ESSAYS IN LABOR ECONOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Evan Nelson Buntrock

May 2015

ESSAYS IN LABOR ECONOMICS

Evan Nelson Buntrock, Ph.D.

Cornell University 2015

The three essays of this dissertation are studies of individual choice and outcomes in labor-economics related problems.

In the first chapter, I use an individual's rank in his coworker-comparison group to predict whether he leaves his job and the amount of earnings growth he will experience over his next few years. Even after controlling for a variety of individual and firm observables and unobservables, I find that an individual's rank is positively correlated with his earnings growth on the current job but negatively correlated with his earnings growth when he changes jobs. The mean reversion of job changers' earnings with respect to rank suggests that rank is a signal of an individual's match productivity with his current firm.

In the second chapter, my co-author and I use a flexible decomposition procedure for job-matching to distinguish changes in job-to-job flows due to structural factors of the labor market from changes due to the evolving composition of workers and firms. We find that the likelihood of workers moving to firms 25-100 miles away from their current firm when changing jobs has increased. This increased integration of local labor markets has gone undetected by other studies of mobility, which focus on interstate and even inter-county job and residential migration.

In the third chapter, I study whether US citizens have become more or less likely over time to marry someone with whom they share a state of birth. Using a variety of descriptive statistics, I find that the proportion of marriages between citizens with different states of birth has increased. Individuals born in later years and those having higher education are generally more likely to marry someone born in a different state.

BIOGRAPHICAL SKETCH

Evan Buntrock was born in Wisconsin, USA in April, 1987, and traversed the midwest before settling in northern Minnesota at age seven. He developed a strong interest in economics in high school when he took Lynn Ellingson's introductory microeconomics and macroeconomics courses. He followed this interest at Notre Dame, where he completed an undergraduate major in Economics. Evan continued his studies in the Cornell economics PhD program under the tutelage of John Abowd. After completing his PhD, he will work as an economist for Amazon.

ACKNOWLEDGEMENTS

This dissertation was written under the supervision of John Abowd, Dan Benjamin, Aaron Bodoh-Creed, and Rick Mansfield. They all made crucial contributions my research and education. John gave me an unparalleled opportunity, unrelentingly demanded excellence, and provided the resources I needed to succeed. Dan reminded me how much I love economics. Aaron taught me how to make my ideas into the reality I wanted. Rick spent an uncountable number of hours teaching me and critiquing my research.

The larger community of professors and students at Cornell also contributed their time and effort, especially the members of John Abowd's and Ted O'Donoghue's research groups. Of these, Michael Strain deserves special mention: I can't imagine making it through graduate school without him. He was generous, savvy, and a comedic genius. Any good deeds I have done at Cornell are at least partly a result of my time with Mike.

I received support that was less research-focused - and therefore more important - from my friends and family. My parents, Chris and Clare Buntrock, reminded me to never give up. My sister Laine encouraged me to take risks. Benjamin Leard rekindled my love of Magic and science fiction. Jeremiah VanScoyoc substantially improved nearly every facet of my existence. Erin Chu taught me to laugh, especially at myself. These three friends have been the best part of going to Cornell.

This list of the muses, hermits, and sylphs who helped me on my journey remains incomplete; I was supported by the kindness of many more friends, acquaintances, and strangers than I can name here. Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

COMPARISON-GROUP POSITION AND JOB SEARCH: INFORMATION OR IRRITATION?

Evan N. Buntrock

Abstract: In this paper, I evaluate how a worker's percentile in the pay distributions of his comparison-groups predicts his labor market behavior and outcomes. Using a dataset that contains earnings information for every worker in America employed in the private, state, or local sectors, I construct within-firm worker comparison-groups and find every worker's percentile in their group's earnings distributions. Workers with lower percentile in their firm comparison-groups are more likely to leave their job but are not more likely to be fired. For workers who stay at their current job, group-position is positively correlated with the raise they receive the next year. Workers of all percentiles see earnings increases when changing jobs, but lower percentile leavers see larger percentage earnings gains than high-percentile leavers. These results suggest that workers' apparent preferences for higher comparison-group position can be explained by pecuniary motives instead of behavioral ones. The closure in the earnings gap between the high and low position workers who switch jobs suggests that a worker's position in the group of similarly educated workers at his firm is in part driven by match effects.

1.1 Introduction

Holding fixed a worker's earnings, how do a worker's positions in his coworker comparison-groups predict his observable outcomes like job changes or future earnings? Being at a high percentile in the earnings distribution of similarly educated coworkers may be a sign that a worker has found an ideal job, but it could also be a sign that he needs to leave the firm to take the next step up the job ladder. Are higher percentile workers in danger of low earnings growth (e.g. from hitting an earnings wall at their current firm), or is their lofty position in the comparison-group a sign of higher earnings growth next year (e.g. from a promotion)? The relationship between comparison-group percentile and future earnings informs provides information about why workers prefer higher group-position. Do the observed correlations between comparison-group percentile and the outcome variables match the behavior of rational, income-maximizing agents or can we only reconcile the patterns with behavioral theories where agents care about position in the group for its own sake? The relationship between a worker's current percentile in her coworker-group earnings distribution and her future earnings will also shed light on the unobservable causes of differences in workers' positions. Are workers in the lower percentiles of their firm-groups' earning distributions as a result of individual qualities that would hamper them at every job, or are these workers suffering from a match to a firm that doesn't value their skills?

Using a data set that contains twenty years of earnings records for 250 million American workers, I find that workers at higher percentiles in the earnings distributions of their coworker comparison-groups are more likely to stay at their current firm. Workers who stay at their firms obtain future earnings gains that grow with the workers' positions in their current comparison-groups. While both high position and low position job-leavers experience significant earnings gains when they change jobs, the percentage and absolute gains from switching are substantially larger for lower percentile workers. Since low-

positioned workers obtain larger earnings gains from job-leaving than higher-positioned workers, a traditional model of rational, income-maximizing agents can explain the behavior relationship between worker position and job changing behavior. The closure in the earnings gap between low and high position workers when they switch jobs cannot be explained by qualities unique to the individual (individual effects). However, this closure would occur if there is a firm-specific reward for a certain skill/quality of the individual; this match effect will increase a worker's earning and position only at a specific firm. In this case, low-positioned, low-earning workers have a lower match effect; changing jobs (on average) help them obtain a better match effect and realize substantial upward income mobility.

For managers and recruiters, the negative correlation between comparison-group position and job changing has substantial implications for recruiting and retention; the lowest paid workers in a group at the firm are the most likely to leave. If there is an excellent performer who you can't lose, paying him marginally more than his peers at the firm might reduce his probability of leaving. Other things, constant, workers in the 75th percentile of the earnings distribution of similarly educated workers are 1.1% less likely to leave than workers in the 25th percentile - this difference is equivalent to the difference in probabilities that arise when one worker earns $10,000 more than another. The underlying causes of workers' different positions also impact the optimal retention and recruiting policies; if a worker being at a low percentile in the earnings distribution indicates a poor match effect, then the relationship between group-position and turnover is one that systematically removes workers with the worst matches. In this instance, the net costs of turnover to the company may be smaller than previously believed. When hiring, the existence of match effects means that workers who did not excel at their prior firm might shine in a new role; their previous low position was (at least partly) driven by an interaction with their former firm instead of some permanent personal flaw.

For workers, group-position can help predict their future earnings whether they stay at their current firm or leave. Other things, constant, workers in the 75th percentile of the earnings distribution of similarly educated workers have 3.5% higher earnings in the next year than workers in the 25th percentile. It is particularly crucial for workers to know whether this association between their percentile and their future earnings is driven by match effects or individual effects. If, other factors constant, group-position reflects unobserved personal qualities, then the information gained from percentile has little value in determining whether another job would be a better fit. However, since a worker's percentile appears to be explained in part by match effects, workers with a low percentile could see large earnings gains from changing jobs; in contrast, workers with a high percentile might consider whether they can expect an even better match effect at their next job.

For policy makers, a worker's percentile has the potential to be a tool that aids income mobility; the relationship between a worker's percentile in the firm-group earnings distribution and earnings gains from switching jobs is substantial. If match effects explain the earnings gains for low percentile workers who change jobs, then measures designed to increase awareness of group-position may lead to higher rates of switching, creating better match effects and increasing surplus for workers and firms. Still, even if everyone has an exceptional match effect, some workers are bound to have low position; position alone could not be persistently informative about match effects in a world with perfect pay transparency.

This paper is the first documentation of the relationship between workers' group-positions, worker behavior, and labor market outcomes that comes from a large, nationally representative data set. Previous attempts to explore the impact of comparison income on worker behavior have been limited by lack of a large, linked employer-employee data set. This forced the authors to use convenience samples and they were unable to observe long

run outcomes. In contrast, I use the Longitudinal Employer-Household Dynamics (LEHD) data, which contains almost twenty years of earnings for over 250 million American workers at 20 million firms. I use this to construct position measures of workers in local, firm-based comparison-groups and tie workers' group-positions to their labor market outcomes; I then disentangle the empirical effects related to group-position from an individual's pay and large set of controls that allows for unobserved individual and firm heterogeneity. Even including a variety of firm and worker control variables, a worker's position in his comparison-group has a significant and large relationship with future earnings: *ceteris paribus*, the differences in future earnings and probability of leaving between workers in the top quartile and bottom quartile are comparable to the expected future earnings differences between two otherwise identical workers with a pay gap of $10,000. Workers who see earnings gains also tend to see large position gains in their comparison-groups.

I provide the first set of basic empirical facts that allow me to address the role of group-position in worker utility functions and the forces that cause differences in worker earnings. My results should not be interpreted as causal, since a worker's percentile in the earnings distribution is not randomly assigned;[1] the differences in group-position between two observationally equivalent workers is driven by differences in the earnings distributions of those workers' comparison-groups. These differences likely reflect different features of the market in which the firms and workers operate and it is reasonable to expect these differences to be systematically correlated with worker outcomes. Nonetheless, my findings suggest that models where workers care directly about group status are not necessary to anticipate job changing behavior; instead, group-position appears to be informative about a workers unobserved match effect with a firm.

This paper is organized as follows. In Section 1.2 I discuss theories about the role of group-position in individual utility and how it might predict behaviors and outcomes in

[1]See Murphy and Weinhardt [2014] for a full discussion of requirements to identify a causal effect of rank on outcomes in a classroom setting.

different theoretical models. In Section 1.3 I describe the data. In Section 1.4, I explain how I will use the data to tease apart the models' predictions and see which one better predicts the data. Section 1.5 contains my results. Section 1.6 provides a summary of the paper and presents my conclusions.

1.2 Theories of Group Position and Relative Income

My goal in this section is to lay out different theories of the importance of position in a comparison-group to individual agents and develop the contrasting predictions they make. I start by covering the literature's (sometimes contradictory) findings concerning group-position, relative income and their importance for workers. Next, I outline several different theories about the causes and consequences of differences in workers' percentiles in the earnings distributions of their comparison-groups; I will summarize the implications of these theories for worker outcomes in Table 1.1, which is the basis for the empirical tests later in the paper.

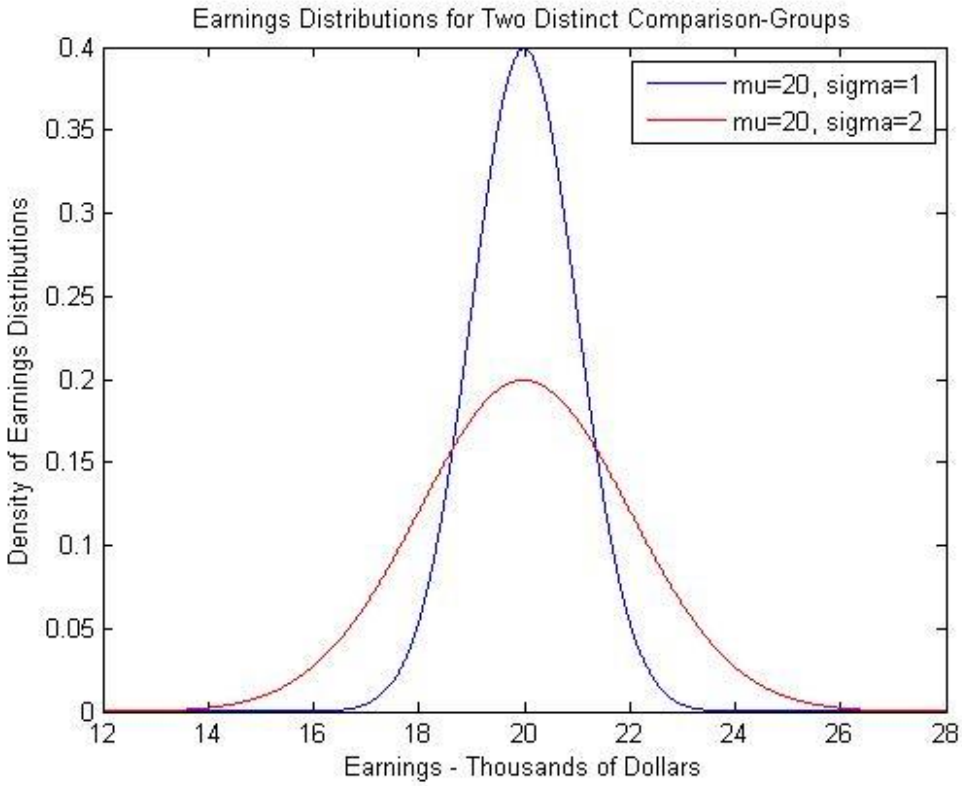
It has been well established that workers care about their comparison-group position and their relative income; the positive correlation between position or relative income and subjective wellbeing reported by Ferrer-i-Carbonell [2005] and Luttmer [2005] are some of the more recent confirmations of the theories of Easterlin [1974, 1995, 2001], Duesenberry [1949], Pigou [1932], Veblen [1965], and Smith [1976]. These preferences for higher position extend into the workplace; evidence linking relative pay to job satisfaction has been found by Clark and Oswald [1996], Hamermesh [2001] and Brown et al. [2008]. Card et al. [2012] even find evidence of a causal relationship between exposure to information about relative income, group-position and self reported job satisfaction of employees in the University of California system.

There is only a small body of evidence linking a worker's group-position or relative income to his actions, particularly the decision to find a new job. Experimental evidence

from Cohn et al. [2014] suggests lowering a workers relative income causes a large decrease in his provision of effort. While Baker et al. [1994] find no evidence that low relative income relates to turnover at a single American financial firm, Galizzi and Lang [1998] examine a small Italian data set and finds workers with lower relative income, as compared to mean earnings at the firm, are less likely to quit their job. These results contrast with the results of Kwon and Meyersson Milgrom [2014], who find that a worker's position in the wage distribution of his within firm-occupation before a merger or acquisition as well as his "expected" post-merger position in that distribution are inversely correlated with his decision to stay at or leave the (newly formed) firm. Rege and Solli [2013] report that when Norwegian workers received an information shock about their co-workers' incomes, workers with low relative income were more likely to leave for new jobs. Evidence from Haltiwanger and Vodopivec [2003] and Brown et al. [2008] could also be interpreted as linking relative income to turnover.

Identifying the precise reasons *why* workers care about and respond to position in the comparison-group is difficult; as noted by Heffetz and Frank [2011], the evidence that individuals prefer higher status is also compatible with other hypotheses. The work of Card et al. [2012] provides evidence that position is a direct utility component. Frank [1984] suggests that workers with higher group-position and relative income are paid well below their marginal value product. Along with Kwon and Meyersson Milgrom [2014], who find evidence supporting the notion that high relative income is associated with less rapid wage growth, this is suggestive of a "market for status" of the kind suggested by Frank [1984] and Becker et al. [2005]. However, there might also be strong pecuniary reasons for group-position to concern individuals. In settings where wages approximate marginal productivity, group-position may provide valuable information. Holding observables like education, experience, and tenure constant, group-position may tell workers about their "unobserved" individual effects or match effects (first described by Jovanovic [1979a,b,

Figure 1.1: Variation in Group-Position, Fixing Absolute and Relative Income



Notes: This figure shows the earnings distributions of two distinct comparison-groups; suppose these are the earnings distributions of two different firms. The distributions have the same mean—$20,000—but the distribution at the red firm has twice the variance of the earnings distribution at the blue firm; this means that an individual making $20,000 at the blue firm is in a higher percentile of the distribution than an individual making $22,000 at the red firm. This is true in spite of the fact that the two individuals have the same absolute income and relative income.

1984]), or at the very least, information about their firm's opinion of these unobservables. While a worker's relative income provides information about the absolute value of his match effect, his group-position provides information about where he falls in the firm's (or within firm peer-group's) distribution of match effects. This variation can be seen in Figure 1.1.

Both group-position and relative income can be considered measures of performance, but they can have different implications. Other factors equal, relative income may be a

sign of how productive (in absolute terms) the individual is relative to the group-average. Other factors (including relative income) equal, group-position might be predictive of a worker's future payoff if he is employed in the tournament settings described by Lazear and Rosen [1981]. For example, if we imagine a modification to the model discussed in Gibbons and Waldman [2004] where there are an infinite number of type 1 jobs at the firm but only a finite number of type two (higher level) jobs. In such a world, the firm will give the type two jobs to workers with the highest (expected) productivity; these workers are generally the highest paid (highest-positioned) workers in earlier periods. In such a world, workers compete for one of a finite number of promotions to the type two jobs and a worker's group-position in the first period is highly correlated with the probability he gets such a job. There is some empirical evidence of this; in a study of workers at a large financial firm, Baker et al. [1994] finds that workers who were had higher group-position (compared to other employees at the firm in their occupation) were more likely to get the next promotion from that group and that promotions were crucial to maintain earnings growth.[2]

If group-position is informative of firm beliefs about worker productivity, ability, and future earnings or individual utility, it should be predictive of the individual's decision to leave the firm, the firm's decision to fire the worker, and the future earnings of that worker. However, the role of group-position in worker preferences and firm beliefs has different implications for the probability that worker leaves his job, the probability that he is fired, his future earnings if he leaves, and his future earnings if he stays. Although there are many versions of the following theories, and the may predictions vary with each specific version, Table 1.1 summarizes the basic economic intuition behind three broad classes theories: theories where group-position is a direct utility component of workers, theories where group-position is a signal of an unobserved worker/individual effect, and

[2]For the rest of this paper I focus on a worker's position in his comparison-groups. The primary reason for this is that once we control for group-position, the coefficients on relative income become statistically and/or practically insignificant.

Table 1.1: Models' Predicted Signs on Comparison-Group Position Coefficients

Theory	**Probability of Leaving**		**Future Earnings**	
	Voluntarily	**Involuntarily**	**Stay**	**Leave**
Utility Only	-	null	null	tradeoff
Signal of Individual Effects	null	null/-	+	+
Signal of Match Effects	-	null/-	+	-

Notes: The entries in this table represent the predicted signs on the coefficient in a regression of the dependent variable in the columns on a worker's comparison-group position, while controlling for his age, experience, tenure, absolute income, relative income, and the firm's size, the percentage employment change at the firm during that time, the firm's industry, and the firm's age. "-" means less than zero, "null" means equals 0, and "+" means greater than 0. "Tradeoff" refers to the fact that the worker's future earnings could be lower (or higher) than those at his current job, but that under this theory, the worker will be compensated with higher (or lower) group position. Otherwise, the worker would not voluntarily leave for the new job.

As a thought experiment, consider the situation where we have two identical workers at two almost identical firms—the only difference between the two workers and firms is the pay distribution within the peer comparison group. Because of the differences in the pay distribution, the workers will have different positions in their comparison-groups (see Figure 1.1). The questions I ask are: Is the worker with lower group position more likely to leave for a new firm? Is he more likely to be fired? Will his future earnings be lower if he stays? Will his future earnings also be lower if both workers switch jobs?

In the "Utility Only" theory (first row), workers enjoy higher group-position and so are less likely to leave as their position increases, but rank carries no information about productivity or future earnings. Thus, rank does not influence the firm's decision to fire the worker.

In the "Signal of Individual Effects" theory (second row), *ceteris paribus*, a higher group-position is a sign of a higher individual effect. This should be associated with higher future earnings for the worker on any job, and firms should be less likely to fire him (or at least, no more likely to fire him). Since the worker's individual effect follows him across all jobs, it should not change his decision to leave his current job.

In the "Signal of Match Effects" theory (third row), *ceteris paribus*, a higher group-position is a sign of a higher match effect. This should be associated with higher future earnings for the worker on his current job, but thanks to mean reversion, it should be associated with lower future earnings if the worker changes jobs. Workers with a higher match effect should be less likely to voluntarily leave their jobs, and firms should be less (or at least no more) likely to fire these workers.

theories where group-position is a signal of an unobserved match effect between a worker and a firm.

Row 1 of Table 1.1 contains the implications of higher group-position (other factors held equal) for the outcomes of leaving, being fired, and future earnings. In a world where group-position contains no information about individual or match effects, but people prefer higher position, then higher positioned workers should be less likely to leave their jobs. Since position is not informative about unobservable factors related to productivity, position should not predict whether firms fire a worker. Likewise, a worker's position should not be predictive of future earnings if he stays at the firm. However, workers who change jobs may face a tradeoff: they can accept positions where they have high position but are comparatively lower paid, or jobs where they have low position but are compensated with extra income for bearing this negative utility; workers pay a premium for their position, as in Frank [1984].

Row 2 of Table 1.1 contains the implications of higher group-position on labor market outcomes if position is a signal of individual effects. The categorization of all theories focused on "individual effects" into a single bin is an oversimplification. However, there is a common economic intuition to all of these models - an individual's fixed effect will not change when he changes firms. In a world where group-position informs workers only about their own individual effects, then position should not provide any information about how likely workers are to leave their job;[3] changing jobs will not alter it or improve his odds of success on a new job. For firing, there are two possibilities. The first is that firms are less likely to fire individuals with higher position, since these are the most productive workers. The second is that if firms pay workers according to their marginal productivity, there is no need for them to let go of a worker of any position - they simply lower or raise the pay to match the workers marginal productivity. Since an individual's fixed effect

[3]Unless workers, upon discovering they have high individual ability, try to leave for firms with other high ability workers. But I find no evidence of that.

will follow him wherever he goes, *ceteris paribus* today's higher positioned workers should have higher future earnings than today's lower positioned workers regardless of whether they stay or leave their current job.

Row 3 of Table 1.1 contains the implications of higher group-position on labor market outcomes if position is a signal of a match effect that exists between a worker and a firm. Again, there are a multitude of theories that feature "match effects"; categorizing them into a single bin is an oversimplification. However, there is again a common economic intuition to all of these models. In all match effect models, a higher match effect is better - it is associated with higher current pay, longer tenure, and higher future pay. If group-position is positively correlated with a worker's match effect draw, then higher positioned workers will also be less likely to leave a job. For firing, there again are two possibilities. The first is that firms are less likely to fire individuals with higher position, since these are the best-matched (and most productive) workers. The second is that if firms pay workers according to their marginal productivity, there is no need for them to let go a worker of any position - they simply lower or raise the pay to match the workers marginal productivity. However, higher match effects have different implications for future earnings than higher individual effects: if position signals a high match effect, then higher positioned workers should see larger future earnings if they stay on the job than they would if they leave and are forced to draw a new match effect which, on average, will not be so exceptional.

I explore which theories' predictions about individual behavior and outcomes are best supported by the data. First, I see whether a worker's position(s) in his comparison-group(s) is positively or negatively correlated with his decision to voluntarily leave a firm. Next, I look at how his position is correlated with the probability that the worker makes a transition to unemployment. Then, to further separate the theories, I look at a how worker's group-position in year t predicts the sum of his earnings over the next one, three, and five years. Finally, as an alternate test for a market for status, I examine whether

workers who experience larger increases in group-position receive smaller earnings changes.

1.3 Data: The LEHD

1.3.1 Overview of the Data

The Longitudinal Employer-Household Dynamics (LEHD) data provide worker and firm earnings information for every worker in all 50 states, with most states' records going back into the early 1990's. Three files form the core of the LEHD: the Employer Household file (EHF), the Employer Characteristics File (ECF), and the Individual Characteristics File (ICF). These files are described in greater detail in the appendix, but the most comprehensive description of the LEHD data is found in Abowd et al. [2009].

1.3.2 Advantages of the Data

The greatest advantage of LEHD is its wide scope: it contains roughly twenty years of quarterly earnings records for over 250 million unique worker records, which constitute the payroll of over 20 million unique firms - a truly representative sample of US workers and firms. This has several important implications.

With the inclusion of every worker, I can construct more natural and local sub-groups than those used in prior studies. As Kwon and Meyersson Milgrom [2014] note, most papers in the literature choose the comparison-group in an arbitrary fashion; previous choices have included residential neighbors, similarly aged and educated individuals in the same geographic region, or even similarly aged and educated individuals from different time periods. I look at a worker's percentile in the earnings distribution within his firm and his percentile in the earnings distribution of similarly educated workers at his firm.

Knowledge of the worker's firm and education allow me to control for the individual's position in multiple comparison-groups - a first among empirical studies. There are several reasons this is important. The first is that an individual may compare herself to multiple groups: the model of Kőszegi and Rabin [2006] provides one example of how an agent may make decisions based on multiple points of reference. The second reason is that the worker's positions in her comparison-groups may also give information about other unobserved factors in her decision even if her position in the group does not directly enter her utility function. Suppose, for example, suppose that we observe a lawyer's positions in the earnings distribution of her firm and similarly educated workers at her firm. The lawyer may not be concerned about her position in the firm distribution, which includes support staff; she derives utility based purely on where she falls in the earnings distribution of lawyers at the firm. To the econometrician, however, a high position in the firm distribution means the employee is probably a lawyer, whereas a lower paid worker is more likely a member of the support staff. A worker's position within the firm is therefore predictive of her behavior to extent that position correlates with occupation or power within the firm and these factors correlate with behavior.

Since the majority of workers in the LEHD have at least 15 years of coverage of their earnings information, I can follow workers and firms across time, enabling me to control for their idiosyncrasies in a way that is impossible without linked data.[4] Likewise, following firms over time allows me to control for time-varying firm variables as well as unobservable firm characteristics that may influence a worker's transition decision. The linked nature of the data also allows me to observe the long-term earnings or employment consequences for lower positioned individuals and whether they can be predicted by a worker's percentile in the earnings distribution of her current comparison-groups.

Finally, given that the earnings records in the LEHD are from an administrative source, there is less need for concern about measurement error in earnings. I have a worker's job-

[4]In particular, I can include worker fixed effects in my models.

specific earnings, not just his overall labor earnings, which is the ideal measure of earnings for the within-firm comparison-groups.

1.3.3 Sample of Workers and Firms

The sampling universe is defined as follows. I use observations from all workers at the largest 20% of firms between the years 1995-2005 in the states of Illinois, Missouri, and North Carolina. I use only worker-year observations from these states' largest firms so that a worker's position in his comparison-group has a meaningful interpretation; being in the 60th percentile in a group of five does not seem the same as being 60th percentile in a group of 100. The sampling frame in this paper is the sampling universe. There are 21 million unique workers and 230,000 unique firms in my sample, which includes 57% of all workers and 40% of all the worker-years contained in these three states. I select workers and firms from three states to keep the sample size tractable but still satisfy the LEHD restrictions on the results that can be published. I chose these three states because they have comparatively large financial sectors, which proves important to my later robustness tests on an "Ideal Sample."

1.4 Methodology

1.4.1 Assumptions

There are some common assumptions underlying all the regression models that I use to test the theories of coworker comparison-group position presented in Section 1.2. I use a discrete time framework. Individuals are born, acquire education, and then enter into the labor force. While in the labor force, they have one job at any given time period, where they receive a wage that is a function of age, experience, tenure, and education plus some

unobserved random variable, which is itself a combination of an unobserved individual effect, employer effect, match effect, and time-period specific shock. Individuals receive only wages (dollars per hour) as compensation. They use their wage information and information about their coworkers' wages to project their own future earnings, calculate their expected utility, and make job transition decisions. At the end of one time period, individuals choose to stay at their current job, move to a new job, or are forced by their employers to move to unemployment in the next period. I adapt this model to the empirical constraints of administrative data, which introduce difficulties that I can only resolve by making further assumptions. I discuss the assumptions common to all tests now; I will discuss other assumptions as they arise.

The first assumption is that individuals have only one job in any given time period. In the quarterly LEHD data, 91% of worker-quarters are from workers who are at only one job. Since having two jobs in a quarter will often arise naturally (e.g. from changing jobs or payroll overlap), it would seem that many the 7% of worker-quarters where the worker has two jobs still fit into the theoretical model described above. Although it could argued that the 2% of worker-quarters also fit in the model (e.g. there are 2 or more job changes in one quarter or extended payroll overlap), that seems less likely. However, in order to include all workers in the at-risk sample for my analysis, I keep all workers but maintain only one job record per year for each worker. In particular, I keep the job information from the worker's dominant job, that is, the job that gave them the highest earnings in the given calendar year. I will use the subscript i to denote individuals, t to denote the time period, and $J(i,t)$ to denote the firm where individual i has his dominant job in time t. $J(i,t)$ is a function that maps a worker i to the preferred firm J in his choice set - that is, the set of firms willing to hire him - in time t. This notation reflects the fact that the firm and its characteristics are not completely exogenous variables; they are at least partly chosen by the worker.

Despite having quarterly earnings and employment data, I use years as the unit of time in my analyses. Seasonality issues, worker-differences in vacation or sick leave used, or the differential arrival of paychecks across quarters[5] can cause the appearance of earnings changes when in fact individual pay is steady and hours worked are constant. Comparisons across larger units of time reduce this noise from the quarterly data. Thus, I use yearly units of time in my analysis, although I sometimes make use of quarterly information to construct other informative variables; I discuss this in more detail below.

The second assumption is that individuals receive only wages as compensation. I assume this out of necessity. Ideally, I could reduce earnings and benefits to some wage (dollars per unit of effort/time) which would allow for apples to apples compensation comparisons of individuals and the wage distribution of their comparison earnings group in a given time t. However, I do not observe benefits, and so I must also assume that benefits are equivalent across individuals, or at least distributed in a manner that is orthogonal to group-position.[6] I also do not observe hours worked - I only observe quarterly earnings in the LEHD. I treat earnings like wages by assuming that each individual works the same number of hours (say 40 per week) in each quarter. In this scenario, earnings on a job only vary when hours worked on the job vary. Holding hours fixed and equal across individuals, I can assign workers to positions in their groups according to their earnings; since I am dividing them by the same number of hours, my "wages" would preserve the positions generated by earnings.

The last assumption I make common to all specifications is that workers know their position in the firm or coworker group. I can't be certain how much workers know about their co-workers' earnings; many corporations in America forbid workers from discussing their pay despite federal laws that unequivocally prohibit such restrictions. Even if such

[5] I thank Erika McEntarfer for pointing this last item out to me.

[6] It would also not harm my results if benefits were increasing in worker earnings - in that case, better group-position is even more desirable. In contrast, if benefits were decreasing in earnings and/or group-position, workers would have conflicting incentives about pursuing higher position.

discussions were not discouraged by employers, there is no guarantee they would take place, and so it is not clear just how much workers know about where they fall in the earnings distribution of their comparison-group. Nonetheless, it seems likely that workers have some knowledge of their co-workers' incomes; discussions may occur, and even in their absence, the presence of a corner office or an instance of conspicuous consumption would provide workers with clues about absolute and relative incomes.

In an effort to test the importance of knowledge about co-worker earnings, I will also run my tests on an "Ideal Sample" (see below) that includes workers in industries where there exists a high degree of information about other workers' pay. I choose workers at firms in the finance and technology sectors - North American Industry Classification System (NAICS) sectors 51 and 52. Indiviglio [2011] observes that financial workers have unusually high knowledge of their co-workers' pay. There is some quantitative evidence to back this up; the higher frequency (relative to their total employment) with which salaries are reported in finance and technology sectors on Glassdoor.com, suggests that in these industries there is a greater availability of and interest in information about the earnings of one's fellow workers.

1.4.2 Construction of Key Variables

The primary variables of interest that I use in my models are an individual's percentiles in the earnings distributions of his coworker comparison-groups. There are two coworker comparison-groups that I use: workers at the same firm, and workers at the same firm who have the same education as the individual.

In the world of my theoretical model, workers would compare his wages to the wages of others in his group at the firm that year. However, I don't observe wages, and workers may change jobs mid-year. If I compared actual worker on-the-job earnings to assign workers positions in their groups, I would be comparing the job earnings of an employee

who is at the firm for 3 months in a year with the earnings of worker who is at the firm for an entire year. To remedy this, I first obtain an annualized measure of the worker's average full quarter earnings from a job. An employee i is said to have worked a full quarter on a job at firm $J(i,t)$ in quarter t when the employee was employed at that same firm in quarters $t-1$, t, and $t+1$. Individual i's "annualized" (counterfactual) earnings at firm $J(i,t)$ in year t are his average full quarter earnings on the job in t multiplied by four. The advantage of using full quarter earnings is that I do not need to worry that an individual's quarterly earnings were affected by his beginning or ending the job in that quarter; the existence of earnings records in $t-1$ and $t+1$ strongly suggest he was employed at the firm for all of quarter t. Since I assume that workers are full time employees, the most accurate measure of his wages comes when he has worked all weeks in the quarter: at that point, dividing his earnings by the (fixed) number of hours worked in the quarter gives me something like wages. If no full quarter earnings exist for an individual's primary job, I instead take the average quarterly earnings of the employee at that job and multiply it by four.

Once I have a workers' annualized earnings, I construct a distribution of those earnings for every worker at a firm $J(i,t)$ in year t. I also construct firm-year earnings distributions for all workers of a given level of education. I use four levels of education to describe workers: workers without a HS diploma, workers with a HS diploma, workers with some college, and workers with a bachelors or advanced degree. I then find a worker's percentiles in the firm-education-year distribution to which he belongs; thus a worker has a position in two distributions: the firm distribution and the within firm distribution for other workers of his same education level.[7] I use a worker's percentiles to predict job transition, future earnings, and search behavior.

[7]Since I only observe education for 12% of workers, I must use imputed values of education for the other 88% of workers in my sample. My results will present the average coefficients and standard errors from regressions run on data with ten implicates of education. I discuss the multiple imputation methodology more in the appendix.

1.4.3 Comparison-Group Position and the Probability of Leaving

In order to see what theories about comparison-group position are best supported by the data, I will first look at how a worker's position predicts his probability of leaving for a new job. A positive correlation between position and the probability of leaving a job would suggest that workers dislike high status or see it as a signal of an earnings dead end; this would be a categorical rejection of the three models represented in Table 1.1. A negative correlation between a worker's percentile in the earnings distribution of his comparison-group and probability of leaving would suggest that workers prefer directly prefer position or that it is a signal of future monetary rewards. I estimated the linear probability model:[8]

$$\begin{aligned} Prob(JobChange_{(i,t,t+1)}) &= \beta_0 + \theta_i + \psi_{J(i,t)} + f(Exper_{i,t}) + f(Age_{i,t}) \\ &+ \beta_3 f(TotalEarn_{i,t}) + \beta_4 FirmSize_{J(i,t),t} + \beta_5 FirmAvgEarnings_{J(i,t),t} \\ &+ \beta_6 FirmLagEmpChange_{J(i,t),(t-1,t)} + \beta_7 FirmLeadEmpChange_{J(i,t),(t,t+1)} \\ &+ \beta_8 f(Tenure_{i,J(i,t),t}) + \beta_9 f(AnnualJobEarn_{i,J(i,t),t}) \\ &+ \gamma_1 FirmEarnRatio_{i,J(i,t),t} + \gamma_2 FirmEducEarnRatio_{i,J(i,t),t} \\ &+ \delta_1 FirmPerc_{i,J(i,t),t} + \delta_2 FirmEducPerc_{i,J(i,t),t} \end{aligned} \tag{1.1}$$

In words, I am regressing a worker $i's$ decision to leave his year t dominant job at his chosen firm $J(i,t)$ and work at a new firm $J(i,t+1)$ between years t and $t+1$ on worker and firm fixed effects, a cubic function of the worker's experience in year t, a cubic function of the worker's age in year t, a cubic function of the worker's total (all jobs)

[8] A logit model seems a more natural choice, given that we have a binary dependent variable. However, the linear probability model allows the inclusion of many individual and firm fixed effects. I estimate all individual and employer effects using the formulas in Abowd et al. [1999b] with the methods described in Abowd et al. [1999a]. Firm effects for the largest 5,000 firms are exact, while those for smaller firms are approximated based on their firm-size decile in year t and their NAICS sector. The largest 5,000 firms account for 46% of the worker-year observations in my sample.

yearly earnings, the size of the firm, the average earnings of the firm, the employment changes at the firm between $t-1$ and t, the employment change between t and $t+1$, a cubic function of the worker's tenure, a cubic function of the worker's job earnings in year t, and finally, his position in the firm earnings distribution and the earnings distribution of coworkers with his same level of education. $FirmPerc_{i,J(i,t),t}$ and $FirmEducPerc_{i,J(i,t),t}$ are the variables of interest; these are i's position in the earnings distribution of coworkers and similarly educated coworkers at firm $J(i,t)$ in year t; they capture the position of the worker in his comparison-groups.

The controls address various confounding factors. Since I am interested in the impact of the worker's position in his comparison-group, I need to control for the workers on-the-job earnings. Otherwise, workers' positions in their comparison-groups will be correlated with income and the coefficient on the position variables will reflect the impact of income. In a related way, I must control for the worker's age, experience, and tenure. Younger, less experienced, and lower-tenured; workers usually leave their job at higher rates.[9] These variables also indirectly relate to position in the comparison-group; you can imagine that being in the 30th percentile of the income distributions has very different implications for young workers with little experience or tenure as opposed to older workers with greater tenure. Thus, we want to control for age, experience and tenure so that their impact does not distort the coefficients on the group-position variables. Since the effects of these variables, particularly tenure, are non-linear, I use cubic functions of them. Although firm size and average earnings should not directly relate to a worker's position in the within firm comparison-groups, they likely predict systematic differences in turnover and so they should be included as controls. Firm-size (employment) changes must necessarily be correlated with *some* individuals' job transition decisions. One particular concern would be that when a firm shrinks it fires/forces out all workers with low relative income. If we

[9]The importance of tenure to a worker's decision to leave his job has been well documented, e.g. by Topel and Ward [1992]; see Farber [1999] for a more exhaustive review of the role of tenure in job changing.

exclude size changes as a control in this situation, we would find a spurious correlation between relative earnings and a worker's job transition decision. I include both a lead and lagged employment term. The lead term (changes from t to $t+1$) is vital because this contains employment changes at the firm in the same time period that the individual's dependent variable is determined. The inclusion of the lagged employment change (from $t-1$ to t) is to account for the fact that hiring or firing at a firm late in one calendar year (e.g. in November and December) may be related to an employees decision to leave a few months later (in January or February or February of the next year) - it may also reflect unobservable changes in firm policy that have a lagged effect on worker turnover. Finally, the worker and firm fixed effects control for time invariant unobserved factors that may affect the worker's decision to leave the job and/or correlate with his position in the relative income distribution (e.g. the individual is a part time worker).

To better understand the identifying assumptions, consider the case where only the individual's percentile in the earnings distribution of similarly educated workers at the firm is changing. By holding age, experience, tenure, absolute income, and relative income (compared to the group average), it is reasonable to think we are holding fixed the individual's absolute productivity, (as measured by income), its deviation from the group average productivity (relative income), and observable factors that might contribute to his productivity.[10] Controlling for these things, the coefficient on worker's percentile in her firm earnings distribution tells us the degree to which changing her position in the coworker comparison-group alters the probability that she leaves her job.

I also include another position variable: the worker's percentile in the firm earnings distribution. Since individuals tend to compare themselves to others who are most like them,[11] an individual's position in the entire firm seems like a less useful predictor of his

[10] My inclusion of individual fixed effects would also control for unchanging observables like education, sex, and race as well as the individual unobserved component.

[11] The findings of Card et al. [2012] are one example of the evidence for this. For an in-depth discussion of how social comparison-groups are formed and their behavior patterns, see Tajfel [2010].

behavior or outcomes than his position in the distribution of similarly educated workers at the firm. Nonetheless, an individual's position in the earnings distribution of the firm might be construed as a proxy for an individual's importance within the firm; this might have different implications for his decision to leave than his position in the group of workers who most closely resemble him.

With both the percentile variables, an increase in the worker's percentile in the earnings distribution of similarly educated coworkers while holding fixed his position in the earnings distribution of his firm has the following interpretation: a worker's position in the group of workers who are like him is increasing, even as the worker's position the firm remains constant. Put differently, even as the value of the individual relative to the rest of his education group has increased, the relative value of that individual to the firm is unchanged (likely because the firm has lowered the value of workers of that type).

In practice, there is a substantial degree of collinearity between a worker's percentile in both groups. The large sample size prevents the standard errors of my coefficients from being too large, but as a robustness check, I have run all the regressions with only a single comparison-group variable. Whether I use the worker's position in the firm's earning distribution or his position in the distribution of similarly educated workers at the firm, the signs and the significance of the group percentile variables remain unchanged. Predictably, dropping one of the group percentile variables strongly increases the magnitude of the coefficient of the other.

A positive coefficient on the percentile variables would indicate that high position workers are more likely to leave for a new job. This would repudiate every theory in Table 1.1. A negative coefficient on the variables would suggest that workers like higher comparison-group position either for its own sake or because it is a sign of high individual or match effects.

Ideally, one would like to measure whether the separation was voluntary or involuntary. While this is not directly feasible in my data, I can use various observable criteria to categorize a transition as either voluntary or involuntary:

1. Whether a worker holds any job in a year (or for how many quarters in each year he is observed working at least one job)
2. The worker's year to year (or quarter to quarter) earnings changes
3. The employment change at a worker's firm. If I see significant cuts in the number of workers at a firm I could treat all leavers of that firm as involuntary leavers (as a robustness check)

For my initial estimates in columns (1) and (2), I define a job to job transition as voluntary if the person switches employers and has records of employment in each quarter of the years t and $t+1$. If the individual has one or more quarters in those years where I do not observe him as unemployed, I classify the transition as involuntary. As a robustness check, I also categorize as involuntary any job transition where the individual's earnings declined by more than 10% or the firm of the individual's time t dominant job experienced an employment contraction of 30% or more between $(t-1, t)$ or $(t, t+1)$ - a threshold commonly used in the literature for determining whether a firm is undergoing mass layoffs. My results are unchanged regardless of the definition used.

1.4.4 Comparison-Group Position and the Probability of Being Fired

The next step to identifying the theories is to determine the extent to which a worker's comparison-group position is informative about whether or not he will be fired.[12] If position conveys no information about individual or match effects (or if firms simply adjust pay to workers instead of firing them), then it should not be predictive of whether or not a worker is fired; this would validate any of the three rows in Table 1.1. If, on the

[12] See the previous two paragraphs for my definition of an involuntary separation.

other hand, it does contain information about the individual or match effects, then we would assume that workers with higher group-position are less likely to be fired (rows 2 and 3 of Table 1.1). For this test, I estimate the linear probability model below:

$$
\begin{aligned}
Prob(Fired_{(i,t,t+1)}) = \beta_0 + \theta_i + \psi_{J(i,t)} + f(Exper_{i,t}) + f(Age_{i,t}) + \beta_3 f(TotalEarn_{i,t}) \\
+ \beta_4 FirmSize_{J(i,t),t} + \beta_5 FirmAvgEarnings_{J(i,t),t} \\
+ \beta_6 FirmLagEmpChange_{J(i,t),(t-1,t)} + \beta_7 FirmLeadEmpchange_{J(i,t),(t,t+1)} \\
+ \beta_8 f(Tenure_{i,J(i,t),t}) + \beta_9 f(AnnualJobEarn_{i,J(i,t),t}) \\
+ \gamma_1 FirmEarnRatio_{i,J(i,t),t} + \gamma_2 FirmEducEarnRatio_{i,J(i,t),t} \\
+ \delta_1 FirmPerc_{i,J(i,t),t} + \delta_2 FirmEducPerc_{i,J(i,t),t}
\end{aligned}
\tag{1.2}
$$

This regression has the same independent variables as the regression in (1) for similar reasons. Again, the coefficients on the percentile variables are of primary interest. If these coefficients lack a sign or statistical significance, it would suggest that these variables are not informative about who gets fired. This would suggest position is not informative of individual or match effects, or the information obtained is not informative enough to cause firms to incur the costs that come from firing workers.[13] A negative sign would suggest that the individual's higher position signals makes him less likely to be fired, which could be a signal of a higher individual or match effect. A positive sign would be a repudiation of all three theories in Table 1.1.

1.4.5 Comparison-Group Position and Future Earnings

The next test to rule out theories will examine how a worker's comparison-group position helps predict his level of future earnings. If position is correlated with lower future

[13] Although as noted in Section 1.2, firms have other options, such as lowering a worker's pay.

earnings, this could signify that position is noisy and there is mean reversion. It might also suggest a preference for higher position, where workers who have higher position are satisfied with smaller raises as long as they maintain their status (e.g. Frank [1984]) - row 1 of Table 1.1. Conversely, if group-position is correlated with higher future earnings, then workers may only prefer position for its pecuniary implications - but what implications exactly? If higher position signals higher earnings, is it indicative of a high individual effect (row 2 of Table 1.1) or a high match effect (row 3 of Table 1.1)? We would expect a high individual effect to improve the worker's future earnings regardless of which job he is at, whereas benefits caused by a high match effect should dissipate if the individual changes jobs. To test this, I regress a the natural log of a worker's cumulative earnings one, three, and five years out from time t on his time t group-position:

$$\begin{aligned}
E[ln(CumEarn_{(i,t,t+X)})] &= \beta_0 + \theta_i + \psi_{J(i,t)} + f(Exper_{i,t}) + f(Age_{i,t}) \\
&+ \beta_3 f(TotalEarn_{i,t}) + \beta_4 FirmSize_{J(i,t),t} + \beta_5 FirmAvgEarnings_{J(i,t),t} \\
&+ \beta_6 FirmLagEmpChange_{J(i,t),(t-1,t)} + \beta_7 FirmLeadEmpchange_{J(i,t),(t,t+1)} \\
&+ \beta_8 f(Tenure_{i,J(i,t),t}) + \beta_9 f(AnnualJobEarn_{i,J(i,t),t}) \\
&+ \gamma_1 FirmEarnRatio_{i,J(i,t),t} + \gamma_2 FirmEducEarnRatio_{i,J(i,t),t} \\
&+ \delta_1 FirmPerc_{i,J(i,t),t} + \delta_2 FirmEducPerc_{i,J(i,t),t} + \delta_3 JobChange_{(i,t,t+1)} \\
&+ \delta_4 FirmPerc_{i,J(i,t),t} * JobChange_{(i,t,t+1)} \\
&+ \delta_5 FirmEducPerc_{i,J(i,t),t} * JobChange_{(i,t,t+1)}
\end{aligned} \tag{1.3}$$

In words, I am regressing the natural log of a worker i's total future earnings in year $t+X$ on worker and firm fixed effects, a cubic function of the worker's experience in year t, the worker's total (all jobs) yearly earnings, the size of the firm, the average earnings of the firm, the employment changes at the firm between $t-1$ and t and t and $t+1$, a cubic function of the worker's tenure, the log of the worker's job earnings in year t, his percentile in the firm earnings distribution, and his percentile in the earnings distribution of workers

at the firm with his level of education. $FirmPerc_{i,J(i,t),t}$ and $FirmEducPerc_{i,J(i,t),t}$ are the variables of interest, which I will use as measures of i's position in his comparison-group. Their interaction with the decision to change jobs will help me to test whether position indicates a worker's individual effect or his match effect. If group-position indicates a higher than average match effect (row 3 of Table 1.1), then if individuals leave their jobs they should, on average, have a smaller match effect at their next job. This suggests their earnings gains from job changing are relatively smaller than those of low position workers, who on average should have higher match effects on their new jobs. In contrast, if position is only indicative of an individual effect (row 2 of Table 1.1), then the impact of job changing on future earnings should not differ by position.

Age, experience, and tenure are all related to current worker earnings, and these variables correlate strongly with their future values (age linearly so), which are in turn tied to future earnings. These factors also are tied to relative earnings in the sense that without holding them fixed, one's position in the earnings distribution might have very different meanings. Likewise, a firm's size and earnings can be reasonably expected to be correlated with a worker's career path (and therefore future earnings), which we might also expect to be related to the worker's relative earnings. Firm employment change tends to be related to earnings growth (particularly next year's earnings). The worker's current job and total earnings also correlated with future earnings; job earnings will correlate with future income and relative income, and total earnings might as well. High total earnings might signal the worker has high productivity this year, which could lead to promotions or other forms of earnings growth. Firm and worker fixed effects control for (time invariant) unobservables that might related to both the percentile variables and future earnings.

1.4.6 The Market for Status

The third test is another attempt to test the a key hypothesis of Frank [1984]: that status acts as a form of compensation. In essence, this is another attempt to test whether we are in the row 1 of Table 1.1 (prestige, equity) or the information hypotheses in rows 2 or 3. Suppose that status has a social value independent of the monetary value. Then when a worker's status decreases, the firm must offer him more money to keep him from quitting. Likewise, a worker should be willing to accept lower raises that come with growth in status. To test for a tradeoff between status and monetary compensation, I regress the percentage earnings increase obtained by workers between years t and $t+1$ on their position increases/decreases. My specification is:

$$\begin{aligned}
E(PctEarnChange_{(i,t,t+1)}) &= \beta_0 + \theta_i + \psi_{J(i,t)} + \beta_1 Exper_{i,t} + \beta_2 f(Age_{i,t}) \\
&+ \beta_3 f(TotalEarn_{i,t}) + \beta_4 FirmSize_{J(i,t),t} + \beta_5 FirmAvgEarnings_{J(i,t),t} \\
&+ \beta_6 FirmLagEmpChange_{J(i,t),(t-1,t)} + \beta_7 FirmLeadEmpchange_{J(i,t),(t,t+1)} \\
&+ \beta_8 f(Tenure_{i,J(i,t),t}) + \beta_9 f(AnnualJobEarn_{i,J(i,t),t}) \\
&+ \gamma_1 FirmEarnRatio_{i,J(i,t),t} + \gamma_2 FirmEducEarnRatio_{i,J(i,t),t} \\
&+ \delta_1 FirmPerc_{i,J(i,t),t} + \delta_2 FirmEducPerc_{i,J(i,t),t} \\
&+ \delta_3 FirmPercChange_{(i,t,t+1)} + \delta_4 FirmEducPercChange_{(i,t,t+1)}
\end{aligned} \tag{1.4}$$

Again, I include all the controls I had before, since I believe these relate to the growth/decline in a worker's group-position and the change in her earnings. $FirmPerc_{i,J(i,t),t}$, $FirmEducPerc_{i,J(i,t),t}$, $FirmPercChange_{(i,t,t+1)}$, and $FirmEducPercChange_{(i,t,t+1)}$ are the variables of interest. A negative coefficient on these variables would support the hypothesis that comparison-group position substitutes for monetary compensation (row 1 of Table 1.1). In contrast, a positive coefficient of these variables would provide additional evidence against the theory in row 1 of Table 1.1 by suggesting that workers face no tradeoffs between position changes and earnings changes.

1.5 Results

1.5.1 Comparison-Group Position and the Probability of Leaving

Table 1.2 shows the relationship between comparison-group position and the likelihood of changing jobs.

Table 1.2: Worker Transition-Rate by Firm Earnings and Worker Earnings Type

Firm Earnings Decile	Worker Earnings Decile 1	2	3	4	5	6	7	8	9	10
1	0.316	0.296	0.283	0.273	0.257	0.253	0.252	0.242	0.227	0.2
2	0.341	0.316	0.296	0.282	0.27	0.261	0.252	0.244	0.236	0.206
3	0.358	0.325	0.302	0.286	0.271	0.262	0.25	0.238	0.226	0.192
4	0.367	0.322	0.295	0.272	0.254	0.239	0.223	0.208	0.189	0.161
5	0.361	0.31	0.274	0.246	0.223	0.206	0.192	0.177	0.165	0.143
6	0.336	0.278	0.237	0.207	0.185	0.169	0.157	0.141	0.129	0.111
7	0.325	0.248	0.207	0.178	0.157	0.146	0.136	0.125	0.112	0.0944
8	0.32	0.235	0.191	0.163	0.145	0.132	0.124	0.114	0.103	0.088
9	0.297	0.21	0.168	0.145	0.13	0.119	0.112	0.106	0.0974	0.0827
10	0.295	0.203	0.166	0.148	0.135	0.126	0.12	0.112	0.103	0.0866

≥ .3
≥ .25
≥ .2
≥ .15
≥ .1
< .1

Notes: This table shows the fraction of workers employed in each (within firm earnings decile, between firm average earnings decile) cell who change jobs the following year. As firm average pay increases (moving from left to right across the columns) or as worker pay increases (moving from top to bottom of the rows), transition rates fall. The sample comes from all job-year observations from 1995-2005 from a worker's highest paying job in a given year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contained 21 million workers and 120 million worker-years.

Table 1.2 shows that workers are less likely to change jobs as they become higher paid in both absolute and relative terms. The rows categorize firms by the average earnings of workers at those firms; the first row contains only workers in the 10% of firms with the lowest average earnings and the 10th row contains only workers in the 10% of firms with the highest average earnings. The columns categorize where the worker falls within his firm earnings distribution: the first column is for workers who are in the bottom 10% of the earnings distribution of their firm. The 10th column is the workers in the top 10% of their firm earnings distribution. So a worker in (2,7) works at a firm in the 20th percentile (a firm with lower mean wages than 80% of other firms) and she earns more than 70% of her coworkers at that firm. The value of the cell in a (row,column) is the percentage of workers of each (between firm earnings decile, within firm earnings decile) cell who have a job to job transition in the next year.

There is a strong negative relationship between a worker's percentile in the earnings' distributions of the coworker comparison-group and the probability that he leaves for a new dominant job. As average firm earnings go from lowest (firm decile 1) to highest (firm decile 10), workers quit at lower rates. Within firms of a given average earnings decile, as you increase workers' percentile in the distribution (moving from worker decile 1 to worker decile 10) they are also less likely to change jobs.[14]

In Table 1.2, relatively higher firm earnings and relatively higher individual earnings are associated with lower turnover. Recall, that in terms of Table 1.1, this would suggest that workers either have a direct preference for status or that it is positively associated with future earnings. However, in Table 1.2, the higher relative earnings of workers and firms are also associated with higher absolute earnings. To isolate the impact of comparison-group position from other variables, I estimate equation (1.1). The results of the regression are in Table 1.3.

[14]Although not shown here, the same is true even when I look within firm-education-year groups.

Table 1.3: Probability a Worker Leaves Given Her Comparison-Group Position

Variable	**Leave**	**Leave with Fixed Effects**	**Leave with FE, Earn and Emp Criteria**
	(1)	**(2)**	**(3)**
Firm Year Quantile	-0.102***	-0.221***	-0.339***
	(0.00057)	(0.000848)	(0.000684)
	[0.436]	[0.461]	[0.55]
Firm Education Year Quantile	-0.0438***	-0.0211***	-0.0155***
	(0.00071)	(0.000929)	(0.000687)
	[0.626]	[0.56]	[0.564]
R-Squared:	0.123	0.372	0.317
Controls:			
Person Fixed Effects		X	X
Firm Fixed Effects	X	X	X
Industry Effects	X	X	X
Firm Size Effects	X	X	X
*Industry*Size Effects*	X	X	X
Lead Emp Change Pct	X	X	X
Lag Emp Change Pct	X	X	X
Firm Age	X	X	X
Age (Cubic)	X	X	X
Experience (Cubic)	X	X	X
Tenure (Cubic)	X	X	X
All Earnings (Cubic)	X	X	X
Job Earnings (Cubic)	X	X	X
Firm Earnings Ratio	X	X	X
Firm Educ Earn Ratio	X	X	X

Notes: This table shows the results of an OLS regression of the probability a worker leaves on his comparison-group position variables and other controls. Column (2) adds worker fixed effects to the regression performed in Column (1). The sample for these regressions comes from all job-year observations in 1995-2005 at a worker's highest paying job in that year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contains 21 million workers and 120 million worker-years.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. The standard errors, which are corrected for variation due to the imputation of missing education variables, are in parentheses. The missingness ratio, which gives the fraction of variability due to imputation, is in brackets.

The regressions in columns (1) and (2) control for the worker's position in both comparison-groups. I interpret the coefficient $-0.221(+/-0.00008)$ on the firm-year percentile in column (2) as follows: suppose we have two workers who are observationally equivalent (and have the same unobserved individual and firm effects) except for their percentile in the firm earnings distribution: the first worker is in the Xth percentile and the second worker is in the Yth percentile. The difference in the probability the two workers leave is $0.221 * (X - Y)/100 = .11$; that is, an increase in a worker's firm position by 50 percentiles is associated with an 11 percentage point decrease in the probability that the worker leaves the firm. Similarly, I interpret the coefficient $-0.0211(+/-0.0009)$ on the firm-education-year percentile as follows: suppose we have two workers equivalent except that worker one is from the 25th percentile in the firm-education-earnings distribution and worker two is in the 75th percentile. Then worker two has a $0.0211 * .5 = .01$ greater probability of leaving the firm in the next year.

One concern might be that I am miscategorizing a worker as leaving when he is actually fired. As a further test, I run the same regression as equation (1.2) but with another definition of whether or not a worker was fired: if a transition leads to at least one quarter of no recorded employment, or leads to a decrease in the worker's annual earnings by 10% or more, or originates from a firm in a year where the firm shrank by 30 percent or more,[15] then I categorize the transition as involuntary. This should help capture shorter unemployment transitions. The results of this regression are in column 3 of Table 1.3; even with the necessary reduction in the number of "voluntary" transitions, the negative correlation between comparison-group position and the probability of making a job to job transition remains significant. I may still be miscategorizing certain involuntary transitions as voluntary, in which cases the absolute magnitude of my coefficients would be biased upwards.

[15] This is often cited as a cutoff point in the mass-layoff literature.

One concern might be that this specification models the impact of comparison-group position as linear. However, when I assign workers to a comparison-group decile in the earnings distribution of the firm or similarly educated workers at the firm, the coefficients estimated on the deciles are also monotonic and approximately linear. Therefore, for simplicity, I use the (linear) percentile specification of comparison-group position in the rest of the tests in this paper.

These results in Table 1.3 support the notion that workers find higher group-position desirable, either because they directly value it or because it indicates an individual has a higher unobserved individual or match effect.

1.5.2 Comparison-Group Position and the Probability of Being Fired

If comparison-group position is positively correlated with unobserved individual or match effects, firms should be less likely to fire workers with high group position - or at least no more likely to fire those workers. To test this, I estimate equation (1.2). The regression results are in Table 1.4.

The specification without individual fixed effects is in column (1); workers with higher group-position are slightly more likely to get fired. This would refute all of the theories in Table 1.1. However, when the specifications include individual fixed effects (column 2), lower position in the firm is still associated with a lower rate of transition to unemployment, but the results are not statistically significant for the firm education-group. This result contradicts the idea that the negative correlation between group-position and job-job transition is driven by a correlation between lower position and erroneously categorized transitions to unemployment.

As a further test, I run the same regression as (1.2) but with the stronger criteria (based on individual employment records, firm employment changes, and individual earnings

Table 1.4: Probability a Worker Is Fired Given His Comparison-Group Position

Variable	**Fired**	**Fired with Fixed Effects**	**Fired with FE, Earn and Emp Criteria**
	(1)	**(2)**	**(3)**
Firm Year Quantile	-0.182***	-0.12***	-0.00118
	(0.000883)	(0.000964)	(0.00108)
	[0.751]	[0.637]	[0.613]
Firm Education Year Quantile	0.0143***	0.00137	-0.00428*
	(0.00109)	(0.00112)	(0.0012)
	[0.83]	[0.736]	[0.693]
R-Squared:	0.24	0.554	0.497
Controls:			
Person Fixed Effects		X	X
Firm Fixed Effects	X	X	X
Industry Effects	X	X	X
Firm Size Effects	X	X	X
*Industry*Size Effects*	X	X	X
Lead Emp Change Pct	X	X	X
Lag Emp Change Pct	X	X	X
Firm Age	X	X	X
Age (Cubic)	X	X	X
Experience (Cubic)	X	X	X
Tenure (Cubic)	X	X	X
All Earnings (Cubic)	X	X	X
Job Earnings (Cubic)	X	X	X
Firm Earnings Ratio	X	X	X
Firm Educ Earn Ratio	X	X	X

Notes: This table shows the results of an OLS regression of the probability a worker is involuntarily separated on his comparison-group position variables and other controls. Column (2) adds worker fixed effects to the regression performed in Column (1). In Column (3), the dependent variable was also coded as "involuntarily separated" if there was a mass layoff at the worker's period t firm or if his annualized earnings dropped by 10% or morewhen he changed jobs. The sample for these regressions comes from all job-year observations in 1995-2005 at a worker's highest paying job in that year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contains 21 million workers and 120 million worker-years.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. The standard errors, which are corrected for variation due to the imputation of missing education variables, are in parentheses. The missingness ratio, which gives the fraction of variability due to imputation, is in brackets.

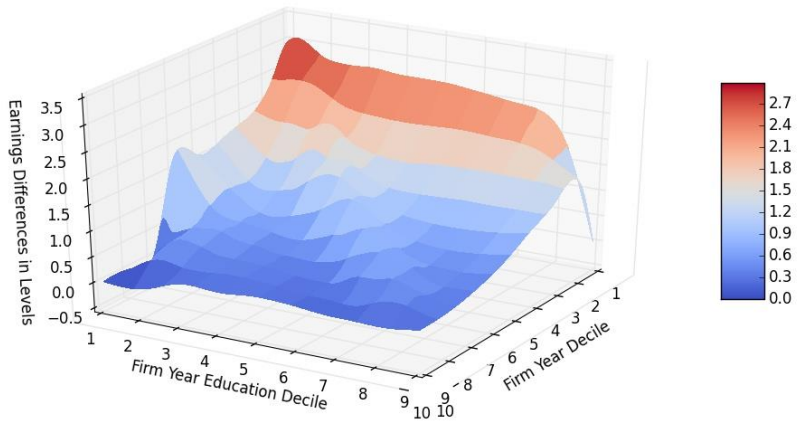
changes) discussed above; by construction this will define more transitions as involuntary. The results are depicted in column 3 of Table 1.4. This time, the correlation between firm education-group position and the probability of getting fired is negative - however, it is not practically significant. Other factors constant, a worker in the lowest possible percentile (0) is only .4 percentage points more likely to leave his job than a worker in the highest possible percentile (1). The absence of a significant coefficient on the percentile variables in this regression supports all three theories depicted in Table 1.1. It is likely the coefficients in column (3) are biased towards zero; if a (low group-position) worker suspects that he is likely to be fired (or that he is first in line when firing takes place), he might instead look for a job, in which case I may categorize him as moving "voluntarily". However, even if the percentile coefficients doubled or tripled in size it would not be practically significant.

1.5.3 Comparison-Group Position and Future Earnings

The negative correlation between a worker's comparison-group position and the probability that a worker leaves his job suggests that workers have a preference for higher positions. To test whether the observed negative correlation between group-position and job-changing is due to a direct preference for group-position or the financial rewards associated with higher group-position, I estimate the correlation between comparison-group position and future earnings. A positive correlation suggests that a model where workers have only financial motives would sufficiently explain the data. A negative correlation suggests that workers enjoy higher position for its own sake and are willing to accept lower raises (presumably as long as their high position is maintained).

If workers prefer higher group-position because it is positively correlated with future earnings, do workers' positions signal their individual effects or their match effects? If high position is primarily the result of a high individual effect, then it should indicate higher future earnings regardless of whether he stays at or leaves his current job. If a high group-position is a sign of a high match effect term, then we should expect relatively lower future earnings if the worker leaves for a new job - a form of mean reversion. Figure 1.2 suggests that match effects are better fit the data.

Figure 1.2: Group-Position, Job Leaving, and Future Earnings Differences



Notes: Each (Firm Year Education Decile, Firm Year Decile) grid point in this figure shows the difference in average earning levels between job-leavers and job-stayers who fall into the (X,Y) pair of comparison-group deciles. The earnings levels of zero in the corners are due to the suppression of thinly populated cells. The colors are used to simplify the reading of the values on the z-axis; they give no information beyond the earnings difference (z-axis value). This sample comes from all job-year observations from 1995-2005 from a worker's highest paying job in a given year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contained 21 million workers and 120 million worker-years.

This figure shows that while all job leavers have higher next-year-earnings than job stayers, the differences in earnings are the greatest for the job leavers who are at the lowest percentiles in both their firm's earnings distribution and earnings distribution of similarly educated coworkers. However, to more rigorously examine the implications of

comparison-group position on future earnings, I estimate equation (1.3). The regression results are in Table 1.5.

The main result from estimating (1.3) with worker fixed effects (columns 4, 5, and 6) is that higher comparison-group position is associated with higher future earnings. Consider column (4). I interpret the coefficient $1.61(+/-0.00603)$ on the comparison-group percentiles as follows: suppose we have two workers in period t, John and Bill, at different firms who are identical to each other in everything except their percentile in the firm-year earnings distribution; John's is at the Xth percentile in his firm's earning distribution whereas Bill is at the Yth percentile in his firm's earnings distribution. If neither of them switch jobs in the next year, then the average difference in earnings levels between John and Bill next year is $1.61 * (X - Y)$.[16] If both of them switch jobs the next year, the average difference in earnings levels is $1.61*(X-Y)-1.2*(X-Y) = .41*(X-Y)$.

The interpretation for column (5) is slightly different. Here, consider John and Bill, where this time the only difference between them is that John is in the Xth percentile of the earnings distribution of similarly educated workers at his firm, whereas Bill is in the Yth percentile of the firm-education-earnings distribution. If neither worker switches jobs in year t (but either or both could possibly switch jobs in $t+1$ or $t+2$), then the difference in John and Bill's level of the cumulative total earnings of each worker over the years $t+1$, $t+2$, $t+3$ will be $0.0284 * (X - Y)$. If both of them switch jobs between t and $t+1$ (and possibly switch again in $t+1$ or $t+2$), the difference between John and Bill in their levels of three-year cumulative earnings is $0.0284*(X-Y)-0.0927*(X-Y) = -0.0643*(X-Y)$.

[16] That is, John's earnings are $e^{[1.61 \ * \ (X-Y)]} * 100\%$ of Bill's earnings.

Table 1.5: Log of a Worker's Total Earnings over the Next 1, 3, or 5 Years

Variable	**Year 1**	**Year 3**	**Year 5**	**Year 1**	**Year 3**	**Year 5**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Firm Year Quantile	2.86***	2.18***	1.98***	1.61***	0.691***	0.406***
	(0.00639)	(0.00494)	(0.00477)	(0.00603)	(0.00329)	(0.00289)
	[0.763]	[0.824]	[0.873]	[0.522]	[0.445]	[0.635]
Firm Education Year	-0.167***	-0.298***	-0.345***	0.0742***	0.0284**	0.0236**
Quantile	(0.008)	(0.00617)	(0.00592)	(0.00615)	(0.00334)	(0.00309)
	[0.845]	[0.885]	[0.916]	[0.553]	[0.478]	[0.69]
Job Change	2.56***	1.53***	1.22***	1.44***	0.567***	0.321***
	(0.00107)	(0.000715)	(0.000587)	(0.00109)	(0.000641)	(0.000457)
	[0.00968]	[0.019]	[0.0205]	[0.0145]	[0.00988]	[0.00998]
FirmQuant*Change	-1.88***	-1.17***	-0.937***	-1.2***	-0.496***	-0.284***
	(0.00755)	(0.00545)	(0.00481)	(0.00833)	(0.00472)	(0.00337)
	[0.109]	[0.242]	[0.346]	[0.318]	[0.266]	[0.272]
FirmEdQuant*Change	-0.503***	-0.328***	-0.286***	-0.225***	-0.0927***	-0.0563***
	(0.00757)	(0.00549)	(0.00485)	(0.00841)	(0.00475)	(0.0034)
	[0.117]	[0.256]	[0.359]	[0.336]	[0.281]	[0.287]
R-Squared:	0.247	0.28	0.316	0.543	0.659	0.756
Controls:						
Person Fixed Effects				X	X	X
Firm Fixed Effects	X	X	X	X	X	X
Industry Effects	X	X	X	X	X	X
Firm Size Effects	X	X	X	X	X	X
*Industry*Size Effects*	X	X	X	X	X	X
Lead Emp Change Pct	X	X	X	X	X	X
Lag Emp Change Pct	X	X	X	X	X	X
Firm Age	X	X	X	X	X	X
Age (Cubic)	X	X	X	X	X	X
Experience (Cubic)	X	X	X	X	X	X

Table 1.5, cont'd: Log of a Worker's Total Earnings over the Next 1, 3, or 5 Years

Variable	**Year 1**	**Year 3**	**Year 5**	**Year 1**	**Year 3**	**Year 5**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Controls:						
Tenure (Cubic)	X	X	X	X	X	X
All Earnings (Cubic)	X	X	X	X	X	X
Job Earnings (Cubic)	X	X	X	X	X	X
Firm Earnings Ratio	X	X	X	X	X	X
Firm Educ Earn Ratio	X	X	X	X	X	X

Notes: This table shows the results of an OLS regression of the log level of a worker's cumlative earnings over 1, 3, or 5 years on his comparison-group position variables, his decision to change jobs and other controls. Columns (4), (5), and (6) add worker fixed effects to the regressions performed in Column (1), (2), and (3). The sample for these regressions comes from all job-year observations in 1995-2005 at a worker's highest paying job in that year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contains 21 million workers and 120 million worker-years.
* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. The standard errors, which are corrected for variation due to the imputation of missing education variables, are in parentheses. The missingness ratio, which gives the fraction of variability due to imputation, is in brackets.

The negative signs on the interaction coefficients in columns (4), (5), and (6) mean that the gain in level of earnings a worker sees from switching jobs is actually decreasing in group-position. Since the dependent variable is the worker's level of cumulative earnings over a given time period, we can say the earnings gap between a higher group-position worker and a lower group-position worker tends to close when these workers both switch jobs. This is supportive of the match effects hypothesis (row 3 of Table 1.1): the worker in the higher position gets a worse match effect draw (on average) at his new job, whereas the worker in the lower position gets a higher draw. The coefficients in this regression are almost certainly biased upward, likely by a substantial amount - workers who have a good outside option are likely to leave the firm whereas those who do not have a good outside option will not leave.[17] Since we only observe the earnings gains of the workers with the best outside offers, we should not assume that randomly assigning a worker to a new job would lead to such a large earnings gain. That the gains are relatively greater for changers when their position is lower is harder to explain away, in part because it has an intuitive interpretation: lower positioned workers have more room to move up and so when they do leave their average earnings gains are greater.

1.5.4 The Market for Status

The third test is another attempt to explore the hypothesis advanced by Frank [1984], namely that status acts as a form of compensation - in essence, this is another attempt to test whether we are in the first row of Table 1.1 (prestige, equity) or the information based models in rows 2 and 3. Suppose that status has a social value independent of the monetary value. Then when a worker's status decreases, the firm must offer him more money to keep him from quitting. Likewise, a worker should be willing to accept lower raises that come with growth in status. For all workers, I run the specification in (1.4) and

[17] Likewise, we can assume that many workers are staying precisely because they face such favorable conditions.

regress the percentage change in their earnings on the change in their comparison-group position between years t to $t+1$. The results of this regression can be found in Table 1.6.

Table 1.6: Worker's Percentage Earnings Change as a Function of Group-Position Change

Variable	**Earnings Change**	**Earnings Change with Fixed Effects**
	(1)	**(2)**
Firm Year Quantile	-6.88***	-19.7***
	(0.144)	(0.531)
	[0.0606]	[0.829]
Firm Education Year Quantile	-2.75***	0.609
	(0.148)	(0.883)
	[0.0855]	[0.94]
Firm Quantile Change	23.8***	20.4***
	(0.148)	(0.157)
	[0.219]	[0.151]
Firm Education Quantile Change	3.18***	3.59***
	(0.145)	(0.154)
	[0.241]	[0.176]
R-Squared:	0.0774	0.331
Controls:		
Person Fixed Effects		X
Firm Fixed Effects	X	X
Industry Effects	X	X
Firm Size Effects	X	X
*Industry*Size Effects*	X	X
Lead Emp Change Pct	X	X
Lag Emp Change Pct	X	X
Firm Age	X	X
Age (Cubic)	X	X
Experience (Cubic)	X	X
Tenure (Cubic)	X	X
All Earnings (Cubic)	X	X
Job Earnings (Cubic)	X	X
Firm Earnings Ratio	X	X
Firm Educ Earn Ratio	X	X

Notes for Table 1.6: This table shows the results of an OLS regression of the percentage earnings gains a worker experienced from year t to $t+1$, his comparison-group position variables, the changes in his comparison-group positions from year t to $t+1$, and other controls. Column (2) adds worker fixed effects to the regression performed in Column (1). The sample for these regressions comes from all job-year observations in 1995-2005 at a worker's highest paying job in that year; the sample was limited to worker-years at firms in the two largest employment size deciles. This sample contains 21 million workers and 120 million worker-years.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. The standard errors, which are corrected for variation due to the imputation of missing education variables, are in parentheses. The missingness ratio, which gives the fraction of variability due to imputation, is in brackets.

Consider column (2), which is the specification that contains worker fixed effects. I interpret the $20.4(+/-0.157)$ coefficient on the firm-year percentile change as saying that other factors held constant, an increase of a worker's group position by X percentiles from years t to $t+1$ is associated with a $20.4 * X\%$ increase in the year-to-year percentage earnings increase obtained by a worker. An increase of a worker's percentile in the firm-education earnings distribution by X from years t to $t+1$ is associated with a $3.59 * X$ increase in the year-to-year percentage earnings increase obtained by a worker. The coefficients are high and are driven by outliers (primarily low earning workers), but with outliers removed, the coefficients are of reasonable size and the signs and significance of the coefficients do not change.

Position changes and earnings changes should be mechanically positively correlated as a consequence of my definition of group-position, so the main result in Table 1.6 should not be surprising. However, this result means there is an absence of evidence that a "market for status" exists, and further tests of behavioral models will not be possible in this paper. Taken together with the results from expected cumulative future earnings, worker's apparent preference for status can be explained by a pecuniary rewards models where comparison-group position is indicative of a high individual and/or match effect, such as those depicted in rows 2 or 3 of Table 1.1. Given the results of the regression

between comparison-group position and future earnings, it would seem that the match-effect hypothesis best fits the data.

1.5.5 Robustness Checks

The Ideal Sample

A natural question to ask about robustness of the results is: Could they be driven by some factor that varies across different sub-populations? To test whether my results rely on the common assumptions I made at the beginning of the methodology section, I will study a sample of workers with minimal confounding factors where my assumptions are most likely to hold. I believe that the group of workers with the fewest confounds will be highly educated men ages 30-40 with income at or above the sample median. Highly educated men will presumably be eligible for more jobs in more places, have better information about the available markets, and greater means to undertake job search than their less educated counterparts. Men are less likely than women to drop out of the labor force due to the birth of a child. By age 30 most men should have decided on a career, and so changing careers should not be a motive for a job transition. At 40, men still have a long career ahead of them, so the cost of searching for a new job and moving should not discourage them, as the potential gains from job-changing are still relatively large. As one can think of status as a form of compensation (Frank [1984]), I would prefer men with above median earnings (relative to the male population); they have more income to give up for status. They are also less likely to experience involuntary transitions to unemployment.[18] These restrictions have the added benefit of creating a subsample of men who are likely working full-time, reducing the difficulties posed by comparing employees who are working different numbers hours. As I noted earlier, there are concerns about workers' knowledge of their own group-position. Thus, I select workers from industries

[18]see Strain [2013] for more on this subject.

where knowledge of others' compensation is well known. I choose the Finance and Technology industries, which are in the NAICS sectors 51 and 52. To minimize the likelihood of contamination from unobservable factors, all of the tests run above were re-run on the ideal sample of 30 to 40 year old men working in finance and information technology who earned yearly income above the sample median. The results that higher group-position correlates with lower probability of leaving and higher future earnings were replicated. There was no evidence of a market for status in the ideal sample.

Single Unit Sample

Another concern is whether I've chosen the appropriate comparison-group. It is intuitively more plausible that a worker would have better information and be more concerned with his position in the establishment group instead of the firm group. Ideally, I would know the worker's establishment. However, for multi-unit firms, the LEHD does not record a worker's assigned establishment(s), only his firm; his establishment is imputed based on a variety of characteristics (distance between the worker's residence and the establishment being the primary factor). To check that using a worker's position among other workers in his firm (instead of his establishment) has not altered the results, I run the above tests on single establishment firms. The results that higher group-position correlates with lower probability of leaving and higher future earnings were replicated. There was also no evidence of a market for status in the single unit sample.

1.6 Conclusion

I find that comparison-group position is a significant predictor of labor market outcomes for a representative sample of US workers and firms. Workers at the bottom of the earnings distribution in their firm comparison-groups are more likely to leave for a new job, but they are not more likely to be fired. The lower a worker's position in his comparison-groups,

the lower his earnings over the next several years. However, lower positioned workers who leave for a new job have greater percentage earnings gains than lower positioned stayers and higher positioned leavers. For all workers, group-position change and earnings change tend to move in tandem; I find no evidence of a "market for status" of the kind proposed by Frank [1984], where a tradeoff exists between increases in group-position and increases in earnings.

My results support a model where comparison-group position is a signal of things to come. Given that lower positioned workers experienced such significant earnings gains when leaving for a new job, group-position appears to capture some information about match quality; low-position leavers see larger gains than low-position stayers because they draw a new match effect, which on average is higher than their old match effect. Likewise, high-position leavers draw a new match effect that, on average, is worse than the previous one (which helped them attain their high position). Position is more informative about a worker's future trajectory at a specific firm than it is about the worker himself.

REFERENCES

J. M. Abowd, H. Finer, and F. Kramarz. Individual and firm heterogeneity in compensation: An analysis of matched longitudinal employer-employee data for the state of Washington. In J. C. Haltiwanger, J. I. Lane, J. R. Spletzer, J. J. M. Theeuwes, and K. R. Troske, editors, *The Creation and Analysis of Employer-Employee Matched Data*, pages 3–24. North-Holland, 1999a.

J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67:251–333, 1999b.

J. M. Abowd, B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. The LEHD infrastructure files and the creation of the quarterly workforce indicators. In T. Dunne, J. B. Jensen, and M. J. Roberts, editors, *Producer Dynamics: New Evidence from Micro Data*, pages 149–230. University of Chicago Press, 2009.

G. Baker, M. Gibbs, and B. Holmstrom. The internal economics of the firm: Evidence from personnel data. *Quarterly Journal of Economics*, 109:881–919, 1994.

G. S. Becker, K. M. Murphy, and I. Werning. The equilibrium distribution of income and the market for status. *Journal of Political Economy*, 113:282–310, 2005.

G. D. A. Brown, J. Gardner, A. J. Oswald, and J. Qian. Does wage rank affect employees well-being? *Industrial Relations*, 47:355–389, 2008.

D. Card, A. Mas, E. Moretti, and E. Saez. Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102:2981–3003, 2012.

A. E. Clark and A. J. Oswald. Satisfaction and comparison income. *Journal of Public Economics*, 61:359–381, 1996.

A. Cohn, E. Fehr, B. Herrmann, and F. Schneider. Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, 12: 877–898, 2014.

J. S. Duesenberry. *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press, 1949.

R. A. Easterlin. Does economic growth improve the human lot? Some empirical evidence. In P. A. David and M. W. Reder, editors, *Nations and Households in Economic Growth*, pages 89–125. Academic Press, 1974.

R. A. Easterlin. Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization*, 27:35–47, 1995.

R. A. Easterlin. Income and happiness: Towards a unified theory. *The Economic Journal*, 111:465–484, 2001.

H. S. Farber. Mobility and stability: The dynamics of job change in labor markets. In O. C. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3B, pages 2439–2483. Elsevier, 1999.

A. Ferrer-i-Carbonell. Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, 89:997–1019, 2005.

R. H. Frank. Are workers paid their marginal products? *American Economic Review*, 74: 549–571, 1984.

R. H. Frank. Interdependent preferences and the competitive wage structure. *The RAND Journal of Economics*, 15:510–520, 1984.

M. Galizzi and K. Lang. Relative wages, wage growth, and quit behavior. *Journal of Labor Economics*, 16:367–391, 1998.

R. Gibbons and M. Waldman. Task-specific human capital. *American Economic Review*, 94(2):203–207, 2004.

J. Haltiwanger and M. Vodopivec. Worker flows, job flows, and firm wage policies: An analysis of Slovenia. *Economics of Transition*, 11:253–290, 2003.

D. S. Hamermesh. The changing distribution of job satisfaction. *Journal of Human Resources*, 36:1–30, 2001.

O. Heffetz and R. H. Frank. Preferences for status: Evidence and economic implications. In J. Benhabib, A. Bisin, and M. O. Jackson, editors, *Handbook of Social Economics*, volume 1A, pages 69–91. North-Holland, 2011.

D. Indiviglio. The case for making wages public: Better pay, better workers. *Atlantic Monthly*, July 2011.

B. Jovanovic. Job matching and the theory of turnover. *Journal of Political Economy*, 87:972–990, 1979a.

B. Jovanovic. Firm-specific capital and turnover. *Journal of Political Economy*, 87:1246–1260, 1979b.

B. Jovanovic. Matching, turnover, and unemployment. *Journal of Political Economy*, 92: 108–122, 1984.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121:1133–1165, 2006.

I. Kwon and E. M. Meyersson Milgrom. How do workers compare their wages in the workplace? Working paper, October 2014.

E. P. Lazear and S. Rosen. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89:841–864, 1981.

R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley, second edition, 2002.

E. F. P. Luttmer. Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics*, 120:963–1002, 2005.

R. Murphy and F. Weinhardt. Top of the class: The effect of ordinal rank. CESifo Working Paper 4815, 2014.

A. C. Pigou. *The Economics of Welfare*. McMillan, 1932. [Originally appeared in 1920.]

M. Rege and I. F. Solli. Lagging behind the Joneses: The impact of relative earnings on job quitting. Working paper, 2013.

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

A. Smith. *The Theory of Moral Sentiments*. Liberty Classics, 1976. [Originally appeared in 1759.]

M. R. Strain. Do volatile firms pay volatile earnings? Evidence using linked worker-firm data. AEI Economic Policy Working Paper, 2013.

H. Tajfel, editor. *Social Identity and Intergroup Relations*. Cambridge University Press, 2010.

R. H. Topel and M. P. Ward. Job mobility and the careers of young men. *Quarterly Journal of Economics*, 107:439–479, 1992.

T. Veblen. *The Theory of the Leisure Class*. A M Kelley, 1965. [Originally appeared in 1899.]

CHAPTER 2

HAVE U.S. LOCAL LABOR MARKETS BECOME LESS LOCAL? APPLYING A NOVEL MATCHING FUNCTION DECOMPOSITION TO U.S. GEOGRAPHIC LABOR MARKET INTEGRATION

Evan Buntrock and Richard Mansfield

Abstract: This paper introduces a novel two-sided decomposition procedure for matching markets that isolates changes in structural factors (relative tastes, costs, or productivities associated with particular types of matches/transitions) from changes in matching patterns that are simply driven by the evolving composition of types on both sides of the market. Exploiting the LEHD database, we use this decomposition (supplemented by regression evidence) to explore the extent to which local labor markets have become more geographically integrated in the fifteen year period (1996-2010 corresponding to the rise of the internet. Our results reveal a consistent and sizeable trend away from job-to-job transitions featuring a short distance (1-10 miles) between origin and destination employers and a corresponding trend towards transitions featuring moderate distance (25-100 miles). These results suggest that integration of local labor markets has primarily occurred at a sub-state and perhaps even sub-county level of geographic aggregation, which has prevented previous inter-state and inter-county analyses of job and residential mobility from detecting the trend.

2.1 Introduction

The advent of internet technology creates the possibility that labor markets are considerably more geographically integrated than they were 20 years ago. The existence of e-mail, company websites, and job-posting websites such as LinkedIn and Monster.com have allowed firms to advertise vacancies and process applications more cheaply. These innovations have also made searching for distant jobs considerably cheaper; even in cases where workers search primarily through a personal network, Facebook and other social media have reduced the cost of maintaining a geographically broader network of contacts. While other papers have examined how the adoption of internet technology affects job search [1], our paper is the first to our knowledge that examines the degree to which the geographic breadth of local labor markets has evolved since the internet era.

Specifically, in this paper we 1) introduce a novel two-sided decomposition procedure for matching markets that isolates changes in structural factors (relative tastes, costs, or productivities associated with particular types of matches/transitions) from changes in matching patterns that are simply driven by the evolving composition of types on both sides of the market, 2) provide descriptive regression evidence documenting a substantial increase of both short distance and 25-100 mile job -transitions between 1996 and 2010, suggesting a change in the spatial matching function over this time, and 3) use our decomposition procedure to produce counterfactual simulations that paint a richer picture of the evolving economic geography of the American labor market, with a particular focus on how the incidence of local economic development policies is likely to have changed over time.

Understanding the geographic scope of labor markets is critical for evaluating the efficacy of a number of state and local economic policies. For example, local development agencies that offer tax breaks to large employers that set up a production site or a

[1] e.g. Kuhn and Mansour [2014] and Kroft and Pope [2014].

headquarters may be motivated by the belief that such employers will increase local employment demand, pushing both local unemployment down and local wages up. However, if a broad geographic recruitment campaign has become cost-effective in recent years, such policies may instead cause workers from elsewhere to be drawn to the area, thereby driving up rent for the local workers the policy was intended to help.[2] Alternatively, perhaps geographic integration has occurred in high skill occupations and industries like information technology and finance, but that search and recruitment in many service and blue-collar occupations such was waiters, plumbers, and construction workers still operate primarily at the local level. In this case, local tax incentives might succeed in boosting the employment prospects of local low skill workers even if they did not create opportunities for high skill workers.

The prospect of differential integration by skill also raises the possibility of brain drain: high skill workers respond to local labor demand shocks by contacting firms recruiting nationally, while low skill workers, concentrated in occupations or industries where firms primarily search locally (e.g. word of mouth or signs posted outside windows), resort to competing for a dwindling number of local jobs.[3] In the face of such differential attrition from local areas, local governments may wish to project the age and skill composition of local workers (and therefore the size of the future tax base) when considering large public investments (new schools, public pension solvency, etc.)

Although there have been a number of studies in the trade literature analyzing geographic integration of product markets, analogous research on labor markets seems to be sparse. In the last few years several papers have used different data sets and approaches to examine various aspects or dimensions of geographic integration. However, the difference

[2]There could be spillovers derived from the services these new employees demand, but knowing how many local workers would obtain jobs may affect the cities' bids for the new establishment.

[3]Notowidigdo [2013] points out that differential mobility of high and low income workers in response to labor demand shocks is partly attributable to the greater relative benefit low income workers get from decreasing house prices (due to a greater fraction of income spent on housing), combined with access to government safety net programs (medicaid and food stamps).

in approaches and findings makes it difficult to form a coherent impression of how labor markets have changed overall.

One strand of the literature is focused exclusively on interstate residential migration. Molloy et al. [2011], using IRS, ACS, and CPS data, document that interstate residential migration has been decreasing over the last three decades. This decline in migration is not explained by changes in demographic composition, the rise of dual earning households, or local housing prices. While the authors do not advance a model to explain this phenomenon, Kaplan and Schulhofer-Wohl [2012] suggest two possible causes of this decline in migration. First, using the same CPS and ACS data, they show that industry and occupation distributions are becoming more homogenous across states over the past thirty years, so that interstate migration is less likely to be necessary for workers seeking their ideal occupation/industry match. This suggests that the decrease in migration may be driven in part by decreases in the benefit to migration rather than increases in costs. Second, the authors posit that the internet, by permitting job search before migration, could decrease mobility if individuals are able to find their ideal employer in one move, rather than testing out various labor markets in search of the best fit.

These explanations are not fully convincing. First, the suggestion that industry homogenization has suppressed mobility seems at odds with the findings of Moretti [2010, 2011]. Using the Decennial Census combined with data from the Bureau of Labor Statistics, he finds real wage differences across states that persist for over twenty years, and shows that these differences are not explained by local amenities or housing prices - they seem to be driven by demand shocks. If the distribution of industries is becoming similar across locations, and some locations offer higher real wages (and seemingly more desirable amenities) than others, one would expect increases in migration flows across states until these wage disparities narrowed.

Second, suppose that workers' skills and job preferences evolve over their career, so that the ideal job fit changes over time. If the distribution of job matches is similar across states, there may be no state where the expected value of job quality is sufficiently high to justify paying the moving cost. However, if particular job offers can be generated without moving in the post-internet era, and job offers arrive infrequently, a particular job offer draw in another state may be sufficiently high to justify a move. Thus, the decreases in search costs driven by internet technology could have potentially increased residential mobility.

While this example suggests that the theoretical impact of decreases in search costs on the rate of mobility is ambiguous, such decreases should have an unambiguous impact on the distribution of destinations. If search costs are high, so that moving-then-searching is optimal, we should see workers with similar characteristics moving almost exclusively to those destinations that offer the most desirable distributions of potential job offers. However, if search costs are low, so that searching-then-moving is optimal, many individuals will happen to find excellent job fits in locations that were ex ante unlikely to produce a good match. Thus, we might expect the distribution of destinations to become more uniform across geographic space if search costs are declining. Motivated by this logic, in this paper we attempt to isolate changes in geographic search and recruiting costs by focusing on *where* individuals move rather than merely *whether* they move.

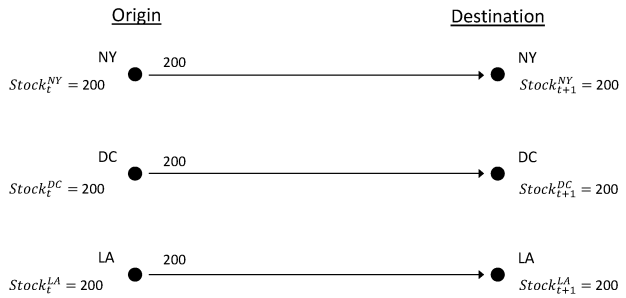
Decreasing geographic search costs produces various predictions about the changes over time in the geographic distribution of job match types. While many of these changes could also be produced using some combination of changes in industry production functions, industry product demand shocks, and changes in worker composition and worker preferences over time, such models of changing product competition or production technology are generally silent on exactly *which* workers leave jobs which locations in year t to join establishments in which locations in year $t+1$. By contrast, any model in which

geographic search and moving costs are changing will make strong predictions not just about changes in the geographic distribution of job match types, but also about the web of worker transitions that facilitate those changes.

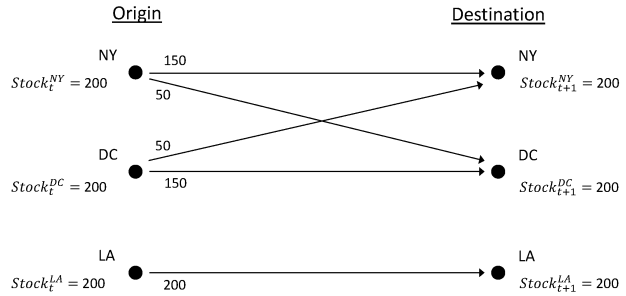
To make this point more concrete, Figures 2.1(a) - 2.1(c) represent the changing labor market as a bipartite graph with the collection of nodes on the left representing time t worker-firm matches and the collection of nodes on the right representing time $t+1$ worker-firm matches. Each node represents a job match type, with the number above it capturing the number of particular worker-firm matches associated with this type. For simplicity, job match types are characterized by the employer's geographic location only (LA, NY, or DC), though we could easily add other worker, firm, or job characteristics to the type definition. The edges represent the particular flows of worker transitions that link the time t and $t+1$ job matches together, with the size of the flow labeled. Figures 2.1(a) - 2.1(c) feature both the same time t and time $t+1$ distributions of worker-firm match types: 200 job matches in each of the three cities. However, the figures illustrate three different collections of worker transitions that are each capable of facilitating the same (lack of) change from the time t job match type distribution to the time $t+1$ job match type distribution. Figure 2.1(a) illustrates a lack of geographic integration. The transition from t to $t+1$ is facilitated by all workers in each city continuing to work in jobs within their city. Figure 2.1(b) illustrates partial integration: the evolution from the t to $t+1$ geographic employment distributions is now generated by offsetting flows of 50 workers between NY and DC (with LA still isolated). Figure 2.1(c) illustrates further integration: the t to $t+1$ transition is generated by offsetting flows of 50 workers between each pair of cities (with 100 stayers in each city). Since the beginning- and end-of-year geographic composition of employment never changes, all of the information about integration in this example is contained in the flows (edges) rather than the geographic distribution of jobs (nodes).

Figure 2.1: A Graphical Depiction of Geographic Integration

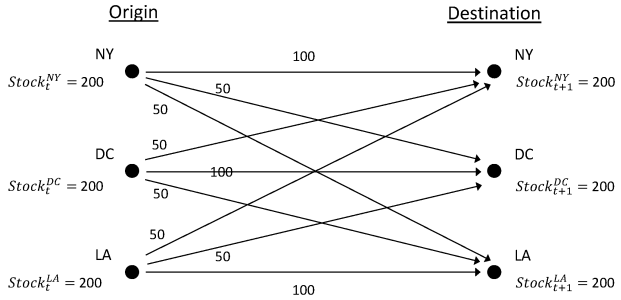
(a) Low Integration



(b) Medium Integration



(c) High Integration



It should be clear that as we increase the number of job matches and job match types to better represent the US labor market, the number of possible transition paths that can facilitate the observed change in job-match type distributions grows exponentially. Since the particular observed path of transitions that facilitates the change in job-type distributions is likely to be especially informative about changes in geographic search and recruiting costs and is less likely to be conflated with other changes in the composition of labor supply and demand, we present evidence of geographic integration that exploits the network of transitions (edges) rather than just the geographic distribution of job matches (nodes).

To do this, we design and implement a decomposition procedure that extracts the part of a network of transitions in a bipartite graph that reflects the relative propensities for particular pairs of job match types to produce transitions between them (akin to odds ratios) while removing the part that is driven by the relative frequencies with which each job match type is observed. This decomposition resembles a discrete version of the decomposition of joint continuous distribution into marginal distributions and a copula, but is designed for matching market types that cannot be ordinally ranked, and is constructed to be consistent with a standard spatial equilibrium model of the labor market (e.g. Rosen [1974]), unlike decompositions based on reweighting either the supply or demand sides of the market.

Importantly, the procedure we develop can be easily applied to any matching market context (marriage, students-to-universities, suppliers-to-retailers). Unlike existing decomposition procedures that rely on cumulative distribution functions[4], our decomposition procedure has the key feature that it accommodates definitions of types on either side of the matching market that do not have a natural ordinal ranking (e.g. races, genders, industries, universities).

[4]e.g. DiNardo et al. [1996], Machado and Mata [2005], and Rothe [2015]. Fortin et al. [2011] provides an overview of decomposition methods.

Our procedure involves the following five step process: 1) We classify jobs (worker-firm matches) into types based on observable worker, firm, and job characteristics, and treat each job-to-job transition as an observed link or edge between particular job match types (nodes or vertices). 2) We construct joint probability mass functions (essentially transition matrices) for each year that capture the empirical probabilities of each particular match type-to-match type transition type. 3) Following Hotz and Miller [1993], we map the observed transition type probabilities in each sample year into a set of parameters capturing the relative propensity for each destination job match type to be filled by hiring workers from each origin job match type. 4) We use the estimated model parameters from earlier years to construct alternative counterfactual joint probability mass functions (market-clearing transition paths) for later years, holding fixed the marginal distributions of both beginning-of-year and end-of-year job match types (nodes) from the later years. This step involves solving a system of equations to find counterfactual wage premia for each job type that would clear the labor market. 5) We integrate over these counterfactual joint probability mass functions to compute various observed and counterfactual mobility statistics that highlight how labor markets have changed in geographic scope (and for which types of workers and which firms they have changed). While these statistics constitute descriptive, non-causal evidence, they can reveal otherwise hard-to-observe patterns that can falsify or corroborate particular hypotheses and guide future research, or they can be used as "cleaner" moments to match in a structural estimation routine.

We conduct our analysis using matched employer-employee data for the years 1996-2010 from the Longitudinal Employer and Household Dynamics (LEHD) database that follows the universe of American employees as they transition between job matches. Importantly, the LEHD includes very detailed geographic coding of employer establishments that allows us to be agnostic about the level of aggregation at which geographic aggregation may be occurring. For example, if the cost of moving residences has increased, the

impact of decreasing geographic search costs may be primarily reflected in an increased likelihood of moving to a job in a neighboring town relative to another job in one's own town (since neither requires a residential move). If labor market integration is occurring within counties or commuting zones, as Manning and Petrongolo [2013] find, then examinations of patterns of geographic mobility at the county or state level (as is observed in the ACS) will be insufficient. We can also examine the degree to which high skill workers, younger workers, or firms of different employment sizes have been disproportionately effected by the technological changes that have occurred in the last two decades.

Our preliminary results suggest that labor markets have indeed become less local since the beginning of the internet era. Interestingly, nearly all of the geographic integration has occurred at the sub-state level, most of it at the sub-county level. Our regression results show that the fraction of job-to-job transitions featuring distances between origin and destination employers of 1-10 miles decreased steadily over the 1996-2010 time period, while the fraction featuring distances of 25-100 miles increased steadily. Counterfactuals based on our two-sided decomposition show that the 1996 search and matching technology would have exhibited about 20 percent (6 percentage points) more transitions of 1-10 miles and about 20 percent (5 percentage points) fewer transitions of 25-100 miles relative to the 2010 search and matching technology given the same (2010) location distributions of both origin and destination jobs. Simulations of plant openings and closings reveal that the geographic incidence of local labor demand shocks is still likely to be quite geographically concentrated.

The rest of the paper is structured as follows. Section 2.2 describes the LEHD data we use. Section 2.3 describes our decomposition and counterfactual-construction methods. Section 2.4 presents empirical results and simulation results based on the counterfactuals. Section 2.5 concludes.

2.2 Data

The Longitudinal Employer-Household Dynamics (LEHD) are the result of a partnership between the federal government and the governments of all 50 states, with most states' records going back into the early nineties. Every state provides the Census Bureau with raw data from the state's Unemployment Insurance (UI) administrative files. The UI wage records report a worker-employer link; when combined with a state supplied extract of the ES-202/QCEW report, these reports uniquely identify the worker, the employer, and the quarterly earnings associated with each job.

The link between the unique worker and employer identifiers means the LEHD data provide a job frame, and since the UI records cover almost all payments to workers in non-farm jobs, the LEHD almost entirely covers the universe of US workers, firms, and jobs. Since the LEHD is a job-frame, we can identify a worker's *dominant job* (the job at which the worker had the highest earnings) for every calendar year. The unique worker and firm identifiers let us follow workers and firms over time and observe their entry, mobility, and exit from the US economy. Therefore, the LEHD also provide a job-transition frame. Crucially, the LEHD coverage spans the period when the internet might have most dramatically affected job-transitions in the US. The earliest states report data from 1990-2010, about half of the states join by 1998, and data from nearly all of the states are available by 2004.

The unique worker and firm identifiers also permit linkages to other datasets, augmenting the core UI data with worker, firm, and establishment characteristics (including geographic location). We will make use of industry and location data from each establishment of a firm in the Quarterly Census of Employment and Wages (QCEW) as well as demographic information provided by records from the Social Security Administration administrative records, decennial census, American Community Survey (ACS), Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP).

For further details about the contents and construction of the LEHD, see Abowd et al. [2009].

Thanks to this richness of the worker and firm information and the totality of the LEHD's coverage, we can construct extremely detailed worker/job flows within and across states, counties, and even census tracts. Once we characterize workers, firms, and jobs by certain characteristics, we are able to examine local and national migration patterns in US labor markets that could not be studied without linked employer-employee data.

We can create a graph in which the nodes are job-types in a given year and the edges are transitions between job types across year. A simple case, where we characterize jobs purely by their geography, has been depicted in Figures 2.1(a) - 2.1(c).[5] This analysis can be repeated and complicated if we characterize jobs by their earnings and industry, or some combination of earnings, industry and geography.

Although we intend to use an even larger sample, we started with a 50% random sample of the following 10 states: CT, IL, IN MD, MN, NJ, NY, PA, RI, WI. We chose these states because of their early availability in the LEHD (they all have records starting in 1996 or earlier) and their relative geographic proximity, which facilitates an analysis of region-level geographic labor market integration. All observations for a job can be characterized as pairs between two years: t and $t+1$. We kept only the pairs where the first year fell within the range of 1996-2010 (inclusive), where we had location information for the job in both years of the pair, and where the tract of the job was adjacent to another tract; the last requirement is imposed for the geographic analyses that measure geographic distance in terms of the pathlength between tracts. Even after these restrictions, we had 30,000,000 unique workers, 3,500,000 unique firms, and 250,000,000 worker-year pairs.

[5] One drawback of the LEHD is that for multi-establishment firms, a worker's location must be imputed. However, the unit to worker impute assigns an establishment to a worker with a probability that decreases in the distance between the worker's residence and that establishment. Thus cases of significant measurement error in true location are by definition less likely to occur. Further, we aggregate types in a geographically compact manner, and so in many cases the worker will be placed in the same type that he would have been had we known his true location.

For each worker in each year, we found the worker's dominant job (the highest paying job held by the worker that year), and identified whether the job changed the next year and if the worker was unemployed for one or more quarters in these two years. We use additional worker and firm information to characterize the job types and the transitions between these types.

2.3 Decomposition Methodology

2.3.1 Overview

Much of the theoretical literature on job search and matching models employs the concept of a "matching function" of the form $P(i, j, \theta, F, G)$, which specifies the probability that a worker of type i either receives an interview with or actually accepts a job with a firm of type j. The matching function may depend on parameters θ that capture matching efficiency (e.g. an index of labor market tightness) as well as the distributions of worker and firm types in the job pool, captured by $F(*)$ and $G(*)$ respectively. To the extent that worker/firm heterogeneity is introduced at all, the worker and firm types in these papers are generally latent, and reflect worker skill or firm productivity (or perhaps levels of search or recruiting intensity).

Structural search models with endogenous search/recruiting decisions tend to define a relatively small set of structural parameters θ_y with clear taste or cost interpretations, and make assumptions about how workers search for firms, how firms recruit, and the relative productivity of different types of workers at different types of firms in order to describe the full equilibrium matching as a function of θ, $F(*)$ and $G(*)$. This approach naturally lends itself to maximum likelihood estimation, and has the advantage that the parameters that are recovered are interpretable, allowing counterfactual simulations in which, say, moving costs are decreased by 50%.

However, the structural approach, by specifying only a small number of parameters, requires untested assumptions about the way job search, recruiting, and production actually operate. Indeed, many alternative assumptions about production and wage determination (e.g. wage posting vs. bargaining), search and recruiting strategy (directed vs. undirected) have been used to close structural search models, and these alternative specifications have generally been estimated on effectively the same match outcome data, often yielding potentially conflicting interpretations of how a particular policy would affect the labor market.

We take a polar opposite approach. We saturate the matching function we estimate with parameters θ that capture different reduced-form functions of the deeper structural parameters, sacrificing (for now) straightforward interpretations of the elements of θ in order to minimize the restrictions we place on preferences and search, recruiting, and production technology. This model flexibility improves our ability to filter out the component of changes in transition or matching patterns that are driven by changes in the marginal distributions (composition) of types on both sides of the market, thus leaving the part of the observed transition patterns that is most likely to be directly informative about the underlying tastes, costs, and productivities of interest. Simulations based on these reduced-form parameters can then reveal patterns in the data (otherwise partially obscured by compositional changes) that a more parsimonious structural model should be trying to fit, thereby helping to guide the exact specification of such a model. Thus, the decomposition method we develop in the rest of this section is designed to provide richer, more informative descriptive evidence that can be used as a first step to guide second-stage structural modeling or causal identification schemes.

2.3.2 Constructing an Empirical Matching Function

In this section, we 1) extend the basic matching function framework to explicitly address transitions of workers between origin and destination job types, and 2) operationalize the framework by making job types (more specifically, worker-establishment match types) observable. Specifically, for each job-to-job transition we classify both the worker-firm origin match and the worker-firm destination match in each period as discrete types based on vectors of observable worker, firm, and match characteristics that are likely to capture underlying (structural) preferences, moving costs, and geographic search costs that determine the relative desirability of the job match for both the worker and the firm.

Including current location as part of the type definitions for worker-firm matches is critical to our goal of characterizing changes in the geographic scope of local labor markets. However, note that there is no natural way of placing geographic locations in an ordinal ranking. Geographic preference is expected to be a horizontal characteristic, and will depend on (among other factors) the initial location of the worker or establishment. Similarly, industries cannot be effectively ranked. Thus, when worker-firm match types are partly determined by geographic location, industry, or any other non-ordinal characteristic, the concept of a cumulative distribution function of types $F(*)$ or $G(*)$ loses meaning. However, the concept of a probability mass function (PMF) for worker-firm match types still retains meaning. Given any arbitrarily chosen ordering of job match types, we let $f(o)$ capture the fraction of all "origin" (i.e. beginning of period) worker-firm matches that have been classified as type o, and we let $g(d)$ capture the fraction of all "destination" (end of period) worker-firm matches that have been classified as type d.[6]

[6]Note that the type space of job matches may be the same for both origin and destination matches, but it need not be; for example, one could choose to classify the destination job matches by a different subset of worker and firm characteristics than the origin matches based on which characteristics of the origin and destination matches one thinks are most critical for determining worker mobility (e.g. "push" factors may be different than "pull" factors).

Thus, the matching function in our context, the joint probability mass function $P(o, d, \theta, f(*), g(*))$, simply captures the probability that a randomly chosen job-to-job transition will consist of a worker from a worker-firm match of type o leaving to create a new worker-firm match of type d. Note that a worker who stays at the same firm can still be included as a degenerate "transition" between the match types o and d associated with their worker-firm match, so focusing exclusively on transitions is without loss of generality.[7]

Given the focus on the possibility of changes in search costs or moving costs over time, in our empirical implementation we will allow for the parameters reflecting preferences, productivities, and costs to be indexed by year: $\theta \rightarrow \theta^y$. Similarly, we will allow for the probability mass functions for origin and destination job match types to evolve over time as worker demographics change or particular industries or locations experience product demand shocks (or as we move slowly toward a new steady-state equilibrium job match distribution), so that $f(*)$ and $g(*)$ will often be indexed by year (e.g. $f^{y'}(*)$ and $g^{y''}(*)$).

2.3.3 Decomposing the Matching Function

Once observable types have been assigned, the matching function for job transitions that existed in year y, $P(*, *, \theta^y, f^y, g^y)$, can be non-parametrically estimated directly by using the fraction of year y job transitions in the LEHD sample that consist of a worker from a worker-firm match of type o leaving to create (or possibly staying to continue) a new worker-firm match of type d.

However, our goal is to decompose changes in the spatial equilibrium over time into origin job match compositional changes, destination job match compositional changes, and changes in tastes/costs that determine which origin job matches become which destination

[7]Similarly, worker nonemployment could also potentially represent a "job" type o or d.

job matches. Thus, we are interested in recovering estimates of $P(*, *, \theta^y, f^{y'}, g^{y''})$ for any combination y, y', y''.

The marginal PMFs $f^{y'}(o)$ and $g^{y''}(d)$ can also easily be calculated for any pair of years (y', y'') simply by identifying the fraction of year $y'-1$ (origin) worker-firm matches that are of each type o and the fraction of year y'' (destination) worker-firm matches that are of type d. Thus, the challenge is to use the observed joint PMFs $P(*, *, \theta^y, f^y, g^y)$ and the marginal PMFs $f^{y'}(o), g^{y''}(d)$ to identify the taste and cost parameters θ_y that determined the spatial equilibrium in each year and use them to construct the joint PMF $P(*, *, \theta^y, f^{y'}, g^{y''})$ that captures the counterfactual network of labor market flows.

Consider a particular transition from origin match type $\tilde{o}$ to destination match type $\tilde{d}$ in a particular year y'. Using the law of total probability, we can represent the probability that a randomly chosen transition in year y' will be of this transition type by $P(\tilde{o}, \tilde{d}, \theta^y, f^{y'}, g^{y'}) = P(\tilde{d}|\tilde{o}, \theta^{y'}, f^{y'}, g^{y'}) * f^{y'}(o')$. One approach to constructing a counterfactual estimate $P(\tilde{o}, \tilde{d}, \theta^y, f^{y'}, g^{y''})$ is to use a simple reweighting estimator that holds the origin marginal $f^{y'}(o')$ at its year y' value, but uses type $\tilde{d}$'s year y conditional choice probability (which can be estimated directly from the data):

$$P^{RW}(\tilde{o}, \tilde{d}) = P(\tilde{d}|\tilde{o}, \theta^y, f^y, g^y) f^{y'}(\tilde{o}) \tag{2.1}$$

This reweighted joint PMF will imply a marginal distribution of destination job types $g(d)$ that does not match the distribution observed in any of the years y, y', or y'': $\sum_{o \in \mathcal{O}} P^{RW}(o, d) \neq g^y(d) \neq g^{y'}(d) \neq g^{y''}(d)$. At first glance this might be construed as a desirable property: it makes predictions as to how the distribution of destination job types would have evolved under a different composition of origin types. However, this counterfactual is implicitly based on an implausible model of spatial equilibrium in which the composition of labor demand is entirely determined by the composition of labor supply: the set of tasks performed by workers (and the locations of those tasks) adjusts to accommodate the change in the origin job types without altering the conditional choice

probabilities. Put differently, within each origin job type, the supply of workers to each destination job type is perfectly inelastic.

To see this, suppose that worker age group partly defines an origin job type, and firm industry partly defines a destination job type. If the workforce ages, then the standard reweighting estimator would predict that workers in each age group would display exactly the same industrial distribution of job matches as before, and that those industries that have historically relied more heavily upon younger workers would shrink proportionately, while those that relied on older workers would grow proportionately. However, the standard supply and demand model would suggest that when youth-intensive industries are faced with a shortfall of supply at the old equilibrium wages, they would begin to raise wages. While labor demand in these industries would begin to fall as higher labor costs eroded profits, we should expect that some slightly older workers would be persuaded by higher wages to remain in the youth-intensive industries for longer. But any non-zero elasticity of supply implies that the conditional choice probabilities $\{P(d|o)\}$ should depend on the marginal distribution of origin job types $f(o)$.

Similarly, an alternative simple reweighting estimator that holds the composition of destination jobs fixed at its year y'' value, $g^{y''}(d')$, could use year y conditional probabilities to estimate the following counterfactual:

$$P^{RW2}(o', d') = P_y(o'|d')g^{y''}(d') \tag{2.2}$$

However, this re-weighting estimator, in addition to imposing that labor demand is perfectly inelastic, will imply a marginal distribution of origin job types that never existed.

Thus, if we want our counterfactual joint probability mass function $P(*, *, \theta^y, f^{y'}, g^{y''})$ to represent a *feasible* transition path from a given origin job type distribution $f^{y'}(*)$ to a given destination job type distribution $g^{y''}(*)$, we need our counterfactual conditional probabilities from (2.2) to integrate to the appropriate origin type marginal distribution from the target year y', $f^{y'}(*)$. We can re-express this condition as the standard

market clearing condition that requires the excess supply/demand vector to equal zero everywhere:

$$\sum_{d \in \mathcal{D}} P(\tilde{o}, d, \theta^y, f^{y'}, g^{y''}) - f^{y'}(\tilde{o}) = 0 \ \forall \ \tilde{o} \in \mathcal{O}$$

While the simple re-weighted counterfactual presented in equation (2.2) violates this market clearing condition, note that the locations of the positive and negative entries in this excess supply/demand vector are informative: they indicate how destination establishments would have needed to adjust their hiring/retention decisions in order for the market for workers in each origin job type to clear. In particular, suppose that given taste/cost parameters θ_y and origin and destination marginal distributions $f^{y'}(o)$ and $g^{y''}(d)$, the simple counterfactual JPMF in (2.2) reveals an excess of demand for workers associated with a particular origin match type $\tilde{o}$. Then we should have expected the offered wage distribution to workers in origin matches of type $\tilde{o}$ to shift up to clear the market, and we should expect a smaller surplus for destination types from hiring workers from origin type $\tilde{o}$, leading some or all destination types to decrease their conditional probabilities of hiring workers from origin match type $\tilde{o}$. This logic motivates the counterfactual probability distribution that we actually present. Specifically, we first propose a flexible multinomial logit model of the destination establishment's choice of which worker from which origin establishment to hire (which could include retaining their own worker, if the job is currently filled). We then estimate the parameters θ_y of this model separately for each year y. Finally, we construct counterfactual joint PMFs $P(*, *, \theta^y, f^{y'}, g^{y''})$ by holding fixed both the estimated parameters θ_y and the destination distribution $g^{y''}(d)$ and finding the requisite wage adjustments necessary to ensure that the market for each type of origin worker-firm match clears, given the targeted origin distribution $f^{y'}(o)$. The next subsection discusses the details of this procedure.

2.3.4 Constructing a Feasible Counterfactual Probability Distribution of Transition Types

Let V_{ijk} represent the present discounted value of the surplus (profit contribution) to a particular destination establishment k of hiring (or retaining) a particular worker i currently employed by a particular origin establishment j. This surplus will incorporate the PDV of the contribution to revenue at firm k produced by worker i, but it will also reflect the recruiting costs for firm k of identifying worker i as well as the PDV of the (equilibrium) compensation necessary for establishment k to lure worker i away from establishment j. Such compensation is likely to depend on worker i's relative preferences for the two establishments, the costs of moving between the two establishments and the value of worker i to the origin establishment j, all of which are structural parameters that are elements of θ.

Denote by $o(i, j)$ the job match type associated with the characteristics of the origin worker-establishment match, and denote by $d(i, k)$ the job match type associated with the characteristics of the destination worker-establishment match. We argue that if appropriate characteristics are chosen to define origin and destination job match types, a substantial component of the surplus V_{ijk} will be common to all job matches formed via a worker transition from a job match of type $o(i, j)$ to a job match of type $d(i, k)$.

For example, moving costs and search costs for a young worker associated with finding and moving to a job in New York from a job in LA may be similar for all such young workers moving between these two particular cities. If both worker age and establishment location are included in the definition of a match type, most of the variation in such costs would exist between rather than within (o, d) transition types. Similarly, highly productive but already highly paid workers may offer little surplus for establishments featuring low-skill jobs (given the high cost of luring the worker), but may offer substantial surplus above and beyond the required compensation for a firm whose output depends critically

on worker skill. If current worker earnings is included in the definitions of origin job types and firm average earnings is included in the definitions of destination job types, much of the variation in relative match quality will exist between rather than within (o, d) transition types.

Thus, we express V_{ijk} as the sum of the mean surplus from the transition type (o, d), $\overline{V}_{od}(\theta)$, and an idiosyncratic deviation from the mean surplus for the particular destination establishment k from hiring the particular worker i from the particular origin establishment j, which we denote ϵ_{ijk}:

$$V_{ijk} = \overline{V}_{od}(\theta) + \epsilon_{ijk} \tag{2.3}$$

ϵ_{ijk} might reflect, for example, the low psychic costs of a particular worker who is moving back to the location where his family lives (thereby lowering the compensation necessary to lure the worker and increasing the surplus), or perhaps particular skill requirements of establishment k that worker i uniquely possesses.

The dependence of $\overline{V}_{od}(\theta)$ on θ highlights the fact that these mean surplus values are likely to be complicated functions of many of the underlying structural taste, cost, and productivity parameters of interest; the set of mean surplus values $\{\overline{V}_{od}(\theta)\}$ capture the variation that we wish to relate to structural changes. By contrast, like many structural models, we treat ϵ_{ijk} as independent of the parameters of interest, relying on a rich type space of transitions (o, d) to capture the part of the heterogeneity in tastes, costs, and productivities that is policy relevant.

However, unlike fully structural models, we do not make explicit how each mean surplus $\overline{V}_{od}(\theta)$ depends on each of the structural parameters in θ. Instead, we simply redefine $\theta_{od} \equiv \overline{V}_{od}$, and treat the mean surplus values $\{\theta_{od}\}$ as the parameters of interest. As noted above, maintaining such a rich parameter space allows us to be agnostic about the particular model of search and firm production that generates the data while capturing all of the variation contained in observed transition patterns that is likely to be most

informative about changes in search and recruiting costs.

Given the additive decomposition of surplus from equation (2.3), Hotz and Miller [1993] demonstrate that an assumption that specifies the distribution of the idiosyncratic component ϵ_{ijk} is both necessary and sufficient to create a one-to-one mapping between observed conditional choice probabilities $\{P(o|d)\ \forall\ o, d\}$ and the transition type surplus means, $\{\theta_{od}\ \forall\ (o, d)\}$, given a scale normalization for the $\{\theta_{od}\}$. To maintain tractability, we follow the dynamic discrete choice literature by assuming that ϵ_{ijk} features a Type 1 extreme value distribution. Since observed job transitions only reflect the *relative* desirability of particular workers from particular origin establishments for the hiring/retaining firm, we normalize $\theta_{1d} = 0\ \forall\ d$. Then the standard logit formula allows us to write the probability that a particular job match type $\tilde{d}$ will be created by luring the worker from an origin job match type $\tilde{o}$ as[8]

$$P(o = \tilde{o}|d = \tilde{d}) = \frac{e^{\theta_{\tilde{o}\tilde{d}}}}{1 + \sum_{o \in \mathcal{O}} e^{\theta_{o\tilde{d}}}}$$

For any origin match type $\tilde{o} > 1$ and any destination match type $\tilde{d}$, we can take the log difference in choice-specific probabilities $ln(P(o = \tilde{o}|d = \tilde{d})) - ln(P(o = 1|d = \tilde{d}))$ to recover $\theta_{\tilde{o}\tilde{d}} - \theta_{1\tilde{d}} \equiv \theta_{\tilde{o}\tilde{d}}$, given our normalization above. In practice, we will use the sample counterparts of $P(o = \tilde{o}|d = \tilde{d})$ to recover estimates of $\hat{\theta}_{\tilde{o}\tilde{d}}$. Note that we can perform this conditional choice probability inversion for any year to recover year-specific estimates of $\hat{\theta}^{y}_{\tilde{o}\tilde{d}}$ for any match-to-match transition $(\tilde{o}, \tilde{d})$.

Our goal is to produce a counterfactual set of conditional choice probabilities that retains the information contained in the base year transition-specific values $\theta^{y}_{\tilde{o}\tilde{d}}$, but also yields a feasible matrix of transition probabilities given the targeted origin and destination

[8]Since our earlier draft, we have realized that the equation needs to account for weighting by the marginal distributions. The correct equation to use would thus be: $P(o = \tilde{o}|d = \tilde{d}) = \frac{f^{y}(\tilde{o})e^{\theta_{\tilde{o}\tilde{d}}}}{\sum_{o \in \mathcal{O}} f^{y}(o)e^{\theta_{o\tilde{d}}}}$ This change would carry through the other equations in this subsection, most notably equation (2.4). Using the form in the body of the paper means that Figure 2.3 was not properly estimated. However, given the nature of the estimates of equation (2.12), we believe that the central findings depicted in Figure 2.3 remain correct. Future drafts of this paper will have (correct) results obtained by using the equation in this footnote.

marginal probability mass functions $f^{y'}(*)$ and $g^{y'}(*)$. We do this by solving for a set of wage premia for workers from particular origin job match types that clear the labor market, given the target marginal distribution of origin types $f^{y'}(o)$. Specifically, we form counterfactual surplus values for destination establishment k for worker i from origin establishment j via:

$$V_{ijk} = \theta^{y}_{o(i,j)d(i,k)} - w_{o(i,j)} + \epsilon_{ijk} \tag{2.4}$$

$w_{o(i,j)}$ represents the part of establishment k's surplus from hiring worker i from establishment j that is lost/gained due to the required increase/decrease in compensation necessary to eliminate the excess/insufficient demand for workers from match type $o(i,j)$ that would have been created based on the original year y compensation levels and preferences/search costs/production technologies $\{\theta^y\}$ and the alternative marginal distributions of origin and destination match types ($f^{y'}(o)$ and $g^{y''}(d)$, respectively).

Equation (2.4) can be transformed into counterfactual conditional choice probabilities of establishment type $\tilde{d}$ hiring a worker from origin match type $\tilde{o}$ given by:

$$P^{cf}(o = \tilde{o}|d = \tilde{d}) = \frac{e^{\theta^{y}_{\tilde{o}\tilde{d}} - w_{\tilde{o}}}}{1 + \sum_{o \in \mathcal{O}} e^{\theta_{o\tilde{d}} - w_o}} \tag{2.5}$$

Note that conditional choice probabilities only depend on the *relative* wage premia of different origin match types, so that only $O-1$ differences in wage premia are identified. Hence, in the equation above we have normalize the counterfactual wage premium for the first origin type to be 0, so that $e^{\theta_{1\tilde{d}} - w_1} = e^0 = 1$.

We can solve for the $O-1$ unknown counterfactual wage premia $\{w_o : o \neq 1 \in \mathcal{O}\}$ by imposing the market clearing condition that total counterfactual demand for each origin type $\tilde{o}$ must equal counterfactual supply. This produces the following system of $O-1$

equations (one for each origin type $o = 2$ to $o = O$):

$$\sum_{d \in \mathcal{D}} P^{cf}(o = 2|d) = \sum_{d \in \mathcal{D}} [\frac{e^{\theta^y_{2d} - w_2}}{1 + \sum_{o \in \mathcal{O}} e^{\theta^y_{od} - w_o}}] g^{y''}(d) = f^{y'}(2)$$

$$\vdots$$

$$\sum_{d \in \mathcal{D}} P^{cf}(o = O|d) = \sum_{d \in \mathcal{D}} [\frac{e^{\theta^y_{Od} - w_O}}{1 + \sum_{o \in \mathcal{O}} e^{\theta^y_{od} - w_o}}] g^{y''}(d) = f^{y'}(O)$$

Given marginal PMFs for the origin and destination types in the target years ($f^{y'}(*)$ and $g^{y''}(*)$) and a matrix of $\{\theta^y\}$ values from a base year, this system can be inverted to recover $\{w_o : o \in [2, O]\}$. Once the wage premia have been solved for, the counterfactual conditional choice probabilities $P^{cf}(o = \tilde{o}|d = \tilde{d})$ from equation (2.5) are fully determined, and can be used to construct the full counterfactual joint PMF over (o, d) transitions via:

$$P(o, d, \theta^y, f^{y'}, g^{y''}) = P^{cf}(o|d) g^{y''}(d)$$

Note that by assuming the additive separability of the mean choice-specific values and a type 1 extreme value distribution for the idiosyncratic surpluses for particular origin-worker-destination combinations, we have effectively pinned down the matrix of wage elasticities of labor demand for each destination establishment match type with respect to each origin match type. However, note that while distributional assumptions for ϵ_{ijk} would yield slightly different counterfactual joint PMFs for transitions, the wage premia are only necessary to eliminate the *excess* demand and supply. Many of the changes over time in relative tastes will be offsetting (e.g. firms in Los Angeles will be more inclined to hire current New York workers than previously, but New York firms will be more inclined to hire Los Angeles workers than previously). Such offsetting changes in conditional probabilities do not create excess demand or supply, and thus would occur without any need for changing wages. Put another way, a substantial part of the identification of the counterfactual joint PMFs will be entirely insensitive to the distributional assumption for ϵ_{ijk}.

2.3.5 Using Counterfactual Joint Probability Mass Functions to Illustrate Changes in the Functioning of Local Labor Markets

We have shown how to recover a counterfactual joint PMF of worker-firm origin and destination matches for arbitrary marginal PMFs of origin and destination match types and for arbitrary sets of taste/technology parameters $\{\theta_{od}\}$. Since our goal is to analyze whether the geographic scope of labor markets have changed, and whether we can find evidence that these changes are driven by changes in geographic search costs or moving costs, a natural application is to compare the observed joint PMF for the most recent year in the data (here, 2010) to a counterfactual based on 2010 marginal distributions of worker and firm types, but taste/technology parameters from the pre-internet era (here, 1996): $P(o, d, \theta^{2010}, f^{2010}(*), g^{2010}(*))$ versus $P(o, d, \theta^{1996}, f^{2010}(*), g^{2010}(*))$. However, comparing differences in the probabilities associated with the observed and counterfactual JPMF provides very little immediate intuition about exactly how the search costs or moving costs have changed (and for which locations, skill levels, firm sizes, age ranges, etc.).

To develop such intuition, we need to integrate over both distributions and compare outcomes on easily understood statistics. For example, if each worker and each firm type were associated with specific locations (perhaps in addition to other observable characteristics), we can easily associate each transition type (o, d) with the distance $D(o, d)$ between the location of the o-th origin job match type and the location of the d-th destination job match type. We could then compare the observed and counterfactual average distance traveled between last year's and this year's jobs:

$$E^{Obs}[D] = \sum_{o \in \mathcal{O}} \sum_{d \in \mathcal{D}} D(o, d) P(o, d, \theta^{2010}, f^{2010}, g^{2010}) \tag{2.6}$$

$$E^{CF}[D] = \sum_{o \in \mathcal{O}} \sum_{d \in \mathcal{D}} D(o, d) P(o, d, \theta^{1996}, f^{2010}, g^{2010}) \tag{2.7}$$

Since the 2010 origin and destination PMFs are used in both JPMFs, this comparison abstracts from any changes in worker mobility between 1996 and 2010 driven by changes in the composition of either origin or destination job locations (e.g. an increasing concentration of job opportunities in urban areas, or in particular metropolitan areas), or even an interaction between the two (e.g. spatial mismatch caused by larger cohorts of new graduates entering the labor market in locations where job opportunities are scarce).

Similarly, since we hypothesized that the changes in geographic search costs may mostly be affecting skilled workers, we might be interested in calculating the average distance traveled between last year's and this year's jobs conditional on belonging to a worker type associated with one of the top two earnings deciles, or conditional on working in the financial industry last year. If each worker type is associated with a single earnings decile and/or 4-digit industry this is easily accomplished by taking the expectation using the conditional joint PMF. Let $\mathcal{R}$ represent the set of types whose workers satisfy the relevant restriction. Then

$$E^{Obs}[D|o \in \mathcal{R}] = \frac{\sum_{o \in \mathcal{R}} \sum_{d \in \mathcal{D}} D(o,d) P(o,d,\theta^{2010}, f^{2010}, g^{2010})}{\sum_{o \in \mathcal{R}} f^{2010}(o)} \tag{2.8}$$

$$E^{CF}[D|o \in \mathcal{R}] = \frac{\sum_{o \in \mathcal{R}} \sum_{d \in \mathcal{D}} D(o,d) P(o,d,\theta^{1996}, f^{2010}, g^{2010})}{\sum_{o \in \mathcal{R}} f^{2010}(o)} \tag{2.9}$$

Restrictions on the destination type, or even on combinations of origin and destination types can be handled in an analogous fashion.

2.4 Results

2.4.1 Regressions

The LEHD allows for examination of the forces that drive worker transition while controlling for composition changes in the characteristics of workers and firms. In three separate

tests, we look for changes over time in the probability that workers change jobs, the average distance traveled, and the relative likelihood of moves of varying distances. We find that job changing is procyclical and the average distance traveled by job changers has not changed over time - but within tract moves and moves of 25-100 miles have became much more likely from 1996 to 2010.

Probability of Changing Jobs

A natural first question is whether workers have become more or less likely to change jobs over time. To answer this, we first estimate the following linear probability model:

$$\begin{aligned} 1(NewJob_{i,j,t}) = \beta_0 + D^{Year}\beta^{Year} + D^{Age}\beta^{Age} + D^{Earn}\beta^{Earn} + D^{FirmSize}\beta^{FirmSize} \\ + D^{FirmEarn}\beta^{FirmEarn} + D^{Industry}\beta^{Industry} + D^{Tract}\beta^{Tract} \end{aligned} \tag{2.10}$$

The outcome, $1(NewJob_{i,j,t})$, is equal to one if worker i at firm j changes dominant jobs (to a new firm) between years t and $t+1$ without an unemployment spell of one or more quarters and equals zero otherwise; thus, the variable is equal to zero if the worker stays at the same firm or is not recorded as employed in the LEHD for one or more quarters between the end of his old job and the beginning of his new one. The independent variables include design matrices of indicator variables capturing the calendar year t, worker i's age category in year t (there are 10 categories), worker i's earnings decile in year t (defined according to the samplewide distribution of inflation-adjusted earnings from the population of dominant jobs), firm j's worker-weighted size decile in year t, firm j's worker weighted earnings decile in year t, the industry of the firm (as determined by NAICS industry codes), the tract of worker i's year t job at firm j.

The coefficients of interest are the $\{\beta^{Year}\}$ associated with the design matrix of year indicators D^{Year}. The signs and magnitudes will reveal any time trends in the incidence of job-changing between the years 1996 and 2010 that are not captured by any of the other

sets of controls. These other design matrices partially address changes in the composition of labor supply or demand that might yield to an increase or decrease in the prevalence of job-to-job mobility even in the absence of any structural changes in the matching technology. D^{Age} is included to mitigate spurious decreases in mobility driven by an aging population, since older workers may always face higher moving costs than young workers. Similarly, D^{Earn} addresses changes in mobility driven mechanically by changes in the earnings distributions (since higher earning workers are less likely to change jobs). The firm size, earnings, and industry design matrices are meant to capture recruiting and retention characteristics of the firm that may relate to a worker's decision to change jobs. The design matrix D^{Tract} capturing the census tract associated with the worker's (origin) establishment is meant to control for geographic differences in employment density and in access to opportunities that exist across different locations; for most professions, there are more easily accessed job opportunities in Manhattan than in Big Sky.

Table 2.1(a) displays the values of $\{\beta^{Year}\}$, while the coefficient vectors associated with the design matrices capturing worker age and earnings and firm size and average earnings are presented in Tables 2.1(b) and 2.1(c). As expected, younger and lower paid workers are more likely to change jobs. Perhaps somewhat surprisingly, there is a quadratic relationship between a firm's size and its probability of retaining employees; workers at mid-size firms (in the fifth and sixth deciles) are more likely to leave. Even after controlling for worker income, it seems workers are more likely to leave low paying firms than high paying firms. There is not a clear time trend in the coefficients of the year variables. Instead, the trend is procyclical: job-changing happens more often in years of economic expansion. Since our focus is on dominant jobs over the course of a year, this does not contradict the findings of Hyatt and Spletzer [2013], who document a decline in the occurrence of single quarter jobs. It does support the findings of Haltiwanger et al. [2014]. Although our results do not primarily distinguish whether our job transitions

come with large earnings increases, Buntrock [2015], which uses the LEHD and employs the same definition of a "job-to-job transition" that we use here, finds that such transitions are associated with large earnings gains on average.

Table 2.1: Trends in the Incidence of Job-to-Job Transitions

Table 2.1(a): Coefficients on Year Indicator Variables $\{\beta^{Year}\}$ from a Regression of a Job-Transition Indicator on Worker, Firm, and Job Characteristics (Equation 2.10)

Variable	**Category**	**Estimate**
Year	1996	-0.00831***
		(0.000135)
Year	1997	0.0147***
		(0.000132)
Year	1998	0.0252***
		(0.000131)
Year	1999	0.0327***
		(0.00013)
Year	2000	0.0323***
		(0.000126)
Year	2001	0.0229***
		(0.000126)
Year	2002	0.0103***
		(0.000126)
Year	2003	0.0102***
		(0.000126)
Year	2004	0.0172***
		(0.000126)
Year	2005	0.0198***
		(0.000125)
Year	2006	0.0227***
		(0.000125)
Year	2007	0.0167***
		(0.000124)
Year	2008	0.0024***
		(0.000125)
Year	2009	-0.00807***
		(0.000126)
Year	2010	

Notes for Table 2.1(a): This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. Standard errors in parentheses.

Table 2.1(b): Coefficients on Earnings Decile and Age Group Indicator Variables ($\{\beta^{Earn}\}$ and $\{\beta^{Age}\}$) from a Regression of a Job-Transition Indicator on Worker, Firm, and Job Characteristics (Equation 2.10)

Variable	Category	Estimate
Worker Earnings Decile	1	0.247***
		(0.000126)
Worker Earnings Decile	2	0.197***
		(0.000125)
Worker Earnings Decile	3	0.156***
		(0.000123)
Worker Earnings Decile	4	0.114***
		(0.00012)
Worker Earnings Decile	5	0.0717***
		(0.000118)
Worker Earnings Decile	6	0.0391***
		(0.000114)
Worker Earnings Decile	7	0.0195***
		(0.00011)
Worker Earnings Decile	8	0.00828***
		(0.000107)
Worker Earnings Decile	9	0.002***
		(0.000105)
Worker Earnings Decile	10	
Worker Age Decile	1	0.00898***
		(0.000121)
Worker Age Decile	2	0.129***
		(0.000115)
Worker Age Decile	3	0.138***
		(0.000114)
Worker Age Decile	4	0.121***
		(0.000113)

Table 2.1(b), cont'd: Coefficients on Earnings Decile and Age Group Indicator Variables ($\{\beta^{Earn}\}$ and $\{\beta^{Age}\}$) from a Regression of a Job-Transition Indicator on Worker, Firm, and Job Characteristics (Equation 2.10)

Variable	Category	Estimate
Worker Age Decile	5	0.108***
		(0.000111)
Worker Age Decile	6	0.0973***
		(0.000111)
Worker Age Decile	7	0.0866***
		(0.000112)
Worker Age Decile	8	0.0737***
		(0.000116)
Worker Age Decile	9	0.0562***
		(0.000125)
Worker Age Decile	10	

Notes: This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

* means significant at the 10% level. ** means significant at the 5% level. ***means significant at the 1% level. Standard errors in parentheses.

Table 2.1(c): Coefficients on Firm Size and Firm-Average Worker Earnings Deciles ($\{\beta^{FirmSize}\}$ and $\{\beta^{FirmEarn}\}$ from a Regression of a Job-Transition Indicator on Worker, Firm, and Job Characteristics (Equation 2.10)

Variable	Category	Estimate
Origin Firm Size Decile	1	0.0152***
		(0.000136)
Origin Firm Size Decile	2	0.00763***
		(0.000128)
Origin Firm Size Decile	3	0.0224***
		(0.000127)
Origin Firm Size Decile	4	0.0309***
		(0.000127)
Origin Firm Size Decile	5	0.0321***
		(0.000126)
Origin Firm Size Decile	6	0.0289***
		(0.000126)

Table 2.1(c), cont'd: Coefficients on Firm Size and Firm-Average Worker Earnings Deciles ($\{\beta^{FirmSize}\}$ and $\{\beta^{FirmEarn}\}$ from a Regression of a Job-Transition Indicator on Worker, Firm, and Job Characteristics (Equation 2.10)

Variable	Category	Estimate
Origin Firm Size Decile	7	0.0268***
		(0.000125)
Origin Firm Size Decile	8	0.0228***
		(0.000124)
Origin Firm Size Decile	9	0.0179***
		(0.000122)
Origin Firm Size Decile	10	
Origin Firm Earnings Decile	1	0.0587***
		(0.000147)
Origin Firm Earnings Decile	2	0.0345***
		(0.000145)
Origin Firm Earnings Decile	3	0.0311***
		(0.000138)
Origin Firm Earnings Decile	4	0.021***
		(0.000132)
Origin Firm Earnings Decile	5	0.0175***
		(0.000128)
Origin Firm Earnings Decile	6	0.015***
		(0.000125)
Origin Firm Earnings Decile	7	0.0098***
		(0.000121)
Origin Firm Earnings Decile	8	0.00832***
		(0.000117)
Origin Firm Earnings Decile	9	0.000834***
		(0.000112)
Origin Firm Earnings Decile	10	

Notes: This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

* means significant at the 10% level. ** means significant at the 5% level. ***means significant at the 1% level. Standard errors in parentheses.

Average Distance Traveled

Although there is no clear time trend in the probability of a worker changing jobs, it is possible that there is a change in observable behavior or outcomes among the set of job changers. As one method to study this, we estimate the following equation:

$$Distance_{i,j,t,k,t+1} = \gamma_0 + D^{Year}\gamma^{Year} + D^{Age}\gamma^{Age} + D^{Earn}\gamma^{Earn} + D^{FirmSize}\gamma^{FirmSize} + D^{FirmEarn}\gamma^{FirmEarn} + D^{Industry}\gamma^{Industry} + D^{Tract}\gamma^{Tract} \tag{2.11}$$

$Distance_{i,j,t,k,t+1}$ is the expected distance a worker i travels when he transitions from working at firm j in year t to firm k in year $t+1$. The right hand side variables are almost identical to (2.10) as many of the confounding factors are the same. The new variable is $FirmSize_{k,t+1}$ that represents the firm size decile of the destination firm k in $t+1$. The size of the destination firm may capture possible recognition effects or advertising/recruiting strategies that may vary with firm size. The location of the origin job as measured by the tract of the worker's establishment is meant to control for differences in access to opportunities that exist across different locations; some tracts may be located in agglomeration economies where job transitions require relatively little travel.

The estimated coefficients on the year indicators from equation (2.11) are presented in Table 2.2(a), while the coefficients on the design matrices capturing worker age and earnings and firm size and average earnings are presented in Tables 2.2(b) and 2.2(c). As expected, younger and richer workers travel further than older or poorer workers. Larger firms seem to recruit over greater distances. Workers at the 30% lowest paying firms seem to move farther than do workers at firms with pay at or above the 80th percentile of firm earnings. However, there is very little change in the year indicators $\{\gamma^{Year}\}$ over the period considered, with the exception of a moderate increase in average distance traveled in the last two years 2009-2010 corresponding to the depth of the Great Recession.

Table 2.2: Trends in the Distance Between Jobs Among Job-to-Job Transitions

Table 2.2(a): Coefficients on Year Indicator Variables $\{\beta^{Year}\}$ from a Regression of Distance Between Origin and Destination Establishments for Job-to-Job Transitions on Worker, Firm, and Job Characteristics (Equation 2.11)

Variable	Category	Estimate
Year	1996	-2.23***
		(0.0531)
Year	1997	-3.01***
		(0.0521)
Year	1998	-2.96***
		(0.0514)
Year	1999	-2.03***
		(0.0505)
Year	2000	-1.93***
		(0.0498)
Year	2001	-1.99***
		(0.0504)
Year	2002	-2***
		(0.0511)
Year	2003	-1.95***
		(0.0511)
Year	2004	-2.22***
		(0.0509)
Year	2005	-2.63***
		(0.0508)
Year	2006	-2.44***
		(0.0507)
Year	2007	-2.6***
		(0.051)
Year	2008	-2.36***
		(0.0522)
Year	2009	-1.2***
		(0.0533)
Year	2010	

Notes: This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year and changed employers; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

Notes for Table 2.2(a), cont'd: * means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. Standard errors in parentheses.

Table 2.2(b): Coefficients on Earnings Decile and Age Group Indicator Variables ($\{\beta^{Earn}\}$ and $\{\beta^{Age}\}$) from a Regression of Distance Between Origin and Destination Establishments for Job-to-Job Transitions on Worker, Firm, and Job Characteristics (Equation 2.11)

Variable	Category	Estimate
Worker Earnings Decile	1	-11.5***
		(0.0558)
Worker Earnings Decile	2	-11.8***
		(0.0561)
Worker Earnings Decile	3	-11.8***
		(0.0567)
Worker Earnings Decile	4	-11.8***
		(0.0575)
Worker Earnings Decile	5	-12***
		(0.0587)
Worker Earnings Decile	6	-11.8***
		(0.06)
Worker Earnings Decile	7	-10.5***
		(0.0611)
Worker Earnings Decile	8	-7.98***
		(0.062)
Worker Earnings Decile	9	-5.46***
		(0.0628)
Worker Earnings Decile	10	
Worker Age Decile	1	3.19***
		(0.0539)
Worker Age Decile	2	7.49***
		(0.0533)
Worker Age Decile	3	5.06***
		(0.0544)
Worker Age Decile	4	3.51***
		(0.0553)
Worker Age Decile	5	1.9***
		(0.0557)
Worker Age Decile	6	0.858***
		(0.0564)

Table 2.2(b), cont'd: Coefficients on Earnings Decile and Age Group Indicator Variables ($\{\beta^{Earn}\}$ and $\{\beta^{Age}\}$) from a Regression of Distance Between Origin and Destination Establishments for Job-to-Job Transitions on Worker, Firm, and Job Characteristics (Equation 2.11)

Variable	Category	Estimate
Worker Age Decile	7	0.6***
		(0.0579)
Worker Age Decile	8	0.685***
		(0.061)
Worker Age Decile	9	0.345***
		(0.0668)
Worker Age Decile	10	

Notes: This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year and changed employers; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. Standard errors in parentheses.

Table 2.2(c): Coefficients on Firm Size and Firm-Average Worker Earnings Deciles ($\{\beta^{FirmSize}\}$ and $\{\beta^{FirmEarn}\}$) from a Regression of Distance Between Origin and Destination Establishments for Job-to-Job Transitions on Worker, Firm, and Job Characteristics (Equation 2.11)

Variable	Category	Estimate
Destination Firm Size Decile	1	-6.38***
		(0.0431)
Destination Firm Size Decile	2	-12.8***
		(0.0444)
Destination Firm Size Decile	3	-11.6***
		(0.0442)
Destination Firm Size Decile	4	-9.71***
		(0.0444)
Destination Firm Size Decile	5	-8.32***
		(0.0446)
Destination Firm Size Decile	6	-6.48***
		(0.0449)
Destination Firm Size Decile	7	-5.99***
		(0.0452)

Table 2.2(c), cont'd: Coefficients on Firm Size and Firm-Average Worker Earnings Deciles ($\{\beta^{FirmSize}\}$ and $\{\beta^{FirmEarn}\}$) from a Regression of Distance Between Origin and Destination Establishments for Job-to-Job Transitions on Worker, Firm, and Job Characteristics (Equation 2.11)

Variable	Category	Estimate
Destination Firm Size Decile	8	-5.27***
		(0.0454)
Destination Firm Size Decile	9	-3.91***
		(0.0461)
Destination Firm Size Decile	10	
Origin Firm Earnings Decile	1	-2.45***
		(0.0544)
Origin Firm Earnings Decile	2	-2.73***
		(0.0564)
Origin Firm Earnings Decile	3	-2.52***
		(0.0561)
Origin Firm Earnings Decile	4	-2.88***
		(0.0558)
Origin Firm Earnings Decile	5	-3.92***
		(0.0561)
Origin Firm Earnings Decile	6	-4.67***
		(0.0563)
Origin Firm Earnings Decile	7	-4.5***
		(0.0568)
Origin Firm Earnings Decile	8	-2.95***
		(0.0571)
Origin Firm Earnings Decile	9	-0.78***
		(0.0565)
Origin Firm Earnings Decile	10	

Notes: This sample comes from all job-year observations from 1996-2010 from a worker's highest paying job in a given year where the worker was employed in at least one quarter of the following year and changed employers; the sample was limited to worker-years where we had location information for both the present and subsequent years and the tract was connected to the largest connected graph of tracts in the state.

* means significant at the 10% level. ** means significant at the 5% level. *** means significant at the 1% level. Standard errors in parentheses.

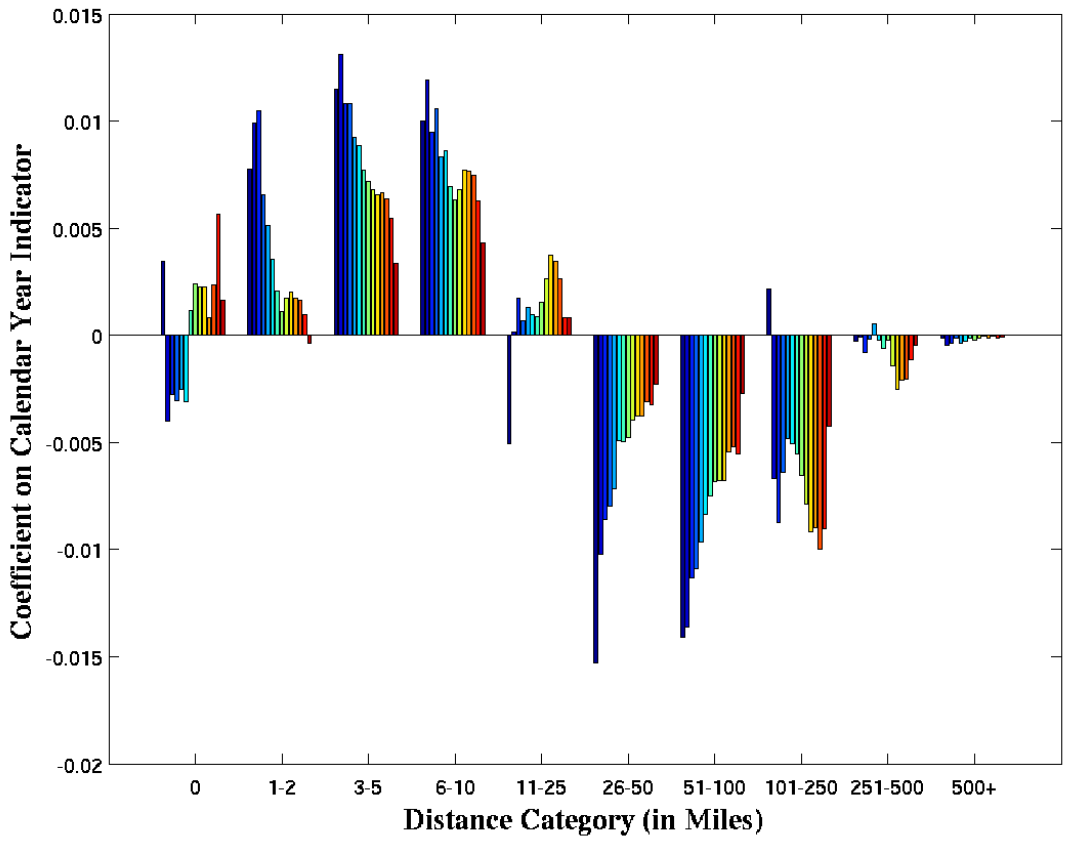
Differences in the Distance Distribution

The lack of change in the average distance traveled may conceal offsetting changes in the distribution of distances traveled by job changers. To examine changes in the relative frequency of moves of different lengths, we run a series of ten regressions that have the following form:

$$Prob(Distance_{i,j,t,k,t+1} \in Bin_l) = \alpha_{l0} + D^{Year}\alpha_l^{Year} + D^{Age}\alpha_l^{Age} + D^{Earn}\alpha_l^{Earn} + D^{FirmSize}\alpha_l^{FirmSize} + D^{FirmEarn}\alpha_l^{FirmEarn} + D^{Industry}\alpha_l^{Industry} + D^{Tract}\alpha_l^{Tract} \tag{2.12}$$

In each of the 10 regressions, the left hand side variable, $Prob(Distance_{i,j,t,k,t+1} \in Bin_l)$ is the probability the distance a worker i travels when he transitions from working at firm j in year t to firm k in year $t+1$ falls in the specified distance range that constitutes bin l and equal to zero otherwise. The ten bins are: 0 miles (a within tract transition), $0 < Distance \leq 2$, $2 < Distance \leq 5$, $5 < Distance \leq 10$, $10 < Distance \leq 25$, $25 < Distance \leq 50$, $50 < Distance \leq 100$, $100 < Distance \leq 250$, $250 < Distance \leq 500$, and $Distance > 500$. The set of independent variables is identical to the set used in equation (2.11). The coefficients of interest are again those on the (categorical) year variables. In order to more easily represent these 140 coefficients, we depict them in graphical form in Figure 2.2, where there are 14 bars for each distance bin, corresponding to the coefficients ($\alpha_l^{1996} - \alpha_l^{2009}$) from the years 1996-2009. The coefficient on the 2010 year indicator in each regression, α_l^{2010}, is set to zero to create an omitted category. Thus, there is an implicit bar for 2010 at the right end of the bar graph for each distance bin.

Figure 2.2: Time Trends in Estimated Coefficients on $Prob(Distance_{i,j,t,k,t+1} \in Bin_l)$



Notes: This figure depicts the coefficients on the year indicator variables from the regressions in (2.12). There are 14 bars for each of the ten distance categories; from left to right, the bars represent the estimated coefficient of a regression of the probability of changing jobs on year indicator-variables from 1996-2009 (2010 is the omitted category) with controls. The sample in this regression was 85,000,000 job pairs for over 20,000,000 unique workers.

The figure reveals a relatively clear trend toward job-transitions where the origin and destination tract are between 25 and 100 miles apart, with a corresponding clear trend away from job transitions where the origin and destination tracts are between 0 and 10 miles apart. However, we also see less consistent trends toward 0 distance (within-tract transitions) and away from long distance (100+ mile) transitions. These latter trends offset the increase in average distance suggested by the first two trends, so that linear distance-based specifications such as (2.11) fail to capture the changes in the scope of

local labor markets.

Taken together, Figure 2.2 suggests that to the extent that geographic labor market integration is occurring, it is occurring within geographic areas corresponding more closely to counties than to states or regions (which prevents these trends from being detected in datasets like the ACS that only record county or state of employment). The lack of evidence for state-level or region-level integration is consistent with Molloy et al. [2011], Molloy et al. [2014], and Hyatt et al. [2015]

2.4.2 Counterfactuals

The regressions presented in the previous section suggest that considerable geographic integration may have taken place at the sub-state level, with job-to-job transitions between jobs 25-100 miles away replacing job-to-job transitions between jobs 1-10 miles away. The regression approach to documenting integration has several advantages. Its minimal computational burden allows us to control for a wider array of worker characteristics and origin and/or destination firm, and job characteristics, allowing us to partially additional dimensions of compositional changes in supply and demand that might otherwise be mistaken for integration (e.g. an aging population that will be less likely to make long distance transitions, all else equal). Its parameters are easily interpreted, and the addition and removal of controls provides guidance as to which compositional changes are of paramount importance. Indeed, the results from the previous subsection suggest that the change in the distance distribution among job-to-job changes is unlikely to be fully attributable to changes in worker age, firm sizes, industry structure, or the earnings distribution. Furthermore, the change in the distance distribution does not appear to be concentrated among a particular part of the age, earnings, or firm size distributions.

However, the time-invariant coefficients on these various controls are not structural parameters; The extent to which a given characteristic predicts the next step in an indi-

vidual's career path depends on the entire distribution of preferences, costs, and match productivities among all current workers and potential employers. Thus, we might expect the conditional correlation between, say, worker age and distance past and current employers to change even if the match-specific productivities, tastes, and search/recruiting costs of each origin-destination job match type have not changed. If there are more older workers and fewer young workers, then we might expect older workers would need to migrate more to fill openings farther away, since more mobile younger workers have become more scarce.

By contrast, as emphasized in Section 2.3, the construction of a counterfactual set of labor market transitions based on an initially observed JPMF can eliminate *any* changes in the composition of worker characteristics, product demand, or firm cost structures that preserve the relative match values among origin job types for each destination job type. Thus, the decomposition approach, by capturing all of the information about relative match values contained in the matrix of job-type-to-job-type transitions in the set of $\{\theta_{od}\}$ parameters, allows us to better distill the part of the evolution of the labor market that represents true changes in relative worker tastes and productivities and, importantly for our purpose, search/recruiting/moving costs.

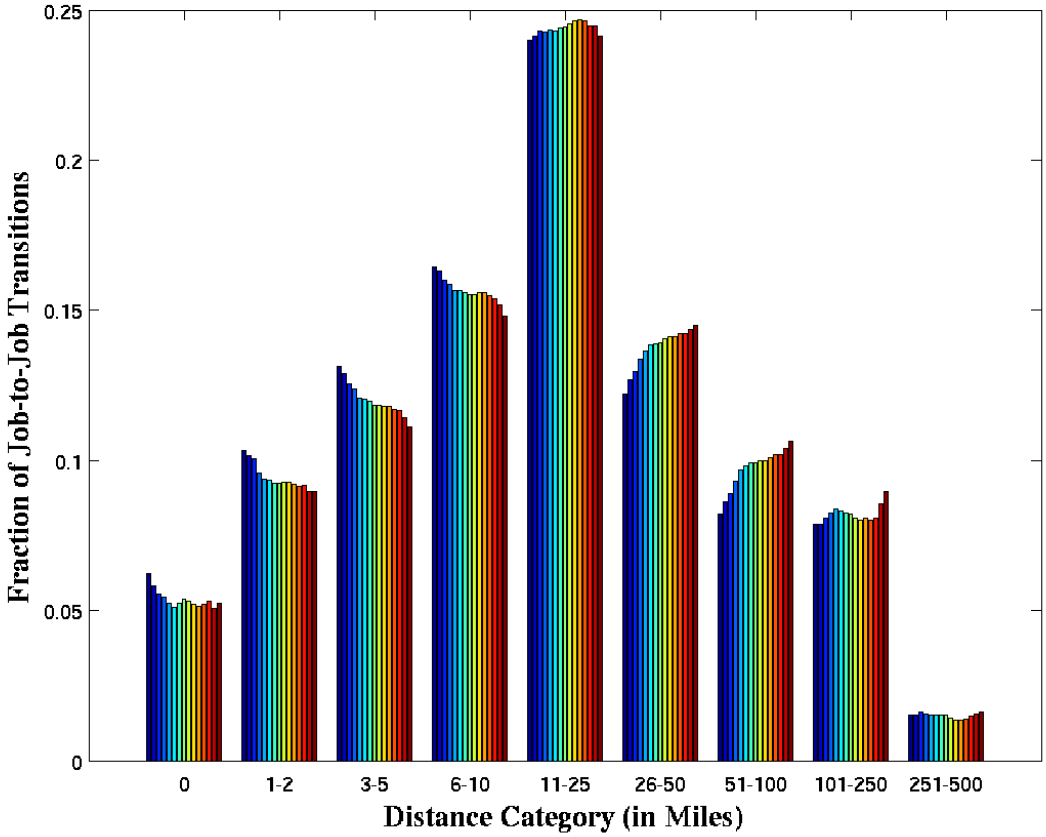
Thus, in this subsection we present results from a set of counterfactuals based on the decompositions described in Section 2.3. Because we are particularly interested in the importance of job locations in determining labor market transitions, we define a job type (either origin or destination) by the Census tract of the establishment. Choosing as small a geographic unit as a Census Tract is vital, since the regressions above suggested that much of the geographic labor market integration that has occurred over the last fifteen years has been not only within state but within county. However, defining location at the tract level already creates a very large type space, and computational feasibility precludes the addition of other characteristics to the type space that might capture differences

in productivity, tastes, or search/moving costs. However, subject to the caveats just mentioned, the regression results above suggest that accounting for changes in the worker earnings and age distribution and in the firm industry and size distribution may not be critical. Furthermore, only within-tract changes in employment composition create concern: if each neighborhood displays roughly the same employment composition over time, but some neighborhoods are rising or falling in employment levels, holding fixed the origin and destination job type marginals will net out the impact of these between-tract compositional changes. Finally, recall that these counterfactuals need not perfectly remove compositional changes in supply and demand; rather, they only need to do a sufficiently good job so that the key underlying trends in structural parameters are revealed.

Given this tract-based type space, we construct counterfactual JPMFs of job-to-job transitions of the form $P_y^{2010,2010}(o,d)$ for each $y \in [1995, 2010]$. In other words, we use the observed JPMFs for 1995 to 2010 to estimate the set of $\{\theta\}^y$ for each year from 1995-2010, then convert them to comparable counterfactual JPMFs by fixing the marginal origin and destination job type distributions for each counterfactual to the distributions observed at the beginning and end of 2010: $f^{2010}(*)$ and $g^{2010}(*)$.

In Figure 2.3 we present a set of bar graphs comparable to those from Figure 2.2 that display the distribution of distances between employers among job-to-job transitions that would have been observed in 2010, given the set of employment locations at the beginning and end of the year, if the relative propensities for employers from each tract to hire existing workers from each alternative tract had stayed as they were in each year, 1996-2010. While the same basic patterns observed in Figure 2.2 appear in Figure 2.3 as well, removing the influence of changes in the tract composition of origin and destination jobs smooths out the results and better reveals the underlying trend toward integration. First, there is a consistent trend toward 10-25 and 25-100 mile transitions. The fraction of transitions that fall into these distance bins increases by about 5 percentage points

Figure 2.3: Counterfactual Estimates of the Distribution of Distances Between Employers Among Job-to-Job Transitions for Each Year 1996-2010 (Holding Composition of Origin and Destination Job Types Fixed at 2010 Marginal Distributions)



Notes: This figure depicts counterfactual estimates of the distribution of distances between employers among job-to-job transitions for each year: 1996-2010. A counterfactual joint PMF capturing the matching function between origin and destination job types was estimated for each year. Each counterfactual joint PMF $P(o, d, \theta^y, f^{2010}, g^{2010})$ employs the marginal distributions of origin and destination job types from 2010, but uses the parameters capturing the relative propensities for certain (origin,destination) transition types $\{\theta^y_{od}\}$ from the chosen year y. See Section 2.3 for details about how the counterfactual joint PMFs are constructed. Each group of bars within the figure corresponds to a different distance bin (indicated on the X-axis) The height of each bar represents the simulated fraction of job-to-job transitions in a given year (left to right: 1996-2010) whose distance between origin and destination employers falls within the chosen bin.

between 1996 and 2010, (a 24% change relative to the 1996 base). Second, there is a corresponding trend away from 1-10 mile transitions. The fraction of transitions that fall into these distance bins decreases by about 5 percentage points (a 13% change relative to the 1996 base). Third, there are only minor and inconsistent trends for within-tract moves and long distance (100+ mile) moves.

2.5 Conclusions

This paper 1) develops a two-sided decomposition procedure for matching markets that distills changes in the relative propensities for particular match types to take place from changes in the composition of types on either side of the matching market, and 2) uses this procedure to provide compelling evidence that the U.S. labor market has become more geographically integrated in the years since the rise of the internet. The manifestation of this integration takes the form of a consistent decreasing trend away from short distance (1-10 mile) job-to-job transitions along with an increasing trend toward medium distance (25-100 mile) job-to-job transitions. Interestingly, we find little evidence of any consistent trend in job-to-job transitions featuring a long distance (100+ mile) between the origin and destination employer's locations. This is consistent with the results on interstate residential and employment mobility uncovered by Molloy et al. [2011] and Hyatt et al. [2015]. Thus, most of the integration seems to be taking place at a county or state scale, rather than a national scale. While these trends are consistent with decreasing geographic search costs for employees and recruiting costs for employers (perhaps combined with substantial moving costs for cross-MSA or cross-county residential moves), note that our decomposition methodology, while perhaps more strongly suggestive than other descriptive evidence, nonetheless does not provide a direct causal link between the adoption of information technology and geographic labor market integration.

As a final note, while our decomposition approach was motivated by and designed to address the problem of identifying integration of local labor markets and analyzing its implications for local development policy, we wish to highlight the generalizability of our methodology to a much wider array of questions in matching markets. For example, if we define types in the marriage market as (gender,education level) or (gender,race) pairs, we can determine whether there has been any change in the underlying propensity for interracial or inter-education group marriages to happen. The dramatic changes in the race and education compositions of both men and women make it very difficult (in the absence of our decomposition) to distill changes in underlying distributions of tastes from changes in marriage patterns that driven purely by changing composition on both sides of the market.

Similarly, suppose one defines types in the college market by the (SAT category, parental income category) cell of the student on one side and by the (Barron's selectivity category, tuition category) cell of the college on the other side. Then one can determine whether students from poorer backgrounds have become more willing or able to attend elite universities, after removing the confounding effects of both changes in the academic ability of students from poorer backgrounds along with changes the distributions of tuitions charged across selectivity categories.

Indeed, our decomposition can be useful in any matching market context in which 1) the compositional changes of principal concern on either side of the matching market can be captured or effectively proxied by observable categorical variables, and 2) there are enough observed matches to properly populate a transition matrix.

REFERENCES

J. M. Abowd, B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. The LEHD infrastructure files and the creation of the quarterly workforce indicators. In T. Dunne, J. B. Jensen, and M. J. Roberts, editors, *Producer Dynamics: New Evidence from Micro Data*, pages 149–230. University of Chicago Press, 2009.

E. N. Buntrock. Comparison-group position and job search: Information or irritation? Working paper, 2015.

J. DiNardo, N. M. Fortin, and T. Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64:1001–1044, 1996.

N. Fortin, T. Lemieux, and S. Firpo. Decomposition methods in economics. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4A, pages 1–102. Elsevier, 2011.

J. Haltiwanger, H. Hyatt, and E. McEntarfer. Cyclical reallocation of workers across employers by firm size and firm wage. Mimeo, 2014.

V. J. Hotz and R. A. Miller. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60:497–529, 1993.

H. Hyatt, E. McEntarfer, K. Ueda, and A. Zhang. The recent decline in residential mobility and job switching. Working paper, 2015.

H. R. Hyatt and J. R. Spletzer. The recent decline in employment dynamics. *IZA Journal of Labor Economics*, 2(5), 2013.

G. Kaplan and S. Schulhofer-Wohl. Understanding the long-run decline in interstate migration. NBER Working Paper 18507, 2012.

K. Kroft and D. G. Pope. Does online search crowd out traditional search and improve matching efficiency? Evidence from craigslist. *Journal of Labor Economics*, 32:259–303, 2014.

P. Kuhn and H. Mansour. Is Internet job search still ineffective? *The Economic Journal*, 124:1213–1233, 2014.

J. A. F. Machado and J. Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20:445–465, 2005.

A. Manning and B. Petrongolo. How local are labor markets? Evidence from a spatial job search model. Working paper, August 2013.

R. Molloy, C. L. Smith, and A. Wozniak. Internal migration in the United States. *Journal of Economic Perspectives*, 25(3):173–196, 2011.

R. Molloy, C. L. Smith, and A. K. Wozniak. Declining migration within the U.S.: The role of the labor market. NBER Working Paper 20065, 2014.

E. Moretti. Local multipliers. *American Economic Review*, 100(2):373–377, 2010.

E. Moretti. Local labor markets. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4B, pages 1237–1313. Elsevier, 2011.

M. J. Notowidigdo. The incidence of local labor demand shocks. Working paper, March 2013.

S. Rosen. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82:34–55, 1974.

C. Rothe. Decomposing the composition effect: The role of covariates in determining between-group differences in economic outcomes. *Journal of Business & Economic Statistics*, 2015. Forthcoming.

CHAPTER 3

CHANGES IN GEOGRAPHIC ENDOGAMY IN THE US: 1970-2000

Evan Buntrock

Abstract: While racial, socioeconomic, and religious endogamy (marriage of two individuals within the same group) have been extensively documented, there has been little attention given to geographic endogamy. I use data from the 1970, 1980, 1990, and 2000 Decennial Censuses of Population and Housing to quantify geographic endogamy in the US and explore factors correlated with the probability that an individual enters into a geographically endogamous marriage. I find that geographic endogamy in the US has declined from 1970-2000, and that an increase in a man's education is correlated with a decrease in his likelihood of marrying someone born in the same state as him.

3.1 Introduction

Are individuals more or less likely to marry someone from their same place of birth than they were thirty years ago? Individuals have become increasingly likely to marry someone from their same socioeconomic, religious, educational, or racial group. To the extent that place of birth is a proxy for desirable cultural features, we might expect a commensurate increase in geographic endogamy. However, the fall in transportation costs and the rise of online dating have made it easier for individuals to find and marry partners from different regions. Using data from the long form of the 1970, 1980, 1990, and 2000 Decennial Censuses, I find that individuals have become increasingly likely to move across states and marry individuals born in different states; this is true even after conditioning on a variety of variables including the individual's education. Further, for those who marry someone born in a different state, the distance between the origin states of couples is growing over time.

In this paper I study the integration of "marriage markets", a concept first advanced in economics by Becker [1973, 1974]. Fraboni [2000] is one of few authors who attempt to formally define the key features of such a market, describing it as, "The place of interaction between the sexes at the moment of the search for a partner: there, each individual neither represents a pure object nor a pure acquirent, but he/she plays both roles at the same time, so that a double choice, a double consent must be verified." A second key feature of the marriage market is that marital surplus is generated from trade between the husband and wife.[1] One party (traditionally the woman) possessed a comparative advantage in household production while the other party (traditionally the man) possessed a comparative advantage in labor market production. Becker theorized that two parties entered into marriage in order to benefit from trading with each other.[2]

[1]For this paper, I will consider only 'traditional marriages' between a man and woman, although most of the principles discussed by Becker could apply to same-sex marriages as well.

[2]Becker argued that love and the production of own children were the core reasons for marriage, since household and labor market products could be traded by unmarried individuals. This observation

This prediction has ample empirical support; as one example, Blau et al. [2000] find that marriage rates fall as women's labor market participation increase (which is generally driven by a rise in the relative wages paid to women and men).

Even with comparative advantage and trade, Becker anticipated endogamy (marriage between two members of the same group)[3] owing to individuals' tastes. Subsequent analysis has supported this prediction; Kalmijn [1998] reviews the ample evidence of endogamy along the lines of race, religion, education, and socio-economic status.

Geographic endogamy remains comparatively under examined. Van De Putte [2003] looks at changes in geographic endogamy in several Flemish cities and Fraboni [2000] studies geographic homogamy in Italy from approximately 1939 to 1995. To my knowledge, no one has attempted to study geographic heterogamy (marriage between two members of different groups) using US data.

It is not immediately clear whether geographic heterogamy in the US would be increasing or decreasing over time. If an individual's place of birth is a proxy for cultural features that he wants his partner to share, then we might expect an increase in endogamy along this dimension comparable to what has been observed in socioeconomic endogamy. Conversely, if birthplace is more important to individuals because of the restrictions it places on the area of a person's search for a partner, then we might expect a decrease in geographic endogamy as individuals become more aware of and able to move to better opportunities. As transportation costs have fallen (Glaeser and Kohlhase [2004]), more people should benefit from moving between marriage markets; therefore, there should be increasing geographic heterogamy in successive cohorts as moving between markets becomes less costly.[4]

anticipates one of the potential confounds discussed later: cohabitation.

[3]In the literature, marriage to someone of similar socio-economic status is usually referred to as homogamy. However, since this also qualifies as endogamy, I will use endogamy from this point forward to refer to marriage of two parties who share any group or characteristic.

[4]Unless the marriage markets are identical in all states; however, the work of Fraboni [2000], Siow [2008], and others suggests this is not the case.

There are many reasons individuals might move into new marriage markets. They might enter a new marriage market only as a consequence of moving to attend school or find a job. However, the work of Fraboni [2000], Van De Putte [2003], and Siow [2008] suggests that conditions in local labor and marriage markets can impact intra-national migration to new marriage markets, the timing of individuals' entry into the marriage market, and how surplus generated by marriages is shared.

While it is difficult to know exactly why individuals migrate to new markets, evidence of increasing geographic heterogamy would be consistent with a model where individuals can select a market to enter and enjoy falling entry costs. In order to quantify the geographic integration of marriage markets, I will use data from the 1970, 1980, 1990, and 2000 Decennial Censuses of Population and Housing to answer the following questions:

1. How much has the percentage of married people who migrated within the US increased?
2. How does the ratio of geographically heterogamous marriages to geographically endogamous marriages change over time?
3. Has the average distance between a couple's places of birth increased over time?
4. What characteristics are positively correlated with geographic heterogamy?

In Section 3.2, I discuss the data. In Section 3.3, I restate my research questions and discuss my methods, and review some potential analytical and empirical problems that are not resolved in this paper. Section 3.4 summarizes my results. Section 3.5 concludes.

3.2 Data

My primary data source is the long form questionnaires of the 1970, 1980, 1990, and 2000 Decennial Censuses of Population and Housing. I use the public samples provided by the

Integrated Public Use Microdata Series (IPUMS). IPUMS provides person-level data on the age, race, education, income, state of residence, state of residence five years prior to the census year, and state of birth of an individual and her spouse. This state of birth information will allow me to determine whether a marriage is geographically endogamous or heterogamous. Combining state of birth with information on the state of residence and state of residence five years prior will let me quantify individual mobility. The other demographic information such as age, race, and education will help me to control for characteristics associated with geographically heterogamous marriages.

My sample is all American born respondents to long forms of the 1970-2000 Decennial Censuses who were between the ages of 15-65 at the time of the Decennial Census in which they were surveyed. All individuals under 15 are unlikely to be "available" for marriage over the period of 1970-2000, and I am concerned that selection out of the sample (or out of marriage) would cause significant bias if I were to include individuals older than 65. Like Fraboni [2000], I exclude any households where a husband or wife is an immigrant (identified when either the husband or wife registers a place of birth outside of the US); they may have faced different marriage market conditions than American born citizens.

Table 3.1 offers a summary of the data from the 1970-2000 Decennial Census long forms for all married couples where both parties were American born citizens between 15 and 65 years old. Using the person weights provided by IPUMS, this represents a total of 13,017,062 married couples.

Table 3.1: Summary Statistics of Married Individuals from 1970-2000

Variable	Mean	Std Dev
Census year	1982.31	62.07206
State (FIPS code)	28.83028	81.44646
Age	41.06355	63.50944
Sex	1.499401	2.664565
Race [general version]	1.164272	3.846807
Birthplace [general version]	31.53757	93.21092
State or Country of Residence 5 years ago	19.99621	165.9885
Age [of spouse]	41.066	63.51763
Sex [of spouse]	1.500599	2.664565
Race [of spouse; general version]	1.164218	3.846433
Birthplace [of spouse; general version]	31.53745	93.20175
State or Country of Residence 5 years ago [of spouse]	19.99566	165.9866
Elementary School Education	0.02492	0.830712
Middle School Education	0.072002	1.377533
High School Education - No Diploma	0.371297	2.57478
High School Diploma or GED	0.128816	1.785237
1-2 Years of College - No Degree	0.173446	2.017783
Associate's Degree	0.030951	0.922923
3-4 Years of College - No Degree	0.094289	1.557335
Bachelor's Degree	0.068101	1.342509
Master's, PhD, or Other Higher Degree	0.036179	0.995142
Elementary School Education (Spouse)	0.024916	0.830643
Middle School Education (Spouse)	0.072001	1.377521
High School Education - No Diploma (Spouse)	0.371284	2.574762
High School Diploma or GED (Spouse)	0.128832	1.785335
1-2 Years of College - No Degree (Spouse)	0.173421	2.017666
Associate's Degree (Spouse)	0.030966	0.923139
3-4 Years of College - No Degree (Spouse)	0.094289	1.557335
Bachelor's Degree (Spouse)	0.068107	1.342568
Master's, PhD, or Other Higher Degree (Spouse)	0.036185	0.995215
Married to Someone of Same Education Block	0.769066	2.245857
White	0.904489	1.566334
White (Spouse)	0.904543	1.565942
Black	0.076418	1.415771
Black (Spouse)	0.07635	1.415186
Native American	0.005316	0.387526
Native American (Spouse)	0.005337	0.388287
Chinese	0.000717	0.142631
Chinese (Spouse)	0.000716	0.142555
Japanese	0.002175	0.248274
Japanese (Spouse)	0.002174	0.248226

Table 3.1, cont'd: Summary Statistics of Married Individuals from 1970-2000

Variable	Mean	Std Dev
Other Asian	0.001545	0.209332
Other Asian (Spouse)	0.001539	0.208896
Two Races	0.002265	0.253353
Two Races (Spouse)	0.002265	0.253348
Three or More Races	0.000172	0.069973
Three or More Races (Spouse)	0.000171	0.069717
Married to Someone Born in the Same State	0.606124	2.603857
Married to Someone Born in a Different State	0.393876	2.603857
Moved in Last Five Years	0.654194	2.534697
Moved in Last Five Years (Spouse)	0.654203	2.534682
Moved Between Birth and Five Years Prior to Census	0.73455	2.3532
Moved Between Birth and Five Years Prior to Census (Spouse)	0.743197	2.328139
Has Moved Between Birth and Census	0.768406	2.2481
Moved Between Birth and Census (Spouse)	0.771135	2.238782
Education Category	4.381454	10.4945
Education Category (Spouse)	4.38154	10.49471
Married to Someone of the Same Race	0.980959	0.728329
Other Race	0.006901	0.441175
Other Race (Spouse)	0.006905	0.441292
Census Closest to Year of Birth	1941.82	85.98534
Census Closest to Year of Birth (Spouse)	1941.82	85.98046
Ten Year Birth Cohort	3.182341	8.598534
Ten Year Birth Cohort (Spouse)	3.182053	8.598046

Notes: This table gives the summary statistics (using person weights from the head of the house) of all married couples in the IPUMS extract of the long forms of the 1970, 1980, 1990, and 2000 Decennial Censuses. With person weights, the sample size is 13,017,062 couples where both the husband and wife are US-born citizens between 15 and 65 years old.

3.3 Methodology

To review time trends in geographic heterogamy in the US, I must concretely define geographic heterogamy. I define a marriage as geographically heterogamous if the married individuals were born in different states. This has three notable drawbacks. First, there is no reason to think that state of birth will be the actual marriage market; a marriage

market might be a city, a metropolitan statistical area (MSA), or an even smaller set of places (e.g. an individual's workplace, church, and neighborhood, etc.). Individuals might not think of entities as geographically large as states as one marriage market that they can search. Second, this definition treats a marriage between a man born in San Francisco and a woman born in Los Angeles as endogamous, but a marriage between a man born in Edison, New Jersey and a woman born in New York City as heterogamous. This seems like a mistake, since the distance separating the former group of partners is larger than that separating the latter group.[5] Third, I lack a complete history of where individuals lived and where they were living when they married. Thus, I am unable to distinguish a couple who meet and marry in their birthplace and two people who were born in the same state, meet in a different state, and return to live in their birthplace. While both marriages are technically homogamous with respect to state of birth, the latter union seems geographically heterogamous - at least more so than the former union.

In spite of these drawbacks, state is the finest level of geographical information I have on an individual's place of birth, and since the Census data don't say where individuals were located when they 'entered the marriage market' (i.e. began searching for a partner) or became married, comparing the states of birth for couples is my best possible measure of geographic heterogamy.

The first way I measure the integration of marriage markets is by quantifying internal migration in the US. For each cohort, I calculate the fraction of married individuals who have moved between their birth and five years before the Census, in the five years prior to the Census, or any time between their birth and the Census. Importantly, I do not have information on individuals' state of residence (and therefore additional movements) in other years. This means I am systematically under measuring the true amount of individuals' interstate movement over any given time period. However, I can see no reason that my measure's difference from the true level of migration would vary across

[5] Fraboni [2000] and Brien [1997] discuss the difficulties of defining a 'marriage market' in more detail.

census years. Thus, I should still be accurately capturing changes in the level of migration over time.

As a second measure of integration, I examine whether or not the ratio of geographically heterogamous to endogamous marriages has increased over time. If inter-state migrations are increasing over time, I would expect more geographically heterogamous marriages. However, it could be that individuals move after they have married, in which case heterogamous marriages would not be increasing even with a rise in inter-state migration.

As a third measure of integration, I measure the distance between geographically heterogamous couples' states of birth. This focuses on the intensive margin (how far individuals move to a new market) rather than the extensive margin (an individual's decision to move or not move). If transport costs decrease over time, it should be easier for people to travel farther when searching for new markets; this would lead to an increase over time in the average and/or median distance between the birth places of heterogamously married individuals.

In order to measure the "marriage distance" of a couple, I must assign specific coordinates to each individual's place of birth. As mentioned above, the Census data only contain the state of an individual's birth, and not the precise latitude and longitude of the location of their birth. I assign individual's coordinates of birth as follows. First, I identify the closest census year to the individual's year of birth. Second, I take the coordinates of the population centroid of the individual's birth state in that census year and assign those coordinates to individual as the latitude and longitude of his birth.[6] Let X_h, Y_h be the assigned latitude and longitude of the husband's birthplace and let X_w, Y_w be the assigned latitude and longitude of the wife's birthplace. Let D_m denote the measure of distance between the couple's birth coordinates. I calculate $D_m = \sqrt{(X_h - X_w)^2 + (Y_h - Y_w)^2}$.

[6]The population centroids of each state by census year are also provided by the US Census Bureau.

My fourth measure of integration controls for composition changes and identifies factors associated with a higher likelihood of entering a geographically endogamous marriage. Again, I examine only married individuals, but this time I include only men ages 20-29 married to women ages 15-35. This is an effort to isolate first marriages in each non-overlapping cohort. I use a logit model where the dependent variable equals one if the person is married to someone born in the same state and equals zero if the person is married to someone born in a different state. The model estimated has the form:

$$\ln \frac{P(SameStateMarr)}{1 - P(SameStateMarr)} = b_0 + b_1 cohort + b_2 educ + b_3 race + b_4 geo + b_5 move \quad (3.1)$$

The variable of interest is the "*cohort*" variable, which describes the individual's ten year birth cohort. The "*educ*" variable is a categorical variable that captures the individual's level of education. The "*race*" variable above is a categorical variable that describes the individual's race. The "*geo*" variable represents three categorical geographic variables: the individual's state of birth, state of residence five years prior to the census, and state of residence at the time of the census. The "*move*" variables are two binary variables that measure whether the individual moved between the time of his birth and five years prior to the census or between the five years prior to the census and the time of the census. Note that for all of these individual variables, I include the same variables for the individual's spouse. The inclusion of spousal controls reflects the joint decision aspect of the marriage.

The coefficients of interest are those on the individual's ten year birth cohorts; if these coefficients are decreasing over time, there is evidence of increasing geographic integration of marriage markets. I control for education and race since endogamy along these dimensions is well documented; a person may move in order to marry endogamously along these lines. I also hypothesize that a person's education may be inversely related to the likelihood an individual enters a geographically endogamous marriage; more educated people

might marry later and have more career options, both of which increase the likelihood an individual may leave her state of birth prior to marrying. I include state of birth to control for the fact that in some states it may be easier for a person to search marriage markets in other states (e.g. searching from Rhode Island compared searching from Montana). I do not control for an individual's income, primarily because once a person is married, their unobserved household arrangements will determine their income. If I could observe each party's income prior to entering the marriage market, I might include that as a control; however, a couple's marriage and individual labor supply decisions can reasonably be assumed to be endogenous.

By focusing only on married individuals in the above analysis, I hope to avoid confounds that have affected marriage in general, e.g. the decline of marriage rates and the rise in the median age at first marriage in the US that have occurred in the last several decades. I measure a trend within the trend of marriage; I study not the choice of whether to marry, but whether to marry someone born in the same state. Still, there are several confounds which could skew my results. First, I cannot perfectly measure the number of marriages that have taken place. I only observe an individual's marital status at the time of the survey, which means I don't observe prior marriages.[7] My results below will misrepresent the change in geographically endogamous marriages if the number of 'hidden' marriages (or the proportion that are geographically endogamous) is changing over time. Second, there is the possibility (indeed, the strong likelihood) of differences across birth cohorts in self-selection into and out of marriages. One example of this is the rise of cohabitation, which probably has significant effects on selection into and out of marriage.[8] Stevenson and Wolfers [2007] note that according to Census data on cohabitation starting in 1990, the rate of cohabitation among adults was 3.5% in 1990 and 5% in 2000.

[7]Although the 1970 and 1980 Censuses identify individuals who were once married but are no longer married at the time of the survey, I have no information on their past spouse. Therefore, I do not use the data on an individual's marital history in this paper.

[8]For a discussion of links between cohabitation and marriage, see Stevenson and Wolfers [2007].

Although I do not examine cohabitation in this paper, cohabiting couples may provide an interesting comparison group (relative to married couples) when studying trends in geographic heterogamy.

3.4 Results

As Table 3.2 makes clear, the number of people who have moved in the five years prior to the census or between their birth and five years prior to the census has increased over the years. The proportion of married adults that have migrated sometime between their birth and the census year has increased from 69% in 1970 to 87% in 1980, a jump of over 25%. In 1990, the proportion of the population who has migrated decreases from its 1980 level but is still significantly above the 1970 level. In 2000, the proportion of people who have migrated is still below the 1980 level but has increased from the 1990 level. Overall, the number of people who have moved across state lines since their birth has increased significantly from 1970 to 2000. I interpret this as evidence of increasing geographic integration.

The increased migration seen in Table 3.2 has occurred at the same time as an increase in geographically heterogamous marriages. In Table 3.3 there is a monotonic increase in the ratio from .57 in 1970 to .71 in 2000, which is almost a 25% increase over the period. The ratio increases even from 1980 to 1990, which is when we see a dip in the proportion of married individuals who moved to new states.

Table 3.2: Measures of Inter-State Migration: U.S. Decennial Census Data 1970-2000

Census Year	N Obs	(1) Moved Between Birth and Census	(2) Moved During the Five Years Prior to Census	(3) Moved Between Birth and Five Years Prior to Census
1970	1,338,848	0.68986	0.55607	0.65087
1980	3,787,244	0.87188	0.81379	0.85363
1990	3,881,204	0.76477	0.64743	0.72964
2000	3,849,724	0.78616	0.67334	0.75672

Notes: Column (1) is the proportion of married individuals who moved away from their birth state any time between their birth and the Census. Column (2) is the proportion of married individuals who moved out of their state of birth in the five years prior to the Census. Column (3) is the proportion of married individuals who moved out of their state of birth more than five years prior to the Census. The information in this table comes from the IPUMS extract of the long forms of the 1970, 1980, 1990, and 2000 Decennial Censuses. With person weights, the sample size is a combined 13,017,062 couples where both the husband and wife are US-born citizens between the ages of 15 and 65 years old.

Table 3.3: Ratios of Geographically Heterogamous to Geographically Endogamous Marriages

Census Year	Ratio of Heterogamous to Endogamous Marriages
1970	0.565007
1980	0.636973
1990	0.706088
2000	0.712273

Notes: This table gives the ratio of geographically heterogamous marriages (i.e. the partners were born in different states) to geographically endogamous marriages (partners born in same state). The information in this table comes from the IPUMS extract of the long forms of the 1970, 1980, 1990, and 2000 Decennial Censuses. With person weights, the sample size is a combined 13,017,062 couples where both the husband and wife are US-born citizens ages 15-65 years old.

The distribution of distances between the birth states of married couples has also shifted to the right. In Table 3.4(a), the mean value of distance between the birth states of all (endogamous and heterogamous) couples goes from roughly 7.0 in 1970 to 9.2 in 2000, which represents a 32% increase. Table 3.4(b) reveals that this increase is not simply an artifact of more people entering geographically heterogamous marriages; within the subset of heterogamous marriages, the distances between couples' states of birth are increasing. Table 3.5 also makes this clear; it shows the decile cutoffs of the distribution of distances between places of birth for all married couples. We can see that at every decile, the cutoff value is (weakly) monotonically increasing over time.

Table 3.4: Mean Distance Between Birth States of Married Couples by Census Year

Table 3.4(a): All Married Couples

Census Year	N Obs	Mean	Std Dev	Minimum	Maximum
1970	1,338,848	7.00672	18.9415	0	89.7962
1980	3,787,244	8.16313	20.3076	0	89.7962
1990	3,881,204	8.82448	20.9811	0	89.7962
2000	3,849,724	9.22695	21.3785	0	89.6785

Table 3.4(b): Heterogamously Married Couples

Census Year	N Obs	Mean	Std Dev	Minimum	Maximum
1970	480,152	19.4073	27.4946	0.39225	89.7962
1980	1,465,372	20.9759	28.1658	0.37216	89.7962
1990	1,560,458	21.86	28.4155	0.34019	89.7962
2000	1,553,512	22.7808	28.6632	0.33378	89.6785

Notes: Table 3.4(a) gives the mean distance between the states of birth of all married couples. Table 3.4(b) gives the mean distance between the states of birth of all geographically heterogamous couples. I calculate distance by assigning each individual in a couple the coordinates of the population centroid (as determined by the US Census Bureau) of their state of birth in the census year closest to the year of their birth. Let X_h, Y_h be the assigned latitude and longitude of the husband's birthplace and let X_w, Y_w be the assigned latitude and longitude of the wife's birthplace. Let D_m be our measure of distance between the couple's birth coordinates. I calculate $D_m = \sqrt{(X_h - X_w)^2 + (Y_h - Y_w)^2}$.

Table 3.5: Decile Cutoffs of Distance Between Birth States of Married Couples

Year	0th-40th Pctl	50th Pctl	60th Pctl	70th Pctl	80th Pctl	90th Pctl	100th Pctl
1970	0	0.034216	0.114257	2.88846	5.760677	13.41498	89.79622
1980	0	0.047222	0.243491	3.590575	7.155262	16.42168	89.79622
1990	0	0.052653	1.434916	4.48846	8.52723	19.88995	89.79622
2000	0	0.060925	1.641283	4.630702	9.276458	22.46118	89.67855

Notes: This table gives the decile cutoffs for the distribution of distances between the birth states of married couples. I calculate distance by assigning each individual in a couple the coordinates of the population centroid (as determined by the US Census Bureau) of their state of birth in the census year closest to the year of their birth. Let X_h, Y_h be the assigned latitude and longitude of the husband's birthplace and let X_w, Y_w be the assigned latitude and longitude of the wife's birthplace. Let D_m be our measure of distance between the couple's birth coordinates. I calculate $D_m = \sqrt{(X_h - X_w)^2 + (Y_h - Y_w)^2}$.

Finally, I turn to the regression results of Table 3.6. In the model estimated by (3.1), the parameters can be interpreted as the change in the log of the odds ratio that a given individual is married to someone born in the same state when the variable on the right hand side changes. This is not a very intuitive way to interpret the results. Rewrite the right hand side of equation (3.1) as $f(x) = bX$. This rearranges to the standard "logit" model:

$$Prob(SameStateMarr) = \frac{e^{f(x)}}{1 + e^{f(x)}} = \frac{e^{bX}}{1 + e^{bX}} \tag{3.2}$$

The coefficients on the individual's birth cohort and education from the vector b are represented in Table 3.6. Using the 1975-1984 birth cohort as a baseline, we can see that, *ceteris paribus*, being born in an earlier cohort is usually associated with a higher probability of being married to someone born in the same state. However, the coefficient on the 1965-1974 cohort is a notable exception: other things equal, this is the cohort correlated with the highest probability of a geographically endogamous marriage. Higher levels of education are correlated with decreasing probability of a geographically endogamous marriage, though again, the pattern is not monotone: individuals with associate's degrees are relatively more likely to enter a geographically endogamous marriage than those with 1-2 years of college. At this time, I offer no explanation for the non-monotonicity of the relationships between education or birth cohort and the probability of geographic endogamy.

Owing to the non-linearity of the logit function, the change in the probability of marrying someone born in the same state in response to a change in one of the independent variables will depend on the value of the other variables. Therefore, I focus more closely on a change in the probability of the event (the "marginal probability") for some particular cases of interest, namely what happens when we change a man's birth cohort or education level.

Table 3.6: Logit Results - Prob(Geographically Endogamous Marriage = 1)

Variable	Categorical Value	Coefficient
10 Year Birth Cohort	*1935-1944*	0.03614***
		(0.004584)
10 Year Birth Cohort	*1945-1954*	0.03184***
		(0.004356)
10 Year Birth Cohort	*1955-1964*	-0.0633***
		(0.003934)
10 Year Birth Cohort	*1965-1974*	0.04017***
		(0.003227)
10 Year Birth Cohort	*1975-1984*	omitted
Level of Education	*Elementary School*	-0.037***
		(0.005776)
Level of Education	*Middle School*	omitted
Level of Education	*HS - No Diploma*	-0.2071***
		(0.002676)
Level of Education	*HS - Diploma or GED*	-0.2197***
		(0.003323)
Level of Education	*1-2 Yrs College*	-0.3471***
		(0.002866)
Level of Education	*Associate's Degree*	-0.237***
		(0.004443)
Level of Education	*3+ Yrs of College*	-0.413***
		(0.003054)
Level of Education	*Bachelor's Degree*	-0.4224***
		(0.003964)
Level of Education	*Master's Degree or Higher*	-0.456***
		(0.006067)
Moved in 5 Years Prior		-0.3394***
		(0.002207)
Moved Between Time of Birth and 5 Years Prior		0.08628***
		(0.00224)

Notes: This table gives the coefficients from a logistic regression where the dependent variable is the log of the odds ratio of the probability that an individual is married to someone born in the same state divided by the probability that an individual is married to someone born in a different state. Additional controls (coefficients not shown) include an individual's and his spouse's race, their states of birth, their states of residence five years prior to the census, their state of residence during the census, as well as the spouse's birth cohort, education, and movement variables. The F-statistics for these controls are highly significant. This regression is performed on US-born males ages 20-29 married to US-born females ages 15-35.

Notes for Table 3.6, cont'd: *** means significant at the 1% level. Standard errors in parentheses.

Consider a handsome young white male from Minnesota named Ivan. Suppose Ivan was born in Minnesota during the years of 1935-1944 and lived there during the five years prior to the 1970 Census and lived in Minnesota at the time of the 1970 Census. Hold the characteristics of his wife fixed. I assume they are the same as his, initially, although I will let her birth cohort change with Ivan's. If Ivan has only a high school education, then the probability he has married someone born in the same state as him (conditional on his being married) is approximately 67.3%. If we held everything about Ivan fixed except his education, which we changed to a bachelor's degree, he would have a 62.7% chance of being married to someone who was also born in Minnesota. If we return to the case where Ivan has a high school education but he is born instead in 1965-1974, his probability of marrying someone born in the same state is 64.3%. If he is born in 1965-1974 and attains a bachelor's degree, the probability he marries someone born in the same state is approximately 59.5%. Thus, both higher levels of education and later birth cohort are associated with a noticeable increase in the likelihood Ivan has married someone from a different state.

On the whole, my results suggest that geographic endogamy is more a consequence of the limitations on individuals' abilities to move between marriage markets rather than a preference for a partner with the same place of birth. As time has passed, moving costs have fallen. The fraction of married individuals who have changed states has generally increased across Censuses, and the distribution of the distance between the origin states of married couples has only shifted rightward. Similarly, the fraction of marriages that are geographically heterogamous strictly increased from 1970-2000.

The results of a logistic regression suggest that the rise of heterogamous marriage is not due strictly to a change in the composition of the characteristics of married individuals;

controlling for demographic characteristics of younger, married men and their partners, we see that men born in later birth cohorts are generally less likely to marry someone with whom they share a state of birth. However, the notable exception of the 1965-1974 birth cohort merits further investigation with different variable definitions (e.g. broader education categories), different data sets (e.g. the ACS), and new comparison groups (e.g. cohabiting couples) to more deeply test whether the results for this cohort are indeed an anomaly within a larger trend.

3.5 Conclusion

In this paper, I explore the change in the geographic integration of US marriage markets. My central idea is that a man (or woman) can choose which marriage market he (she) enters. If moving to new markets is costly, but those costs have fallen over time, I would expect to see an increase in individuals' movement between states (my proxy for marriage markets). I find that inter-state migration has increased over time, as has the ratio of geographically heterogamous and geographically endogamous marriages. I also find that the average distance between states of birth of married couples has increased over time, even when focusing only on couples where both parties were born in different states. As would be expected, increasing levels of education and being in a later birth cohort are associated with a decreasing likelihood of endogamous marriage (conditional on having made the decision to marry). While there are many factors that drive migration and marriage decisions, the data support the idea marriage markets have become more geographically integrated.

REFERENCES

G. S. Becker. A theory of marriage: Part I. *Journal of Political Economy*, 81:813–846, 1973.

G. S. Becker. A theory of marriage: Part II. *Journal of Political Economy*, 82:S11–S26, 1974.

F. D. Blau, L. M. Kahn, and J. Waldfogel. Understanding young women's marriage decisions: The role of labor and marriage market conditions. *Industrial & Labor Relations Review*, 53:624–647, 2000.

M. J. Brien. Racial differences in marriage and the role of marriage markets. *Journal of Human Resources*, 32:741–778, 1997.

R. Fraboni. *Marriage market and homogamy in Italy: An event history approach*. PhD thesis, Università degli Studi di Roma "La Sapienza," 2000.

E. L. Glaeser and J. E. Kohlhase. Cities, regions, and the decline of transport costs. *Papers in Regional Science*, 83:197–228, 2004.

M. Kalmijn. Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, 24:395–421, 1998.

A. Siow. How does the marriage market clear? An empirical framework. *Canadian Journal of Economics*, 41:1121–1155, 2008.

B. Stevenson and J. Wolfers. Marriage and divorce: Changes and their driving forces. *Journal of Economic Perspectives*, 21(2):27–52, 2007.

B. Van De Putte. Homogamy by geographical origin: Segregation in nineteenth-century Flemish cities (Gent, Leuven, and Aalst). *Journal of Family History*, 28:364–390, 2003.

APPENDIX A

APPENDIX TO "COMPARISON-GROUP POSITION AND JOB SEARCH: INFORMATION OR IRRITATION?"

A.1 Overview of the LEHD Data

The EHF is the backbone of the LEHD. For every job covered by unemployment insurance (UI), a state maintains a record containing a person identifier, a firm identifier, the amount of earnings the firm paid the worker, and the year and quarter in which the payments were made. Every state provides the Census Bureau (via the Local Employment Dynamics federal/state partnership) with the raw data from its UI administrative files, and these records form the EHF. In practice, these records exist for all non-farm jobs. Each UI wage record reports a worker-employer link; therefore, by virtue of the LEHD records being based on the existence of a worker-employer pair, the data are a job frame from which we can sample workers or firms.

The ECF provides more information on the firms. It is derived primarily from the ES-202/Quarterly Census of Employment and Wages (QCEW) report, which contains information about the number, location, and industry of a firm's establishments.

The ICF records workers' dates of birth, citizenship status, sex, race, ethnicity, and education.[1] It is derived primarily from demographic information provided by records from the Social Security Administration, decennial census, American Community Survey (ACS), Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP). Place of residence can also be linked from the other administrative data assembled at the Census Bureau; Schmutte [2015] describes this in more detail.

[1]There are some records for which this information is missing, most notably education, which is observed for only 12% of workers in the LEHD. To address the missing nature of the data, I use multiple imputation described by Little and Rubin [2002]. This procedure is described more in later in this appendix.

Because unrestricted access to the detailed data in the LEHD would pose a risk to the confidentiality of the workers and firms it covers, the LEHD data can be accessed only after a researcher writes a Census Bureau-approved research proposal and obtains security clearance; even then, restrictions exist on what results may published. Researchers interested in using the LEHD data products may start with the public Quarterly Workforce Indicators (QWI) data or contact the administrator of a Census-sponsored Research Data Center (RDC) about using the restricted-access LEHD.

A.2 Variable Definition and Construction

Below are the definitions of variables used in my tables, figures, and regressions.

1. θ_i - the individual fixed effect of worker i
2. $\psi_{J(i,t)}$ - the firm fixed effect. However, since calculating worker and firm fixed effects is probably asking too much of SAS, we'll using worker fixed effects for all workers, firm effects for the largest 5000 firms, and then NAICS and firm size variables for the rest of the firms.
3. $FirmSize_{J(i,t),t}$ - I assign a firm $J(i,t)$ into a size decile based on the average number of workers employed at $J(i,t)$ in year t. The average number of workers is obtained by taking the sum of the worker-full-quarter observations and dividing by four.
4. $FirmAvgEarnings_{J(i,t),t}$ is the average earnings decile of the firm $J(i,t)$ in year t, based on the average annualized earnings of workers at the firm.
5. $FirmLagEmpchange_{J(i,t),t-1,t}$ is the percentage employment change (growth or decline) of firm $J(i,t)$'s size from years $t-1$ to t.
6. $FirmLeadEmpchange_{J(i,t),t,t+1}$ is the percentage employment change (growth or decline) at firm $J(i,t)$ from years t to $t+1$.
7. $TotalEarn_{i,t}$ is worker i's total (all-job) earnings in year t. The $f(\cdot)$ denotes a cubic function.
8. $Age_{i,t}$ - The average age in years of the worker i across all quarters in year t. The $f(\cdot)$ denotes a cubic function.
9. $Exp_{i,t}$ - The average experience in years of the worker i across all quarters in year t. The $f(\cdot)$ denotes a cubic function. I calculate experience based on observed records, but I cannot be sure of a worker's experience, particularly for older workers whose first records appear in the first year of a state's data; they have likely worked for many years before I first observed them. As a test, I performed the following

calculation. Let us assume that if the first age at which I observe a worker in the data is 18 years or younger, and that record is in the first year of the states data, I treat that as the first actual record for that worker (I am assuming no truncation). For workers who are over age 18 and first appear in the state's first year of data, I assume their records were truncated. Under these assumptions, 62% percent of workers in the data are age 18 or older when I first observe them in the data.

10. $Tenure_{i,J(i,t),t}$ - The average tenure in years of worker i's tenure at firm $J(i,t)$ across all quarters in year t. The $f(\cdot)$ denotes a cubic function. I can only know a worker's tenure precisely if he arrives at a firm after the first year and quarter that a state has appeared in the data - if he arrives before that time, it is impossible for me to know long he has been at the firm. Fortunately, I can know tenure (that is, the first worker record on that job is at least one year after the state's appearance in the LEHD) for over 92% of the worker-year-dominant-job observations in the data. Furthermore, as time progresses, the importance of the difference between my observed tenure and actual tenure declines; put differently, my predicted probability that a worker quits will be further off in the case where I have observed one year of tenure but actual tenure is six years than it will be in the case where observed tenure is six years and actual tenure is 11 years.

11. $LogAnnualJobEarn_{i,J(i,t),t}$ - The natural log of worker i's annualized earnings at firm $J(i,t)$ in year t. I use annualized earnings so that I can compare worker's who have worked at firm for two (full) quarters to workers who have worked at the firm for a year (or longer) - it is meant to allow for an apples to apples earnings comparison across workers. To create annualized measures of earnings I take a worker's average full quarter earnings at the firm in year t and multiply them by four.

12. $JobChange_{(i,t,t+1)}$ is a binary variable that is equal to 1 if I observed that the SEIN of worker i's employer at his dominant job changed from years t to $t+1$ and this change meets my criteria for being voluntary - namely, that a worker had a registered employer in the data in all quarters of t and $t+1$.

13. $E(CumEarn_{(i,t,t+X)})$ is the expected cumulative earnings to date a worker i in year t can expect to have in year $t+X$, that is, it is the total of his earnings over the years $t+1, t+2, ..., t+X-1, t+X$.

14. $E(PctEarnChange_{(i,t,t+1)})$ is the expected percentage earnings change for a worker i at his dominant job between years t and year $t+1$ (using earnings in t as the base). Note that the worker can be at different jobs in t and $t+1$; I am not just considering stayers or leavers.

15. $FirmEarnRatio_{i,J(i,t),t}$ The ratio of worker i's earnings to the average earnings of workers at firm $J(i,t)$ in year t.

16. $FirmEducEarnRatio_{i,J(i,t),t}$ The ratio of worker i's earnings to the average earnings of similarly educated workers at firm $J(i,t)$ in year t.

17. $FirmPerc_{i,J(i,t),t}$ is worker i's position in the annualized job earnings distribution of all worker's at firm $J(i,t)$ in year t. I estimate the rank variables first as deciles

to test for possible non-linearities in the effect of percentile, but as the effect seems linear, I switch to a continuous measure. The lowest value is $1/N$, where N is the number of workers at firm $J(i,t)$ in year t. The highest value is 1.

18. $FirmEducPerc_{i,J(i,t),t}$ is worker i's position in the annualized earnings distribution of workers at firm $J(i,t)$ in year t who have the same level of education as i. I estimate the rank variables first as deciles to test for possible non-linearities in the effect of percentile, but as the effect seems linear, I switch to a continuous measure. The lowest value is $1/N$, where N is the number of workers at firm $J(i,t)$ in year t with the same education as worker i. The highest value is 1.
19. $FirmPercChange_{(i,t,t+1)}$ is the change in firm position that worker i experiences between years t and $t+1$. It equals $FirmPerc_{i,J(i,t+1),t+1} - FirmPerc_{i,J(i,t),t}$
20. $FirmEducPercChange_{(i,t,t+1)}$ is the change in firm-education-group position that worker i experiences between t and $t+1$. It equals $FirmEducPerc_{i,J(i,t+1),t+1} - FirmEducPerc_{i,J(i,t),t}$.

A.3 Multiple Imputation

Since I only observe education for roughly 12% of the workers in my sample, I must use the LEHD education imputes for the other 88% of workers. However, since education is missing completely at random (MCAR), the conditions for the accuracy of multiple imputation are perfectly met. The results presented in the paper come from running ten regressions, each regression using a different implicate for education, and combining the results. The procedure, which I closely template from Chapter 5 of Little and Rubin [2002], is as follows.

1. Let $Q(Y)$ represent the statistic(s) of interest we are trying to estimate from data Y.
2. Let $Q_m(Y^m)$ be the statistic estimated from the mth implicate.
3. Let M be the number of implicates. In my analysis, $M = 10$.
4. Let $\bar{Q} = \frac{\sum_{m=1}^{M=10} Q_m(Y^m)}{M}$ be the average estimand
5. Let $V_m(Y^m)$ be the covariance matrix of $Q_m(Y^m)$ from the mth implicate
6. Let $\bar{V} = \frac{\sum_{m=1}^{M=10} V_m(Y^m)}{M}$ be the average covariance matrix
7. Let $B = \frac{\sum_{m=1}^{M=10}(Q_m(Y^m)-\bar{Q})(Q_m(Y^m)-\bar{Q})^T}{M}$ be the between implicate variation of the different $Q_m(Y^m)$ estimands. Let its individual elements be denoted b_{ij}.

8. Let $T = \bar{V} + (1 + \frac{1}{M})B$ denote the corrected covariance matrix; this accounts for the missing data's contribution to variance. Let its individual elements be denoted t_{ij}.
   Then the missingness ratio of any statistic i of the vector of estimands Q can be calculated by $(1 + \frac{1}{M}) * \frac{b_{ii}}{t_{ii}}$. The missingness ratio measures the proportion of total variance that is due to the between implicate variance.
   In my regression tables (Tables 1.3, 1.4, 1.5, and 1.6) I report $\bar{Q}$, $\sqrt{t_{ii}}$, and the missingness ratio.