3.7E-05	3.7E-05	3.6E-05	3.6E-05	0.254	0.257	0.249	0.251
1 State Away	1.1E-05	1.1E-05	1.1E-05	1.1E-05	0.073	0.074	0.074	0.074
2+ States Away	4.2E-06	4.2E-06	4.3E-06	4.3E-06	0.141	0.143	0.144	0.146
Out of Sample	4.8E-06	4.8E-06	4.9E-06	5.0E-06	0.379	0.383	0.394	0.398

Notes: See Table 3 for expanded definitions of the row labels. The first four columns capture the average change in the probability of being employed in the destination year attributable to a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. The last four columns capture the share of all stimulus-driven employment gains accruing to workers in each distance bin. Different columns consider average employment impacts from stimuli featuring jobs with establishments from different combinations of firm size quartile and firm average worker earnings quartile in the respective nationwide establishment-level distributions. Each column averages results across 8 stimulus packages featuring jobs with establishments in the same firm size quartile/firm average pay quartile combination but in different industry supersectors (as well as simulated 500 simulations for each stimulus package specification featuring different target census tracts). "Sm./Low": The 500 stimulus jobs are generated by establishments whose employment levels place them in the smallest quartile of firms and whose average worker pay levels place them in the 2nd smallest quartile of firms. "Lg./Low": The 500 stimulus jobs are generated by establishments whose employment levels place them in the largest quartile of firms and whose average worker pay levels place them in the 2nd smallest quartile of firms. "Sm./Hi": The 500 stimulus jobs are generated by establishments whose employment levels place them in the smallest quartile of firms and whose average worker pay levels place them in the highest quartile of firms. "Lg./Hi": The 500 stimulus jobs are generated by establishments whose employment levels place them in the largest quartile of firms and whose average worker pay levels place them in the highest quartile of firms.

Table 14: Expected Change in Utility and Share of Nationwide Utility Gains from New Stimulus Positions by Distance from Focal Tract: Stimuli Consist of 500 New Positions in Alternative Combinations of Firm Size Quartile/Firm Average Pay Quartile (Averaged Across Industry Supersectors)

Distance from	Avg. Welfare Change ($)				Share of Welfare Gains			
Focal Tract	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi	Sm./Low	Lg./Low	Sm./Hi	Lg./Hi
Target Tract	1223	1167	1201	1132	0.009	0.009	0.009	0.009
1 Tct Away	408	364	384	352	0.016	0.014	0.015	0.014
2 Tcts Away	221	198	222	197	0.022	0.020	0.021	0.020
3+ Tcts w/in PUMA	100	98	99	98	0.056	0.054	0.054	0.053
1 PUMA Away	63	64	64	64	0.051	0.051	0.050	0.050
2 PUMAs Away	41	42	41	42	0.098	0.098	0.096	0.096
3+ PUMAs w/in State	21	22	22	22	0.396	0.405	0.390	0.397
1 State Away	3	3	3	3	0.100	0.100	0.099	0.099
2+ States Away	0.4	0.4	0.4	0.4	0.078	0.078	0.079	0.078
Out of Sample	0.4	0.4	0.4	0.4	0.173	0.172	0.186	0.185

Notes: See Table 3 for expanded definitions of the row labels. See Table 13 for expanded definitions of column labels. The first four columns capture the average change in job-related welfare (scaled to be equivalent to $ of 2012 annual earnings) in the destination year attributable to a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. The last four columns capture the share of all stimulus-driven welfare gains accruing to workers in each distance bin. Each column averages results across 8 stimulus packages featuring jobs with establishments in the same firm size quartile/firm average pay quartile combination but in different industry supersectors (as well as simulated 500 simulations for each stimulus package specification featuring different target census tracts).

Table 15: Heterogeneity in Change in P(Employed) and Share of Total Employment Gains by Distance from Focal Tract Across Alternative Subpopulations of Focal Tracts

Panel A: Urbanicity and # Jobs within 5 Miles

Distance from	Change in P(Employed)					Share of Employment Gains				
Focal Tract	All	Rural	Urban	Low	High	All	Rural	Urban	Low	High
Target Tract	0.002	0.004	0.002	0.007	0.001	0.005	0.009	0.003	0.010	0.003
1 Tct Away	0.001	0.001	2.2E-04	0.002	1.6E-04	0.009	0.014	0.003	0.015	0.004
2 Tcts Away	3.7E-04	0.001	1.6E-04	0.001	1.3E-04	0.012	0.018	0.007	0.017	0.008
3+ Tcts w/in PUMA	2.3E-04	2.8E-04	1.4E-04	3.1E-04	1.2E-04	0.031	0.037	0.022	0.037	0.026
1 PUMA	1.7E-04	1.7E-04	1.1E-04	1.9E-04	1.1E-04	0.029	0.031	0.024	0.035	0.025
2 PUMAs Away	1.0E-04	1.1E-04	7.9E-05	1.1E-04	8.5E-05	0.056	0.061	0.047	0.064	0.050
3+ PUMAs w/in State	3.6E-05	3.6E-05	3.4E-05	3.6E-05	3.5E-05	0.253	0.141	0.397	0.135	0.340
1 State Away	1.1E-05	1.2E-05	8.8E-06	1.2E-05	1.0E-05	0.074	0.092	0.042	0.093	0.056
2+ States Away	4.2E-06	4.6E-06	3.7E-06	4.5E-06	4.0E-06	0.144	0.166	0.114	0.164	0.127
Out of Sample	4.9E-06	5.4E-06	4.3E-06	5.4E-06	4.5E-06	0.389	0.431	0.342	0.430	0.361

Panel B: Two-Bedroom Apartment Rent and Poverty Rate

Distance from	Change in P(Employed)					Share of Employment Gains				
Focal Tract	All	Cheap	Expen.	Low	High	All	Cheap	Expen.	Low	High
Target Tract	0.002	0.005	0.001	0.002	0.002	0.005	0.007	0.004	0.004	0.005
1 Tct Away	0.001	0.002	2.1E-04	3.5E-04	0.001	0.009	0.015	0.004	0.005	0.010
2 Tcts Away	3.7E-04	0.001	1.9E-04	2.9E-04	4.1E-04	0.012	0.016	0.009	0.011	0.012
3+ Tcts w/in PUMA	2.3E-04	4.2E-04	1.3E-04	1.8E-04	2.8E-04	0.031	0.040	0.024	0.029	0.033
1 PUMA	1.7E-04	2.4E-04	9.8E-05	1.3E-04	2.0E-04	0.029	0.042	0.023	0.027	0.031
2 PUMAs Away	1.0E-04	1.5E-04	7.1E-05	9.1E-05	1.2E-04	0.056	0.064	0.044	0.055	0.055
3+ PUMAs w/in State	3.6E-05	5.4E-05	3.2E-05	3.4E-05	4.0E-05	0.253	0.111	0.408	0.293	0.225
1 State Away	1.1E-05	1.3E-05	8.6E-06	1.1E-05	1.0E-05	0.074	0.100	0.034	0.065	0.076
2+ States Away	4.2E-06	4.4E-06	3.7E-06	4.2E-06	4.3E-06	0.144	0.169	0.113	0.137	0.150
Out of Sample	4.9E-06	5.5E-06	4.2E-06	4.7E-06	5.0E-06	0.389	0.435	0.338	0.373	0.403

Notes: See Table 3 for expanded definitions of the distance bins captured by the row labels. The first five columns provide the estimated change in the probability of employment in the destination year caused by a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. The next five columns provide the share of total stimulus-driven employment gains that accrue to workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. Each column displays the average welfare outcome by distance bin among a subset of simulations featuring focal census tracts whose characteristics align with the column label. "All": An average of all 500 target census tracts chosen as sites of simulated stimulus packages. "Rural"/"Urban": An average over the 100 census tracts featuring the lowest/highest residential density (residents per square mile) among the full 500 target tracts simulated. "Low"/"High": In Panel A (B), an average over the 100 census tracts featuring the smallest/largest number of jobs within 5 miles (poverty rate) among the full 500 target tracts simulated. "Cheap"/Expen.": An average over the 100 census tracts featuring the cheapest/most expensive rent for a two-bedroom apartment among the full 500 target tracts simulated.

Table 16: Heterogeneity in Average Welfare Gain and Share of Total Welfare Gains by Distance from Focal Tract Across Alternative Subpopulations of Focal Tracts

Panel A: Urbanicity and # Jobs within 5 Miles

Distance from	Avg. Welfare Gain ($)					Share of Welfare Gains				
Focal Tract	All	Rural	Urban	Small	Large	All	Rural	Urban	Small	Large
Target Tract	1181	2765	818	3029	558	0.009	0.019	0.005	0.020	0.005
1 Tct Away	377	933	79	1126	68	0.015	0.025	0.005	0.026	0.007
2 Tcts Away	210	408	62	467	48	0.021	0.032	0.011	0.030	0.014
3+ Tcts w/in PUMA	99	137	45	144	41	0.054	0.068	0.035	0.067	0.046
1 PUMA	64	64	35	69	35	0.051	0.056	0.037	0.063	0.042
2 PUMAs Away	41	45	25	45	30	0.097	0.114	0.072	0.119	0.085
3+ PUMAs w/in State	22	22	12	22	15	0.397	0.245	0.583	0.234	0.511
1 State Away	3	4	2	4	2	0.099	0.133	0.047	0.133	0.071
2+ States Away	0.4	0.4	0.4	0.4	0.4	0.078	0.088	0.066	0.086	0.072
Out of Sample	0.4	0.5	0.3	0.5	0.3	0.179	0.221	0.141	0.223	0.149

Panel B: Two-Bedroom Apartment Rent and Poverty Rate

Distance from	Avg. Welfare Gain ($)					Share of Welfare Gains				
Focal Tract	All	Cheap	Expen.	Low	High	All	Cheap	Expen.	Low	High
Target Tract	1181	2472	737	927	1142	0.009	0.015	0.005	0.007	0.010
1 Tct Away	377	1061	112	161	430	0.015	0.027	0.007	0.009	0.017
2 Tcts Away	210	531	82	112	298	0.021	0.029	0.014	0.019	0.022
3+ Tcts w/in PUMA	99	187	47	70	122	0.054	0.072	0.037	0.050	0.059
1 PUMA	64	95	32	52	78	0.051	0.077	0.033	0.047	0.054
2 PUMAs Away	41	60	22	35	47	0.097	0.119	0.067	0.092	0.098
3+ PUMAs w/in State	22	31	10	18	26	0.397	0.201	0.592	0.451	0.363
1 State Away	3	4	2	3	3	0.099	0.149	0.038	0.087	0.099
2+ States Away	0.4	0.4	0.4	0.4	0.4	0.078	0.082	0.068	0.077	0.082
Out of Sample	0.4	0.6	0.3	0.4	0.5	0.179	0.230	0.139	0.161	0.196

Notes: See Table 3 for expanded definitions of the distance bins captured by the row labels. The first five columns provide the estimated gain in expected welfare (scaled in $ of annual earnings) in the destination year caused by a 500 job stimulus package for workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. The next five columns provide the share of total stimulus-driven welfare gains that accrue to workers whose distance between their origin jobs and the census tract receiving the stimulus package place them in the distance bin indicated in the row label. Each column displays the average welfare outcome by distance bin among a subset of simulations featuring focal census tracts whose characteristics align with the column label. "All": An average of all 500 target census tracts chosen as sites of simulated stimulus packages. "Rural"/"Urban": An average over the 100 census tracts featuring the lowest/highest residential density (residents per square mile) among the full 500 target tracts simulated. "Low"/"High": In Panel A (B), an average over the 100 census tracts featuring the smallest/largest number of jobs within 5 miles (poverty rate) among the full 500 target tracts simulated. "Cheap"/Expen.": An average over the 100 census tracts featuring the cheapest/most expensive rent for a two-bedroom apartment among the full 500 target tracts simulated.

Table 17: Regressions Predicting Local Employment and Welfare Incidence Using Standardized Tract Characteristics: Stimulus Packages Adding 500 Positions at High-Paying Manufacturing Firms

		All Target PUMA Workers				Low-Paid Target PUMA Workers				All Low-Paid U.S.	
Variable	Mean (S.D.)	Emp. Gain	Emp. Share	Wel. Gain	Wel. Share	Emp. Gain	Emp. Share	Wel. Gain	Wel. Share	Emp. Share	Wel. Share
Pop. Density	5071	-9.8E-06	-0.007	-36	-0.015	-1.2E-05	-0.004	-25	-0.003	0.003	0.001
	(6452)	(7.6E-06)	(0.001)	(16)	(0.001)	(1.4E-05)	(4.8E-04)	(12)	(3.7E-04)	(4.0E-04)	(4.2E-04)
Rent (Two-Bed)	1090	-8.6E-05	-0.011	-237	-0.027	-1.5E-04	-0.006	-204	-0.008	0.009	-0.001
	(478)	(9.0E-06)	(0.001)	(19)	(0.002)	(1.6E-05)	(0.001)	(14)	(4.3E-04)	(4.7E-04)	(4.9E-04)
Poverty Rate	0.156	3.7E-05	0.002	52	0.003	7.3E-05	0.002	42	0.001	-5.7E-05	-0.001
	(0.112)	(7.0E-06)	(0.001)	(14)	(0.001)	(1.2E-05)	(4.4E-04)	(11)	(3.3E-04)	(3.7E-04)	(3.8E-04)
Job Density	2779	8.1E-06	0.001	30	0.004	1.4E-05	0.001	18	0.001	-0.002	-0.001
	(7671)	(5.9E-06)	(0.001)	(12)	(0.001)	(1.0E-05)	(3.7E-04)	(9)	(2.8E-04)	(3.1E-04)	(3.2E-04)
Median Income	57950	-5.2E-05	-0.001	-108	-0.004	-7.8E-05	-2.4E-04	-81	-0.002	0.001	-0.002
	(27120)	(1.0E-05)	(0.001)	(21)	(0.002)	(1.8E-05)	(0.001)	(16)	(4.9E-04)	(0.001)	(0.001)
Jobs w/in 5 Mi.	119800	-1.2E-04	-0.002	-224	5.0E-04	-2.0E-04	-0.002	-176	-0.002	-0.001	-0.003
	(144600)	(7.6E-06)	(0.001)	(16)	(0.001)	(1.4E-05)	(4.8E-04)	(12)	(3.6E-04)	(4.0E-04)	(4.1E-04)
% College Grad.	0.280	4.8E-05	0.004	132	0.013	8.6E-05	0.002	107	0.003	-0.005	-1.2E-04
	(0.1859)	(7.8E-06)	(0.001)	(16)	(0.001)	(1.4E-05)	(4.9E-04)	(12)	(3.7E-04)	(4.1E-04)	(4.2E-04)
Outcome Mean	–	3.6E-04	0.049	754	0.095	0.001	0.032	592	0.025	0.664	0.260
R^2	–	0.298	0.229	0.314	0.250	0.277	0.209	0.344	0.338	0.238	0.079
N	–	3277	3277	3277	3277	3277	3277	3277	3277	3277	3277

Notes: This table reports regression coefficients and their accompanying standard errors (in parentheses) from tract-level regressions based on 2500 simulated stimulus packages creating 500 new positions at large, high-paying manufacturing firms in different randomly chosen focal tracts. Simulated employment and welfare outcomes listed in the column label are regressed on standardized versions of the tract characteristics associated with the focal tract that are listed in the row labels. Tract characteristics were collected by Chetty and Hendren (2018). The first four columns consider as regressands mean outcomes and shares of aggregate gains accruing workers initially in the focal PUMA receiving the stimulus, while the next four display the same regressands computed for the low-paid subset of focal PUMA workers (initially in the bottom two earnings quartiles). The final two columns display shares of employment and welfare gains accruing to low-paid workers nationally (rather than high-paid or initially unemployed workers). "Pop. Density": The focal tract's number of residents per square mile. "Rent (Two-Bed)": The focal tract's average monthly rent for a two-bedroom apartment. "Poverty Rate": The focal tract's share of households below the federal poverty line. "Job Density": The focal tract's employment per square mile. "Median Income": The focal tract's household median income. "Jobs w/in 5 Mi.": The number of jobs within 5 miles of the focal tract. "% College Grad.": The share of the focal tract's adult residents who are college graduates.

Table 18: Regressions Predicting Local Employment and Welfare Incidence Using Standardized Tract Characteristics: Stimulus Packages Adding 500 Positions at Low-Paying Retail Firms

		All Target PUMA Workers				Low-Paid Target PUMA Workers				All Low-Paid U.S.	
Variable	Mean (S.D.)	Emp. Gain	Emp. Share	Wel. Gain	Wel. Share	Emp. Gain	Emp. Share	Wel. Gain	Wel. Share	Emp. Share	Wel. Share
Pop. Density	5071	-1.2E-05	-0.007	-40	-0.015	-1.9E-05	-0.005	-36	-0.005	0.003	3.2E-04
	(6452)	(7.9E-06)	(0.001)	(13)	(0.001)	(1.6E-05)	(0.001)	(17)	(0.001)	(3.6E-04)	(3.0E-04)
Rent (Two-Bed)	1090	-8.8E-05	-0.011	-222	-0.027	-1.8E-04	-0.007	-268	-0.011	0.011	-4.5E-05
	(478)	(9.2E-06)	(0.001)	(16)	(0.002)	(1.9E-05)	(0.001)	(20)	(0.001)	(4.2E-04)	(3.5E-04)
Poverty Rate	0.156	2.8E-05	0.002	27	0.001	6.7E-05	0.002	43	0.001	-5.8E-05	-5.5E-05
	(0.112)	(7.2E-06)	(0.001)	(12)	(0.001)	(1.5E-05)	(0.001)	(15)	(0.001)	(3.3E-04)	(2.8E-04)
Job Density	2779	1.1E-05	0.002	25	0.004	2.4E-05	0.001	29	0.001	-0.001	-1.2E-04
	(7671)	(6.0E-06)	(0.001)	(10)	(0.001)	(1.3E-05)	(4.7E-04)	(13)	(4.2E-04)	(2.8E-04)	(2.3E-04)
Median Income	57950	-5.3E-05	-0.001	-84	-0.004	-9.9E-05	-0.001	-102	-0.002	0.001	-0.002
	(27120)	(1.1E-05)	(0.001)	(18)	(0.002)	(2.2E-05)	(0.001)	(23)	(0.001)	(4.8E-04)	(4.0E-04)
Jobs w/in 5 Mi.	119800	-1.2E-04	-0.001	-172	0.001	-2.3E-04	-0.001	-225	-0.002	2.5E-04	-0.001
	(144600)	(7.8E-06)	(0.001)	(13)	(0.001)	(1.6E-05)	(0.001)	(17)	(0.001)	(3.6E-04)	(3.0E-04)
% College Grad.	0.280	4.6E-05	0.003	102	0.011	1.0E-04	0.002	134	0.005	-0.007	-8.3E-05
	(0.186)	(8.0E-06)	(0.001)	(14)	(0.001)	(1.7E-05)	(0.001)	(17)	(0.001)	(3.7E-04)	(3.1E-04)
Outcome Mean	–	3.6E-04	0.049	605	0.081	0.001	0.038	777	0.035	0.695	0.326
R^2	–	0.281	0.205	0.319	0.278	0.256	0.178	0.308	0.294	0.366	0.024
N	–	3277	3277	3277	3277	3277	3277	3277	3277	3277	3277

Notes: This table reports regression coefficients and their accompanying standard errors (in parentheses) from tract-level regressions based on 2500 simulated stimulus packages creating 500 new positions at large, low-paying retail firms in different randomly chosen focal tracts. Simulated employment and welfare outcomes listed in the column label are regressed on standardized versions of the tract characteristics associated with the focal tract that are listed in the row labels. Tract characteristics were collected by Chetty and Hendren (2018). The first four columns consider as regressands mean outcomes and shares of aggregate gains accruing workers initially in the focal PUMA receiving the stimulus, while the next four display the same regressands computed for the low-paid subset of focal PUMA workers (initially in the bottom two earnings quartiles). The final two columns display shares of employment and welfare gains accruing to low-paid workers nationally (rather than high-paid or initially unemployed workers). "Pop. Density": The focal tract's number of residents per square mile. "Rent (Two-Bed)": The focal tract's average monthly rent for a two-bedroom apartment. "Poverty Rate": The focal tract's share of households below the federal poverty line. "Job Density": The focal tract's employment per square mile. "Median Income": The focal tract's household median income. "Jobs w/in 5 Mi.": The number of jobs within 5 miles of the focal tract. "% College Grad.": The share of the focal tract's adult residents who are college graduates.

Table 19: Comparing the Impact of Plant Closings and Openings at Different Scales and Distances from Focal Tract Across Several Outcomes

Panel A: Employment Outcomes

	Change in P(Employed)						Share of Employment Gains or Losses					
Distance from	Plant Opening			Plant Closing			Plant Opening			Plant Closing		
Focal Tract	125	250	500	125	250	500	125	250	500	125	250	500
Target Tract	8.5E-05	1.7E-04	3.3E-04	-0.001	-0.003	-0.006	0.004	0.004	0.004	0.059	0.064	0.071
1 Tct Away	5.9E-05	1.2E-04	2.3E-04	-5.5E-05	-1.1E-04	-2.1E-04	0.007	0.007	0.007	0.007	0.007	0.006
2 Tcts Away	4.7E-05	9.4E-05	1.8E-04	-4.3E-05	-8.4E-05	-1.6E-04	0.009	0.009	0.009	0.008	0.008	0.007
3+ Tcts w/in PUMA	3.8E-05	7.5E-05	1.5E-04	-3.6E-05	-7.0E-05	-1.3E-04	0.023	0.023	0.023	0.022	0.021	0.021
1 PUMA Away	3.2E-05	6.3E-05	1.3E-04	-3.0E-05	-5.9E-05	-1.1E-04	0.020	0.020	0.020	0.019	0.019	0.018
2 PUMAs Away	2.1E-05	4.2E-05	8.4E-05	-2.0E-05	-4.0E-05	-7.8E-05	0.039	0.039	0.039	0.037	0.037	0.036
3+ PUMAs w/in State	9.0E-06	1.8E-05	3.6E-05	-8.7E-06	-1.7E-05	-3.4E-05	0.228	0.228	0.228	0.222	0.220	0.217
1 State Away	2.5E-06	4.9E-06	9.9E-06	-2.4E-06	-4.7E-06	-9.3E-06	0.040	0.040	0.040	0.038	0.038	0.037
2+ States Away	1.2E-06	2.3E-06	4.6E-06	-1.1E-06	-2.2E-06	-4.3E-06	0.172	0.172	0.172	0.161	0.160	0.160
Out of Sample	1.4E-06	2.9E-06	5.8E-06	-1.3E-06	-2.7E-06	-5.3E-06	0.458	0.459	0.459	0.428	0.427	0.426

Panel B: Welfare Outcomes

	Change in E[Welfare] (in $)						Share of Welfare Gains or Losses					
Distance from	Plant Opening			Plant Closing			Plant Opening			Plant Closing		
Focal Tract	125	250	500	125	250	500	125	250	500	125	250	500
Target Tract	83	162	310	-3807	-7114	-12300	0.011	0.011	0.010	0.290	0.286	0.274
1 Tct Away	55	109	211	-45	-86	-160	0.016	0.016	0.016	0.012	0.012	0.012
2 Tcts Away	33	66	129	-29	-57	-109	0.019	0.019	0.019	0.013	0.013	0.013
3+ Tcts w/in PUMA	20	40	79	-19	-37	-71	0.047	0.047	0.047	0.035	0.035	0.035
1 PUMA Away	14	28	56	-13	-26	-50	0.041	0.041	0.041	0.030	0.030	0.030
2 PUMAs Away	10	20	39	-9	-18	-35	0.077	0.077	0.077	0.057	0.057	0.058
3+ PUMAs w/in State	5	10	20	-5	-9	-18	0.377	0.377	0.377	0.289	0.289	0.291
1 State Away	0.5	1.0	2	-0.5	-0.9	-2	0.050	0.050	0.050	0.036	0.036	0.037
2+ States Away	0.1	0.3	0.5	-0.1	-0.2	-0.4	0.099	0.099	0.100	0.065	0.065	0.068
Out of Sample	0.2	0.3	0.6	-0.1	-0.3	-0.6	0.262	0.263	0.263	0.173	0.176	0.183

Notes: See Table 3 for expanded definitions of the distance bins captured by the row labels, as well as definitions of the outcome measures used in both panels. The column subheadings "125", "250", and "500" indicate the number of jobs in the focal tract that were either added in "plant opening" simulations or removed in "plant closing" simulations whose incidence is summarized in the chosen column. Each "plant opening" or "plant closing" adds positions to or removes positions from large, high-paying manufacturing establishments.

Table 20: Change in Probability of Unemployment and Average Welfare Loss From a Plant Closing Removing 500 Positions at Large High-Paying Manufacturing Firms in the Target Tract Among Workers at Different Combinations of Initial Earnings Quartile (or Unemployment) and Distance from Target Tract

Panel A: Change in P(Unemployed)

Distance Bin	UE $\leq$ 25	UE $>$ 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Target Tract	0.001	0.002	0.006	0.006	0.007	0.006
1 Tct Away	3.3E-04	0.001	3.0E-04	1.7E-04	1.1E-04	7.8E-05
2 Tcts Away	2.8E-04	0.001	2.3E-04	1.3E-04	8.1E-05	5.5E-05
3+ Tcts w/in PUMA	2.3E-04	0.001	1.7E-04	1.0E-04	6.4E-05	4.9E-05
1 PUMA Away	2.0E-04	0.001	1.5E-04	8.4E-05	5.0E-05	4.1E-05
2 PUMAs Away	1.4E-04	3.5E-04	9.7E-05	5.5E-05	3.4E-05	2.8E-05
3+ PUMAs w/in State	6.7E-05	1.6E-04	4.0E-05	2.2E-05	1.4E-05	1.1E-05
1 State Away	1.7E-05	4.4E-05	1.2E-05	6.2E-06	3.7E-06	3.2E-06
2+ States Away	8.2E-06	2.1E-05	5.4E-06	2.8E-06	1.7E-06	1.5E-06
Out of Sample	2.3E-05	2.6E-05	8.1E-06	3.6E-06	2.3E-06	2.2E-06

Panel B: Average Welfare Loss

Distance Bin	UE $\leq$ 25	UE $>$ 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Target Tract	-189	-248	-2955	-5804	-11320	-17200
1 Tct Away	-114	-108	-119	-152	-181	-207
2 Tcts Away	-88	-74	-86	-102	-116	-148
3+ Tcts w/in PUMA	-58	-50	-56	-66	-76	-97
1 PUMA Away	-44	-37	-43	-48	-53	-61
2 PUMAs Away	-33	-26	-31	-34	-37	-43
3+ PUMAs w/in State	-19	-14	-17	-18	-19	-20
1 State Away	-2	-1	-2	-2	-2	-2
2+ States Away	-0.4	-0.2	-0.4	-0.4	-0.5	-0.6
Out of Sample	-0.7	-0.1	-0.5	-0.6	-0.6	-0.6

Notes: See Table 3 for expanded definitions of the distance bins represented by the row labels. See Table 6 for expanded definitions of the origin employment status/earnings quartiles indicated by the column labels. Each entry gives the change in the probability of employment (Panel A) or the expected welfare loss (Panel B) among workers whose initial location falls into the distance bin associated with the row label and whose initial employment status/earnings quartile falls into the bin associated with the column sublabel due to a simulated plant closing in which 500 positions are removed at either large, high-paying manufacturing firms. The average is taken across 200 simulations featuring different target census tracts.

Table 21: Heterogeneity by Plant Composition in Incidence Among Workers at Different Initial Earnings Quartiles (or Unemployed): Comparing Plant Openings and Closings Featuring High-Paying Manufacturing Positions vs. Low-Paying Retail Positions

Panel A: Share of All Gains/Losses

	Share of Emp. Change				Share of Wel. Change			
	Opening		Closing		Opening		Closing	
Earn Cat.	Manu.	Ret.	Manu.	Ret.	Manu.	Ret.	Manu.	Ret.
UE $\leq$ Age 25	0.096	0.126	0.087	0.103	0.044	0.079	0.031	0.053
UE $>$ Age 25	0.317	0.325	0.293	0.298	0.040	0.049	0.028	0.038
1st Quartile	0.247	0.252	0.241	0.280	0.164	0.209	0.131	0.228
2nd Quartile	0.149	0.138	0.151	0.152	0.214	0.223	0.180	0.239
3rd Quartile	0.097	0.084	0.108	0.090	0.244	0.222	0.238	0.226
4th Quartile	0.095	0.076	0.121	0.079	0.294	0.217	0.392	0.216

Panel B: Expected Employment/Welfare Gain or Loss among Focal Tract Workers

	Change in P(Employed)				Change in E[Welfare]			
	Opening		Closing		Opening		Closing	
Earn Cat.	Manu.	Ret.	Manu.	Ret.	Manu.	Ret.	Manu.	Ret.
UE $\leq$ Age 25	0.001	0.001	-0.001	-0.001	232	347	-189	-208
UE $>$ Age 25	0.003	0.001	-0.002	-0.001	417	132	-248	-121
1st Quartile	3.8E-04	3.3E-04	-0.006	-0.016	174	124	-2955	-5772
2nd Quartile	2.3E-04	1.4E-04	-0.006	-0.007	251	103	-5804	-5572
3rd Quartile	1.5E-04	6.6E-05	-0.007	-0.003	312	93	-11320	-4582
4th Quartile	8.5E-05	3.9E-05	-0.006	-0.001	354	84	-17200	-3810

Notes: Panel A displays the shares of all employment and welfare gains or losses (in columns labeled "Share of Emp. Change" and "Share of Wel. Change", respectively) generated by the simulated plant openings or closings that accrue to all workers nationally that were assigned in the initial employment status bin indicated by the row label. Panel B displays the expected change in employment probability and job-related welfare (scaled in $ of 2012 annual earnings) from these openings and closings that accrue to local workers (those initially employed (or unemployed) in the focal tract) who belonged to the initial employment status bin indicated by the row label. See Table 6 for expanded definitions of the origin employment status/earnings quartiles indicated by the row labels. The column subheadings "Opening" and "Closing" indicate whether the results displayed in the chosen column reflect simulated plant openings featuring the creation of 500 jobs from the focal tract or plant closings featuring the removal of 500 jobs. The column subheadings "Manu" and "Ret." indicate whether the results displayed in the chosen column reflect the creation or destruction of 500 positions at large, high paying manufacturing firms or large, low-paying retail firms, respectively.

Table 22: Change in Probability of Destination Employment (or Nonemployment) at Different Distances from Focal Tract after a Plant Closing Removing 500 Positions at either Manufacturing or Retail Firms for Workers Initially Employed in the Focal Tract by Initial Earnings Quartile (or Nonemployment)

Panel A: Large High-Paying Manufacturing

Distance Bin	Overall	UE ≤ 25	UE > 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Unemployment	0.006	0.001	0.002	0.006	0.006	0.007	0.006
Target Tract	-0.040	-0.003	-0.004	-0.028	-0.036	-0.052	-0.060
1 Tct Away	0.002	-3.4E-05	2.0E-05	0.001	0.002	0.002	0.002
2 Tcts Away	0.002	2.6E-05	5.6E-05	0.001	0.002	0.002	0.002
3+ Tcts w/in PUMA	0.004	1.8E-04	2.7E-04	0.003	0.004	0.005	0.005
1 PUMA Away	0.003	2.4E-04	2.4E-04	0.002	0.003	0.004	0.004
2 PUMAs Away	0.004	3.5E-04	3.5E-04	0.003	0.004	0.006	0.006
3+ PUMAs w/in State	0.011	0.001	0.001	0.008	0.011	0.014	0.017
1 State Away	0.001	8.9E-05	8.9E-05	4.0E-04	0.001	0.001	0.002
2+ States Away	0.002	1.8E-04	1.8E-04	0.001	0.001	0.002	0.004
Out of Sample	0.006	8.5E-05	1.5E-04	0.002	0.004	0.009	0.012

Panel B Large Low-Paying Retail

Distance Bin	Overall	UE ≤ 25	UE > 25	1st Q.	2nd Q.	3rd Q.	4th Q.
Unemployment	0.006	0.001	0.001	0.016	0.007	0.003	0.001
Target Tract	-0.034	-0.005	-0.002	-0.084	-0.047	-0.022	-0.010
1 Tct Away	0.001	1.2E-04	-2.4E-05	0.003	0.002	0.001	3.6E-04
2 Tcts Away	0.001	1.5E-04	-1.5E-05	0.003	0.002	0.001	3.6E-04
3+ Tcts w/in PUMA	0.003	4.2E-04	4.6E-05	0.007	0.004	0.002	0.001
1 PUMA Away	0.002	3.7E-04	7.5E-05	0.005	0.003	0.001	0.001
2 PUMAs Away	0.004	0.001	1.6E-04	0.010	0.006	0.003	0.001
3+ PUMAs w/in State	0.012	0.002	0.001	0.031	0.016	0.007	0.004
1 State Away	0.001	1.2E-04	4.5E-05	0.001	0.001	3.9E-04	2.0E-04
2+ States Away	0.001	2.6E-04	1.0E-04	0.002	0.001	0.001	3.8E-04
Out of Sample	0.003	9.1E-05	8.9E-05	0.005	0.005	0.003	0.002

Notes: See Table 3 for expanded definitions of the distance bins represented by the row labels. See Table 6 for expanded definitions of the origin employment status/earnings quartiles indicated by the column sublabels. Each entry gives the change in the probability of employment at a location whose distance falls into the distance bin associated with the row label among workers whose initial employment status/earnings quartile falls into the bin associated with the column sublabel for workers initially working (or most recently working) in the focal census tract. The changes in employment probability are due to a simulated plant closing in which 500 positions are removed at either large, high-paying manufacturing firms (Panel A) or large, low-paying retail firms (Panel B) . Each entry represents an average over 200 simulations featuring different target census tracts. The entries in the row labeled "Unemployment" provides the change in the share of workers who stay or become unemployed due to the plant closing.

Table 23: Assessing Robustness to Model Assumptions: Employment and Welfare Incidence from Plant Opening Simulations for Alternative Models

Panel A: Employment Outcomes

	Change in P(Employed)				Share of Employment Gains			
Distance from Focal Tract	Base Spec.	Job Mult.	Endo. Vac.	Choo-Siow	Base Spec.	Job Mult.	Endo. Vac.	Choo-Siow
Target Tract	0.001	0.002	0.001	0.001	0.004	0.003	0.003	0.003
1 Tct Away	5.0E-04	0.001	4.5E-04	3.9E-04	0.007	0.007	0.006	0.006
2 Tcts Away	3.1E-04	0.001	2.9E-04	2.5E-04	0.010	0.010	0.010	0.008
3+ Tcts w/in PUMA	2.0E-04	3.8E-04	1.8E-04	1.6E-04	0.027	0.030	0.026	0.023
1 PUMA	1.5E-04	2.7E-04	1.4E-04	1.4E-04	0.026	0.027	0.026	0.024
2 PUMAs Away	9.5E-05	1.7E-04	9.2E-05	9.2E-05	0.051	0.053	0.050	0.049
3+ PUMAs w/in State	3.6E-05	6.2E-05	3.5E-05	3.5E-05	0.247	0.252	0.250	0.244
1 State Away	1.1E-05	1.9E-05	1.1E-05	1.1E-05	0.074	0.074	0.073	0.074
2+ States Away	4.3E-06	7.4E-06	4.2E-06	4.7E-06	0.147	0.146	0.147	0.160
Out of Sample	5.1E-06	8.6E-06	5.0E-06	5.1E-06	0.408	0.398	0.408	0.411

Panel B: Welfare Outcomes

	Avg. Welfare Gain ($)				Share of Welfare Gains			
Distance from Focal Tract	Base Spec.	Job Mult.	Endo. Vac.	Choo-Siow	Base Spec.	Job Mult.	Endo. Vac.	Choo-Siow
Target Tract	1049	1254	910	369	0.008	0.006	0.007	0.002
1 Tct Away	333	489	282	245	0.014	0.012	0.013	0.005
2 Tcts Away	198	316	171	203	0.020	0.019	0.019	0.008
3+ Tcts w/in PUMA	99	173	89	169	0.053	0.056	0.051	0.022
1 PUMA	65	108	62	169	0.050	0.051	0.050	0.024
2 PUMAs Away	42	72	40	152	0.095	0.099	0.095	0.042
3+ PUMAs w/in State	22	38	21	126	0.391	0.405	0.391	0.210
1 State Away	3	5	3	113	0.096	0.100	0.097	0.071
2+ States Away	0.4	0.7	0.4	111	0.078	0.076	0.079	0.187
Out of Sample	0.5	0.7	0.5	111	0.196	0.178	0.198	0.430

Notes: See Table 3 for expanded definitions of the distance bins captured by the row labels, as well as definitions of the outcome measures used in both panels. The mean outcomes displayed for each of four alternative models are averages over 500 simulations with different focal tracts featuring the creation of 500 positions at large, high-paying manufacturing firms. "Base. Spec.": The baseline assignment model described in Sections 2, 3, and 5; "Job Mult.": the baseline assignment model is augmented with a job multiplier process in which the original 500 manufacturing positions spawn additional service-sector jobs throughout the target PUMA, using a high-tech manufacturing multiplier of 1.71 from Bartik and Sotherland (2019); "Endo. Vac.": the baseline assignment model is augmented by allowing nearby firms to endogenously adjust the number of positions they wish to fill in response to stimulus-induced increases in required pay per efficiency unit of labor. Final equilibrium is determined by the convergence of a fixed point. "Choo-Siow": the baseline assignment model is altered to mimic Choo and Siow (2006) by replacing the single idiosyncratic surplus component ϵ_{ik} with the sum of two components $\epsilon^1_{lk} + \epsilon^2_{if}$.

Table 24: Model Validation Results: Dissimilarity Index Values Comparing Forecasted and Actual Worker Reallocations Following Large Local Shocks Using Alternative Transition Group Definitions and Methods for Generating Forecasts

	All U.S.					Target PUMA Only				
Level of Group Aggregation	Two-Sided Matching	Param. Logit	Raw CCP	Smoothed CCP	Choo-Siow	Two-Sided Matching	Param. Logit	Raw CCP	Smoothed CCP	Choo-Siow
Full Group Space	0.062	0.383	0.074	0.060	0.056	0.341	0.433	0.370	0.344	0.324
	(4.7E-04)	(0.003)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)
Sm. Dist. Bins	0.045	0.376	0.057	0.044	0.041	0.193	0.399	0.224	0.192	0.175
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Sm. Dist. Bins & No Firm Char.	0.010	0.186	0.018	0.016	0.010	0.042	0.282	0.062	0.060	0.044
	(4.5E-04)	(0.002)	(0.001)	(0.001)	(4.7E-04)	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)
Lg. Dist. Bins & No Firm Char.	0.009	0.178	0.016	0.013	0.009	0.037	0.275	0.056	0.053	0.039
	(4.5E-04)	(0.002)	(0.001)	(0.001)	(4.7E-04)	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)
E-NE & NE-E Only & All Loc.	0.009	0.186	0.020	0.020	0.009	0.027	0.235	0.077	0.084	0.034
	(4.4E-04)	(0.002)	(4.9E-04)	(4.8E-04)	(4.5E-04)	(0.001)	(0.004)	(0.002)	(0.003)	(0.001)
E-NE & NE-E Only & Lg. Dist. Bins	0.009	0.173	0.016	0.013	0.009	0.013	0.195	0.036	0.036	0.016
	(4.5E-04)	(0.002)	(0.001)	(0.001)	(4.7E-04)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)

Notes: This table examines the fit of model-based predicted worker reallocations to the actual reallocations that occurred following a set of local employment shocks to particular census tracts in particular years spanning 2003-2012. See Section A8 for a detailed description of the model validation exercise. Each row of the table considers a different metric for measuring model fit, while each column considers a different combination of model and target population. Columns 1-5 examine the job reallocation fit among all U.S. citizens in my 19 state LEHD sample, while columns 6-10 consider the fit only among workers initially working in the same PUMA as the tract receiving the shock. Each entry averages the fit metric across all 180 local shocks identified. For each shock, predictions are based on parameters estimated using local data from the year before the shock occurred. "Two-sided Matching" refers to the preferred two-sided matching model presented in this paper. "Param. Logit" refers to a one-sided parametric conditional logit model (See A8 for a list of the predictor variables). "Raw CCP" refers to a prediction that holds the previous year's conditional choice probability (CCP) distribution constant for each destination type, but updates the destination type marginal distribution to reflect the shock, while "Smoothed CCP" does the same but smooths the CCPs across similar destination types before constructing the predicted reallocation. None of the three alternative models impose market clearing. "Choo-Siow" uses Choo and Siow (2006)'s version of the assignment model to generate predicted allocations. This model replaces the idiosyncratic surplus component ϵ_{ik} with the sum of two components $\epsilon^1_{lk} + \epsilon^2_{if}$. "Full Group Space" evaluates model fit using the index of dissimilarity between the actual and predicted distribution across groups in the transition group space. "Sm. Dist. Bins", "Sm. Dist. Bins & No Firm Char" and "Lg. Dist. Bins & No Firm Char" evaluate the index of dissimilarity on aggregated group spaces in which origin and destination locations are each aggregated to small or large distance bins relative to the focal tract, and, in the latter two cases, destination types featuring the same distance bin but different non-location characteristics are combined. "E-to-UE and UE-to-E Only (All Loc.)" calculates the index of dissimilarity only among transition groups featuring employment-to-unemployment and unemployment-to-employment transitions, while "E-to-UE and UE-to-E Only (Lg. Dist. Bins)" does the same but aggregates locations to large distance bins relative to the focal census tract.

Online Appendix

A1 Proof of Proposition A1

Proposition A1:

Let $|l|$ and $|g_k|$ denote, respectively, the number of workers classified as worker type l and the number of workers whose job match would be classified as group g (either stayers or new hires among those in l) if hired by position k (a subset of the workers in $l(g)$). In addition, let $n(l)$ denote the share of all workers assigned to worker type l, so that $|l| = n(l)I$. Further, define C_l as the mean value of $e^{-\frac{r_i}{\sigma}}$ for a given worker type l. Define $S_{g|l,k}$ as the share of workers of worker type l who would be assigned to group g if they filled position k (i.e. potential stayers if $z(g) = 1$, movers if $z(g) = 0$), and define $\overline{S}_{g|l,f}$ to be the mean of $S_{g|l,k}$ among all k assigned to position type f. Suppose the following assumptions hold:

$$\textit{Assumption 1:}\quad \frac{1}{|g_k|}\sum_{i:g(i,k)=g} e^{-\frac{r_i}{\sigma}} \approx \frac{1}{|l|}\sum_{i:l(i)=l(g)} e^{-\frac{r_i}{\sigma}} = C_{l(g)}\ \forall (g,k) \tag{14}$$

$$\textit{Assumption 2:}\quad S_{g|l,k} \approx \overline{S}_{g|l,f}\ \forall\, k,\ \forall\, g \tag{15}$$

Then the equilibrium aggregate group-level choice probabilities can be written as follows:

$$P(g|f) = \frac{e^{\frac{\theta_g}{\sigma}}\overline{S}_{g|l,f}n(l)C_l}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)} e^{\frac{\theta_{g'}}{\sigma}}\overline{S}_{g'|l',f}n(l')C_{l'}} \tag{16}$$

Proof: First, note that the law of total probability implies:

$$P(g|f) = \sum_{k\in f} P(g|f,k)P(k|f) = \frac{1}{|f|}\sum_{k\in f} P(g|k) = \frac{1}{|f|}\sum_{k\in f}\sum_{i:g(i,k)=g} P(i|k)$$
$$= \frac{1}{|f|}\sum_{k\in f}\sum_{i:g(i,k)=g} \frac{e^{\frac{\theta_g - r_i}{\sigma}}}{\sum_{i'\in\mathcal{I}} e^{\frac{\theta_{g'}-r_{i'}}{\sigma}}} = \frac{1}{|f|}\sum_{k\in f} \frac{(e^{\frac{\theta_g}{\sigma}})(\sum_{i:g(i,k)=g} e^{\frac{-r_i}{\sigma}})}{\sum_{i'\in\mathcal{I}} e^{\frac{\theta_{g'}-r_{i'}}{\sigma}}}, \tag{17}$$

where $|f|$ captures the number of positions k assigned to position type f.

Assumption 1 imposes that the mean exponentiated worker utility values $e^{\frac{-r_i}{\sigma}}$ vary minimally across groups g featuring the same worker type $l(g)$. Given the characteristics used to define l and g in the empirical application, this states that existing employees (potential stayers) and non-employees of each establishment have approximately the same mean value of r_i among workers whose initial jobs were in the same local area and pay category.[47] In other words, the payoffs that

[47] Recall that the only characteristic z that distinguishes match groups featuring the same combination of worker and position types (l, f) is an indicator for whether the worker i was already employed by k in the previous period, so that a given (l, f) pair contains at most two groups, potential stayers and potential new hires.

workers in the same skill class require in equilibrium will not differ systematically across establishments within a small local area. This becomes a better approximation as more characteristics and categories are used to define a worker type $l(i)$.

Assumption 2 imposes that the share of potential stayers vs. new hires among workers from each worker type l is common across establishments within position type f. In the chosen context, this means that establishments in the same geographic area, industry supersector, and establishment size and average pay categories have roughly the same number and past pay composition of employees.

These assumptions are necessary because the aggregate mean of a non-linear function of a random variable (in this case $e^{\frac{-r_i}{\sigma}}$) depends on its entire distribution. Essentially, the probability of filling a position with an existing employee depends on how many employees one already has, so that the group average depends on the establishment size distribution among firms who are at risk of creating a match that could be classified into g. We are essentially hoping that Jensen's inequality is close to equality ($f(E[X]) \approx E[f(X)]$) after conditioning on the characteristics that define the worker and position types (most notably establishment size category).

Note first that Assumption 2 implies that $|g_k| \equiv S_{g|l,k} n(l(g))I \approx \overline{S}_{g|l,f} n(l(g))I$. Thus, Assumptions 1 and 2 together imply:

$$\sum_{i:g(i,k)=g} e^{-\frac{r_i}{\sigma}} \approx \overline{S}_{g|l(g),f(g)} n(l(g))(I)C_{l(g)}. \tag{18}$$

Applying this result to the last expression in (17), one obtains:

$$P(g|f) = \sum_{k\in f}(\frac{1}{|f|})\frac{e^{\frac{\theta_g}{\sigma}}\sum_{i:g(i,k)=g} e^{-\frac{r_i}{\sigma}}}{\sum_{i'\in\mathcal{I}} e^{\frac{\theta_{g'}-r_{i'}}{\sigma}}} = \sum_{k\in f}(\frac{1}{|f|})\frac{e^{\frac{\theta_g}{\sigma}}\sum_{i:g(i,k)=g} e^{-\frac{r_i}{\sigma}}}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)}\sum_{i':g(i',k)=g'} e^{\frac{\theta_{g'}-r_{i'}}{\sigma}}}$$

$$= \sum_{k\in f}(\frac{1}{|f|})\frac{e^{\frac{\theta_g}{\sigma}}\overline{S}_{g|l,f} n(l)(I)C_l}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)} e^{\frac{\theta_{g'}}{\sigma}}\overline{S}_{g'|l',f} n(l')(I)C_{l'}}$$

$$= \frac{e^{\frac{\theta_g}{\sigma}}\overline{S}_{g|l,f} n(l)(I)C_l}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)} e^{\frac{\theta_{g'}}{\sigma}}\overline{S}_{g'|l',f} n(l')(I)C_{l'}}\sum_{k\in f}(\frac{1}{|f|}) = \frac{e^{\frac{\theta_g}{\sigma}}\overline{S}_{g|l,f} n(l)C_l}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)} e^{\frac{\theta_{g'}}{\sigma}}\overline{S}_{g'|l',f} n(l')C_{l'}} \tag{19}$$

This concludes the proof.

A2 Proof of Proposition 1

Proposition 1:

Define the set $\Theta^{D-in-D} \equiv \{\frac{(\theta_g-\theta_{g'})-(\theta_{g''}-\theta_{g'''})}{\sigma}\,\forall\,(g,g',g'',g''') : l(g)=l(g''), l(g')=l(g'''), f(g)=f(g'), f(g'')=f(g''')\}$. *Given knowledge of* Θ^{D-in-D}, *a set* $\tilde{\Theta}=\{\tilde{\theta}_g\ \forall\ g\in\mathcal{G}\}$ *can be constructed such that the unique group level assignment* $P^{CF}(g)$ *that satisfies the market-clearing conditions (13) using* $\theta^{CF}_g=\tilde{\theta}_g\,\forall\,g$ *and arbitrary marginal PMFs for worker and position types* $n^{CF}(*)$

and $g^{CF}(*)$ will also satisfy the corresponding market-clearing conditions using $\theta_g^{CF} = \theta_g \,\forall\, g \in \mathcal{G}$ and the same PMFs $n^{CF}(*)$ and $g^{CF}(*)$. Furthermore, denote by $\tilde{\mathbf{C}}^{\mathbf{CF}} \equiv \{\tilde{C}_1^{CF}, \ldots, \tilde{C}_L^{CF}\}$ and $\mathbf{C}^{\mathbf{CF}} \equiv \{C_1^{CF}, \ldots, C_L^{CF}\}$ the utility vectors that clear the market using $\theta_g^{CF} = \tilde{\theta}_g$ and using $\theta_g^{CF} = \theta_g$, respectively. Then $\tilde{\mathbf{C}}^{\mathbf{CF}}$ will satisfy $\tilde{C}_l^{CF} = C_l^{CF} e^{\frac{-\Delta_l}{\sigma}} \,\forall\, l \in \mathcal{L}$ for some set of worker type-specific constants $\{\Delta_l : l \in [1, L]\}$ that is invariant to the choices of $n^{CF}(*)$ and $g^{CF}(*)$.

Proof: We prove Proposition 1 by construction.

Let $z(i,k) = 1(m(j(i)) = m(k))$ represent an indicator that takes on the value of 1 if the firms associated with positions $j(i)$ and k are the same, and 0 otherwise. Recall also that all job matches assigned to the same match group g share values of the worker and establishment characteristics that define the worker and position types l and f, respectively, as well as the value of the indicator $z(i,k)$. Thus, one can write $l(g)$, $f(g)$ and $z(g)$ for any group g. Let the worker types be ordered (arbitrarily) from $l = 1 \ldots l = L$, and let the position types be ordered (arbitrarily) from $f = 1 \ldots f = F$. Let $g(l,f,z)$ denote the group associated with worker type l, position type f, and existing worker indicator z. Assume that the set $\Theta^{D-in-D} = \{\frac{(\theta_g - \theta_{g'}) - (\theta_{g''} - \theta_{g'''})}{\sigma} \,\forall\, (g, g', g'', g''')\}$ is known, since a consistent estimator for each element of the set can be obtained via adjusted log odds ratios, as described in Section 3. Consider defining the set of alternative group-level joint surplus values $\tilde{\Theta} = \{\tilde{\theta}_g\}$ as follows:

$$\tilde{\theta}_{g'} = 0 \,\forall\, g' : (l(g') = 1 \text{ and/or } f(g') = 1) \text{ and } z(g') = 0 \tag{20}$$

$$\tilde{\theta}_{g'} = \frac{(\theta_{g'} - \theta_{g(1,f(g'),0)}) - (\theta_{g(l(g'),1,0)} - \theta_{g(1,1,0)})}{\sigma} \,\forall\, g' : (f(g') \neq \text{ and } l(g') \neq 1) \text{ and/or } z(g') \neq 0 \tag{21}$$

Under the definitions in (20) and (21), we have:

$$\frac{(\tilde{\theta}_g - \tilde{\theta}_{g'}) - (\tilde{\theta}_{g''} - \tilde{\theta}_{g'''})}{\sigma} = \frac{(\theta_g - \theta_{g'}) - (\theta_{g''} - \theta_{g'''})}{\sigma}$$
$$\forall\, (g, g', g'', g''') : l(g) = l(g''), l(g') = l(g'''), f(g) = f(g'), f(g'') = f(g''') \tag{22}$$

Thus, the appropriate difference-in-differences using elements of $\tilde{\Theta}$ match their analogues among the true surpluses in Θ^{D-in-D}, so that all the information about Θ in the identified set Θ^{D-in-D} is retained. And unlike the true set Θ, the construction of $\tilde{\Theta}$ only requires knowledge of Θ^{D-in-D}.

Next, note that the elements of $\tilde{\Theta}$ can be written in the following form:

$$\tilde{\theta}_g = \theta_g + \Delta^1_{l(g)} + \Delta^2_{f(g)} \,\forall\, g \in \mathcal{G}, \text{ where} \tag{23}$$

$$\Delta^1_{l(g)} = \theta_{g(l(g),1,0)} - \theta_{g(1,1,0)} \quad \text{and} \quad \Delta^2_{f(g)} = \theta_{g(1,f(g),0)} \tag{24}$$

where $\mathcal{G}$ is the set of all possible match groups. In other words, each alternative surplus $\tilde{\theta}_g$ equals the true surplus θ_g plus a constant $(\Delta^1_{l(g)})$ that is common to all groups featuring the same worker type and a constant ($\Delta^2_{f(g)}$) that is common to all groups featuring the same position type.

Next, recall that there exists a unique aggregate assignment associated with each combination of marginal worker and position type distributions $n^{CF}(l)$ and $h^{CF}(f)$ and set of group-level surpluses, including $\tilde{\Theta}$. Let $\tilde{P}^{CF}(*) \equiv P^{CF}(*|\tilde{\Theta}, \tilde{C}^{CF}_2, \ldots, \tilde{C}^{CF}_L))$ represent the assignment that results from combining arbitrary marginals $n^{CF}(l)$ and $h^{CF}(f)$ with $\tilde{\Theta}$. $\tilde{\mathbf{C}}^{\mathbf{CF}} = [1, \tilde{C}^{CF}_2 \ldots \tilde{C}^{CF}_L]$ denotes the vector of mean exponentiated utility values for each worker type l (with $\tilde{C}^{CF}_1$ normalized to 1) that solves the system of excess demand equations below, and thus yields $\tilde{P}^{CF}(g) \; \forall \; g \in \mathcal{G}$ when plugged into equation (10) along with the elements of $\tilde{\Theta}$, n^{CF} and $\overline{S}^{CF}_{g'|l(g'),d}$:

$$\begin{aligned}
&\sum_{f \in \mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=2} P^{CF}(g|f, \tilde{\Theta}, \tilde{\mathbf{C}}^{\mathbf{CF}})) = n^{CF}(2) \\
&\vdots \\
&\sum_{f \in \mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=L} P^{CF}(g|f, \tilde{\Theta}, \tilde{\mathbf{C}}^{\mathbf{CF}})) = n^{CF}(l)
\end{aligned} \tag{25}$$

We wish to show that $\tilde{P}^{CF}(*) \equiv P^{CF}(*|\tilde{\Theta}, \tilde{\mathbf{C}}^{\mathbf{CF}})$ will be identical to the alternative unique counterfactual equilibrium assignment $P^{CF}(*|\Theta, \mathbf{C}^{\mathbf{CF}})$ that combines the same arbitrary marginal distributions $n^{CF}(l)$ and $h^{CF}(f)$ with the set Θ instead of $\tilde{\Theta}$. Here, $\mathbf{C}^{\mathbf{CF}} = [1, C^{CF}_2 \ldots C^{CF}_L]$ denotes a vector of l-type-specific mean exponentiated utility values that clears the market by satisfying the following alternative excess demand equations:[48]

$$\begin{aligned}
&\sum_{f \in \mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=2} P^{CF}(g|f, \Theta, \mathbf{C}^{\mathbf{CF}})) = n^{CF}(2) \\
&\vdots \\
&\sum_{f \in \mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=L} P^{CF}(g|f, \Theta, \mathbf{C}^{\mathbf{CF}})) = n^{CF}(l)
\end{aligned} \tag{26}$$

Since all other terms are shared by the systems (25) and (26), it suffices to show that $P^{CF}(g|f, \tilde{\Theta}, \tilde{\mathbf{C}}^{\mathbf{CF}}) = P^{CF}(g|f, \Theta, \mathbf{C}^{\mathbf{CF}})) \; \forall \; g \in \mathcal{G}$ for some vector $\mathbf{C}^{\mathbf{CF}}$. Consider the following vector $\mathbf{C}^{\mathbf{CF}}$:

$$C^{CF}_l = \tilde{C}^{CF}_l e^{\frac{\Delta^1_l}{\sigma}} \; \forall \; l \in [2, \ldots, L] \tag{27}$$

where Δ^1_l is as defined in (24). For an arbitrary choice of g, we obtain:

$$P^{CF}(g|f(g), \tilde{\Theta}, \tilde{\mathbf{C}}^{\mathbf{CF}}) = \frac{e^{\frac{\tilde{\theta}^{CF}_g}{\sigma}} \overline{S}^{CF}_{g|l(g),f(g)} n^{CF}(l(g)) \tilde{C}^{CF}_l}{\sum_{l' \in \mathcal{L}} \sum_{g' \in (l',f)} e^{\frac{\tilde{\theta}^{CF}_{g'}}{\sigma}} \overline{S}^{CF}_{g'|l'(g'),f(g)} n^{CF}(l') \tilde{C}^{CF}_{l'}}$$

[48]Note that we have suppressed the dependence of $P^{CF}(*|\Theta, \mathbf{C}^{\mathbf{CF}}, n^{CF}(l), h^{CF}(f), \overline{S}_{g|l,f})$ on $n^{CF}(l)$, $h^{CF}(f)$, and $\overline{S}_{g|l,f}$ because these are held fixed across the two alternative counterfactual simulations.

$$
\begin{aligned}
&= \frac{e^{\frac{(\theta_g^{CF}+\Delta^1_{l(g)}+\Delta^2_{f(g)})}{\sigma}} \overline{S}^{CF}_{g|l(g),f(g)} n^{CF}(l(g)) C_l^{CF} e^{\frac{-\Delta^1_l}{\sigma}}}{\sum_{l'\in\mathcal{L}} \sum_{g'\in(l',f)} e^{\frac{(\theta_{g'}^{CF}+\Delta^1_{l(g')}+\Delta^2_{f(g')})}{\sigma}} \overline{S}^{CF}_{g'|l'(g'),f(g)} n^{CF}(l') C_{l'}^{CF} e^{\frac{-\Delta^1_{l'}}{\sigma}}} \\
&= e^{\frac{\Delta^1_{l(g)}}{\sigma}} e^{\frac{\Delta^2_{f(g)}}{\sigma}} e^{\frac{-\Delta^1_{l(g)}}{\sigma}} \frac{e^{\frac{\theta_g^{CF}}{\sigma}} \overline{S}^{CF}_{g|l(g),f(g)} n^{CF}(l(g)) C_l^{CF}}{e^{\frac{\Delta^2_{f(g)}}{\sigma}} \sum_{l'\in\mathcal{L}} e^{\frac{\Delta^1_{l(g')}}{\sigma}} e^{\frac{-\Delta^1_{l(g')}}{\sigma}} \sum_{g'\in(l',f)} e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}^{CF}_{g'|l'(g'),f(g)} n^{CF}(l') C_{l'}^{CF}} \\
&= \frac{e^{\frac{\theta_g^{CF}}{\sigma}} \overline{S}^{CF}_{g|l(g),f(g)} n^{CF}(l(g)) C_l^{CF}}{\sum_{l'\in\mathcal{L}} \sum_{g'\in(l',f)} e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}^{CF}_{g'|l'(g'),f(g)} n^{CF}(l') C_{l'}^{CF}} = P^{CF}(g|f,\Theta,\mathbf{C^{CF}}) \qquad (28)
\end{aligned}
$$

This proves that $P^{CF}(g|f,\Theta,\mathbf{C^{CF}})$ also satisfies the market clearing conditions (26) above, and will therefore be the unique group-level assignment consistent with marketwide equilibrium and stability. Thus, we have shown that the counterfactual assignment that is recovered when using an alternative set of surpluses $\tilde{\Theta}$ derived from the identified set Θ^{D-in-D} will in fact equal the counterfactual assignment we desire, which is based on the true set of joint surplus values Θ. Furthermore, while worker-type specific mean utility values $\mathbf{\tilde{C}^{CF}}$ that clear the market given $\tilde{\Theta}$ will differ for each worker type from the corresponding vector $\mathbf{C^{CF}}$ based on the true surplus set Θ, these differences are invariant to the marginal worker type and position type distributions $n^{CF}(l)$ and $h^{CF}(f)$ used to define the counterfactual. This implies that differences in utility gains caused by alternative counterfactuals among worker types are identified, permitting comparisons of the utility incidence of alternative labor supply or demand shocks. This concludes the proof.

A3 Proof of Proposition A2

Proposition A2:

Suppose the following assumptions hold:

1') *The assumptions laid out in sections 2.2 and 3 continue to hold. Namely, each joint surplus π_{ik} is additively separable in the group-level and idiosyncratic components, the vector of idiosyncratic components ϵ_{ik} is independently and identically distributed, and follows the type 1 extreme value distribution, and Assumptions 1 and 2 hold.*

2') *The set of destination positions $k \in \tilde{\mathcal{K}}$ that will be filled in the stable counterfactual assignment are known in advance, and the set of destination positions $k \in \tilde{\mathcal{K}}$ that will remain unfilled in the stable counterfactual assignment are ignorable, in the sense that their existence does not change the assignment nor the division of surplus among the remaining set of positions $\mathcal{K}$ and set of workers $\mathcal{I}$.*

3') $\frac{1}{|g_i|}\sum_{k:g(i,k)=g} e^{-\frac{q_k}{\sigma}} \approx \frac{1}{|f|}\sum_{k:f(k)=f(g)} e^{-\frac{q_k}{\sigma}} = C_{f(g)}\ \forall (g,i)$.

4') $P(g|i,f(g)) \approx P(g|l(g),f(g))\ \forall (g,i)$.

Then the group-level assignment $P^{CF}(g)$ that satisfies the following $L-1$ excess demand equations represents the unique group-level equilibrium assignment $P^{CF^*}(g)$ consistent with the unique worker/position level stable matching μ^{CF}:

$$\sum_{f\in\mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=2} P^{CF}(g|f,C_2^{CF},\ldots,C_L^{CF}) = n^{CF}(2)$$
$$\vdots$$
$$\sum_{f\in\mathcal{F}} h^{CF}(f)(\sum_{g:l(g)=L} P^{CF}(g|f,C_2^{CF},\ldots,C_L^{CF}) = n^{CF}(L) \tag{29}$$

where $P^{CF}(g|f,C_2^{CF},\ldots,C_L^{CF})$ is given by:

$$P^{CF}(g|f) = \frac{e^{\frac{\theta_g^{CF}}{\sigma}}\overline{S}^{CF}_{g|l(g),f}n^{CF}(l(g))C_l^{CF}}{\sum_{l'\in\mathcal{L}}\sum_{g'\in(l,f)} e^{\frac{\theta_{g'}^{CF}}{\sigma}}\overline{S}^{CF}_{g'|l(g'),d}n^{CF}(l')C_{l'}^{CF}}\ \forall\ f\in[1,\ldots,F] \tag{30}$$

Proof: Proposition A2 states that assignment $P^{CF}(g)$ implied by the vector of mean utility values $\mathbf{C^{CF}} = [1, C_2, \ldots, C_L^{CF}]$ that solves the system of equations (29) in fact represents the unique group-level stable (and equilibrium) assignment $P^{CF^*}(g)$.

First, note that if unfilled positions are ignorable for the counterfactual assignment, then we can focus on finding a stable assignment of a restricted version of the assignment game in which only remaining K positions need to be considered. As discussed in section 3.2, Assumption 2' implicitly requires that no position that remains unfilled is ever the second-best option for any worker who takes a job in the destination period.

Furthermore, Assumption 2' imposes that each of the remaining positions will be filled in any stable matching. Recall that stability in the individual-level matching μ^{CF} requires:

$$\mu_{ik}^{CF} = 1 \text{ iff } k \in \arg\max_{k\in\tilde{\mathcal{K}}\cup 0} \pi_{ik} - q_k^{CF} \text{ and } i \in \arg\max_{i\in\tilde{\mathcal{I}}\cup 0} \pi_{ik} - r_i^{CF} \tag{31}$$

Assumption 2' allows us to replace $i \in \arg\max_{i\in\tilde{\mathcal{I}}\cup 0} \pi_{ik} - r_i^{CF}$ with $i \in \arg\max_{i\in\tilde{\mathcal{I}}} \pi_{ik} - r_i^{CF}$. In other words, we assume in advance that the individual rationality conditions that any proposed match yield a higher payoff to the position than remaining vacant, $\pi_{ik} - r_i > \pi_{0k}$ when $\mu_{ik} = 1$, are satisfied and can be ignored. Implicitly, this requires that the joint surpluses to workers and firms from matching up are sufficiently large relative to both workers' and firms' outside options.[49]

[49]This implicitly requires that the unobserved draws ϵ_{0k} for position vacancy values are taken from a bounded distri-

Imposing Assumption 2' may cause utility losses among local workers from negative local labor demand shocks to be overstated, since some workers would likely find jobs at positions that were not willing to hire at the original wage level but would enter the labor market at lower wage levels. Conversely, gains to local workers from positive shocks may be understated, since some local firms that filled positions at the original wage levels might choose to remain vacant (or move to other locations) when competition for local workers becomes more fierce.

In our applications the number of positions that will be filled is greater than the number of workers seeking positions (I). In order to be able to consistently allocate workers to match groups, even when they move to (or remain in) nonemployment, we define a "nonemployment" position type as the last position type F. Because the number of workers who end up nonemployed is assumed to be known, we allocate enough "nonemployment" positions within type F, $h^{CF}(F)$, so that the number of workers I equals the number of "positions" K, once K includes the dummy nonemployment positions. We then normalize this common number of workers and firm positions (assumed to be very large) to be 1, and reinterpret $n^{CF}(l)$ and $h^{CF}(f)$ as probability mass functions providing shares of the relevant worker and position populations rather than counts.

As discussed in section 3, Assumption 1', when combined with the stability conditions (31), implies that the probability that a given position k will be filled by a particular worker i is given by the logit form (9). When combined with Assumptions 1 and 2 (also cited by Assumption 1'), this implies that the group-level conditional choice probability $P(g|f)$ takes the form (30) for any position types f that are composed of positions k (as derived in section 3).

However, note the statement of Proposition A2 makes it clear that the form (30) also holds for the last type F, which contains the "dummy" nonemployment positions whose "choices" will be workers moving to nonemployment. The stability conditions (31) do not provide any justification for why these dummy nonemployment positions should be filled via the same logit form as the other position types that consist of actual positions at firms. Thus, the inclusion of these dummy positions, and the assumption that the probability distribution over alternative groups representing different worker and job match characteristics $(l(g), z(g))$ follows the logit form, are mere computational devices to calculate the equilibrium assignment. That this computational device in fact yields the unique stable assignment for the proposed counterfactual labor market is the primary reason Proposition A2 requires a proof.

However, the stability conditions and Assumption 1' imply that the probability that a given worker i will choose a particular position k (where $k = 0$ represents nonemployment) is also given by the logit form (Decker et al. (2013)):

$$P^{CF}(k|i) = \frac{e^{\frac{\theta_g^{CF} - q_k^{CF}}{\sigma}}}{\sum_{k' \in \mathcal{K} \cup 0} e^{\frac{\theta_{g'}^{CF} - q_{k'}^{CF}}{\sigma}}} \tag{32}$$

bution rather than the Type 1 extreme value distribution.

This can then be aggregated (using the same steps as in section A1) to provide an expression for the probability that a randomly chosen worker from a given worker type l matches with a position that yields a transition assigned to group g:

$$P^{CF}(g|l) = \frac{1}{|l|}\sum_{i\in l}\frac{(e^{\frac{\theta_g^{CF}}{\sigma}})(\sum_{k:g(i,j(i),k)=g} e^{\frac{-q_k^{CF}}{\sigma}})}{\sum_{k'\in\mathcal{K}\cup 0} e^{\frac{\theta_{g'}^{CF}-q_{k'}^{CF}}{\sigma}}} \tag{33}$$

Assumptions 3' and 4', which are analogues to Assumptions 1 and 2 in section 3, allow us to simply this expression to the following:

$$P^{CF}(g|l) = \frac{e^{\frac{\theta_g^{CF}}{\sigma}}\overline{S}^{CF}_{g|l(g),d}h^{CF}(f(g))\tilde{C}_f^{CF}}{\sum_{f'\in\mathcal{F}}\sum_{g'\in(l,f')} e^{\frac{\theta_{g'}^{CF}}{\sigma}}\overline{S}^{CF}_{g'|l(g'),d}h^{CF}(f')\tilde{C}_{f'}^{CF}}\ \forall\, l\in[1,\dots,L] \tag{34}$$

Assumption 3' states that the discounted profits of alternative positions k of the same position type f are roughly the same. This implies that the profit share that workers must provide to the position in a stable matching is approximately the same for their existing positions as for other positions in the same local area with the same industry and establishment size and establishment average pay categories, and can be summarized by a parameter C_f^{CF} that is defined at the position type level.

Taken literally (given the characteristics we use to define groups), Assumption 4' states that every worker assigned to the same worker type starts the year in firms with the same number of destination positions, which clearly does not hold. More broadly, though, Assumptions 3' and 4' allow us to replace the term $\sum_{k:g(i,k)=g} e^{\frac{-q_k^{CF}}{\sigma}}$ that depends on the individual i with an expression $P^{CF}(g|l,f(g))h^{CF}(f(g))\tilde{C}^{CF}_{f(g)}$ that depends on only group and destination-type level terms. Essentially, we are assuming that ignoring within-worker type variation in the number of positions at which they would be stayers (due to different establishment sizes of initial job matches) when aggregating is not generating significant bias in the counterfactual assignment and incidence estimates.

Under Assumptions 1' through 4', the group-level stable matching must satisfy the following market clearing conditions, which specify that supply must equal demand for each position type f:

$$\sum_{l\in\mathcal{L}} n^{CF}(l)(\sum_{g:f(g)=2} P^{CF*}(g|l,\mathbf{\tilde{C}^{CF}}) = h^{CF}(2) \tag{35}$$

$$\vdots \tag{36}$$

$$\sum_{l\in\mathcal{L}} n^{CF}(l)(\sum_{g:f(g)=F} P^{CF*}(g|l,\mathbf{\tilde{C}^{CF}}) = h^{CF}(F) \tag{37}$$

where $\mathbf{\tilde{C}^{CF}}$ represents the $F-1$ length vector = $[1,\tilde{C}_2^{CF},\dots,\tilde{C}_F^{CF}]$ and each conditional probability $P^{CF*}(g|l,\mathbf{\tilde{C}^{CF}})$ takes the form in (34).

Assumption 2' allows us to ignore the possibility that supply might exceed demand for some position types (implying some vacant positions). In this alternative position-side system of equations, the expressions for each conditional probability $P^{CF^*}(g|l)$ do in fact stem directly from the necessary stability conditions. And all of the feasibility conditions for a stable matching are incorporated into the zero-excess demand equations (since $P^{CF^*}(g|l)$ sum to 1 by construction, the assignment $P^{CF^*}(g)$ that satisfies this system necessarily sums to the worker-type PMF $n^{CF}(l)$). Thus, one can apply the proof by Decker et al. (2013) that there exists a unique group-level assignment that satisfies all of the group-level feasibility and stability conditions (and is thus consistent with a stable matching in the assignment game defined at the level of worker-position matches).

If one wished, one could directly compute the unique group-level counterfactual assignment $P^{CF^*}(g|l)$ by finding a $F-1$ length vector $\tilde{\mathbf{C}}^{\mathbf{CF}}$ that solved this system, and constructing the implied assignment by plugging this vector into the conditional probability expressions (34). However, when $F \gg L$, solving this system is considerably more computationally burdensome than solving the worker-side counterpart (29), which features $L-1$ equations. Thus, the remainder of this proof is devoted to showing that any assignment $P^{CF}(g)$ implied by a solution to (29) must equal the assignment $P^{CF^*}(g)$ implied by a solution to (37). And since we know that the latter solution represents the unique group-level matching consistent with stability in the assignment game, the former solution must also be unique, and must also represent the group-level matching consistent with stability in the assignment game. Essentially, this amounts to showing that the device of adding "dummy" nonemployment positions present in (29) appropriately incorporates the surpluses π_{i0} that workers obtain from staying single.

Consider an L length vector $\mathbf{C}^{\mathbf{CF}} = [1, C_2^{CF}, \ldots, C_L^{CF}]$ that solves (29) and yields assignment $P^{CF}(g)$. We will show that one can use $\mathbf{C}^{\mathbf{CF}}$ to construct an alternative F length vector $\tilde{\mathbf{C}}^{\mathbf{CF}} = [1, \tilde{C}_2^{CF}, \ldots, \tilde{C}_F^{CF}]$ that solves (37), and that the assignment it generates, $P^{CF^*}(g)$, equals $P^{CF}(g)$.

We propose the following vector $\tilde{\mathbf{C}}^{\mathbf{CF}}$:

$$\tilde{C}_f^{CF} = \frac{\sum_{l=1}^{L} \sum_{g':(l(g'),f(g'))=(l,F)} e^{\frac{\theta_{g'}}{\sigma}} n^{CF}(l) \overline{S}_{g'|l,F} C_l^{CF}}{\sum_{l=1}^{L} \sum_{g':(l(g'),f(g'))=(l,f)} e^{\frac{\theta_{g'}}{\sigma}} n^{CF}(l) \overline{S}_{g'|l,f} C_l^{CF}} \; \forall \; f \in [1, \ldots, F] \tag{38}$$

Here, the numerator captures the inclusive value (as defined by Menzel (2015)) associated with the nonemployment position type F, while the denominator captures the inclusive value for the chosen position type f. This implies that $\tilde{C}_F^{CF} = 1$. While any position type could be chosen as the one whose mean exponentiated profit value is normalized, normalizing the nonemployment type is particularly appealing, since it implies "profit" values of 0 for the dummy nonemployment position type F ($\tilde{C}_F^{CF} = e^{\overline{q}_F} = e^0 = 1$).

To conserve notation, let λ represent the inclusive value of the nonemployment position type F, the numerator in (38):

$$\lambda = \sum_{l=1}^{L} \sum_{g':(l(g'),f(g'))=(l,F)} e^{\frac{\theta_{g'}}{\sigma}} n^{CF}(l) \overline{S}_{g'|l,F}^{CF} C_l^{CF} \tag{39}$$

Note that λ is independent of position type. We begin by showing that the assignments implied by the vectors $[C_1^{CF}, \dots, C_L^{CF}]$ and $[C_1^{CF}, \dots, \tilde{C}_F^{CF}]$ are identical: $P^{CF}(g) = P^{CF^*}(g)$.

Since $\mathbf{C}^{CF}$ solves the worker-side system of excess demand equations (29), we know that

$$\sum_{f' \in \mathcal{F}} h^{CF}(f') \sum_{g' \in (l,f')} \frac{e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}_{g'|l,f'}^{CF} n^{CF}(l) C_l^{CF}}{\sum_{l'=1}^{L} \sum_{g':(l(g'),f(g'))=(l',f)} e^{\frac{\theta_{g'}}{\sigma}} n^{CF}(l') \overline{S}_{g'|l',f} C_{l'}^{CF}} = n^{CF}(l) \; \forall \, l \in [1, L]$$

$$\Rightarrow \sum_{f' \in \mathcal{F}} \sum_{g' \in (l,f')} \frac{e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}_{g'|l,f'}^{CF} h^{CF}(f')}{\sum_{l=1}^{L} \sum_{g':(l(g'),f(g'))=(l,f)} e^{\frac{\theta_{g'}}{\sigma}} n^{CF}(l) \overline{S}_{g'|l,f} C_l^{CF}} = \frac{1}{C_l^{CF}} \; \forall \, l \in [1, L]$$

$$\Rightarrow \sum_{f' \in \mathcal{F}} \sum_{g' \in (l,f')} \frac{e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}_{g'|l,f'}^{CF} h^{CF}(f')}{\frac{\lambda}{\tilde{C}_{f'}^{CF}}} = \frac{1}{C_l^{CF}} \; \forall \, l \in [1, L]$$

$$\Rightarrow \sum_{f' \in \mathcal{F}} \sum_{g' \in (l,f')} e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}_{g'|l,f'}^{CF} h^{CF}(f') \tilde{C}_{f'}^{CF} = \frac{\lambda}{C_l^{CF}} \; \forall \, l \in [1, L] \tag{40}$$

We can now proceed:

$$\begin{aligned}
P^{CF^*}(g) &= n^{CF}(l) P^{CF^*}(g|l) = n^{CF}(l) \frac{e^{\frac{\theta_g^{CF}}{\sigma}} \overline{S}_{g|l,f}^{CF} h^{CF}(f) \tilde{C}_f^{CF}}{\sum_{f' \in \mathcal{F}} \sum_{g' \in (l,f')} e^{\frac{\theta_{g'}^{CF}}{\sigma}} \overline{S}_{g'|l,f'}^{CF} h^{CF}(f') \tilde{C}_{f'}^{CF}} \\
&= \frac{n^{CF}(l) e^{\frac{\theta_g^{CF}}{\sigma}} \overline{S}_{g|l,f}^{CF} h^{CF}(f) \tilde{C}_f^{CF} C_l^{CF}}{\lambda} \\
&= h^{CF}(f) \frac{e^{\frac{\theta_g^{CF}}{\sigma}} n^{CF}(l) \overline{S}_{g|l,f}^{CF} \lambda C_l^{CF}}{\lambda \sum_{l'=1}^{L} \sum_{g':(l(g'),f(g'))=(l',f)} e^{\frac{\theta_{g'}}{\sigma}} f^{CF}(l') \overline{S}_{g'|l',f} C_{l'}^{CF}} \\
&= h^{CF}(f) P^{CF}(g|f) = P^{CF}(g)
\end{aligned} \tag{41}$$

It remains to show that the chosen $\tilde{\mathbf{C}}^{\mathbf{CF}}$ vector (38) solves (37). Consider the left-hand side of the excess demand equation for an arbitrary position type f in the system (37). One can write:

$$\begin{aligned}
&\sum_{l=1}^{L} \sum_{g:(l(g),f(g))=(l,f)} n^{CF}(l) P^{CF^*}(g|l, \Theta^{CF}, \tilde{\mathbf{C}}^{CF}) \\
&= \sum_{l=1}^{L} \sum_{g:(l(g),f(g))=(l,f)} h^{CF}(f) P^{CF}(g|f, \Theta^{CF}, \mathbf{C}^{CF}) \\
&= h^{CF}(f) \sum_{l=1}^{L} \sum_{g:(l(g),f(g))=(l,f)} P^{CF}(g|f, \Theta^{CF}, \mathbf{C}^{CF})
\end{aligned}$$

$$= h^{CF}(f) \sum_{g:f(g)=f} P^{CF}(g|f, \Theta^{CF}, \mathbf{C}^{CF})$$
$$= h^{CF}(f) \tag{42}$$

where the last line imposes that $P^{CF}(g|f)$ is a (conditional) probability distribution and thus sums to one. Since we have proved that the implied "demand" by workers for positions of an arbitrary position type equals the "supply" $h^{CF}(f)$, we have proved that $\tilde{C}^{CF}$ solves the system (37).

Notice that the expression for the proposed equilibrium mean ex post profit vector (38) has value beyond its use in proving proposition A1. Once the L-vector of mean ex post utilities $\{C_l^{CF}\}$ for each worker type have been computed, one can use (38) to directly calculate the mean ex post profit vector for each position type f without having to solve a system of $F-1$ equations. This is quite valuable when $F \gg L$, as it is in our application. Of course, the equivalent mapping can be inferred by symmetry for the opposite case where $L \gg F$:

$$C_l^{CF} = \frac{\sum_{f=1}^{F} \sum_{g':(l(g'),f(g'))=(L,f)} e^{\frac{\theta_{g'}}{\sigma}} h^{CF}(f) \overline{S}_{g'|L,f} \tilde{C}_f^{CF}}{\sum_{f=1}^{F} \sum_{g':(l(g'),f(g'))=(l,f)} e^{\frac{\theta_{g'}}{\sigma}} h^{CF}(f) \overline{S}_{g'|l,f} \tilde{C}_f^{CF}} \; \forall \, l \in [1, \ldots, L] \tag{43}$$

In section 3.2 we showed that these vectors are sufficient to determine both the worker and position type-level incidence of any counterfactual shocks to the composition or spatial distribution of labor supply and/or labor demand. Thus, at least in cases where the proposed model is a reasonable approximation of the functioning of the labor market (and housing supply is sufficiently elastic and agglomeration effects and other product market spillovers are second order), a proper welfare analysis of such shocks only requires solving at most $\min\{L, F\}$ non-linear excess demand equations. Since an analytical Jacobian can be derived and fed as an input to non-linear equations solvers, relatively large scale assignment problems featuring thousands of types on one side of the market (and perhaps more on the opposite side) can be solved within a matter of minutes.

A4 Estimating the Value of σ

We attempt to estimate σ, the standard deviation of the unobserved match-level component ϵ_{ik}, by exploiting the evolution in the composition of U.S. worker and position types $n^y(l)$ and $h^y(f)$ across years y. Specifically, we estimate the set of group-level surpluses $\{\theta_g^{2003}\}$ from the observed 2003-2004 matching. Then, holding these surplus values fixed, we combine $\{\theta_g^{2003}\}$ with $n^y(l)$ and $h^y(f)$ from each other year $y \in [2004, 2012]$ to generate counterfactual assignments and changes in scaled mean (exponentiated) utility values $\{C_l^{CF}\}$ for each worker type. These counterfactuals predict how mean worker utilities by skill/location combination would have evolved given the observed compositional changes in labor supply and demand had the underlying surplus values $\{\theta_g\}$ been constant and equal to $\{\theta_g^{2003}\}$ throughout the period.

To the extent that most of evolution in the utility premia enjoyed by workers in particular locations and skill categories was due primarily to changes in supply and demand composition rather than changes in the moving costs, recruiting costs, tastes, and relative productivities that compose the joint surplus values $\{\theta_g\}$, these counterfactual predictions will be reasonable approximations of the realized evolution of ex post utility over time by worker type. Recall that $C_l^{CF} \approx \frac{1}{|l|}\sum_{i:l(i)=l} e^{\frac{-r_i^{CF}}{\sigma}}$. Thus, if ex post utility r_i^{CF} does not vary too much across individuals within a worker type, so that Jensen's inequality is near equality and $\frac{1}{|l^y|}\sum_{i:l(i)=l} e^{\frac{-r_i^{CF,y}}{\sigma^y}} \approx e^{\frac{\overline{r}_l^{CF,y}}{\sigma^y}}$, then taking logs yields $ln(C_l^{CF,y}) \approx \frac{\overline{r}_l^{CF,y}}{\sigma^y}$.

Next, we form the corresponding changes in observed annual earnings for each worker type in each year, $\overline{Earn}_l^{y+1} - \overline{Earn}_l^{y}$.[50] We then run the following regression at the l-type level for each year $y \in [2004 - 2012]$:

$$\overline{Earn}_l^{y+1} - \overline{Earn}_l^{y} = \beta_0^y + \beta_1^y(ln(C_l^{CF,y+1}) - ln(C_l^{CF,y})) + \nu_l^y \tag{44}$$

Recall that the $\overline{r}_l^{CF,y}$ values represent predicted money metric utility gains, and are thus denominated in dollars. However, even if the surplus values $\{\theta_g\}$ are time invariant over the chosen period (and the other assumptions of the assignment model specified above all hold, including the approximations just described), dollar-valued mean utility gains would not equal mean annual earnings gains for a given worker type if its workers systematically moved to jobs featuring better or worse amenities, avoided more moving/recruiting training costs, or moved to jobs featuring better or worse continuation values. However, if such changes in other sources of utility nearly cancel out among workers assigned to the same worker type (for all worker types), then $\overline{r}_l^{CF,y+1} - \overline{r}_l^{CF,y}$ should approximately equal $\overline{Earn}_l^{y+1} - \overline{Earn}_l^{y}$. This implies that $\beta_1^y \approx \sigma^y$.

As noted in Section 5.1, the worker type space depends on which location is considered the target location for the shock, with the geographic units that partially define worker types becoming more aggregated farther from the shock. To address this issue, in practice we constructed separate true and counterfactual earnings changes and estimated equation (44) for the collapsed worker type spaces associated with each possible target PUMA among the sample states, and averaged the estimates of β_1 across all regressions satisfying a minimum R^2 threshold of .1 to obtain $\hat{\beta}_1^y$.[51] The estimates of $\hat{\beta}_1^y$ are fairly consistent across years, so we use the mean estimate across all years, $\overline{\sigma} = 18,400$, to produce dollar values for all the results relating to utility gains presented in the paper.

Clearly, given the additional strong assumptions required, this approach represents a relatively crude attempt to calibrate σ. Indeed, further efforts could conceivably be taken to exclude worker types l' whose surplus values $\{\theta_g : l(g) = l'\}$ were known to be changing over the chosen time pe-

[50]Note that while worker earnings in initial job matches were used to assign workers to skill categories, to this point we have not used observed worker earnings in destination positions to identify any other parameters.

[51]A few PUMAs and states experienced relatively little year-to-year change in the distribution of employment across position types, so that the counterfactual earnings forecasts predicted true earnings changes poorly. In this case, the R^2 from the regression was very low, and β_1^y was badly identified. The results become far more stable across the remaining alternative type spaces when a minimum R^2 was imposed to eliminate the few badly identified estimates, which tended to produce outliers.

riod, or to allow θ_g to evolve in a particular parametric fashion. In fact, GS discuss how a vector of σ values associated with different types or combinations of types based on observed characteristics might potentially be jointly estimated with other model parameters (thereby allowing heteroskedasticity across types in the idiosyncratic match component). Since the focus in this paper is primarily on examining relative incidence across different worker types from shocks featuring different changes in labor demand composition, we opted for the simpler, more transparent approach.

A5 Using Transfers to Decompose the Joint Surpluses $\{\theta_g\}$

This appendix examines whether observing equilibrium transfers, denoted w_{ik}, allows the identification of additional parameters of interest. In CS's assignment model, the unobserved match-level heterogeneity is assumed to take the form $\epsilon_{ik} = \epsilon^1_{l(i)k} + \epsilon^2_{if(k)}$, so that aggregate surplus is left unchanged when two pairs of job matches (i, k) and (i', k') belonging to the same group g swap partners. The elimination of any true (i, k) match-level surplus component implies that equilibrium transfers cannot vary among job matches belong to the same group g, so that $w_{ik} = w_{g(i,k)} \; \forall \, (i, k)$.[52] GS show that under this assumption, observing the (common) group-level transfers w_g would be sufficient to decompose the group-level mean joint surplus θ_g into the worker and position's respective pre-transfer payoffs, which we denote θ^l_g and θ^f_g, respectively.

Because the model proposed in section 2.2 does not impose the additive separability assumption $\epsilon_{ik} = \epsilon^1_{l(i)k} + \epsilon^2_{if(k)}$, equilibrium transfers will in general vary among (i, k) pairs within the same group g. Given the substantial earnings variance within observed groups g regardless of the worker, position, and job match characteristics used to define g, the CS restriction on the nature of unobserved match-level heterogeneity would be strongly rejected in the labor market context.

However, one can still consider whether the observed transfers $\{w_{ik}\}$ identify additional objects. From section 2.1, equilibrium transfers are related to equilibrium worker and position payoffs via:

$$w_{ik} = \pi^f_{ik} - q_k \tag{45}$$

$$w_{ik} = r_i - \pi^l_{ik} \tag{46}$$

Next, recall from equation (11) that under Assumptions 1 and 2 in Proposition A1 the log odds that a randomly chosen position from arbitrary position type f will choose a worker whose hire would be assigned to group g_1 relative to g_2 are given by:

$$\ln(\frac{P(g_1|d)}{P(g_2|f)}) = \ln(P(g_1|f)) - \ln(P(g_2|f)) = \frac{\theta_{g_1}}{\sigma} + \ln(\overline{S}_{g_1|l(g_1),f}) + \ln(n(l(g_1))) + \ln(C_{l(g_1)}) - \frac{\theta_{g_2}}{\sigma} - \ln \overline{S}_{g_2|l(g_2),f}) - \ln(n(l(g_2))) - \ln(C_{l(g_2)}) \tag{47}$$

[52] If $w_{ik} > w_{i'k'}$ for any two matched pairs (i, k) and $(i'k')$ such that $g(i, k) = g(i', k')$, then (i', k) would form a blocking pair by proposing a surplus split between them featuring a transfer between w_{ik} and $w_{i'k'}$, thus undermining the stability of the proposed matching.

Since $\ln(\overline{S}_{g_1|l(g_1),f})$, $\ln(\overline{S}_{g_2|l(g_2),f})$, $\ln(n(l(g_1)))$, and $\ln(n(l(g_2)))$ are all observed (or, if a large sample is taken, extremely precisely estimated), one can form adjusted log odds:

$$\ln(\frac{\hat{P}_{g_1|f}/(\overline{S}_{g_1|l(g_1),f})n(l(g_1)))}{\hat{P}_{g_2|f}/(\overline{S}_{g_2|l(g_2),f}n(l(g_2)))}) = (\frac{\theta_{g_1}-\theta_{g_2}}{\sigma}) + (\ln(C_{l(g_1)}) - \ln(C_{l(g_2)})) \tag{48}$$

Under Assumption 1, C_l is the mean of exponentiated (and rescaled) equilibrium utility payoffs owed to workers $i : l(i) = l$:

$$C_l = \frac{1}{|l|}\sum_{i:l(i)=l(g)} e^{-\frac{r_i}{\sigma}} \approx \sum_{\frac{1}{g_k}} \sum_{i:g(i,k)=g} e^{-\frac{r_i}{\sigma}} \;\forall\, k \tag{49}$$

Plugging (46) into (49) and then (49) into (48) yields:

$$\begin{aligned}
&\ln(\frac{\hat{P}_{g_1|f}/(\overline{S}_{g_1|l(g_1),f}n(l(g_1)))}{\hat{P}_{g_2|f}/(\overline{S}_{g_2|l(g_2),f}n(l(g_2)))}) \\
&= (\frac{\theta_{g_1}-\theta_{g_2}}{\sigma}) + (\ln(\frac{1}{|l|}\sum_{i:l(i,j(i))=l(g_1)} e^{-\frac{w_{ik}+\pi^l_{ik}}{\sigma}}) - \ln(\frac{1}{|l|}\sum_{i:l(i,j(i))=l(g_2)} e^{-\frac{w_{ik}+\pi^l_{ik}}{\sigma}}))
\end{aligned} \tag{50}$$

It is not immediately obvious how to use equation (50) to recover parameters of interest. Only when one adds further assumptions that are at odds with the structure of the model can one recover an expression that mirrors the one in CS. Specifically, suppose the following assumptions hold:

$$\begin{aligned}
&r_i \approx r_{l(i)} \;\forall\, i : l(i) = l \;\forall l \in \mathcal{L} \\
&\pi^l_{ik} = \pi^l_{g(i,k)} \equiv \theta^l_g \;\forall\, (i,k) : g(i,k) = g \;\forall\, g \in \mathcal{G} \\
&w_{ik} = w_{g(i,k)} \forall\, (i,k) : g(i,k) = g \;\forall\, g \in \mathcal{G}
\end{aligned} \tag{51}$$

These assumptions are extremely unlikely to hold in any stable matching if there is meaningful variance in ϵ_{ik} among the (i,k) pairs within the same group g. Nonetheless, they yield:

$$\begin{aligned}
&\ln(\frac{\hat{P}_{g_1|f}/(\overline{S}_{g_1|l(g_1),f}n(l(g_1)))}{\hat{P}_{g_2|f}/(\overline{S}_{g_2|l(g_2),f}n(l(g_2)))}) = (\frac{\theta_{g_1}-\theta_{g_2}}{\sigma}) + (\ln(e^{-r_{l(g_1)}}) - \ln(e^{-r_{l(g_2)}})) \\
&= (\frac{\theta_{g_1}-\theta_{g_2}}{\sigma}) + \frac{-r_{l(g_1)} + r_{l(g_2)}}{\sigma} = (\frac{\theta_{g_1}-\theta_{g_2}}{\sigma}) + (\frac{-(w_{g_1}+\theta^l_{g_1}) + (w_{g_2}+\theta^l_{g_2})}{\sigma}) \\
&= \frac{\theta^f_{g_1} - \theta^f_{g_2} + (w_{g_2} - w_{g_1})}{\sigma}
\end{aligned} \tag{52}$$

Given an estimate of σ based on multiple markets (as described in Appendix A4) and data on mean annual earnings for each match group $g \in \mathcal{G}$, one could identify the difference in the position component of the joint surplus for arbitrary groups g_1 and g_2. This provides information about the relative profit contributions of different types of workers for each type of position before such workers salaries are considered. Note that one could still not separate the training cost, recruiting cost, current revenue contribution, and continuation value components of θ_g^f without additional data.

A similar progression using adjusted log odds based on the worker side conditional probabilities $P(g_1|l_1)$ and $P(g_2|l_1)$ would yield an estimate of the corresponding difference in the worker components of the joint surplus $\theta_{g_1}^l - \theta_{g_2}^l$ for any two groups featuring the same worker type. Since one such group could represent nonemployment, this approach would provide estimates of the desirability of working at various types of firms in various locations for zero pay relative to nonemployment. These values identify the reservation salary necessary to convince each worker type to take (or continue) a position of each type. Again, one could not disentangle the moving cost, search cost, non-wage amenity value, and continuation value components of the surplus without further data.

Because 1) we deem the assumptions (51) to be antithetical to the spirit of the model and at odds with the data, and 2) other than estimating σ, the use of transfers is not necessary to fulfill the primary aim of the paper, evaluating the utility and profit incidence across worker and position types of alternative local labor demand shocks, we do not make further use of the observed annual earnings distributions in the destination period.

A6 Imputing Missing Transitions Involving Unemployment and Missing Match Group Characteristics

Recall that nonemployed workers are only included in the sample in a given year if they are observed resuming work in a future year. This requirement is imposed so as to better distinguish unemployed workers from those exiting the labor force, but it creates the likely possibility of undercounting employment-to-unemployment (E-to-U) and unemployment-to-unemployment (U-to-U) transitions toward the end of the sample, when high shares of unemployment spells are right-censored due to data availability.[53] In addition, the inability to observe the characteristics of those working in states that did not approve the use of their LEHD data creates a further need for imputation for employment-to-employment (E-to-E) and unemployment-to-employment (U-to-E) transitions originating in out-of-sample states. This appendix describes how data from the harmonized American Community Survey (hereafter ACS) series created by IPUMS along with official unemployment statistics from the Bureau of Labor Statistics (hereafter BLS) were used to address these problems.

A6.1 E-to-U and U-to-U Transitions

Note first that a match count must be generated for each match group g classified as an E-to-U transition, which consists of a combination of origin location and earnings quartile (since there is a single unemployment position type). Because the 1% ACS sample is too small to generate accurate E-to-U counts at even the PUMA level, we construct population-weighted E-to-U ACS counts by initial state and earnings group, and re-scale each count so that the aggregate stock of E-

[53] Since nearly all states enter the sample well before the years used for this analysis, the analogous risk of undercounting unemployment-to-employment transitions is negligible.

to-U transitions matches the count of workers unemployed less than 52 weeks from the BLS for the chosen year.[54] For in-sample initial states, we impute a match group g for each implied individual from the rescaled ACS counts by combining the ACS characteristics with a draw from the observed conditional empirical distribution of origin tracts given origin state among E-to-U transitions in the LEHD. E-to-U counts of transitions from the ACS that originate out of sample are aggregated across states, leaving four groups corresponding to the four initial earnings quartiles.

For U-to-U transitions, we begin with age-specific counts of long-term unemployment (> 52 weeks) from the BLS, and distribute them across origin and destination states according to the joint distribution of state pairs among U-to-U counts in the ACS. We then impute an origin tract for each U-to-U transition that originates from an in-sample state by drawing from the conditional empirical distribution of origin tracts given origin state and age group among the combined pool of E-to-U and E-to-E LEHD transitions that end in the observed state, so as to ensure appropriately full support among origin tracts.

A6.2 E-to-E and U-to-E Transitions

Note that full match group counts are observed for all E-to-E transitions among in-sample states. Since we aggregate out-of-sample destination positions to a single type, in-sample to out-of-sample E-to-E match counts are also fully observed (by combining the absence of observed earnings with the provided indicator for non-zero earnings somewhere in the U.S.). E-to-E match counts among out-of-sample states require an initial earnings quartile to be assigned. We draw this using the distribution of initial earnings quartiles among LEHD in-sample observations. E-to-E match counts from out-of-sample to in-sample states are completed similarly, except that the distribution of earnings quartiles also conditions on destination state, industry, firm size, and firm average pay as well as on being a state switcher.

U-to-E transitions in a fashion analogous to that of E-to-E transitions, except that an initial location must be imputed as well. If the worker has worked in-sample previously, we use the most recently observed employer tract as the worker's initial location. For those without previously observed employers (mostly young new entrants to the labor market), we use the same method for drawing origin tracts that was detailed for U-to-U transitions in the previous subsection.

A7 Smoothing Procedure

In this appendix we describe how we smooth the empirical distribution of job matches across groups, $\hat{P}(g)$, prior to estimation in order to generate accurate estimates of the set of identified joint surplus difference-in-differences Θ^{D-in-D}. We smooth for two reasons. First, such smoothing serves as a "noise infusion" technique that removes the risk that individual or establishment identities could be

[54] Because we use a 50% random sample of LEHD transitions, we multiply estimated E-to-U counts (and U-to-U counts) by .5.

revealed by any estimates presented in the paper, as required of all research results generated from confidential microdata in Federal Statistical Research Data Centers (FSRDCs). Second, smoothing is necessary because there are sufficiently few observations per match group such that many match groups are rarely (or never) observed in a given matching despite substantial underlying matching surpluses simply due to sampling error. Essentially, $\hat{P}(g)$ is only a consistent estimator of $P(g)$ as the number of observed job matches per group I/G approaches infinity.

We overcome this sampling error problem by assuming that the underlying frequency $P(g)$ with which a job match belongs to a particular match group is a smooth function of the observed characteristics that define group g (following Hotz and Miller (1993) and Arcidiacono and Miller (2011)). This permits the use of a kernel density estimator that computes a weighted average of the empirical probabilities $\hat{P}(g')$ of "nearby" groups g' that feature "similar" vectors of characteristics to generate a well-behaved approximation of $P(g)$ from the noisy empirical distribution $\hat{P}(g)$.

Such smoothing introduces two additional challenges. First, excessive smoothing across other match groups erodes the signal contained in the data about the degree of heterogeneity in the relative surplus from job matches featuring different combinations of worker characteristics, establishment characteristics, and origin and destination locations. Since highlighting the role of such heterogeneity in forecasting the incidence of labor market shocks is a primary goal of the paper, decisions about the appropriate bandwidth must be made with considerable thought. The second, related challenge consists of identifying which of the worker and position characteristics that defines other groups makes them "similar", in the sense that the surplus $\{\theta_{g'}\}$ is likely to closely approximate the surplus θ_g whose estimate we wish to make more precise.

Recall that each group $g \equiv g(l, f, z)$ is a combination of 1) the origin establishment location (which we denote $loc(l)$) and workers' initial earnings quartile (or unemployment status) at the origin establishment (denoted $earn(l)$); 2) the destination establishment's location ($loc(f)$), establishment size category ($f\_size(f)$), establishment average earnings category ($f\_earn(f)$), and industry supersector ($ind(f)$); and 3) the indicator $z(i, k)$ for whether establishment $j(i)$ and establishment k are the same, so that worker i is a job stayer rather than a mover (denoted $stayer(g)$).

Given the goal of accurately characterizing incidence at a very high spatial resolution, we wish to preserve as accurately as possible any signal in the data about the structure of spatial ties between nearby local areas. Thus, wherever possible the kernel estimator should place non-zero weight only on alternative groups g' that share the same origin and destination locations ($loc(l(g)) = loc(l(g'))$ and $loc(f(g)) = loc(f(g'))$). Similarly, we posit that an establishment's combination of size, average worker pay, and industry is likely to be more important than its location in determining the skill category of worker that generates the most surplus. To develop a smoothing approach that embodies these principles, we exploit the fact that $P(g)$ can be decomposed via:

$$\begin{aligned} P(g) &= P(g|f(g))h(f(g)) = P([l(g), f(g), z(g)]|f)h(f(g)) \\ &= P([loc(l(g)), earn(l(g)), stayer(g)]|f)h(f(g)) \end{aligned}$$

$$
\begin{aligned}
&= P(loc(l(g))|earn(l(g)), stayer(g), f)P([earn(l(g)), stayer(g)]|f)h(f(g)) \\
&= 1(stayer(g)=1)P(loc(l(g))|earn(l(g)), 1(stayer(g)=1), d)P([earn(l(g)), 1(stayer(g)=1)]|f)h(f(g)) \\
&+ 1(stayer(g)=0)P(loc(l(g))|earn(l(g)), 1(stayer(g)=0), d)P([earn(l(g)), 1(stayer(g)=0)]|f)h(f(g)) \\
&= 1(stayer(g)=1)1(loc(l(g))=loc(f(g)))P([earn(l(g)), 1(stayer(g)=1)]|f)h(f(g)) \\
&+ 1(stayer(g)=0)P(loc(l(g))|earn(l(g), 1(stayer(g)=0), d)P([earn(l(g)), 1(stayer(g)=0)]|f)h(f(g))
\end{aligned}
\tag{53}
$$

where the first two lines use the law of total probability and the set of characteristics that define $l(g)$ and $z(g)$, the third line uses the fact that $z(g) \equiv stayer(g)$ only takes on two values (0 for job movers and 1 for stayers), and the last line uses the fact that $P(loc(l(g))|earn(l(g)), 1(stayer(g) = 1), f) = 1(loc(l(g)) = loc(f(g)))$, since a potential stayer associated with a particular position type must have already been working at the same location in the origin period (since we treat establishments that switch locations as different establishments for computational reasons). We use separate kernel density estimator procedures to estimate each of $P(loc(l(g))|earn(l(g)), 1(stayer(g) = 0), f(g))$, $P(earn(l(g)), 1(stayer(g) = 0)|f(g))$, and $P(earn(l(g)), 1(stayer(g) = 1)|f(g))$.

Consider first the estimation of $P(loc(l(g))|earn(l(g), 1(stayer(g) = 0), f(g))$, the conditional probability that a particular new hire would be originally located at location $loc(l)$, given the hired worker's initial earnings category and the destination position's type f. Let $K^{dist}(g, g')$ denote the metric capturing how similar an alternative group g' is to g for the purpose of estimating the propensity for establishments of type f to hire workers from a particular location (conditional on skill level). As discussed above, wherever possible we only assign non-infinite distance $K^{dist}(g, g') < \infty$ (corresponding to non-zero weight) to empirical conditional probabilities $P(loc(l(g'))|earn(l(g')), 1(stayer(g') = 0), f(g'))$ of alternative groups g' that feature both the same origin location $loc(l(g')) = loc(l(g))$ *and* destination location $loc(f(g')) = loc(f(g))$.[55]

$K^{dist}(g, g')$ assigns the smallest distance to alternative groups g' that also feature the same position type ($f(g') = f(g)$), so that g and g' only differ in the skill category of hired workers. The closer $earn(l(g'))$ is to $earn(l(g'))$, the smaller is the assigned distance $K^{dist}(g, g')$, but the profile flattens so that all groups g' that differ from g' only due to $earn(l(g'))$ contribute to the weighted average. $K^{dist}(g, g')$ assigns larger (but still noninfinite) distance to groups g' featuring position types that also differ on establishment size, avg. pay, or industry dimensions. The more different the establishment composition of the group, the smaller is its weight, with the profile again flattening so that all groups g' featuring the same origin and destination locations receive non-zero weight. Thus, groups with less similar worker and establishment characteristics receive non-negligible weight only when there are too few observations from groups featuring more similar worker and establishment characteristics to form reliable estimates. The weight assigned to a particular alternative group g' also depends on the number of observed new hires made by $f(g')$ at a particular skill level $earn(l(g'))$, denoted $N^{dist}(g')$ below, since this determines the signal strength of the empirical

[55]There are a very small number of worker and position types that are never observed in any job match. By necessity, we put positive weight on groups featuring nearby origin or destination locations in such cases.

CCP $P(loc(l(g')|earn(l(g')), 1(stayer(g') = 0), f(g'))$. Thus, we have:

$$P(loc(l(g))|earn(l(g)), 1(stayer(g) = 0), f(g)) \approx$$
$$\sum_{g'}\left(\frac{\phi(K^{dist}(g', g)N^{dist}(g'))}{\sum_{g''}\phi(K^{dist}(g'', g)N^{dist}(g''))}\hat{P}(loc(l(g'))|earn(l(g')), 1(stayer(g') = 0), f(g')\right) \quad (54)$$

where $\phi(*)$ is the normal density function (used as the kernel density), and $\frac{\phi(K^{dist}(g',g)N^{dist}(g'))}{\sum_{g''}\phi(K^{dist}(g'',g)N^{dist}(g''))}$ represents the weight given to a particular nearby match group g'.[56]

Next, consider the estimation of $P(earn(l(g)), 1(stayer(g) = 1)|f)$ and $P(earn(l(g)), 1(stayer(g) = 0)|f)$, the conditional probabilities that either a job stayer or mover originally paid at a particular earnings quartile (or possibly unemployed for movers) will be hired to fill a position of position type f. Let $K^{earn/move}(g, g')$ and $K^{earn/stay}(g, g')$ represent the metrics capturing how similar alternative groups g' are to g for the purpose of estimating the propensity for firms of type f to hire (or retain) workers at particular skill levels.

$K^{earn/move}(g, g')$ and $K^{earn/stay}(g, g')$ each assign infinite distance (i.e. zero weight) to groups g' featuring different combos of establishment size, average pay, and industry than the target group g. $K^{earn/move}(g, g')$ ($K^{earn/stay}(g, g')$) assigns small distances to the conditional probabilities for groups g' representing hiring new (retaining) workers from the same initial earnings (or nonemployment) category $earn(l(g)) = earn(l(g'))$ among firms from the same position type $f(g) = f(g')$ but who are hiring nearby workers. The distance metric increases in the tract pathlength between $loc(l(g'))$ and $loc(l(g'))$, but flattens beyond a threshold distance, so that groups featuring all origin locations (but shared values of other characteristics) contribute to the estimate.

Larger (but finite) distance values for $K^{earn/move}(g, g')$ and $K^{earn/stay}(g, g')$) are assigned to conditional probabilities from groups g' that feature different (but nearby) destination locations (so $f(g) \neq f(g')$ but the same combination of establishment size and average earnings quartiles and industry supersector. Again, the distance metric increases in the pathlength between $loc(f(g))$ and $loc(f(g'))$, but eventually flattens at a large but non-infinite value. As before, the weight given to a group g' also depends on the precision of its corresponding number of total hires made by firms of the position type $f(g')$, which is proportional to $h(f(g'))$.

Again, the motivation here is that targeted skill level and the retention/new hire decision (conditional on the utility bids required by workers in different locations) is likely to be driven more by an establishment's production process (proxied by size, mean pay, and industry) than by its location. Since there still may be spatially correlated unobserved heterogeneity in production processes conditional on the other establishment observables, we place greater weight on the skill/retention decisions of proximate firms. More distant firms receive non-negligible weight only when too few local observations exist to form reliable estimates. The estimators for $P(earn(l(g)), 1(stayer(g) =$

[56] A standard deviation of 10 was used as the bandwidth choice for both this and the kernel densities presented below. The results were insensitive to moderate changes in bandwidth choice, though choosing a very small bandwidth resulted in very volatile simulation estimates across target tracts, highlighting the need for smoothing.

$1)|f)$ and $P(earn(l(g)), 1(stayer(g) = 0)|f)$ can be expressed via:

$$P(earn(l(g)), 1(stayer(g) = 0)|f(g)) \approx$$
$$\sum_{g'} \left(\frac{\phi(K^{earn/move}(g', g)h(f(g')))}{\sum_{g''} \phi(K^{earn/move}(g'', g)h(f(g'')))}\hat{P}(earn(l(g')), 1(stayer(g') = 0)|f(g')) \right. \tag{55}$$
$$P(earn(l(g)), 1(stayer(g) = 1)|f(g)) \approx$$
$$\sum_{g'} \left(\frac{\phi(K^{earn/stay}(g', g)h(f(g')))}{\sum_{g''} \phi(K^{earn/stay}(g'', g)h(f(g'')))}\hat{P}(earn(l(g')), 1(stayer(g') = 1)|f(g')) \right. \tag{56}$$

Bringing the pieces together, this customized smoothing procedure has a number of desirable properties. First, by requiring the same origin and destination locations as a necessary condition for non-zero weight when estimating the propensity for particular position types to hire workers from each location, one can generate considerable precision in estimated CCPs without imposing assumptions about the spatial links between locations. Second, at the same time, one can still use information contained in the hiring and retention choices of more distant establishments to learn about the propensity for establishments of different sizes, pay levels, and industries to retain and hire workers at different skill levels and from unemployment. Third, the procedure places non-trivial weight on match groups featuring less similar worker and establishment characteristics only when there are too few observed hires/retentions made by establishments associated with groups featuring very similar characteristics to yield reliable estimates. Fourth, overall the estimated probabilities $P(g|f)$ place weight on many groups, so that no element of the resulting smoothed distribution contains identifying worker or establishment information, eliminating disclosure risk.

A8 Model Validation

The simulations consider relatively large, locally focused labor demand shocks, but the estimated surplus parameters $\hat{\Theta}^{D-in-D}$ that underlie them are identified from millions of quotidian job transitions driven by small firm expansions/contractions and worker retirements and preference or skill changes over the life cycle that generate considerable offsetting churn in the U.S. labor market. Thus, one might reasonably wonder whether parameters governing ordinary worker flows are capable of capturing the response to sizable, locally focused positive or negative shocks. To address this concern, in this section we describe and present results from a model validation exercise in which surplus parameters estimated on pre-shock ordinary worker flows were used to forecast the reallocation of workers after actual local economic shocks observed in the LEHD sample.

Specifically, 180 shocks to employment in a census tract were identified in the LEHD sample that satisfied the following criteria: 1) the shock occurred in a sample state during the years 2003 - 2012; 2) at least 100 more or 100 fewer positions (and at most 3000) were filled in the chosen census tract than the year before; 3) the change in the number of positions constituted at least 10% and at most 100% of the total number of filled positions in the chosen census tract in the prior year;

4) The chosen tract featured at least 200 positions in the year prior to the shock; 5) no other tract in the same PUMA experienced an offsetting shock more than 50% as large as the shock to the chosen tract; and 6) less than 50% of the change in number of positions filled in the year of the shock was offset by a shock to the same tract in the opposite direction the following year.

These criteria were chosen to ensure that a sufficient number of states would be reporting data in both the shock year and the prior year to properly capture any worker reallocation, that the shock was big enough to represent a meaningful disruption to both the chosen tract and the surrounding area, and that the shock was sufficiently persistent that the possibility of a reporting error by a large firm in the unemployment insurance data was unlikely to cause a spurious "shock".

To create a forecast of the worker reallocations that a given shock occurring in year y would engender, the full set of model parameters was estimated based on the nationwide sample of worker transitions between years $y-2$ and $y-1$, using the same procedures for smoothing and aggregating types featuring distant locations described in Section 5.1. A counterfactual allocation was then generated by holding fixed the estimated surplus parameters but imposing the marginal distributions of origin and position types from the pair of years capturing the shock, $f^{y-1}(l)$ and $h^y(f)$. Since the exact composition of the shock (as reflected in $h^y(f)$) is built into the forecast, the test of the model is the degree to which the particular flows of workers of different worker types to particular destination position types that resulted from the shock can be predicted.

We assess the accuracy of the forecast using the index of dissimilarity, which measures the percentage of predicted job matches that would need to be reassigned to a different match group in order to perfectly match the distribution of actual job matches across groups. It sums the absolute differences across all match groups g in the share of all matches assigned to g both in the forecast and in the actual data and multiplies by one-half: $\sum_g \frac{1}{2}|P(\hat{g}) - P(g)|$.

To help understand the sources of improvements and shortfalls in model fit, we also compute the index of dissimilarity between the true allocation and four alternative forecasts. The first is a standard parametric conditional logit specification, in which the probability that a random position of type f is filled by a worker whose match would be assigned to group g is given by $P^y(g|f) = \frac{e^{X_g^y \lambda}}{\sum_{g'} e^{X_{g'}^y \lambda}}$, where X_g^y includes a substantial set of regressors constructed for year y that capture the kinds of predictors of joint surplus that researchers often use, and λ is the corresponding vector of parameters estimated from the relationship between the previous year's data, $P^{y-1}(g|f)$ and X_g^{y-1}. The regressors include full sets of dummies for the following categorical variables: origin-destination distance bins using tract pathlength within PUMA, PUMA pathlength within state, and state pathlength between states, initial earnings quartile $\times$ supersector dummies, initial earnings $\times$ firm size quartile dummies, and initial earnings $\times$ firm average pay quartile dummies. The regressors also include an indicator for whether the group g is associated with job movers or stayers ($1(z(g) = 1)$), the worker type frequency $n(l(g))$ interacted with the geographic category of the position type associated with g (tract, PUMA, or state), an interaction between $n(l(g))$ and an indicator for whether $f(g)$ represents the "nonemployment" position type, and dummies for whether the origin and position types associated with match group g share a PUMA and share a state.

The second alternative forecast simply imposes that the CCPs that existed between $y-2$ and $y-1$ also hold during the shock year, so that $P^y(g) = \hat{P}^{y-1}(g|f)h^y(f)$. The third alternative forecast mimics the second, except that the smoothing procedure described in Section A7 is applied to the $y-2$ data prior to constructing $\hat{P}^{y-1}(g|f)$. Like much research on either worker job search or firm job filling, all these alternative forecasts ignore the problem's two-sided nature, and thus do not impose that the proposed allocation satisfies the marginal distribution of worker types, $n^{y-1}(l)$. The fourth alternative forecast is based on Choo and Siow (2006)'s version of the assignment model, in which the idiosyncratic job match-level joint surplus component ϵ_{ik} is replaced by two terms capturing surplus interactions between worker and position type and worker type and position rather than between worker and position: $\epsilon^1_{if(k)} + \epsilon^2_{l(i),k}$. This comparison is useful for understanding the importance of assumptions about correlation structure among unobserved components in driving predictions about counterfactual assignments.

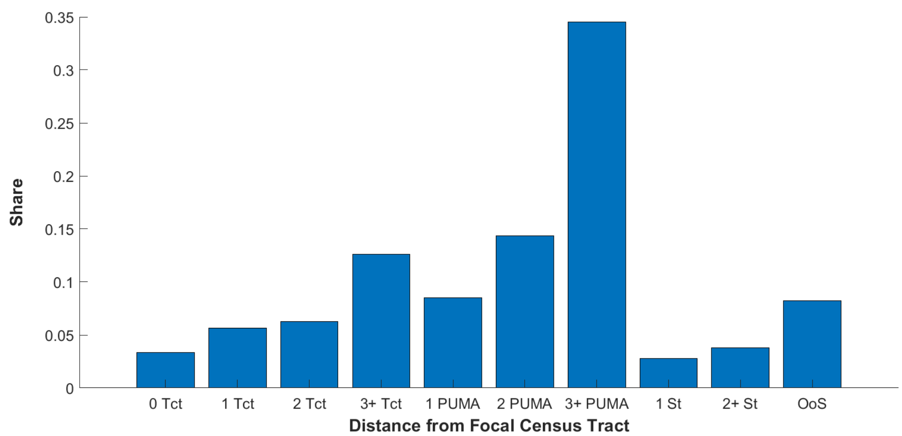
Table 24 contains the results of this exercise. All entries consist of averages across the 514 shocks considered. The first five columns in Row 1 form the index of dissimilarity over all groups g in the 19 state sample, while the second five columns only consider the allocation among groups g featuring origin worker types from the same PUMA as the tract receiving the shock, so as to hone in on the local area most disrupted by the shock. The two-sided matching model, with parameters estimated from the previous period, would only need 6.2% of all job matches in the country to be reallocated to different match groups to perfectly match the data, although 34.1% of the workers originally in the relevant PUMA were misallocated. However, predicting the exact joint distribution of origin tract and initial skill category among workers hired separately for positions defined by tract/size/avg. pay/industry combinations is quite a tall order. Comparing across columns, we see that the parametric logit, despite over 100 regressors, performs considerably worse: 38.3% of all U.S. transitions and 43.3% of transitions starting in the relevant PUMA must be reallocated to a different match group to match the actual post-shock allocation. Holding fixed the full prior year CCP distribution (cols. 3 and 8) performs slightly worse than the two-sided estimator within the target PUMA (37.0% misallocated), while smoothing the CCPs improves the fit to 34.4%. The Choo-Siow model slightly performs the current model by this metric, with only 32.4% misallocated.

For many purposes, however, forecasting exactly the right origin and destination tracts of transitions may be less important than correctly assessing the degree to which the disruption dissipates farther from the shock. To this end, row 2 reports results in which groups are combined that feature the same worker and establishment characteristics as well as origin and destination locations that belong to the same distance bin (using 42 bins), so that the dissimilarity index is computed over a somewhat coarser set of match groups. Only 4.5% of national job matches and 19.3% of matches within the target PUMA are now misallocated by the two-sided forecast, with the two CCP forecasts following suit, suggesting that a substantial share of "incorrect" predictions might nonetheless be sufficiently accurate for most purposes. Furthermore, row 3 shows that combining groups featuring the same distance bins and worker earnings category but different establishment size, avg. pay, and industry categories reduces the index of dissimilarity to 4.2% for workers originating in the targeted PUMA, and to 1% nationally. Furthermore, the two-sided model significantly outperforms the simpler smoothed and unsmoothed CCP models at this level of aggregation (4.2% vs. 6.2% and 6.0%, respectively, within PUMA), and slightly outperforms the Choo-Siow model (4.4%). This suggests that the two-sided matching model matches well the locations of job movers and stayers, but is slightly less effective at matching small differences in the establishment characteristics of the jobs to which workers move. Aggregating from 42 to 17 larger distance bins (row 4) provides a slight improvement, showing again that many "incorrect" predictions are nonetheless fairly accurate.

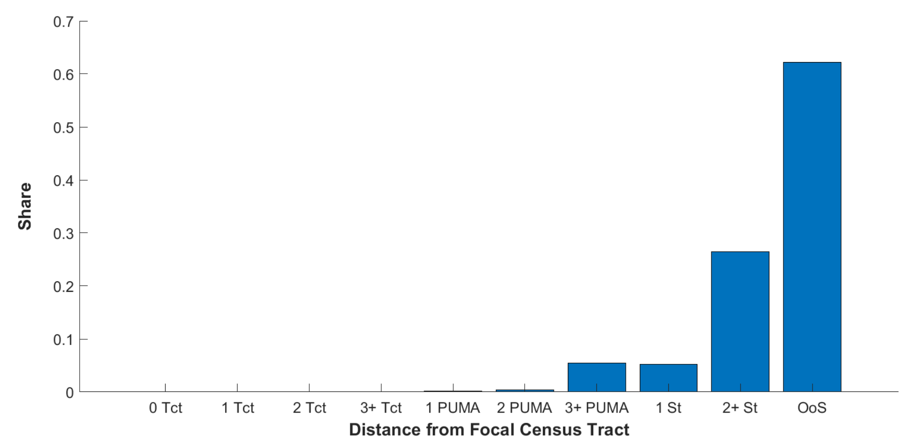
For other purposes, the primary goal of a forecast might be to properly predict the geographic and skill incidence of unemployment. To this end, row 5 computes the index of dissimilarity exclusively over the set of groups featuring workers entering or exiting unemployment, so that the exercise is to predict the location and initial earnings quartile of those losing jobs and the firm composition of those finding jobs (separately by worker initial location). Using the full set of locations, the worker or firm types of only 0.9% of workers entering or exiting unemployment would need to be altered in order for the two-sided prediction to match the allocation that actually occurred. Focusing on only the workers originally working (or most recently working) within the target PUMA increases this value to 2.7%. The two-sided estimator easily outperforms the CCP estimators both nationally and within the target PUMA (both estimators are around 2% and 8%, respectively), and slightly outperforms the Choo-Siow model within the target PUMA (3.4%) Aggregating locations into 42 distance bins shows that the two-sided predictions only badly predicts origin and destination locations for 1.3% of unemployment entrants originating in the PUMA, suggesting that it predicts quite well the geographic and skill incidence of changes in unemployment following the shocks considered. Taken together, the model does quite a good job of predicting the reallocation of workers across job types and particularly across employment/unemployment status that follows major local labor market shocks.

Figure A1: Key Distributions

(a) Empirical Distribution of 2012-2013 Job Transitions



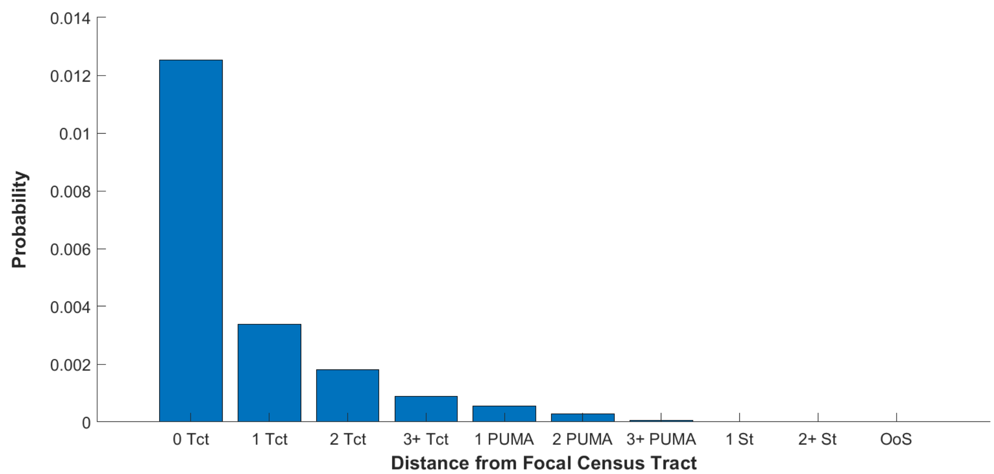
(b) Distribution of the Distance Between Workers' Origin Position and the Census Tract Targeted by the Simulated Stimulus Package: Average across All Simulated Stimuli



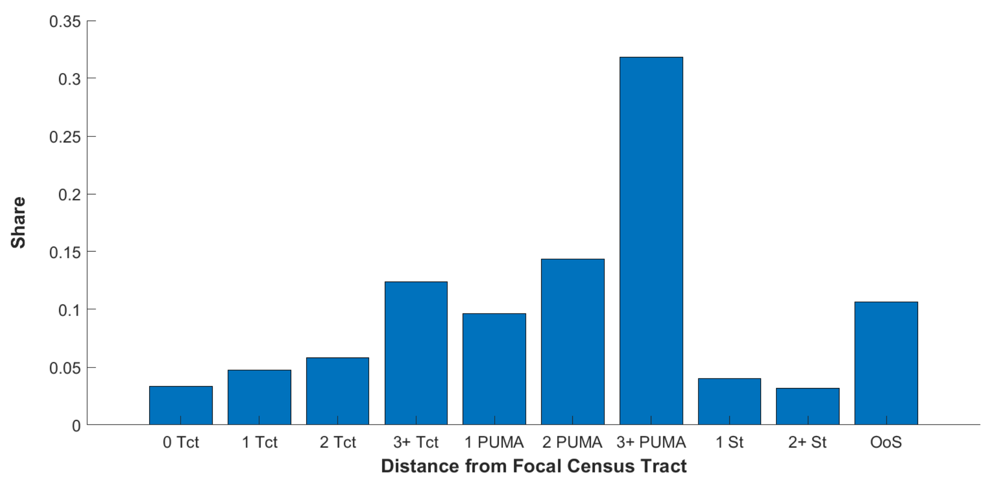
Notes: The bar heights in Figure A1a capture the shares of all worker transitions between dominant positions in 2012 and 2013 in which the geographic distance between these positions' establishments fell into the distance bins indicated by the bar labels. The bar heights in Figure A1b capture the shares of all workers for whom the geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the labeled distance bins (computed separately for each target tract, then averaged across all 500 target tracts). "0/1/2/3+ Tct" indicates that the two establishments (or, for Figure A1b, the establishment and the targeted tract) were in the same tract or one, two, or 3+ tracts away (by tract pathlength) within the same PUMA. "1/2/3+ PUMA" and "1/2+ State" indicate the PUMA pathlength (if within the same state) and state pathlength, respectively. "OoS" indicates that the worker's position was in an out-of-sample state.

Figure A2: Probability of Obtaining a Stimulus Job and Share of All Stimulus Jobs Obtained by Distance From Focal Tract: Average across All Simulated Stimuli

(a) Probability of Obtaining a Stimulus Job



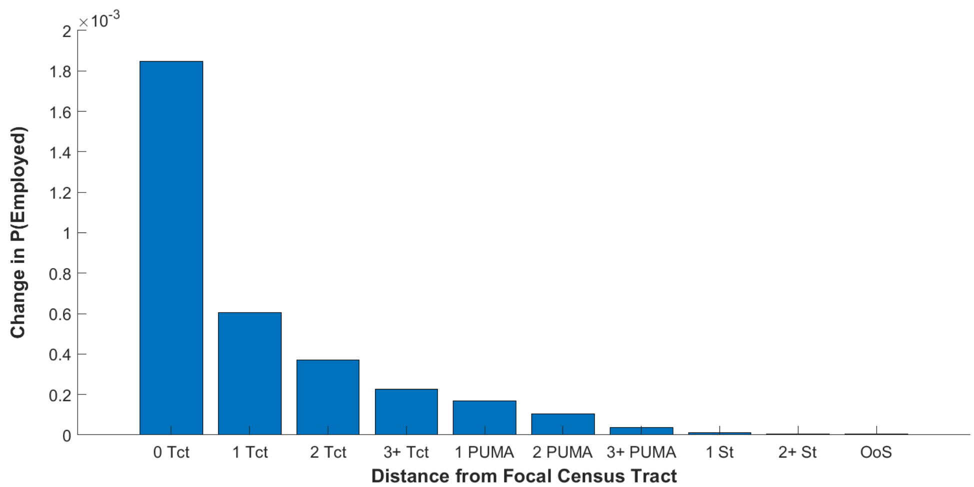
(b) Share of All Stimulus Jobs



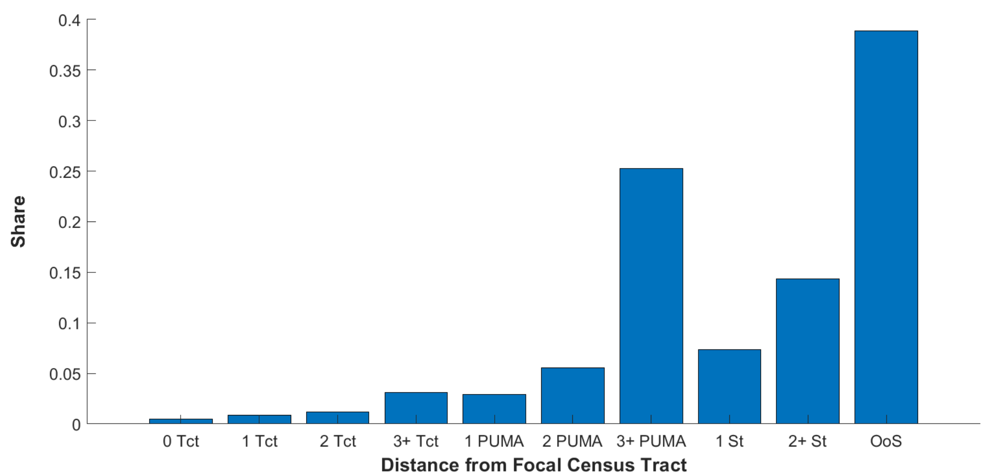
Notes: The bar heights in Figure A2a capture the average probability of obtaining a stimulus job among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins indicated by the bar labels. These probabilities average across different initial earnings categories and across stimulus packages featuring different firm compositions. Figure A2b displays the share of all stimulus jobs that redounds to workers in the chosen distance bin. "0/1/2/3+ Tct" indicates that the origin establishment was in the target tract or was one, two, or 3 or more tracts away (by tract pathlength) within the same PUMA. "1/2/3+ PUMA" and "1/2+ State" indicate the PUMA pathlength (if within the same state) and state pathlength (if in different states), respectively. "OoS" indicates that the worker's position was in an out-of-sample state.

Figure A3: Change in P(Employed) and Share of Additional Employment Obtained by Distance From Focal Tract: Average across All Simulated Stimuli

(a) Change in P(Employed)



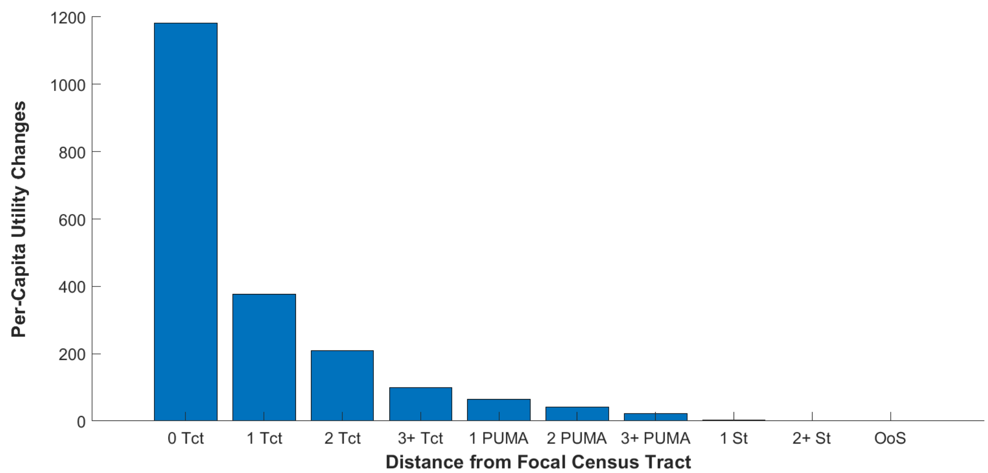
(b) Share of All Additional Employment



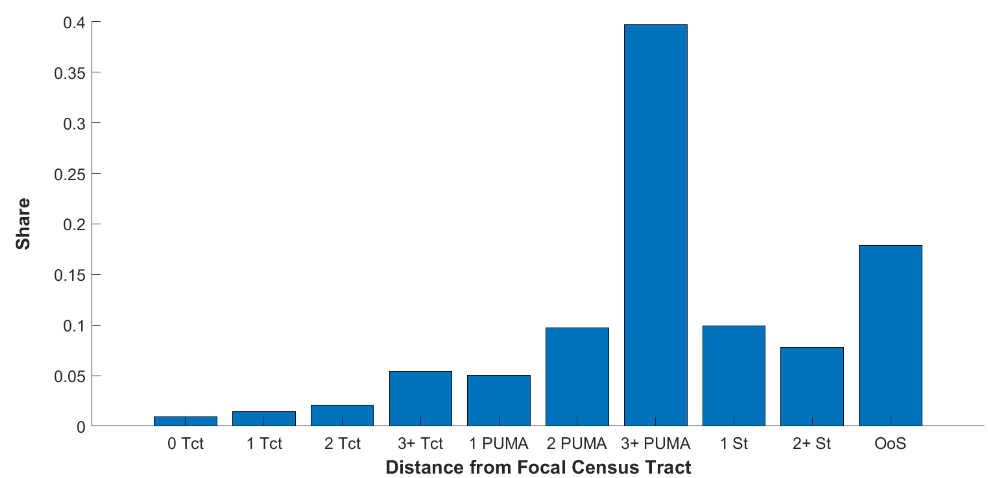
Notes: The bar heights in Figure A3a capture the average change in employment probability among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins indicated by the bar labels. These probabilities average across different initial earnings categories and across stimulus packages featuring different firm compositions. Figure A3b displays the share of additional employment generated by the stimulus that redounds to workers in the chosen distance bin. "0/1/2/3+ Tct" indicates that the origin establishment was in the target tract or was one, two, or 3 or more tracts away (by tract pathlength) within the same PUMA. "1/2/3+ PUMA" and "1/2+ State" indicate the PUMA pathlength (if within the same state) and state pathlength (if in different states), respectively. "OoS" indicates that the worker's position was in an out-of-sample state.

Figure A4: Expected Welfare Changes and Share of Total Welfare Gains by Distance From Focal Tract: Average across All Simulated Stimuli

(a) Expected Welfare Changes ($)

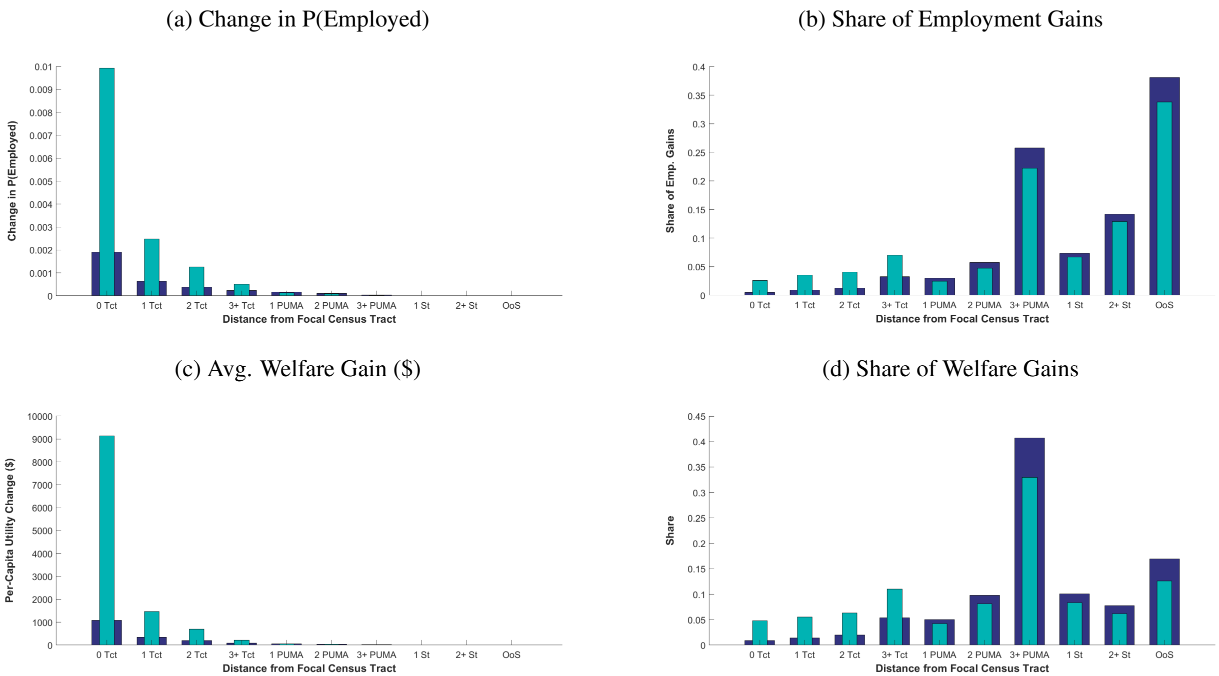


(b) Share of Total Welfare Gains



Notes: The bar heights in Figure A4a capture the average welfare gain (scaled in $ of annual earnings) among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins indicated by the bar labels. Averages are taken across different initial earnings categories and across stimulus packages featuring different firm compositions. Figure A4b displays the share of all welfare gains generated by the stimulus that redounds to workers in the chosen distance bin. “0/1/2/3+ Tct” indicates that the origin establishment was in the target tract or was one, two, or 3 or more tracts away (by tract pathlength) within the same PUMA. “1/2/3+ PUMA” and “1/2+ State” indicate the PUMA pathlength (if within the same state) and state pathlength (if in different states), respectively. “OoS” indicates that the worker’s position was in an out-of-sample state.

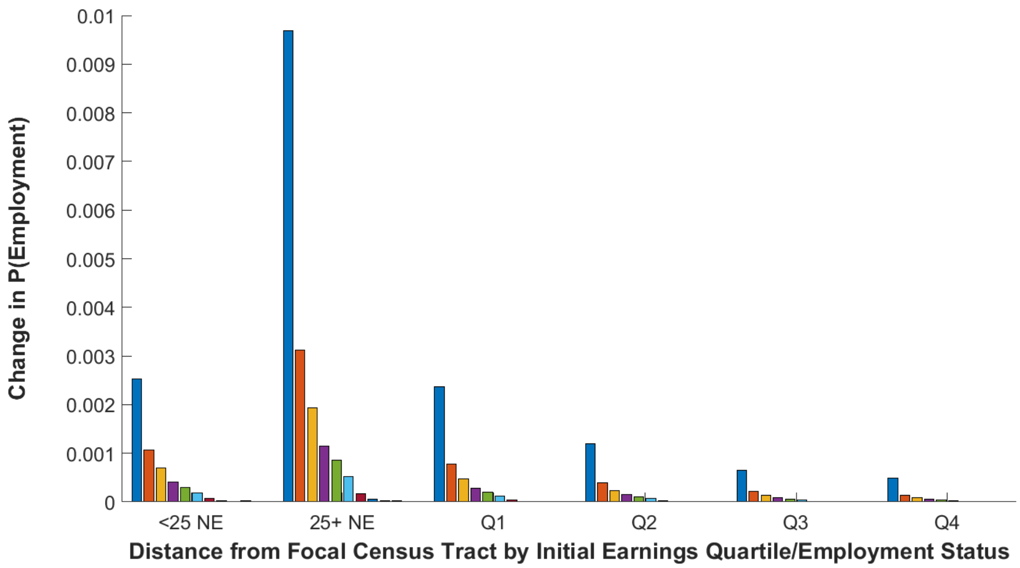
Figure A5: Assessing the Value of Restricting Stimulus Jobs to Fill Positions Within the Target PUMA: Spatial Employment and Welfare Incidence for Restricted and Unrestricted 500-Position Stimulus Packages



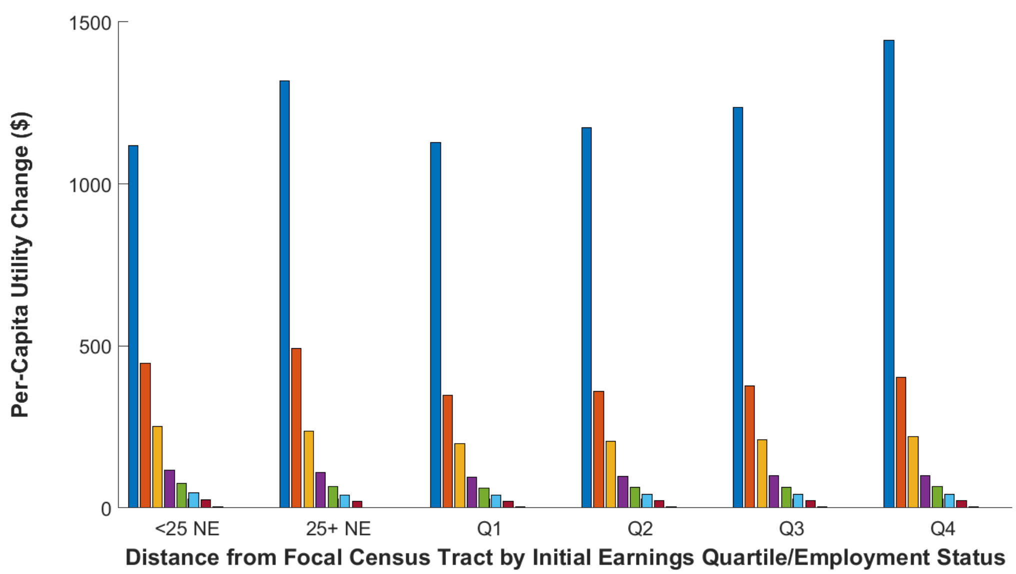
Notes: The bar heights capture the average measure of stimulus incidence associated with the chosen figure from a 500 person stimulus package among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins indicated by the labels. The thin, light blue bars capture the case in which the new positions are restricted to be filled by existing workers within the targeted PUMA, while the wide, dark blue bars capture the case in which new positions can be filled by any worker. Each bar represents an average over 500 simulations featuring different target census tracts as well as over 32 packages for each these 500 simulations featuring different firm composition (combinations of industry supersector, firm size quartile, and firm avg. pay quartile). “0/1/2/3+ Tct” indicates that the origin establishment was in the target tract or was one, two, or 3 or more tracts away (by tract pathlength) within the same PUMA. “1/2/3+ PUMA” and “1/2+ State” indicate the PUMA pathlength (if within the same state) and state pathlength (if in different states), respectively. “OoS” indicates that the worker’s position was in an out-of-sample state.

Figure A6: Change in P(Employed) and in Expected Utility Among Workers Initially Employed at Different Combinations of Initial Earnings/Employment Status and Distance from Focal Tract: Average across All Simulated Stimuli

(a) Change in P(Employed)



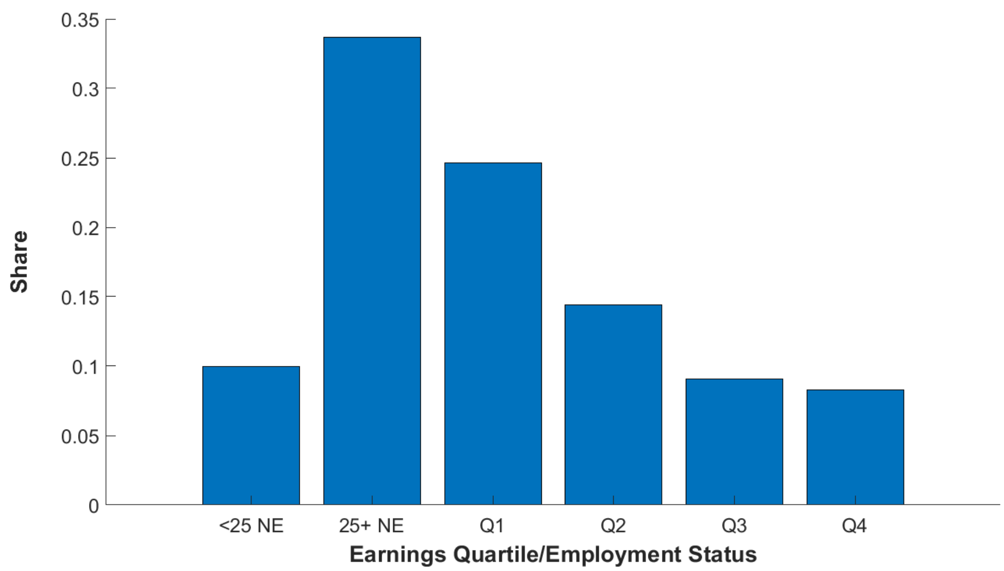
(b) Change in Expected Utility



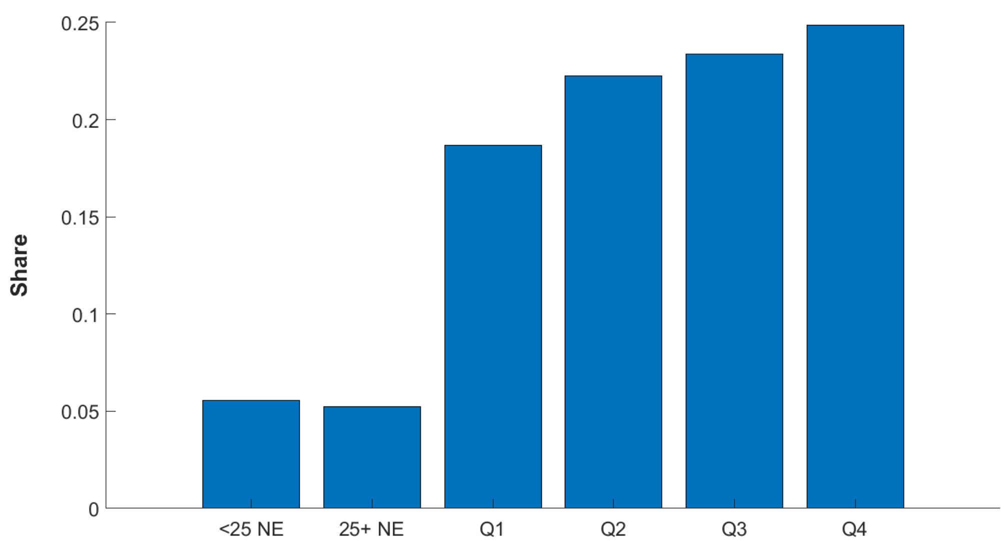
Notes: The bar heights within a particular group in Figures A6a and A6b capture the average change in employment probability and expected job-related utility, respectively, among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins defined in Figure A2. Each group of bars displays average outcomes across distance bins for groups of workers defined by their origin employment status/earnings category. Averages are taken across stimulus packages featuring different firm supersector/size/avg. pay compositions, as well as across 500 simulations featuring different targeted census tracts for each firm composition. "<25 NE/25+ NE": Workers who were not employed in the origin year who were younger than 25 years of age/at least 25 years of age. "Q1/Q2/Q3/Q4": Workers whose pay at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th Quartile of the national earnings distribution.

Figure A7: Shares of Additional Employment and of Total Utility Gains among Workers Initially Employed (or Nonemployed) at Different Initial Earnings Quintiles (or Nonemployment): Average across All Simulated Stimuli

(a) Share of Additional Employment



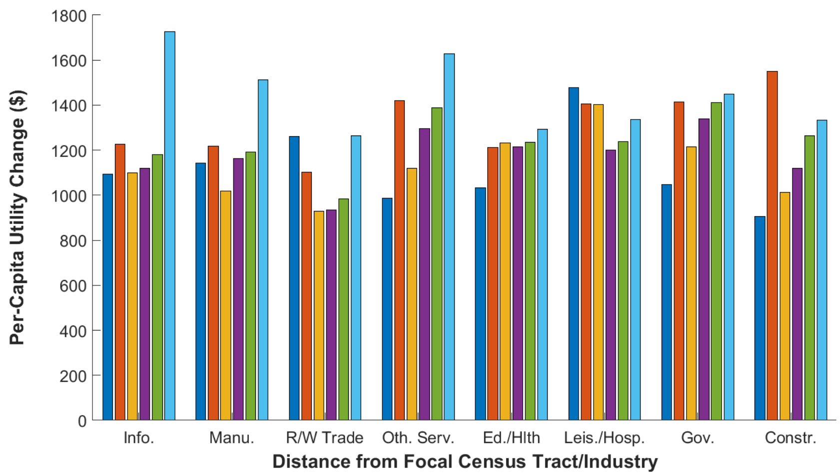
(b) Share of Total Utility Gains



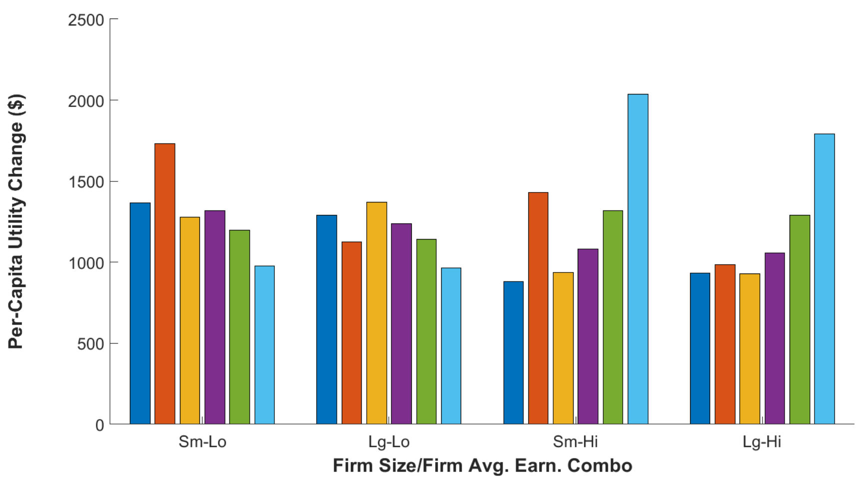
Notes: The bar heights in Figure A7a and A7b capture the average share of all employment gains and of all welfare gains, respectively, among workers whose origin employment status fell into the employment status/earnings quartiles indicated by the bar labels. Averages are taken across different bins capturing the distance between the workers initial (or most recent) employment and the targeted tract receiving the stimulus, across stimulus packages featuring different industry supersector/firm size/firm avg. pay compositions, as well as across 500 simulations featuring different targeted census tracts for each stimulus composition. "<25 NE/25+ NE": Workers who were not employed in the origin year who were younger than 25 years of age/at least 25 years of age. "Q1/Q2/Q3/Q4": Workers whose pay at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th Quartile of the national earnings distribution.

Figure A8: Expected Welfare Gains (in $) Among Workers Originally Working in the Targeted Tract by Initial Earnings/Employment Status: By Industry Supersector or Combination of Firm Size Quartile/Firm Average Pay Quartile

(a) Expected Welfare Gain by Industry Supersector



(b) Expected Welfare Gain by Firm Size Quartile/Firm Avg. Pay Quartile Combination

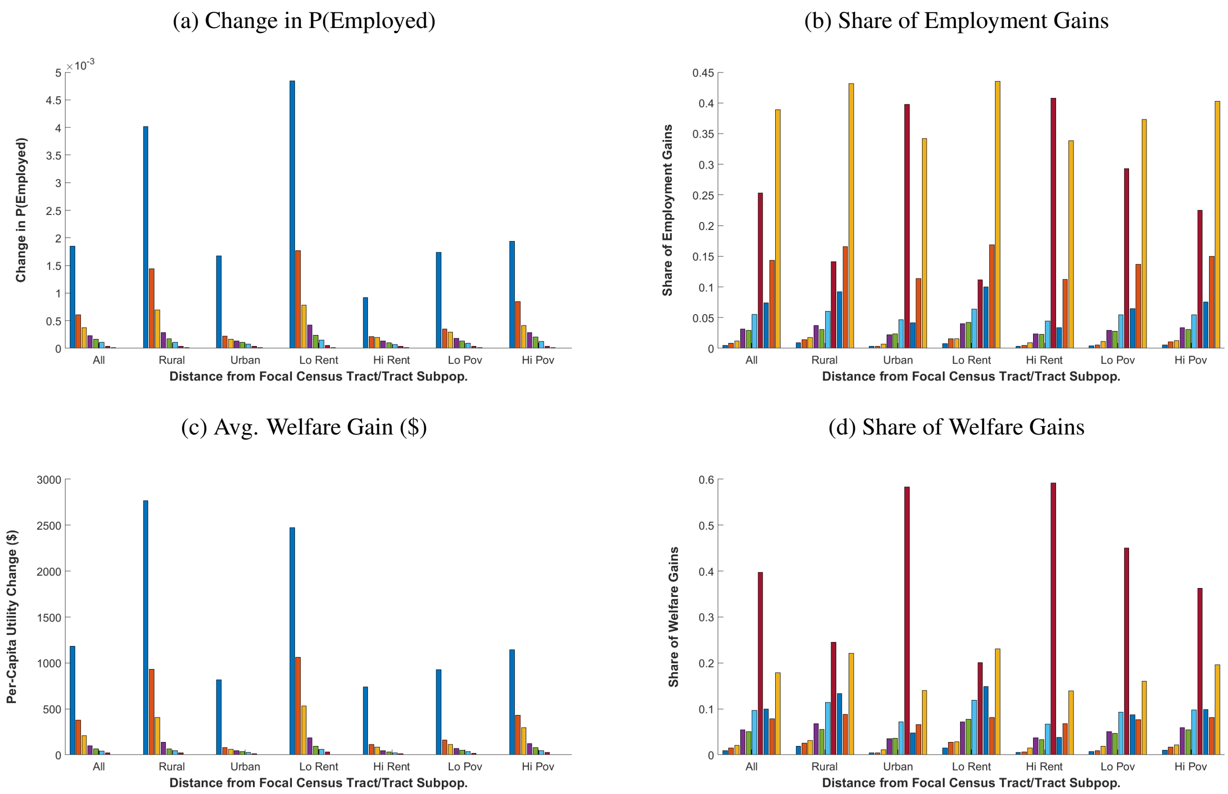


Notes: The bar heights within a particular group in Figures A8a and A8b capture the average welfare gain from a 500 person stimulus package among workers whose origin employment status/earnings quartile fell into the bins defined in Figure A7. Each group of bars displays average outcomes among simulated stimulus packages featuring positions within the particular industry supersector (in Figure A8a) or particular firm size/firm avg. earnings quartile combination (in Figure A8b) given by the group's label. Averages are taken across different initial distance from the focal tract of the shock, across stimulus packages featuring different industry supersector compositions (in Figure A8b) or different firm size/firm pay compositions (in Figure A8a), as well as across 500 simulations featuring different targeted census tracts for each supersector/firm size/firm avg. pay combo. "R/W Trade" = Trade, Transportation and Utilities. "Other Services" includes repair and maintenance firms, personal and laundry services, and religious/civic/professional organizations and private households. "Sm-Lo": Establishments in the 1st quartile of establishment size (based on employment) and 2nd quartile of average pay. "Sm-Hi": 1st size quartile, 4th pay quartile "Lg-Lo": 4th size quartile, 2nd pay quartile. "Lg-Hi": 4th size quartile, 4th pay quartile.

Figure A9: Expected Utility Changes Among Workers Originally Working in the Targeted Tract by Initial Earnings/Employment Status: All Stimulus Packages

Notes: Each line traces the expected welfare gain among focal tract workers generated by a stimulus package featuring 500 positions among firms with a particular supersector/firm size quartile/firm pay quartile combination across alternative origin nonemployment or earnings quartile categories. 32 different lines corresponding to 32 different firm supersector/size/pay level compositions are displayed. Averages are taken across 500 simulations featuring different targeted census tracts for each supersector/firm size/firm avg. pay combo. "<25 NE/25+ NE": Workers who were not employed in the origin year who were younger than 25 years of age/at least 25 years of age. "Q1/Q2/Q3/Q4": Workers whose pay at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th Quartile of the national earnings distribution.

Figure A10: Heterogeneity in the Geographic Concentration of Several Incidence Measures Across Various Subsets of Focal Tracts



(a) Change in P(Employed)

(b) Share of Employment Gains

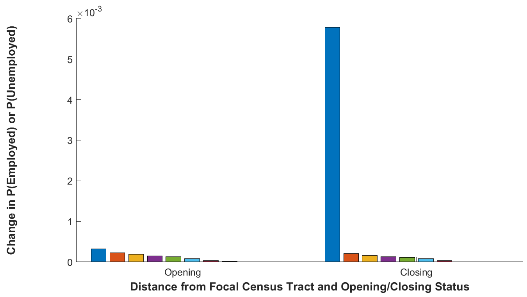
(c) Avg. Welfare Gain ($)

(d) Share of Welfare Gains

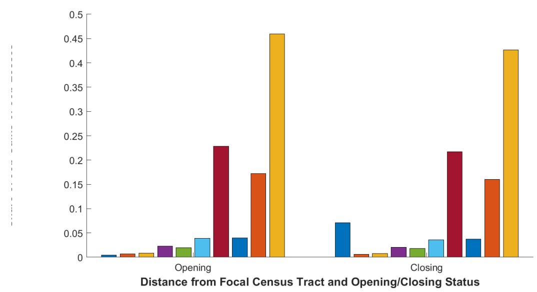
Notes: The bar heights within a particular group in Figures A10a-A10d capture the average measure of stimulus incidence associated with the chosen figure from a 500 person stimulus package among workers whose geographic distance between their origin establishments and the census tract receiving the simulated stimulus package fell into the distance bins defined in Figure A2. Each group of bars displays this incidence distribution across distance bins for a particular subset (indicated by the group's label) of the 500 simulations featuring different focal tracts that were performed. In addition to averaging over the simulations featuring different target tracts within the chosen subset, the displayed results also average over different stimuli featuring the same target census tract but different firm compositions. "All": Average is taken among all 500 target tracts. "Rural"/"Urban": Average is taken among the 100 target tracts with the smallest/largest residential population density. "Lo Rent"/"Hi Rent": Average is taken among the 100 target tracts with the lowest/highest rent for a two bedroom apartment. "Lo Pov"/"Hi Pov": Average is taken among the 100 target tracts with the lowest/highest household poverty rate.

Figure A11: Asymmetry in Employment and Welfare Incidence from Plant Openings and Closings of Equivalent Magnitude

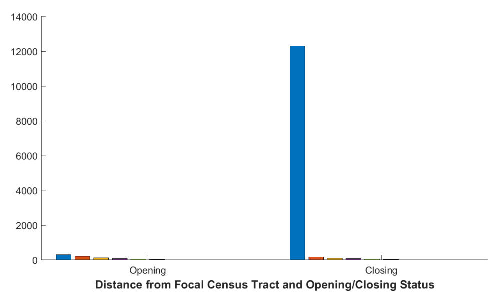
(a) Change in P(Employed)



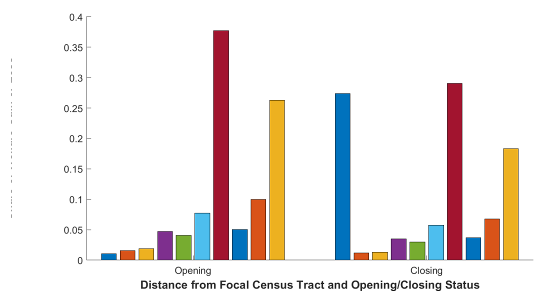
(b) Share of Employment Gains



(c) Avg. Welfare Gain ($)



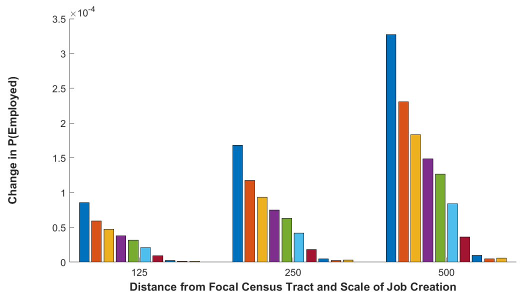
(d) Share of Welfare Gains



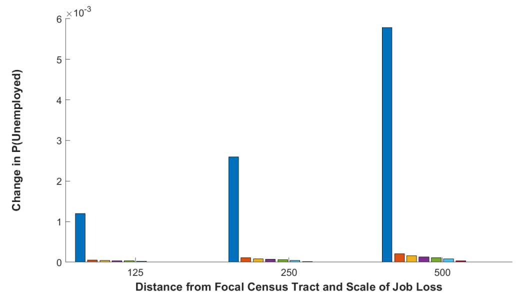
Notes: The bar heights within a particular group in Figures A11a-A11d capture the average value of the incidence measure associated with the figure from pairs of simulated plant openings and closings among workers whose geographic distance between their origin establishments and the census tract experiencing the disaster fell into the distance bins defined in Figure A2. Each opening or closing is associated with the creation or removal of 500 positions at large, high paying manufacturing firms in the focal tract. For each opening or closing, averages are taken across 200 simulations featuring different targeted census tracts.

Figure A12: Employment and Welfare Incidence from Plant Openings and Closings of Different Magnitudes: 125, 250, and 500 Jobs Created or Removed

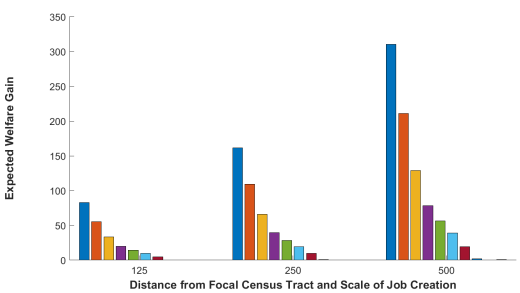
(a) Change in P(Employed): Openings



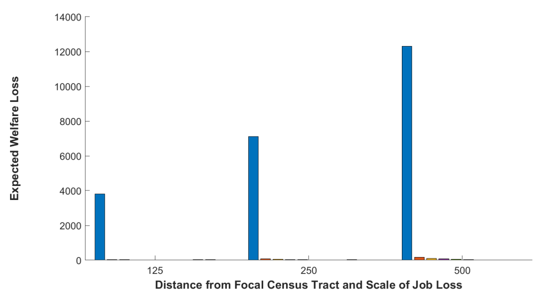
(b) Change in P(Unemployed): Closings



(c) Avg. Welfare Gain ($): Openings



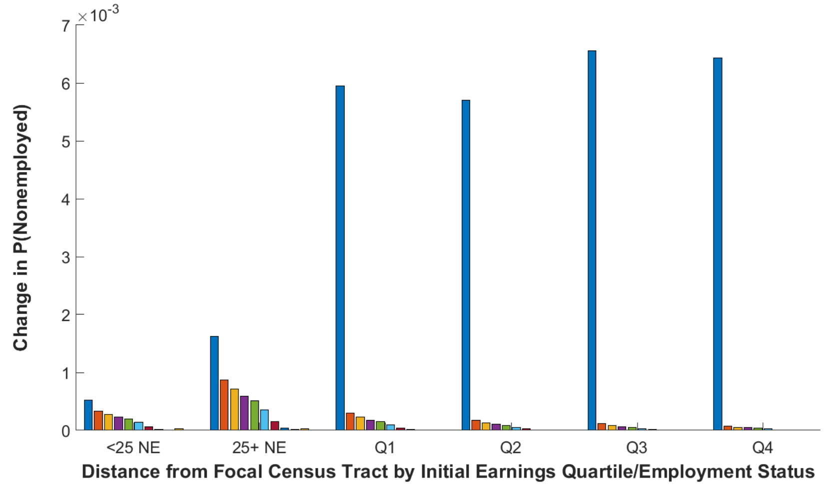
(d) Avg. Welfare Loss ($): Closings



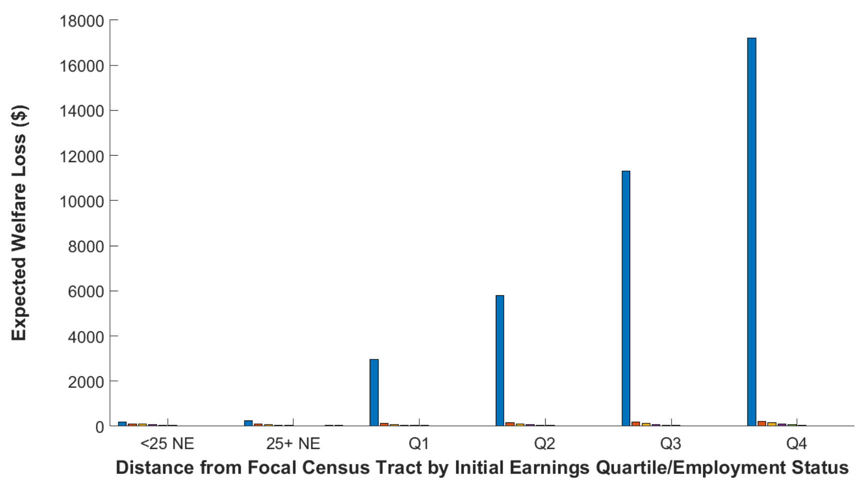
Notes: The bar heights within a particular group in Figures A12a-A12d capture the average value of the incidence measure associated with the figure from pairs of simulated plant openings and closings among workers whose geographic distance between their origin establishments and the census tract experiencing the disaster fell into the distance bins defined in Figure A2. Each opening or closing is associated with the creation or removal of 125, 250, or 500 positions at large, high paying manufacturing firms in the focal tract. For each opening or closing of each scale, averages are taken across 200 simulations featuring different targeted census tracts.

Figure A13: Changes in the Probability of Unemployment and Expected Welfare Produced by the Removal of 500 Positions at Large, High Paying Manufacturing Firms for Randomly Chosen Workers with Different Combinations of Initial Earnings/Employment Status and Distance from Focal Tract

(a) Change in P(Unemployed)



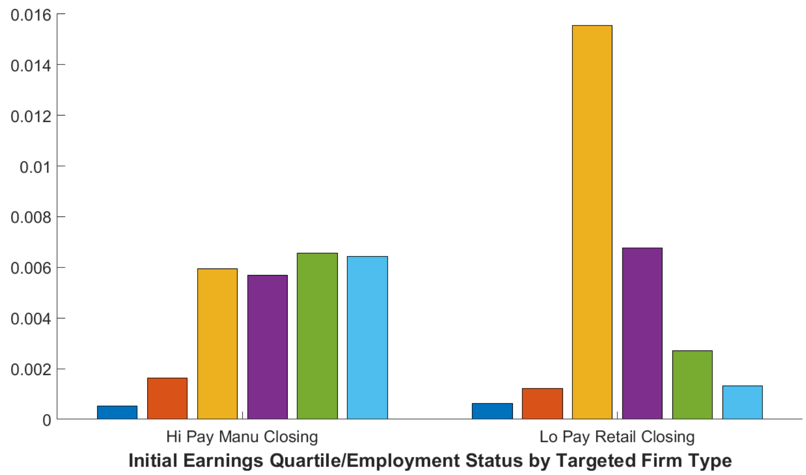
(b) Change in Expected Welfare (in $)



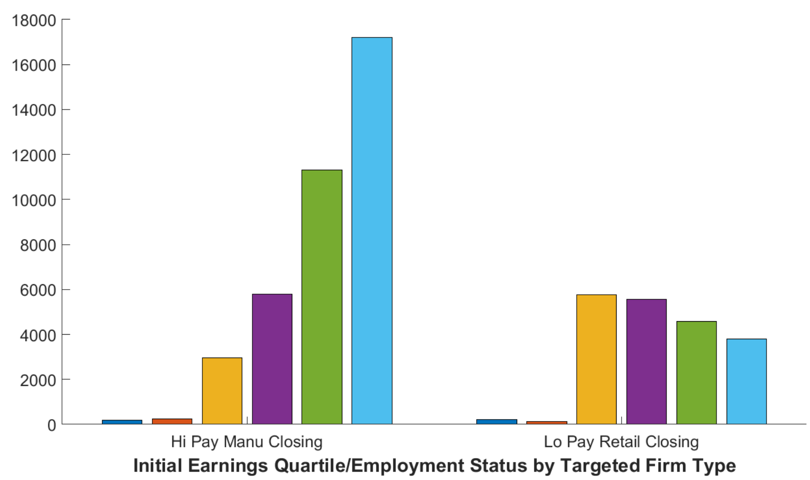
Notes: The bar heights in Figure A13a capture the average change in unemployment probability from a "plant closing" eliminating 500 positions at a large, high paying manufacturing firm among workers whose geographic distance between their origin establishments and the census tract experiencing the closing fell into the distance bins defined in Figure A2. Each group of bars displays the distribution of employment losses across distance bins for groups of workers defined by their origin employment status/earnings category. Figure A13b displays the corresponding set of distributions of expected job-related welfare losses (scaled in $ of 2012 annual earnings). For each outcome, averages are taken across 500 plant closing simulations featuring different targeted census tracts. "<25 NE/25+ NE": Workers who were not employed in the origin year who were younger than 25 years of age/at least 25 years of age. "Q1/Q2/Q3/Q4": Workers whose pay at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th quartile of the national earnings distribution.

Figure A14: Changes in Unemployment Rates and Expected Job-Related Utility for Workers Initially Employed (or Nonemployed) in the Focal Tract Produced by Plant Closings Featuring either High-Paying Manufacturing Positions or Low-Paying Retail Positions by Initial Earnings Quintile (or Nonemployment)

(a) Change in P(Unemployed)



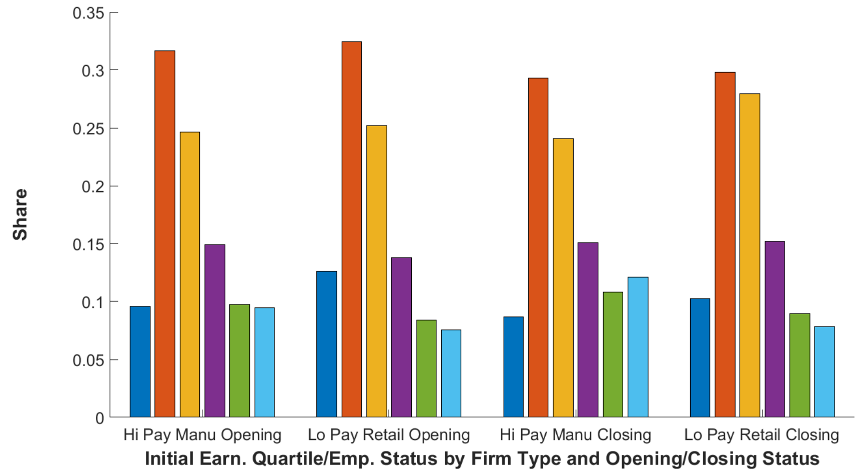
(b) Expected Utility Loss (in $)



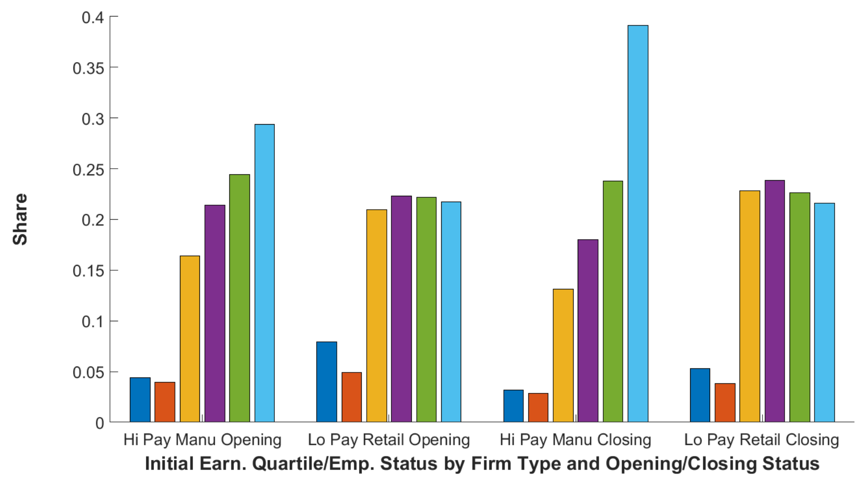
Notes: The bar heights within a particular group in Figures A14a and A14b capture the change in unemployment rate and expected utility, respectively, from a set of simulated plant closings among workers who were initially employed (or nonemployed) in the focal tract and whose origin employment status/earnings quartile fell into the bins defined in Figure A7. For each outcome, the left group of bars depicts the incidence of the removal of 500 positions at large, high paying manufacturing firms, while the right group depicts the corresponding incidence of the removal of 500 positions at large, low-paying retail firms. For each type of plant closing, averages are taken across 500 simulations featuring different targeted census tracts.

Figure A15: Expected Shares of Additional Unemployment and Welfare Losses Produced by Plant Closings Featuring either High-Paying Manufacturing Positions or Low-Paying Retail Positions among Workers Initially Employed (or Nonemployed) at Different Initial Earnings Quartiles (or Nonemployment)

(a) Share of Additional Unemployment



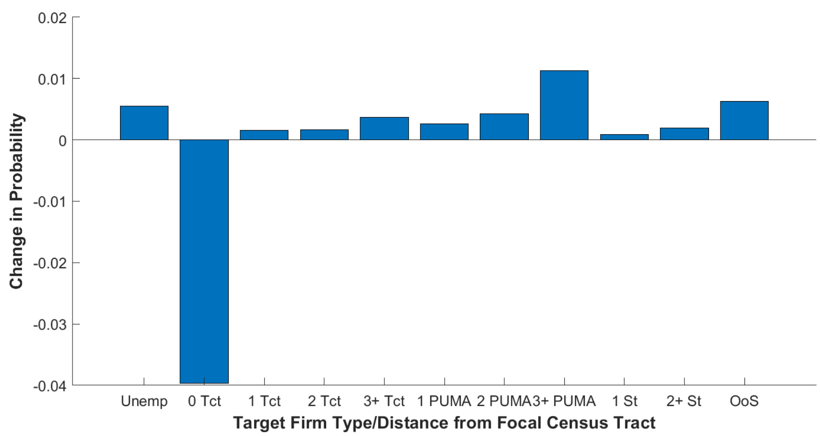
(b) Share of Total Welfare Losses



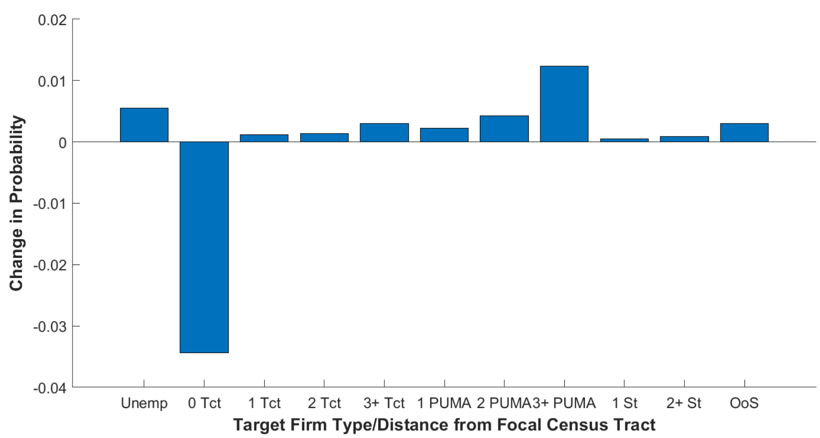
Notes: The bar heights within a particular group in Figures A15a and A15b capture the average share of additional employment or unemployment and welfare gains or losses, respectively, from a set of simulated plant openings and closings among workers whose origin employment status/earnings quartile fell into the bins defined in Figure A7. For each outcome, the left group of bars depicts the incidence of the creation of 500 positions at either large, high paying manufacturing firms or large, low-paying retail firms, while the right group depicts the incidence of removing 500 positions with corresponding firm compositions. For each type of plant opening or closing, averages are taken across 500 simulations featuring different targeted census tracts.

Figure A16: Comparing Changes in the Distribution of Employment Locations (or Unemployment) for Workers Initially Employed in the Focal Tract after Plant Closings that Remove 500 Positions from either Large High-Paying Manufacturing Firms or Large Low-Paying Retail Firms (Averaged Across Initial Earnings/Employment Status Categories)

(a) High-Paid Manufacturing



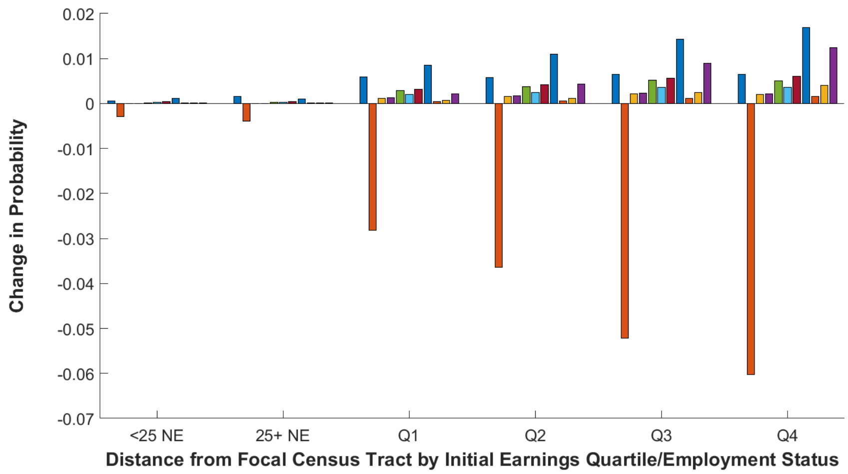
(b) Low-Paid Retail



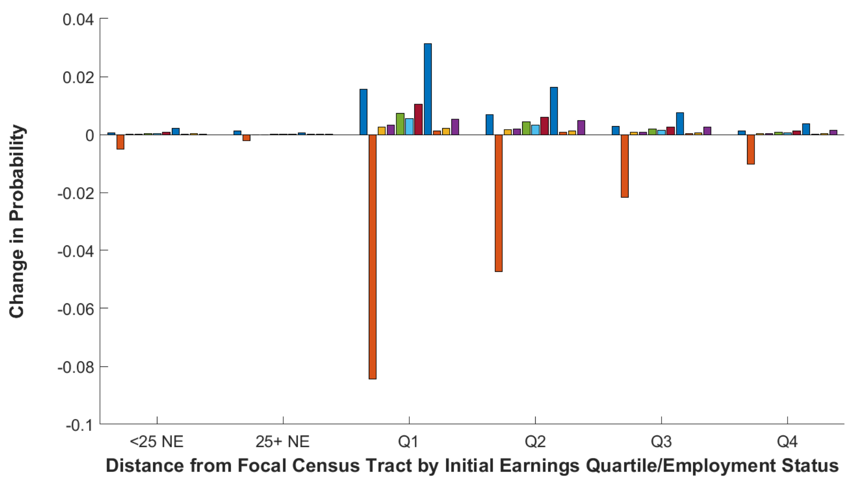
Notes: The bar heights in Figures A16a and A16b capture the impact of experiencing a plant or store closing, respectively, that removes 500 positions on the probability that a worker initially employed (or most recently employed) in the targeted tract would be employed at a position whose geographic distance from the census tract experiencing the closing fell into the distance bins defined in Figure A2 (or become/remain unemployed, the leftmost bar in each group). For both plant and store closings, averages are taken across 500 simulations featuring different targeted census tracts.

Figure A17: Sensitivity in the Change in the Distribution of Employment Locations (or Nonemployment) following Plant Closings to the Match between Workers' Initial Earnings/Employment Status and the Closing Firm's Sector and Pay Level

(a) High-Paid Manufacturing: Change in Distribution of Destinations by Initial Earnings/Employment Status



(b) Low-Paid Retail: Change in Distribution of Destinations by Initial Earnings/Employment Status



Notes: The bar heights within a particular group in Figures A17a and A17b capture the impact of experiencing a plant closing or store closing that removes 500 jobs in the focal tract on the probability that a worker initially employed (or most recently employed) in the targeted tract would be employed at a position whose geographic distance from the census tract experiencing the disaster fell into the distance bins defined in Figure A16a (or become/remain unemployed, the leftmost bar in each group). Each group of bars captures the change in destination employment probabilities among workers from the initial earnings/employment status given by the label. "<25 NE/25+ NE": Workers who were not employed in the origin year who were younger than 25 years of age/at least 25 years of age. "Q1/Q2/Q3/Q4": Workers whose pay at their dominant job in the origin year placed them in the 1st/2nd/3rd/4th Quartile of the national earnings distribution. Figure A17a considers a plant closing that removes 500 positions from large, high-paying manufacturing firms, while Figure A17b considers a store or mall closing that removes 500 positions from large, low-paying retail firms. For both plant and store closings, averages are taken across 500 simulations featuring different targeted census tracts.