RANKING FIRMS USING REVEALED PREFERENCE

Isaac Sorkin

Abstract

This paper estimates workers’ preferences for firms by studying the structure of employer-to-employer transitions in U.S. administrative data. The paper uses a tool from numerical linear algebra to measure the central tendency of worker flows, which is closely related to the ranking of firms revealed by workers’ choices. There is evidence for compensating differentials when workers systematically move to lower-paying firms in a way that cannot be accounted for by layoffs or differences in recruiting intensity. The estimates suggest that compensating differentials account for over half of the firm component of the variance of earnings. JEL Codes: C63, J31, J42.

*An earlier version of this paper was the first chapter of my dissertation at the University of Michigan: thanks to Matthew D. Shapiro, John Bound, Daniel Ackerberg and Josh Hausman for patient advising and support. Thanks also to Lawrence Katz, anonymous referees, John Abowd, Audra Bowlsus, Charles Brown, Jediphi Cabal, Varanya Chaube, Raj Chetty, Tim Conley, Cynthia Doniger, David Dorn, Matthew Fiedler, Eric French, Matt Gentzkow, Paul Goldsmith-Pinkham, Henry Hyatt, Gregor Jarosch, Lisa Kahn, Patrick Kline, Pawel Krolikowski, Margaret Levenstein, Ilse Lindenlaub, Kristin McCue, Erika McEntarfer, Andreas Mueller, Michael Mueller-Smith, Matt Notowidigdo, Luigi Pistaferri, Giovanni Righi, Justin Wolfers, Mary Wootters, Eric Zwick and numerous seminar and conference participants for helpful comments and conversations. Thanks to Giovanni Righi for research assistance, Kristin McCue for help with the disclosure process, and David Gleich for making Matlab BGL publicly available. This research uses data from the U.S. Census Bureau’s Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. This research was supported in part by an NICHD training grant to the Population Studies Center at the University of Michigan (T32 HD007339) and the Robert V. Roosa Dissertation Fellowship. This research was also supported by the CenHRS project, funded by a Sloan Foundation grant to the University of Michigan, and by the Michigan Node of the NSF-Census Research Network (NSF SES 1131500). Work on this paper took place at the Michigan, Chicago and Stanford Federal Statistical Research Data Centers. Part of the work on this paper was completed while I was employed by the Federal Reserve Bank of Chicago. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Chicago, the Federal Reserve System, or the U.S. Census Bureau. All results have been reviewed to ensure no confidential information is disclosed. Correspondence: Isaac Sorkin, Stanford University, Department of Economics, 579 Serra Mall, Stanford, CA. Email: sorkin@stanford.edu.
Dating back to at least [Smith (1776/2003, Book 1, Chapter 10)] (see also [Rosen (1986)]), economists have argued that differences in the nonpay characteristics of jobs explain some of earnings inequality. To find evidence for these compensating differentials, the literature has typically taken a bottom-up, hedonic approach. In the classic hedonic approach, the researcher considers a cross-sectional regression of earnings on one (or a few) nonpay characteristics and interprets the coefficient on each nonpay characteristic as the market price of that characteristic. For stark case studies such as fatality risk or whether or not a PhD scientist has control over their research agenda, this approach has identified compensating differentials. But the difficulties of estimating compensating differentials outside these case studies have led some (e.g., [Hornstein et al. (2011, pg. 2883)]) to conclude that compensating differentials are not likely to prove important for understanding overall earnings inequality.

This conclusion is potentially unwarranted because the hedonic approach can lead to an incomplete picture of the importance of compensating differentials for at least two reasons. First, it assumes that a researcher knows—and can measure—all the nonpay characteristics that workers value. Even among the characteristics a researcher can measure, if the unobserved characteristics are negatively correlated with the observed characteristics, then estimated prices can be biased down. Second, it assumes that the labor market is perfectly competitive and so utility is equalized across jobs. As emphasized by [Mortensen (2003)], if there is dispersion in utility, then higher-paying jobs might also have more desirable nonpay characteristics, also biasing estimates down.

This paper develops and implements an empirical framework to measure the role of compensating differentials that addresses these two critiques via two building blocks. First, the framework uses a revealed preference argument. As opposed to measuring and valuing one nonpay characteristic at a time, revealed preference takes a top-down approach and relies on worker choices to tell the researcher which bundle of characteristics they value. Second, the framework allows for differences in utility across jobs. As opposed to assuming that the labor market is perfectly competitive, the framework quantifies the extent of utility dispersion across jobs.

To see how these two building blocks could lead to an estimate of the role of compensating differentials, suppose there are two firms: A and B. Suppose that the firms do not tailor their offers to specific workers, and workers have common preferences (up to an idiosyncratic utility draw). Suppose also that both firms are initially the same size and make the same number of offers to

---

1 For recent work on fatality risk see [Lavetti and Schmutte (2017)] and [Lavetti (2017)]. For PhD scientists and their research agenda, see [Stern (2004)] and [Mas and Pallais (2017)] for an interesting recent study of alternative work arrangements.
workers at the other firm at random. If more workers accept A’s offer than B’s offer, then we can infer that workers prefer firm A to firm B. If it also turns out that B is higher-paying than A, then we infer that B offers worse nonpay characteristics than A (since workers prefer A to B despite the lower pay). Hence, compensating differentials explains why B pays more than A. This example incorporate the two building blocks as follows. It relies on revealed preference because it uses the information in workers’ choices between A and B. And it entertains the possibility of utility dispersion because it allows us to conclude that, from the workers’ perspective, one firm is better.

In the data I use, rather than two firms, there are about half a million firms, which poses the computational challenge of how to aggregate the flows in an economically interpretable way. To do so, I develop a structural interpretation of Google’s PageRank algorithm. I begin by interpreting the flow data as arising from binary choices between firms where workers perceive a common value of firms and an idiosyncratic utility draw. These assumptions imply a simple recursive definition of good: “good firms hire from other good firms and have few workers leave.” A similar recursive definition underlies Google’s PageRank algorithm which aggregates the link structure of the web: “good webpages are linked to by other good webpages.” Compared to billions of webpages, a matched employer-employee dataset with a half million firms is almost “small data” and it is computationally quite cheap to solve this recursion.

I then incorporate a few other explanations for the structure of flows besides differences in the values of firms: differences in size, offers, and the possibility that workers were laid-off. First, a large firm will naturally have more workers moving away from it than a small firm. I account for this because I observe firm size. Second, a firm that makes a lot of offers will naturally have more workers moving towards it. I account for this because I estimate the offer distribution using information in nonemployment-to-employer flows. By jointly estimating the offer distribution and the value of nonemployment, I allow nonemployed workers to reject offers. Finally, to identify workers who were laid off (and so who could not choose to stay at their current employer), I use information in what a worker’s coworkers were doing at the time of the separation. In the spirit of the displaced worker literature (Jacobson et al. [1993]), if the firm is contracting and an unusually high share of coworkers are also separating, then a firm-level shock caused the separations, and there is a high probability that any given worker was laidoff.

\footnote{With exactly two firms, this idea will find that compensating differentials explains either all or none of the pay gap. With three or more firms, however, this idea can find that it is a mix: suppose the ranking based on choices is A then B then C, while the ranking based on pay is B then A then C. Then the A and B pay gap is compensating differentials, while the B and C pay gap is not.}
Combined, these pieces give me an estimate of the value of working at a firm, which I compare to a measure of firm-level pay to get an estimate of the role of compensating differentials in firm-level pay differences. There are a couple reasons to focus on firm-level pay differences. First, to focus on the common preference, I want to aggregate over idiosyncratic factors affecting workers’ choices and observe multiple workers facing similar choices. Aggregating to the firm level provides some hope of doing this. Second, it lets me build on recent work emphasizing the role of firms in pay setting in the labor market (e.g., Abowd et al. (1999) (AKM), Andersson et al. (2012), Card et al. (2013), Song et al. (2016), Barth et al. (2016), Card et al. (2016), Goldschmidt and Schmieder (2017), Engbom and Moser (2017), and Abowd et al. (2017)). Hence, part of the goal of this paper is to open up the black box of what the AKM firm effects represent by asking to what extent higher-paying firms are more desirable firms. Specifically, I interpret the extent to which higher-paying firms are more desirable firms as evidence of rents, while the extent to which this is not the case as evidence of compensating differentials. Naturally, by focusing on firm effects, I only have something to say about the portion of the variance of earnings that is at the firm level. Through the lense of the AKM decomposition—which decomposes pay into a firm effect, a worker effect, covariates and a residual—I leave out any compensating differentials which would be reflected in components besides the firm effect.

Once we allow for utility dispersion, nonpay characteristics can be both compensating and augmenting and so interpreting the comparison of firm-level pay and values in terms of compensating differentials is subtle. In the classic hedonic setting of Rosen (1986), utility is equalized across jobs (at the margin). Hence, all variation in nonpay characteristics is offset by compensating variation in pay. In contrast, in the presence of utility dispersion (implied by frictional models, and discussed in Mortensen (2003)), nonpay characteristics can contribute to utility dispersion by augmenting variation in pay. Thus, for any given nonpay characteristic it is not obvious whether it is compensating or augmenting (and this might differ by firm). By focusing on revealed preference, I partially sidestep this ambiguity. I develop a model of a firm’s posting decision, where firms post a compensation package consisting of both pay and nonpay characteristics. There are two sources of firm heterogeneity: heterogeneity in desired utility (the “Mortensen” motive), which generates augmenting variation, and heterogeneity in the marginal cost of the provision of nonpay characteristics (the “Rosen” motive), which generates compensating variation. I show that the variation in pay conditional on overall value maps into the pure Rosen motive, or the part of

3In this paper, I use the word firm and employer interchangeably.
the Rosen motive that is orthogonal to the Mortensen motive. Hence, this comparison identifies a theoretically coherent concept of compensating differentials. In contrast, I cannot identify the importance (or presence) of the augmenting nonpay characteristics.

I estimate the model on the U.S. Census Bureau’s Longitudinal Employer Household Dynamics (LEHD) dataset and develop three main findings. First, the framework finds that compensating differentials explain about two-thirds of the variance of firm-level earnings. Aggregated, this finding implies that compensating differentials explain at least 15% of the variance of earnings. Second, if the estimated nonpay characteristics were removed and earnings changed to compensate workers, then earnings inequality would decline. This reduction comes mainly from the lower tail of the income distribution shifting up. Finally, the finding of a large role for compensating differentials helps resolve the puzzle emphasized by Hornstein et al. (2011) that benchmark search models cannot generate the extent of observed residual earnings inequality. Workers act as if a large share of the variance of firm-level earnings does not reflect variation in value.

Numerous supplementary analyses build the plausibility of the results. First, I aggregate the firm-level estimates to the sector level and the ranking of sectors is intuitively plausible, as is the implied distribution of nonpay characteristics. For example, education has good nonpay characteristics, while many blue-collar sectors, such as mining and manufacturing, have bad nonpay characteristics. Second, the finding of a large role for compensating differentials rests on a conservative interpretation of the underlying patterns in the data: I only interpret 40% of moves to lower-paying firms as being explained by more desirable nonpay characteristics, with the remaining moves explained by a combination of layoffs and negative idiosyncratic shocks. Third, the moves to lower-paying firms are not offset by future earnings increases, which suggests that some nonpay characteristic drives the move. Fourth, the basic result is robust across subgroups defined by age, gender, worker effects, geography and industry.

Nevertheless, numerous caveats related to both the data and the framework remain. In terms of data, I do not observe hours and so it could be that all the variation in nonpay characteristics I estimate is hours. Such a finding would perhaps be reassuring about the validity of the framework, but might make the results less novel. I provide some suggestive evidence that hours variation is not the dominant source of compensating differentials by looking at sectoral-level variation in hours and find that it explains about 15% of the sectoral compensating differentials. In addition, I have no measures of working conditions to compare to my estimates.
In terms of caveats related to the framework, as the two firm example makes clear, it is quite stylized. Specifically, it omits many mechanisms that have been discussed in the literature. For example, it omits worker heterogeneity, screening, systematic forms of preference heterogeneity, mobility costs, and counteroffers. Moreover, it assumes that workers only ever face binary choices. The absence of worker heterogeneity precludes sorting between firms and workers. It is typically hard, however, to tell a story of how these simplifications lead to biased estimates. The reason is that it is not enough that any one move be for a reason outside the model, because the idiosyncratic utility draw allows for unmodeled reasons. Instead, it is necessary to explain how this omission generates a systematic pattern of mobility that is towards lower-paying firms. One example that generates an overstatement of the role of compensating differentials is if all voluntary mobility to lower-paying firms is because workers were laid-off. Another example is if the nonemployed and the employed search from very different distributions. This difference could scramble the values and lead to a weaker relationship between values and pay. In addition, noise in the estimates of the values and pay leads to an overestimate of the role of compensating differentials, though I present Monte Carlo evidence that such bias is quantitatively small. Going the other way, suppose that in order to work at a high-paying firm a worker needs experience at a low-paying firm. Then workers systematically move from low-paying to high-paying firms, but the high pay reflects their time-varying skills and not rents.

**Literature:** This paper builds on a number of literatures. The idea that earnings cuts identify nonpay characteristics is shared with a few papers (e.g., Becker (2011), Nunn (2013), Sullivan and To (2014), Hall and Mueller (2017), and Taber and Vejlin (2016)). The most closely related paper is Taber and Vejlin (2016), which also uses matched employer-employee data and a revealed preference argument. Relative to this paper, Taber and Vejlin (2016) attempt to explain all of the variance of earnings and not just the firm-level component. This ambition, however, means that the mapping to the data is less straightforward. Relative to the remaining papers (which all use individual level data), I focus on the firm component, which averages out idiosyncratic nonpay aspects of job value.

This paper uses how workers move across firms in a new way relative to existing literature. Bagger and Lentz (2017) also emphasize patterns in worker reallocation across firms, but do not allow for nonpay characteristics and do not exploit the complete structure of employer-to-employer

---

4 The actual model that would generate this is slightly more complicated, because in order to get this to be a steady state this implies that the layoffs shocks are non-neutral with respect to worker type, and it is not obvious how this propagates through estimation.

5 There is also literature (e.g., Gronberg and Reed (1994), Dey and Flinn (2005), Bonhomme and Jolivet (2009), Aizawa and Fang (2015) and Jarosch (2015)) which estimates the value of specific amenities in a search environment.
moves. Similarly, Moscarini and Postel-Vinay (2016) and Haltiwanger et al. (2017) explore worker flows and ask whether these are consistent with a job ladder defined by a particular observable characteristic (e.g., size or wages). I invert the approach in these papers and instead construct the job ladder implied by worker flows.

The estimation approach applies conditional choice probability estimation (Hotz and Miller (1993)) to matched employer-employee data, which allows gross worker flows between firms to exceed net flows. Other papers exploit similar modeling insights to study situations where gross flows exceed net flows; e.g., Kline (2008) and Artuc et al. (2010).

Finally, this paper echoes some themes in the interindustry wage differential literature, and I discuss this relationship further in the concluding section.

Roadmap: This paper unfolds as follows. Section I introduces the data. Section II shows that firms play a large role in explaining the variance of earnings, and documents some simple summary statistics which suggest the importance of nonpay motives in driving mobility: over a third of EE moves come with earnings cuts and about 40% of EE moves are to lower paying firms. Section III shows that these EE moves to lower-paying firms reflect a systematic pattern of mobility (and shows how to measure the systematic pattern of mobility). Section IV takes up the task of how to rank firms using mobility data. Section V presents the main results of the paper. Section VI presents the earnings inequality counterfactual and discusses the relationship to the Hornstein et al. (2011) puzzle. Section VII shows that the main results are robust across a variety of subgroups. Finally, section VIII relates this paper to the interindustry wage differential literature, and discusses some caveats and promising avenues for future work.

I Matched employer-employee data

I use the U.S. Census Bureau’s Longitudinal Employer Household Dynamics data, which is a quarterly dataset that is constructed from unemployment insurance records. The LEHD is matched employer-employee data and so allows me to follow workers across employers. The notion of an employer in the dataset is a state-level unemployment insurance (UI) account. For a firm with a single establishment, this notion is equivalent to a firm. For firms that operate in multiple states, or have multiple UI accounts in a state (as might happen if the firm has operations in multiple indus-

\[^6\text{See Abowd et al. (2009) for details.}\]
tries), this notion is “smaller” than a firm[7]. Working conditions are probably more similar within establishments than within employers, so having a “smaller” notion of an employer is desirable from the perspective of measuring compensating differentials.

I look at the worker’s annual dominant employer: the employer from which the worker made the most money in the calendar year. To facilitate coding transitions, I require that the worker had two quarters of employment at the employer and that the second quarter occurred in the calendar year[8]. I also restrict attention to workers aged 18-61 (inclusive) and, following [Card et al. (2013)], require that the annualized real earnings exceed $3,250. Earnings are annualized by adjusting earnings for the number of quarters a worker was at a particular employer. Throughout the paper, nominal earnings are converted to 2011 dollars using the CPI-U.

To understand more about the transition, I use the quarterly detail of the LEHD to code transitions as employer-to-employer or employer-to-nonemployment-to-employer. Specifically, following [Bjelland et al. (2011) and Hyatt et al. (2014)], I code a transition as employer-to-nonemployment-to-employer if between the annual dominant employers there is a quarter when the worker is nonemployed or has very low earnings. This definition introduces several sources of measurement error. One source of measurement error is if a worker moves to a job outside the set of 27 states I use, then they would appear to make an employment-to-nonemployment transition, even though in reality they made an employer-to-employer transition. Another source of measurement error is if a worker moves to a job that is not covered by the unemployment insurance system (such as self-employment or contract work, which are increasingly prevalent as documented by [Katz and Krueger (2016), Jackson et al. (2017) and others]), then I would code this pattern as an employment-to-nonemployment, rather than as an employer-to-employer transition. To mitigate these sources of error, I only count transitions where workers subsequently appear in the dataset. Hence, if the worker leaves the 27 states and never returns, or enters self-employment or contract work and never returns to employment covered by UI, then I do not count the worker as having made an employer-to-nonemployer transition. See Appendix A for details on dataset construction (including how earnings are annualized).

---

[7]Firms with multiple SEINs in a state is a rare occurrence: according to personal communication from Henry Hyatt (dated June 12, 2014): “the employment weighted fraction of firmids with multiple SEINs [state employer identification number] in a given state is about 1.5%, and...this fraction is actually lower in some of the larger states.”

[8]Reduction to one observation per person per year is common. See [Abowd et al. (1999) (France), Abowd et al. (2003) (US), Card et al. (2013) (Germany), and Card et al. (2016) (Portugal). Even outside of estimating statistical wage decompositions, [Bagger et al. (2014)] also reduce to one such observation per year.
Two features of the LEHD should be kept in mind when interpreting the results. First, the unemployment insurance system measures earnings, but not hours. Thus, variation in hours as well as in benefits will be included in my measure of compensating differentials. I provide some evidence on the extent to which hours variation explains the results in section V.D. Second, only employers that are covered by the unemployment insurance system appear in the dataset. Overall, in 1994 the unemployment insurance system covered about 96% of employment and 92.5% of wages and salaries (BLS 1997 pg. 42).

I pool data from 27 states from the fourth quarter of 2000 through the first quarter of 2008. Pooling data means that I keep track of flows between as well as within these states.

I impose three restrictions to eliminate the smallest firms where it would be hard to plausibly estimate a firm effect. The first restriction is a minimum size threshold. Specifically, I eliminate firms where there are strictly fewer than 90 non-singleton person-years at the firm (or 15 per year), where a singleton person-year is one where I never observe the worker again. The second restriction is that I look at the strongly connected subset of those firms (I define strongly connected in section IV.B). Third, within this set of firms, I look at the strongly connected set of firms that also hire from nonemployment, and that appear in a sufficient number of bootstrap replications. Table I shows that the first restriction eliminates about 92% of employers, 14% of people, and 19% of person-years, and the remaining restrictions have relatively small effects on sample size.

II Earnings

I now document that conditional on person fixed effects, firms play a large role in explaining the variance of earnings and that there are many moves to lower-paying firms.

---

9The notion of earnings captured by UI records is as follows: “gross wages and salaries, bonuses, stock options, tips and other gratuities, and the value of meals and lodging” (BLS 1997 pg. 44)). This omits the following components of compensation: “employer contributions to Old-age, Survivors, and Disability Insurance (OASDI); health insurance; unemployment insurance; workers’ compensation; and private pension and welfare funds” (BLS 1997 pg. 44)).

10This restriction results in the exclusion of certain sectors of the economy. In particular, small nonprofit (or- ganizations employing fewer than four workers), domestic, self-employed, some agricultural and federal government (but not state and local government) workers are excluded. For more complete discussions, see Kornfeld and Bloom (1999 pg. 173), BLS (1997 pg. 43) and http://workforcesecurity.doleta.gov/unemploy/pdf/UILawCompare/2012/coverage.pdf.

11I use the following states: CA, FL, GA, HI, ID, IL, IN, KS, MD, ME, MN, MO, MT, NC, ND, NJ, NM, NV, PA, OR, RI, SC, SD, TN, VA, WA, and WI. See Figure A1 in Appendix L for a map.

12For example, all observations in 2007 count as singleton person-years because it is the last year in the dataset.

13An additional motivation to impose a minimum size threshold is to minimize variation in the identified set of firms across the bootstrap resamples.
II.A Firms play an important role in earnings determination

To measure firm-level earnings, I use the following equation for log earnings:

\[
\log \text{earnings} = \alpha_i \text{person effect} + \Psi_{J(i,t)} \text{firm effect} + x'_{it} \beta + r_{it} \text{residual/error term},
\]

where \(y_{it}\) is log earnings of person \(i\) at time \(t\), \(\alpha_i\) is a person fixed effect, \(\Psi_{J(i,t)}\) is the firm fixed effect at the employer \(j\) where worker \(i\) is employed at time \(t\) (denoted by \(J(i,t)\)), \(r\) is an error term, and \(x\) is a set of covariates including higher-order polynomial terms in age.

I quantify the role of firms in earnings using two different decompositions. The first decomposition is what Card et al. (2017) call the “ensemble” decomposition:

\[
\text{Var}(y_{it}) = \text{Cov}(\alpha_i, y_{it}) + \text{Cov}(\Psi_{J(i,t)}, y_{it}) + \text{Cov}(x'_{it}\beta, y_{it}) + \text{Cov}(r_{it}, y_{it}).
\]

In this decomposition, the share of the variance of earnings accounted for by firms is \(\frac{\text{Cov}(\Psi_{J(i,t)}, y_{it})}{\text{Var}(y_{it})}\).

The third portion of Table I shows that according to this decomposition firms account for about 21% of the variance of earnings.

The second decomposition is known as the Abowd et al. (1999) decomposition:

\[
\text{Var}(y_{it}) = \text{Var}(\alpha_i) + \text{Var}(\Psi_{J(i,t)}) + \text{Var}(x_{it}\beta) + 2\text{Cov}(\alpha_i, \Psi_{J(i,t)}) + 2\text{Cov}(\alpha_i + \Psi_{J(i,t)}, x_{it}\beta) + \text{Var}(r_{it}).
\]

The fourth portion of Table I reports the results of this decomposition. It shows results that are quite similar to results for an identical time period in the US using a different dataset. In particular, relative to Song et al. (2016, Table C3, column (8)), which covers the US for 2001-2007 using Social Security Administration data, I find a similar role of workers (51% v. 52%), a slightly larger role for firm effects (14% vs. 12%), and a slightly larger role for the covariance of firm and worker effects (10% vs. 7%). Putting the pieces together, I find a larger correlation between firm effects and worker effects than Song et al. (2016, Table C2, column 8) (0.19 vs. 0.08). I also find a 6 percentage

---

14Because I only use seven years of data, the linear terms in the age-wage profile are highly correlated with the person fixed effects and thus, following Card et al. (2013), are omitted. Following Card et al. (2013), I assume that earnings are flat at age 40, and include quadratic and cubic terms in age. See Card et al. (2017, pg. 10) for further discussion of this point. I also include a gender dummy interacted with the type of earnings observation used: “continuous” or “full” (see Appendix A for details).
point increase in the adjusted $R^2$ from including match effects, which is nearly identical to that found by Song et al. (2016, Table C2, column 8).

II.B Earnings declines are an important feature of the data

This section begins to build the empirical case that something besides the pursuit of higher pay explains some employer-to-employer moves. I show that earnings declines are widespread, are captured by the firm effects, and are not offset by future earnings increases.

Individual-level earnings declines are widespread in the data. Panel A of Table II shows that 43% of transitions between annual dominant employers see earnings declines, while 37% of such EE transitions see earnings declines (in nominal terms these shares are naturally smaller: 40% and 34%). Hence, many employer-to-employer transitions cannot be explained by pursuit of higher-pay. By revealed preference, there must be some good nonpay characteristics that justify these earnings cuts. The nonpay characteristics, however, might be idiosyncratic to the firm-worker match and would not generate compensating differentials because such factors are not necessarily priced in the labor market.

I now present evidence that the earnings declines are captured by the firm effects. Panel B of Table II shows that 52% of the EE transitions to lower-paying firms have earnings declines, while only 27% of the such transitions to higher-paying firms have earnings declines. Figure I plots the change in firm effects against the probability of an earnings decline on all transitions (top panel) and EE transitions (bottom panel). The probability of an earnings decline on an EE transition decreases from 75% for the largest downward moves to 10% for the largest upward moves.

One explanation for moves to lower-paying firms is that workers trade-off the level of pay for the promise of more rapid earnings growth. But the earnings declines captured by moving to lower-paying firms are not offset by future earnings increases. Following Abowd et al. (1999), I estimate firm-specific earnings slopes using the wage growth of the stayers. When workers move to lower-paying firms, Figure II shows that they do not move to firms offering steeper slopes in earnings.

---

15 The share of earnings declines is quantitatively consistent with evidence from survey datasets where researchers are able to calculate changes in hourly earnings. Jolivet et al. (2006, Table 1) find that in the Panel Study of Income Dynamics 23.1% of job-to-job transitions come with an earnings cut. For the Survey of Income and Program Participation, Tjaden and Welschmied (2014, Table 2) find 34%. And for the National Longitudinal Survey of Youth 1997, Sullivan and To (2014, Table 1) find 36%.

16 In Appendix B I report an additional exercise where following Chetty et al. (2014a) I regress earnings changes on the change in firm effects and show that the magnitudes of earnings changes line up well. I also show that this fit is not mechanical by simulating data from a structural model that emphasizes comparative advantage (e.g., Eeckhout and Kircher (2011), Lopes de Melo (2016) and Hagedorn et al. (2017)) and so violates the assumptions of the AKM model.
if anything, they move to firms offering shallower slopes (the coefficient on all moves is 0.000 and on EE moves is 0.002). Similarly, the firm effects in the intercept are positively correlated with the slope when estimated in the same regression (the correlation is 0.033).

III Systematic pattern of mobility

While many employer-to-employer transitions are to lower-paying firms, Panel B of Table II shows that workers are more likely to transition to higher-paying firms than to lower-paying firms: 53% of all moves are to higher paying firms. Hence, the moves to lower-paying firms might best be explained by idiosyncratic shocks and not be evidence of compensating differentials. To interpret these moves as evidence of compensating differentials, I want to show that they cannot be explained by idiosyncratic shocks.

III.A Aggregating moves

I now develop a way of averaging out idiosyncratic shocks and extracting the systematic pattern of mobility. I present the method in the context of a rank-aggregation problem where I view the set of EE transitions as generated from an equal-number of workers facing the choice between any pair of firms. In section IV I introduce additional notation and assumptions that maps this approach more tightly to the empirical context of flows between firms where firms might differ in the number of workers, in their probability of making an offer, and some of the flows might reflect exogenous shocks.

To introduce the rank aggregation problem, suppose I observe $N$ workers choosing between firms $k$ and $j$. Out of these $N$ workers, $M_{kj}^o$ workers choose $k$ and $M_{jk}^o = N - M_{kj}^o$ choose $j$, where the $o$ superscript means observed (in section IV I introduce a distinction between observed and endogenous moves). Two comments on this set-up and notation are in order. First, this notation records both the choice set and choice in the subscripts; i.e., $M_{kj}^o$ records the number of workers choosing between $k$ and $j$ who chose $k$. This parsimony is feasible because I assume there are only binary choices. Second, because it is not clear how many workers choosing between $j$ and $j$ choose $j$ ($M_{jj}^o$), this set-up leaves the diagonal elements undefined. I show below that the statistic I compute is invariant to the value of the diagonal elements, so this conceptual ambiguity is irrelevant.

To produce a single ranking that best-rationalizes the data, suppose that the common value of firm $k$ is $\tilde{V}_k^{EE}$. When choosing between firm $k$ and $j$, workers take into account the common value
as well as an idiosyncratic draw, \( \iota \), which is distributed type I extreme value with scale parameter 1. This idiosyncratic utility draw captures preference heterogeneity, where part of the preference heterogeneity might be that workers receive a negative shock to the value of the match. This distributional assumption implies that the probability of choosing firm \( k \) over \( j \) is \( \frac{\exp(\tilde{V}_{EE}^k)}{\exp(\tilde{V}_{EE}^k) + \exp(\tilde{V}_{EE}^j)} \). A simple estimator of the relative common values is the ratio of the empirical choice probabilities:

\[
\frac{M_{kj}/N}{M_{jk}/N} = \frac{M_{kj}^0}{M_{jk}^0} = \frac{\exp(\tilde{V}_{EE}^k)}{\exp(\tilde{V}_{EE}^j)}.
\]

While it would be possible to take equation (4) directly to the data, doing so would run into two problems. First, for many pairs of firms there are not flows in both directions so this approach would not yield well-defined values of employers. Second, there is no guarantee that this approach would yield consistent valuations of employers. For example, there might be Condorcet cycle-like cases where combining the comparisons of employer A with employer B and employer B with employer C would give different relative valuations of employer A and employer C than the direct comparison of employer A and employer C.

To estimate the firm-level value, I relax the pair-wise restrictions embedded in the model and instead impose only one restriction per firm, which addresses the two problems mentioned above. First, by reducing to one restriction per firm, this model reduces to an exactly identified set of equations and so there exists a unique set of firm-level values that best explain all the flows and do so in a way that is computationally feasible (unlike MLE). Second, the condition for uniqueness, which I discuss formally later in this section, is a restriction on the pattern of zeros in the flows that is much weaker than that each pair of firms have flows in both directions. (In Appendix D I present an over-identifying test which relies on asking whether the model estimates satisfy the pair-wise comparisons.)

Now let \( \mathcal{E} \) be the set of employers. Cross-multiplying (4) gives

\[
M_{kj}^0 \exp(\tilde{V}_{EE}^j) = M_{jk}^0 \exp(\tilde{V}_{EE}^k), \quad \forall j \in \mathcal{E},
\]

\[\text{17Because the idiosyncratic utility draw includes negative shocks to the value of the match, I do not interpret moves to lower-value firms that are rationalized by the idiosyncratic draw as being utility-increasing.}\]
where the “for all” holds because \([4]\) holds for all pairs of employers. Summing across all employers on both sides gives

\[
\sum_{j \in E} M^o_{kj} \text{ exp}(\tilde{V}^{EE}_j) \text{ \# entering } k = \sum_{j \in E} M^o_{jk} \text{ exp}(\tilde{V}^{EE}_k) \text{ \# exiting } k.
\]

Dividing by the sum on the right-hand side gives

\[
\frac{\sum_{j \in E} M^o_{kj} \text{ exp}(\tilde{V}^{EE}_j)}{\sum_{j \in E} M^o_{jk} \text{ \# exiting } k} = \text{ exp}(\tilde{V}^{EE}_k).
\]

Equation \((6)\) implies one linear restriction per firm. The equation generates a recursive definition of employer quality: a good firm is chosen over other good firms, and has few workers not choose it. Formally, this equation is closely related to the recursion underlying Google’s PageRank approach to ranking webpages, which says that a good webpage is linked to by other good webpages. I now exploit this connection to show how to estimate \(\tilde{V}^{EE}\).

To solve for the values, create the matrix version of equation \((6)\). Specifically, define a diagonal matrix \(S^o\) with the \(k^{th}\) diagonal entry being \(S^o_{kk} = \sum_{j \in E} M^o_{jk}\). Define \(M^o\) to be the matrix where the \((j,k)\) entry is \(M^o_{j,k}\). Then letting \(\text{ exp}(\tilde{V}^{EE})\) be the \(|E| \times 1\) vector that contains the firm-level \(\text{ exp}(\tilde{V}^{EE}_k)\) yields the following:

\[
S^{o,-1}M^o \begin{bmatrix} \text{ exp}(\tilde{V}^{EE}) \end{bmatrix} = \begin{bmatrix} \text{ exp}(\tilde{V}^{EE}) \end{bmatrix}.
\]

This equation allows me to solve for \(\text{ exp}(\tilde{V}^{EE})\). Intuitively, \(\text{ exp}(\tilde{V}^{EE})\) is the fixed point of the function \(S^{o,-1}M^o : \mathbb{R}^{|E|} \rightarrow \mathbb{R}^{|E|}\). In many settings in economics, fixed points can be found by starting with an initial guess and repeatedly applying the function to the resulting output until it converges. Despite the very high-dimensionality of the function, the same idea applies here.

I now discuss when \(\text{ exp}(\tilde{V}^{EE})\) exists. To show when the \(\text{ exp}(\tilde{V}^{EE})\) vector exists, note that in the context of a linear system, the fixed point is an eigenvector corresponding to an eigenvalue of 1. The technical condition is that \(S^{o,-1}M^o\) has an eigenvalue of 1 and this eigenvalue is the largest
one. Moreover, in order for the values to be interpretable, the \( \exp(\tilde{V}^{EE}) \) vector needs to be all positive so that \( \tilde{V}^{EE} \) is defined (the log of a negative number is not defined).

**Result 1** Let \( S^{n-1}M^o \) be matrices representing the set of flows across a set of employers and be defined as above. If the adjacency matrix associated with \( M^o \) represents a set of strongly connected employers, then there exists a unique-up-to-multiplicative-factor vector of the same sign \( \exp(\tilde{V}^{EE}) \) that solves the following set of equations:

\[
S^{n-1}M^o \exp(\tilde{V}^{EE}) = \exp(\tilde{V}^{EE}).
\]

**Proof** See Appendix E (also for graph theory definitions).

This result shows that I can estimate the value of employers in the *strongly connected* set. Strongly connected is a restriction on the pattern of zeros in the \( M^o \) matrix. To be in the strongly connected set, an employer has to both hire a worker from *and* have a worker hired by an employer in the strongly connected set. This result is intuitive. The information used to estimate values is relative flows. If an employer either never hires, or never has anyone leave, then we cannot figure out its relative value. To see this, consider equation (6). If a firm never hires, then its value is mechanically zero. Alternatively, if a firm has no workers leave, then the denominator is zero and the value of the firm is infinite. This result is related to the identification result in Abowd et al. (2002) who show that the employer fixed effect in Abowd et al. (1999) can only be estimated in the *connected* set of employers. To be in the connected set, an employer has to either hire a worker from *or* have a worker hired by an employer in the connected set.

The analogy to Abowd et al. (2002) is helpful in understanding the data requirements to estimate \( \tilde{V}^{EE} \). As in Abowd et al. (2002), it is not necessary for there to be flows between every pair of employers to estimate the relative values (or relative pay). All that is required is that there is enough information in the non-zero entries to learn about each firm. It is possible to construct examples where in the limit as the number of firms grow the share of non-zero entries goes to zero, but the values remain well-estimated.[18]

**Remark on Result [1]** Because the discrete choice setting implies that the \( S^o \) matrix is different than in standard applications, the novelty in result [1] is showing that the top eigenvalue is 1 (the

---

[18]Suppose that there are flows (in both directions) between firms 1 and 2, 2 to 3, and \( N-1 \) to \( N \) (assuming \( N \) is even). Then the number of non-zero entries in \( M^o \) is proportional to \( N(2N-2) \), while the number of entries is proportional to \( N^2 \). Hence, the share of non-zero entries is proportional to \( \frac{1}{N} \) and goes to zero as \( N \) grows large. But if the number of observations going into each of these comparisons grows, then the estimates of \( \tilde{V}^{EE} \) converge.
Perron-Frobenius theorem is used to show that the top eigenvector is unique. $S^o$ divides the $i^{th}$ row of $M^o$ by the $i^{th}$ column sum of $M^o$. In other applications (e.g., Pinski and Narin (1976), Page et al. (1998) (Google’s PageRank), and Palacios-Huerta and Volij (2004)), the normalizing matrix instead divides the $i^{th}$ column of $M^o$ by the $i^{th}$ column sum. This normalization makes the resulting matrix a transition matrix and standard results imply that the top eigenvalue is 1. With the alternative normalization implied by the discrete choice model, standard results do not apply.

The diagonal entries in $M^o$ are not defined using the discrete choice setting (or the search model defined below). The following result shows that because of the normalization, the top eigenvector of $S^o-1M^o$ is invariant to the value of the diagonal entries in $M^o$.

**Result 2** Suppose that $\exp(\breve{V}^{EE})$ is a solution to $\exp(\breve{V}^{EE}) = S^o-1M^o\exp(\breve{V}^{EE})$ for a particular set of $\{M^o_{k,k}\}_{k \in E}$. Pick arbitrary alternative values of the diagonal: $\{M^o'_{k,k}\}_{k \in E} \neq \{M^o_{k,k}\}_{k \in E}$. Let $S^o'$ and $M^o'$ be the natural variants on $S^o$ and $M^o$. If $\exp(\breve{V}^{EE})$ solves the equation $\exp(\breve{V}^{EE}) = S^o-1M^o\exp(\breve{V}^{EE})$, then it also solves the equation $\exp(\breve{V}^{EE}) = S^o'-1M^o'\exp(\breve{V}^{EE})$.

**Proof** See Appendix E.

**Remark on Result 2**: A natural statistic to compute to assess the noise in the estimation of $\breve{V}^{EE}$ is the spectral gap, or the difference between the first and second eigenvalues. This result shows that the spectral gap is not pinned down by the data because, in general, the second eigenvalue depends on the diagonal entry in the matrix.

### III.B Relationship between mobility and pay

I now use the results of the previous section to study the relationship between mobility and pay. The $(k,j)$ entry in $M^o$ is the total number of workers making an EE transition from firm $j$ to $k$ over the sample period, where an EE transition means that there is a quarter where a worker has earnings at both employers. In treating the data this way, the assumption is that a worker moving from $j$ to $k$ faced a binary choice between $j$ and $k$. I then use equation (7) (and the definition of $S^o$ immediately preceding it in the text) to solve for $\breve{V}^{EE}$.

Panel C of Table II shows that by picking a ranking to fit the pattern of EE mobility, it is (perhaps unsurprisingly) possible to fit the pattern of mobility better than ranking firms based on

---

19To see this point in a simple example, consider the following matrix: $M^o = \begin{bmatrix} x & y \\ y & x \end{bmatrix}$, where $y > 0$, and $x \in (-\infty, +\infty)$. The eigenvalues of the normalized matrix are $\{1, \frac{x+y}{y} \}$. If we restrict attention to $x \geq 0$, then this second eigenvalue ranges from $[-1, 1)$. If we allow $x < 0$, then the range of the second eigenvalue is $(-\infty, 1]$.

---
pay. Specifically, while 57% of EE moves are to higher paying firms, 66% of EE moves are to higher $\tilde{V}^{EE}$ firms. The remaining rows of Panel C show that while $\tilde{V}^{EE}$ fits the moves better than pay, it is related to pay: a move to a higher-paying firm is more likely to yield an increase in $\tilde{V}^{EE}$ than a move to a lower-paying firm. Panel A of Figure III shows this pattern more generally.

Nevertheless, 30% of the moves with the biggest decreases in firm effects see an increase in $\tilde{V}^{EE}$. The revealed preference logic suggests that the presence of a systematic pattern of moves to lower-paying firms implies nonpay characteristics which outweigh the pay cuts. Panel E of Table II confirms that on average workers do move to higher-paying firms, but the systematic pattern of mobility is not perfectly explained by pay: the correlation coefficient between $\tilde{V}^{EE}$ and $\Psi$ is 0.43. An alternative implementation of the revealed preference logic used in this paper would be to assume a frictionless labor market and use market share (size) as a marker of utility. Compared to this marker, $\tilde{V}^{EE}$ is much more tightly aligned with pay. For example, the rank correlation coefficient between size and $\Psi$ is only 0.07.

**Moves to lower-paying firms only partly reflect layoffs:** A natural concern is that moves to lower-paying firms might not reflect workers’ preferences. Instead, workers might have been laid off. Note that the computation of $\tilde{V}^{EE}$ already allows for one form of layoffs through the idiosyncratic shocks. Specifically, because $\tilde{V}^{EE}$ captures the average pattern of mobility, it allows for any given move to be value decreasing. An additional way of addressing this concern is to follow a tradition starting with Jacobson et al. (1993) and use negative firm growth rates as a proxy for negative firm-level shocks. I restrict to firm-years whose annual growth rates are in the interval $[-10\%, +20\%]$. Panel E of Table II shows that when I estimate $\tilde{V}^{EE}$ in this restricted sample, the correlation rises to 0.57, but is still far from 1.

**Moves to lower-paying firms do not reflect differences in offer intensity (or size):** The data-generating process described above relied on the implausible assumption that all firms are the same size and make the same number of offers. In the next section, I write down a search model that nests the approach developed in this section. This model implies that under the assumption that all workers search from the same offer distribution, the appropriate functional form to adjust for offers and size is $\exp(\tilde{V}^{EE}) \frac{g}{\pi}$, where $g$ is the size of the firm and $f^o$ is a measure of the share of offers. The intuition for this functional form is that large firms will have more workers leaving them than small firms, but this does not mean that they are less desirable firms. Similarly, firms that make lots of offers will hire lots of workers, but this does not mean that they are desirable—in the context of the model, they are simply hunting for good draws of idiosyncratic utility.
To implement this correction, I need a measure of the share of offers made by a firm, or \( f^o \) (helpfully, \( g \) is observed in the data). The feature of the data that I use to measure the offer distribution is the share of workers hired from nonemployment hired by a particular firm. Using where nonemployed workers are hired to measure the offer distribution facing employed workers relies on two assumptions. The first assumption is that employed and nonemployed workers search from the same offer distribution. The second assumption is that the nonemployed workers do not reject offers and so where they are hired reveals the offer distribution.

Panel E of Table II shows that when I adjust \( \tilde{V}^{EE} \) by adjusting for offers and size the correlation between the adjusted \( \tilde{V}^{EE} \) and the firm effect, \( \Psi \), rises to 0.57, but is still far from 1.

The next section develops an economic model that nests equation (6) in a data generating process that more closely matches the labor market context rather than the static discrete choice context of the rank aggregation problem. An additional feature of the model is that it incorporates the features of the data discussed in this section. Moreover, it makes precise the (strong) assumptions needed to interpret aggregate firm-level outcomes as revealing information about values and nonpay characteristics.

IV Ranking firms using revealed preference

IV.A A model with utility-posting firms

The model is a partial equilibrium, posting, random, on-the-job search model with exogenous search effort and homogeneous workers in the spirit of Burdett and Mortensen (1998) where firms post utility offers. This class of models is sometimes described through the metaphor of a job ladder, where there is a common ranking of firms and workers try to climb the ladder through employer-to-employer mobility. Partial equilibrium and posting mean that I do not model the source of firm heterogeneity, and instead treat firms as mechanical objects. Posting means that there is no bargaining. Random means that search is not directed. Random search, jointly with the assumption that nonemployed and employed workers search from the same offer distribution, gives enough structure to estimate the offer distribution. On-the-job search is where the “revealed preference” action in the model is: sometimes workers get outside offers and decide whether or not to accept them. Exogenous search effort means that the arrival rate of offers does not depend on a worker’s firm. Finally, the assumption of homogeneous workers means that the world looks the same to all
workers—specifically, they have the same search parameters, search from the same offer distribution and—up to an i.i.d. draw—value all firms the same.

Types of separations

Before getting into the details of the model, I discuss at a conceptual level the types of separations that occur in the model, the motivation for these definitions, and what these are in the data.

As the rank aggregation exercise in section III highlights, the dataset I want to approximate is one with workers’ choice sets and the choices they made. It is conceptually clear what this dataset means in the case of unemployed workers who are looking for a job.\footnote{The reason is that in these cases we can difference out the value of unemployment and have a cleaner comparison of the value of the multiple job offers.} One example of such a paper is \textcite{Stern2004} in the context of PhD scientists looking for their first job, where he collects the set of offers the scientists received and their ultimate decision. Similarly, \textcite{HallMueller2017} have information on offers and acceptances for unemployed workers in New Jersey. In the empirical context of this paper of studying transitions between employers, it is hard to imagine a dataset that perfectly captures worker choice sets. One way to try to figure out whether workers had the choice of staying at their incumbent employer is to ask whether the worker separated in a quit or layoff, and then only use information in quits on the assumption that workers who report a quit had the choice of staying. This approach, while appealing, has some limitations. A long tradition in economics has discussed whether the quit-layoff distinction is meaningful in the presence of efficient turnover.\footnote{See, for example, \textcite{Beckeretal1977,McLaughlin1991}.} Concretely, this distinction is ambiguous because workers can quit when they see the writing on the wall, or can be laid-off when they decide that they want to quit and so reduce effort. Moreover, this distinction may not map into the choice set logic because quits respond to both push and pull factors. For example, \textcite{Flaenetal2017} document that the probability of separating and reporting a “quit” in survey data increases when employers contract. This finding suggests that survey-reported quits respond to both the pull factor of an outside offer, and the push factor of the change in the value of a current match.

Rather than trying to approximate the quit-layoff distinction, the model features two classes of separations: endogenous and exogenous. In the model, an endogenous separation comes from a maximizing choice, while an exogenous separation is one where the worker does not make a maximizing choice. In terms of the choice set logic, I only interpret the endogenous moves as potentially revealing preferences, while I interpret the exogenous moves as being like layoffs. The
maximizing choice in the endogenous moves takes into account both the common value of employers as well as an idiosyncratic utility shock. Because the idiosyncratic utility shock can be negative, a given EE move can reflect a mix of the pull factor of the desirability of the outside offer, and the push factor of the decline in the desirability of the current firm. In terms of the logic of survey-responses, were we to ask a worker whether a given endogenous move was a quit, many of them might respond “no” because the separation was in response to a negative shock at the incumbent employer.

In the data, I operationalize the distinction between endogenous and exogenous as follows. The model has a theory of the probability of an endogenous separation at each employer in each time period, which is given by the combination of the value of the employer, the randomness of offers, and the idiosyncratic utility draw. I view separations in excess of this probability at contracting employers as exogenous. The motivation for this distinction is that these excess separations are likely due to a firm-level shock. While a richer model would feature explicit firm-level shocks and then model how these shocks get propagated through the employment relationship and into separations, doing so is well-beyond the scope of this paper.

Here are some examples of what I have in mind when I refer to exogenous job destruction (EN) and reallocation (EE) shocks. An example of an exogenous job destruction shock is a mass layoff as in Jacobson et al. (1993). And an example of an exogenous reallocation shock are the EE transitions resulting from the increase in search activity in advance of mass layoff suggested by, for example, Bowlus and Villhuber (2002), where the idea is that workers know their firm is about to undergo a mass layoff and so take jobs that they would not have in the absence of the impending mass layoff.

Employers

What does an employer do in the model? An employer $j$ posts a flow payoff $v_j$, share of offers denoted by $f_j$ and employs a share of workers denoted by $g_j$. Finally, employers also differ in their exogenous separation rates $\delta_j$ and $\rho_j$, where $\delta_j$ is the probability of an exogenous job destruction shock that sends a worker to nonemployment, and $\rho_j$ is the probability of an exogenous reallocation shock. While my notation allows these shocks to be firm-specific, in estimation they vary across 20 sectors. Combined, the forward-looking value of being at firm $j$ is denoted by $V_e(v_j, \delta_j, \rho_j)$, which I abbreviate as $V_e^j$. This value includes both pay and nonpay components, $V_e^j = \omega(\Psi_j + a_j)$, where $a_j$ is the nonpay characteristic at firm $j$ and $\omega$ is the (unknown) unit conversion from log dollar units to forward-looking values.
Were I to impose steady state, then there would be a mechanical relationship between \( f_j \) and \( g_j \) (given all other parameters of the model). In estimation, I do not impose steady state so that I allow firms to grow and shrink. Because I do not impose steady state, one might think that growing firms would be mechanically better. In fact, the correlation between firm growth rates and the estimated value is negative: \(-0.10\).

**Workers**

What does a worker do in the model? The following Bellman equation summarizes how the model looks to a worker. A worker at employer \( j \) has the following value function:

\[
V^e(v_j, \delta_j, \rho_j) = v_j + \beta \mathbb{E} \left\{ \delta_j \int_{t_1} \{ V^n + \iota_1 \} dI \right\}
\]

\[
+ \rho_j (1 - \delta_j) \sum_k \mathbb{E} \left\{ \delta_j \int_{t_2} \{ V^e_k + \iota_2 \} dI \tilde{f}_k \right\}
\]

\[
+ (1 - \rho_j)(1 - \delta_j) \left[ \lambda_1 \sum_k \int_{t_3} \int_{t_4} \max \{ V^e_k + \iota_3, V^e_j + \iota_4 \} dI dI \tilde{f}_k \right]
\]

\[
+ (1 - \lambda_1) \int_{t_5} \int_{t_6} \max \{ V^n + \iota_5, V^e_j + \iota_6 \} dI dI \tilde{f}_j \}
\]

(8)

Reading from left to right, a worker employed at \( j \) has value \( V^e(v_j, \delta_j, \rho_j) = V^e_j \). This value consists of the deterministic flow payoff, \( v_j \), and the continuation value, which she discounts by \( \beta \).

The continuation value weights the expected value of four mutually exclusive possibilities. Two possibilities generate exogenous separations. With probability \( \delta_j \), a worker is hit with a job destruction shock and ends up in nonemployment, where she receives value \( V^n + \iota_1 \), where \( V^n \) is the common component of the value, and \( \iota_1 \) is an idiosyncratic draw from a type I distribution. With probability \( \rho_j \), a worker is hit with a reallocation shock and is forced to make an employer-to-employer move. The density of such offers from firm \( j \) is \( \tilde{f}_j \), and the summand is over the set of all

---

\(^{22}\) The fact that the idiosyncratic shock shows up on the forward-looking values, rather than in the flow payoff, may look odd but is standard in the conditional choice probability literature. See [Arcidiacono and Ellickson (2011), pg. 368](#).
employers. If this happens, the worker receives value $V^e_j + \iota_2$, where $V^e_j$ is the common component of the value, and $\iota_2$ is an idiosyncratic draw from a type I distribution.

Two possibilities generate endogenous separations. If neither of the exogenous separations happen, then with probability $\lambda_1$ a worker receives an offer from another firm, the probability the offer is from any particular firm is $f_k$ (note that this distribution is distinct from the distribution following reallocation shocks), and the worker makes a maximizing choice of whether to accept or reject the offer. The worker compares $V^e_k + \iota_3$ and $V^e_i + \iota_4$ and makes a maximizing decision. Note that the $\iota$s are drawn independently. Finally, if she does not receive an outside offer, then the worker makes a maximizing choice of whether to “quit” to nonemployment.

In equation (8), the model contains ingredients to address each of the three issues mentioned above. First, when a worker considers an offer from another firm she considers both the value of the firm, the $V^e_j$, as well as the idiosyncratic utility draw, $\iota$. By taking into account the $\iota$, the model allows workers to make off-setting moves and, more generally, for gross flows to exceed net flows. Second, the $\rho_i$ shocks allow the possibility that workers move between firms in an exogenous way. Third, the fact that different firms can make a different share of offers (i.e., it might be the case that $f_j \neq f_k$) means that patterns in mobility can reflect differences in offer rates rather than differences in value.

A nonemployed worker has the Bellman equation:

$$
V^n = b + \beta \sum_j \mathbb{E}\{ \lambda_0 \int_{\iota_7}^{\iota_8} \max\{ V^e_j + \iota_7, V^n + \iota_8 \} dI dI_f^j \} + (1 - \lambda_0) \int_{t_9}^{t_4} \{ V^n + \iota_9 \} dI \}.
$$

(9)

Reading from left to right, a nonemployed worker receives a total value of nonemployment of $b$, which includes both unemployment benefits as well as the value of nonmarket time and household production. Then each period two things might happen. She might receive an offer from an employer, in which case she decides whether or not to accept it. Or, nothing might happen. In this case, she receives a new idiosyncratic draw associated with nonemployment. Because of the assumption that the nonemployed and employed search from the same distribution—that is, the $f$
in the second row of equation (9) is the same $f$ in the third row of equation (8)—which firms hire nonemployed workers is informative about the offer distribution facing employed workers.

**Relationship between the search model and AKM**

Given that I ultimately compare the $V^e$ to $\Psi$, I now discuss how the assumptions in the search model and AKM relate. The take-away is that the mobility assumptions in the search model are both more and less restrictive than the mobility assumptions in Abowd et al. (1999) (AKM). In the search model, mobility depends on $V^e_j$ and $\iota$. On the one hand, this assumption is more restrictive than those in AKM because in AKM mobility decisions can depend also on the worker effect ($\alpha_i$), the covariates ($X_i \beta$) and the history (past and future) of firms ($\Psi_j$). That is, in AKM high-paid workers can have different mobility patterns than low-paid workers, or older workers can have different mobility patterns than young workers. This dependence of mobility patterns on $\alpha_i$ allows AKM to generate sorting of workers to firms. While the search model does allow mobility patterns to depend on the identity of the current firm, this dependence is more restrictive than what AKM assumes. Moreover, the search model does not allow mobility to depend on $\alpha_i$ and so cannot generate sorting of workers to firms. On the other hand, the assumption on mobility is less restrictive than those in AKM because I can relax the strict exogeneity assumption that mobility is independent of all past and future realizations of the residual ($r_{it}$). Specifically, in the context of the search model, mobility can depend on the idiosyncratic utility draw ($\iota$). I can selection correct the earnings equation and allow mobility to depend on the error term. See Appendix C for details.

**IV.B Estimating the utility levels that firms post**

This section shows how to estimate the utility levels that firms post. There are three steps of estimation: first, I summarize the systematic pattern of worker flows, which summarizes the information in accepted offers by averaging out idiosyncratic shocks; second, I estimate the offer distribution using information in nonemployment to employment flows, which allows for some of the patterns in accepted offers to reflect differences in recruiting intensity; and third, I measure exogenous separations as excess separations at contracting firms. The three steps are interdependent and I perform them in a loop.

---

23It is possible to include a persistent worker-firm match effect in AKM, which is not possible in the search model.
Summarizing the systematic pattern of worker flows

This section rewrites the model so that it gives rise to the same expression used in section III to measure the systematic pattern of worker flows. Relative to the derivation in section III, the point here is to show how the systematic pattern combines both the underlying value of the firm, the $V_j^e$, as well as differences in layoff rates ($\{\delta_j, \rho_j\}$) as well as differences in size ($g$) and the offer rates ($f$). Then the remaining parts of this subsection discuss how to estimate these components.

Recall that the point of the model is to find values that rationalize the structure of flows between employers. Record the endogenous flows between employers in a mobility matrix, denoted by $M$. The $(j,k)$ entry in $M$ is the number of endogenous flows to employer $j$ from employer $k$. In the model, workers receive one offer at a time and, therefore, only ever make binary choices. Because I adopt the standard continuum assumption in discrete choice models, such flows from employer $k$ to employer $j$ are given by

$$M_{jk} = \frac{g_k W (1 - \delta_k)(1 - \rho_k) \lambda_1 f_j \Pr(j > k)}{f_k g_j (1 - \delta_j)(1 - \rho_j) \Pr(k > j)},$$

where $W$ is the number of employed workers. To interpret this equation, note that there are $g_k W$ workers at employer $k$ and $(1 - \delta_k)(1 - \rho_k)$ share of them do not undergo exogenous separations. These workers get an offer from $j$ with probability $\lambda_1 f_j$ and accept the offer with probability $\Pr(j > k)$.

To connect this model to the expressions in section III the model implies a simple expression for the flow-relevant value of an employer and nonemployment. To derive this expression, consider relative flows between pairs of employers, which are given by:

$$\frac{M_{jk}}{M_{kj}} = \frac{f_j g_k (1 - \delta_k)(1 - \rho_k) \Pr(j > k)}{f_k g_j (1 - \delta_j)(1 - \rho_j) \Pr(k > j)}.$$

The type I extreme value distribution assumption simplifies $\Pr(j > k)$:

$$\frac{M_{jk}}{M_{kj}} = \frac{f_j}{f_k} \frac{g_k (1 - \delta_k)(1 - \rho_k)}{g_j (1 - \delta_j)(1 - \rho_j)} \frac{\exp(V^e(v_j))}{\exp(V^e(v_k))}.$$ 

Relative flows (accepted offers) are directly related to relative values, but multiplied by relative offers and effective size. Why do offers, and effective size matter? Offers matter because workers flow towards a firm that makes lots of offers. These flows do not reveal that the firm is more
desirable, but that it is hunting for a good idiosyncratic draw. Effective size matters because more
workers flow away from a large firm than a small firm. These flows do not reveal that the firm is
less desirable, but that there are more chances for a negative idiosyncratic draw.

Now introduce notation which defines the flow-relevant firm-level value that summarizes the
determinants of relative flows:

\[
\exp(\tilde{V}_j) \equiv \frac{f_j \exp(V_j^e)}{g_j (1 - \delta_j) (1 - \rho_j)} \times \frac{\text{offer } \times \text{value}}{\text{effective size}}.
\]

\(\exp(\tilde{V}_j)\) is the flow-relevant value of an employer. It combines differences in the underlying value of
an employer, as well as differences in (effective) size and the offer rate\(^{24}\). Combining (12) and (13) gives relative flows between employers in terms of \(\exp(\tilde{V}_i)\):

\[
\frac{M_{jk}}{M_{kj}} = \frac{\exp(\tilde{V}_j)}{\exp(\tilde{V}_k)}.
\]

This expression is the same as that in section [III]. There are two differences: first, I use the flows to
and from nonemployment. Second, the model now shows the appropriate functional form to adjust
for differences in layoff rates, size and offers.

### The offer distribution

As discussed in section [III] the main feature of the data that I use to pin down the offer distribu-
tion is where nonemployed workers are hired. Relative to the intuitive approach in section [III] I
allow nonemployed workers to reject offers. Two assumptions jointly allow estimation of the offer
distribution, while allowing nonemployed workers to reject offers. The first assumption is common
to section [III] and is that nonemployed and employed workers search from the same distribution. In

\[^{24}\text{For flows between employers and the nonemployed state, an analogous derivation implies:} \]

\[
\frac{M_{nj}}{M_{jn}} = \frac{(1 - \lambda_1) g_j W (1 - \delta_j) (1 - \rho_j) \Pr(n > j)}{\lambda_0 f_j U \Pr(j > n)} = \frac{\exp(\tilde{V}_n)}{\exp(\tilde{V}_j)} \times \frac{\text{offers } \times \text{value}}{\text{size}},
\]

where

\[
\exp(\tilde{V}_n) = \frac{(1 - \lambda_1) W \exp(V^n)}{\lambda_0 U}.
\]

\(U\) is the number of nonemployed workers, \(\lambda_0\) is the probability of an offer when nonemployed, and \(n\) is the nonemployed state.
the Bellman equations, the employed and nonemployed draw offers from a common $F$. The second assumption is new relative to section [III] and is that it is only when workers do not receive an outside offer that workers consider endogenous quits to nonemployment. In the Bellman equation (equation (8)), the rate of “offers” of nonemployment is $1 - \lambda_1$. The first assumption is important because it reduces the number of parameters to estimate and allows me to incorporate another dimension of the data.

The second assumption is important because it adds an additional moment: the level of EE flows. The parameter that matches the number of EE transitions implies how often workers get offers from nonemployment. Knowing how often employed workers get offers from nonemployment then distinguishes between two explanations for relative flows between employment and nonemployment: the relative value of nonemployment and the offer rate of nonemployment.

Formally, solving for the offer distribution combines information in three equations. The first equation defines $f^o_j$, or the share of workers hired from nonemployment hired by firm $i$. The fact that this is informative about the offer distribution reflects the first assumption:

$$f^o_j \equiv \frac{M_{jn}}{\sum_{k \in E} M_{kn}}.$$ (15)

The second equation captures information both in employment-to-nonemployment and nonemployment to employment flows:

$$\sum_{j \in E} M_{jn} e^{\tilde{V}_n} = \sum_{j \in E} \frac{\lambda_0 U f^e_j e^{V^e_j}}{e^{V^e_j} + e^{V^e_j}} (1 - \lambda_1) W e^{V^n} \lambda_0 U.$$ (16)

Finally, the third equation captures the level of employer-to-employer flows and pins down $\lambda_1$:

$$\frac{\sum_{k \in E \setminus \{j\}} \sum_{j \in E} M_{kj}}{W \sum_{j \in E} g_j (1 - \delta_j) (1 - \rho_j)} = \lambda_1 \sum_j g_j (1 - \delta_j) (1 - \rho_j) \sum_k f^e_k \frac{e^{V^e_k}}{e^{V^e_k} + e^{V^e_j}}.$$ (17)

**Identifying exogenous separations**

The revealed preference interpretation of an employer-to-employer transition relies on believing that workers had the option of staying at their original firm. Relative to the intuitive approach in section [III] where I threw out separations from firms that were contracting by “too much,” here I downweight separations when the firms are contracting. This approach has the benefit of preserving more of the information in the data.
The central idea is that there is some level of separations from the firm that we expect when there are no shocks to the firm, and any excess separations when the firm is contracting are due to a firm level shock. The expected level of separations are the endogenous separations in the model that arise from outside offers that are accepted and quits to nonemployment. Since these separations might happen because of negative idiosyncratic shocks, I do not interpret all of these as utility-increasing moves.

Mechanically this approach finds the level of endogenous separations based on what happens at expanding firms, and then views separations above that level when the firm contracts as the exogenous separations. To see this approach graphically and using aggregated data, consider Figure IV. The figure shows EE and EN separation probabilities as a function of quarterly employer growth. The approach takes the average separation rate on the right hand side of the graph—where the employer is expanding—and assumes that this rate is the endogenous separation rate. Then when the employer is contracting—on the left hand side of the graph—if the separation rate is higher than the endogenous rate then these excess separations are viewed as exogenous. That is, when the employer is contracting there are some workers who would have separated even if the employer was not contracting and some who separate because the employer is contracting. In practice, I do not know which worker falls into which group, and so I assign a probability that the separation was exogenous to each worker who separates at a contracting firm and add these up across years to get the exogenous separation rate at the firm.

I now write down equations that correspond to this idea for the EE separations and analogous equations apply for the EN separations. Let $EE(j,t)$ be a random variable equal to 1 if a worker at employer $j$ in year $t$ separates in an EE transition (and equal to 0 otherwise), and let $EE(j,t)^{endog}$ and $EE(j,t)^{exog}$ be random variables equal to 1 if the EE separation is exogenous or endogenous (and equal to 0 otherwise). Because a given separation is either exogenous or endogenous, $EE(j,t) = EE(j,t)^{endog} + EE(j,t)^{exog}$. When these probabilities are time-invariant, I suppress the $t$ argument.

The model’s theory of the endogenous EE separation probability at employer $j$—which corresponds to the EE separation rate at expanding firms on the right hand side of Figure IV—is given by

$$\Pr(EE(j)^{endog} = 1) = (1 - \delta_j)(1 - \rho_j) \left[ \lambda_1 \sum_{k \in \mathcal{E}} \frac{f_k}{\exp(V_k^e) + \exp(V_j^e)} \right],$$

$$\tag{18}$$

This figure is inspired by Davis et al. (2012, Figure 6).

This approach is conceptually related to that in Flaaen et al. (2017).
where this is the probability that a worker is not laid off, gets an offer, and accepts it. I suppress
the $t$ subscript in this expression because this endogenous probability is time-invariant.

Now develop notation to measure the separation rate when the firm is contracting. Let $C(t, j)$
be an indicator variable for firm $j$ being on the right hand side of Figure [V] and contracting in
terms of the number of workers in period $t$. The following expression records the probability of an
EE transition in a year $t$ when firm $j$ is contracting: $\Pr(EE(j, t) = 1 | C(t, j) = 1)$.

Finally, combine the previous two expressions to generate an expression for the probability that
a separation at firm $j$ when it contracts in year $t$ was “expected anyway” and so counts as an
endogenous separation:

$$\Pr(EE(j, t)^{endog} = 1 | EE(j, t) = 1, C(t, j) = 1) = \min \left\{ \frac{\Pr(EE(j)^{endog} = 1)}{\Pr(EE(j, t) = 1 | C(t, j) = 1)}, 1 \right\},$$

while the exogenous probability is the complementary probability: $\Pr(EE(j, t)^{exog} = 1 | EE(j, t) = 1, C(t, j) = 1) = 1 - \Pr(EE(j, t)^{endog} = 1 | EE(j, t) = 1, C(t, j) = 1)$. In terms of the figure (and
the EN measure), the endogenous probability is the thick solid line relative to the total length of
the line, while the exogenous probability is the thin dashed line relative to the total length of the
line. When a firm is expanding, I assume that the exogenous probability is 0 and the endogenous
probability is 1. Replacing EE with EN in the above gives analogous expressions to decompose EN
separations at contracting firms into exogenous and endogenous components.

Now I use these weights to construct two objects that feature in the model: $M_{ji}$, or endoge-

uous mobility, and the exogenous shocks $\{\delta_j, \rho_j\}$. This amounts to reweighting the observed
flows (things with $o$ superscripts) by the probability that the flow was endogenous. For all $j, k \in \mathcal{E}$
$M_{kj} = \sum_t M_{kj,t}^o \Pr(EE(j, t)^{endog} = 1 | EE(j, t) = 1)$, and, for flows to nonemployment
($n$), and for all $k \in \mathcal{E}$ $M_{nj} = \sum_t M_{nj,t}^o \Pr(EN(t, j)^{endog} = 1 | EN(t, j) = 1)$. I assume that
flows from nonemployment are all endogenous: $M_{jn} = \sum_t M_{jn,t}^o$. Finally, the two shocks come
from summing over all the exogenous moves: $\rho_j = \frac{\sum_t \sum_{k \in \mathcal{E}} \sum_{i \in \mathcal{E}} M_{kj,t}^o \Pr(EE(t, j)^{exog} = 1 | EE(t, j) = 1)}{\sum_t \sum_{k \in \mathcal{E} \cup n} \sum_{i \in \mathcal{E}} M_{kj,t}^o}$, and
$\delta_j = \frac{\sum_t \sum_{n \in \mathcal{E}} \sum_{j \in \mathcal{E}} M_{kj,t}^o \Pr(EN(t, j)^{exog} = 1 | EN(t, j) = 1)}{\sum_t \sum_{k \in \mathcal{E} \cup n} \sum_{i \in \mathcal{E}} M_{kj,t}^o}$.

To implement this approach, I need an estimate of the values of firms, which then depends on
how I divide the transitions between exogenous and endogenous. Hence, I perform this step in a
loop where I update my model-predicted separation probabilities at each step. I initialize the loop
using the average EE and EN separation rates in expanding years. In practice, it takes only a few
iterations for the loop to converge.
While it is possible to estimate firm-specific values of $\delta$ and $\rho$, to limit the number of parameters to estimate I aggregate to the sectoral level in two places. Specifically, in equation (18) I take the sectoral average (weighting firms by the number of person-years) of the model-implied separation probabilities, and then use this coarser counterfactual throughout the equations in this section. In addition, when I substitute into equation (13) to unravel for the firm-specific value, I use the sectoral average of the $\delta$ and $\rho$.

Summary

Appendix G provides details on how the pieces of estimation fit together. Given a sample, there are two loops and a grid search. The outer loop is the loop to separate the exogenous and endogenous transitions, which is conditional on a guess for the firm values ($V^e_j$). The inner loop is the fixed point problem to compute the $\exp(\tilde{V})$, which depends on the division between exogenous and endogenous transitions. The grid search is the step of finding a $\lambda_1$ which generates the extent of EE transitions. A byproduct of an estimate of $\lambda_1$ is an estimate of the offer distribution $F$. Since size, $g$, is observed, this allows me to unravel the $\exp(\tilde{V})$ and recover a new guess of the firm value, which I use to generate an updated division of separations into exogenous and endogenous (and estimates of the $\delta$ and $\rho$) and then repeat the three steps.

IV.C Measurement error

As detailed in section V, the empirical exercise in this paper reduces to computing the $R^2$ between the values and earnings and so noise in either term biases me towards finding a larger role for compensating differentials. I take three complementary approaches to addressing this possibility. First, I use an empirical Bayes approach and shrink the estimates. Second, I split the sample in half on the basis of workers, which lets me generate two independent estimates of the values and pay at each firm. The correlation between these two estimates is thus informative about the amount of noise. Third, I consider a subset of very large firms (1000 or more non-singleton observations per year) and re-estimate the values and earnings in this subsample. Appendix H provides more details and Appendix I provides Monte Carlo evidence. Needless to say, there is an active literature exploring how to do estimation and inference in matched employer-employee datasets and the literature has not reached a consensus on best practices. Hence, these various checks should be viewed as suggestive rather than definitive.
To provide more intuition about the model, Appendix J shows that it is also possible to rank firms based on properties of just the worker inflows or properties of just the worker outflows. The inflow-based measure follows Bagger and Lentz (2017) and ranks firms based on the share of hires on (endogenous) employer-to-employer moves relative to all hires. The reason this statistic ranks firms in the model is that “better firms hire from better firms,” where nonemployment is viewed as an exceptionally bad firm. The outflow-based measure follows a tradition in the inter-industry wage differential literature and ranks firms based on the (endogenous) quit rate. The reason this statistic ranks firms in the model is that “workers are less likely to leave better firms.” Estimating the full model has the benefit of providing a model-consistent way of separating transitions into endogenous and exogenous, incorporating information in both inflows and outflows, providing a way of measuring how important the common component of job value is relative to the idiosyncratic part, and incorporating the feature of the data that is most directly about revealed preference: the structure of employer-to-employer transitions.

Before describing the main results of the paper, I now present some results internal to the model and discuss features of the data the model does not match.

Table III shows the estimated model parameters, as well as the split of separations into EE and ENE. The top rows of Table III shows that the annual separation probability is about a quarter and only two-fifths of these are employer-to-employer. About three-quarters of these EE transitions are endogenous, while only two-thirds of the ENE transitions are endogenous. Combining the exogenous weights and the separation probabilities gives an annual exogenous employer-to-nonemployment, or job destruction, rate ($\delta$) of 0.04. Combining the exogenous weights and the separation probabilities, gives an annual exogenous employer-to-employer, or reallocation, rate ($\rho$) of 0.03. The feature of the data that drives the estimates of $\delta$ and $\rho$ is the extent to which the EE and EN separation probabilities are higher at contracting than at expanding employers, and the probability of employers contracting. Finally, the annual probability of receiving an offer is a little under one-fifth. The feature of the data that drives this estimate is matching the level of EE transitions given the estimated offer distribution and the relative values of employers. I discuss the probability of accepting an offer from nonemployment in section VI.
While, the model matches many features of the structure and level of transitions, it does not perfectly fit the data. Here are two features of the data the model has no mechanism to match. First, the model has no mechanism to generate sorting of workers to firms because in the model all workers are ex-ante identical. Second, the model does not generate the extent of the decline in the separation rate by job-tenure because separation rates are constant within each firm.

V Main results

So far I have shown how to estimate a value of each employer as revealed by worker choices, as well as the earnings at each firm. This section shows how to combine these two measures to decompose the variance of firm earnings into rents and compensating differentials. I find that about 70% of the firm component of the variance of earnings reflects compensating differentials.

V.A Measuring compensating differentials and rents

To decompose firm-level pay into rents and compensating differentials, I confront the basic challenge that the theory of compensating differentials is typically cast in a frictionless labor market where utility is equalized across jobs (at the margin). In contrast, this paper’s framework allows the possibility that there are frictions so that utility is not equalized across jobs and workers earn rents. In this setting, nonpay characteristics play two roles: first, they can augment differences in pay and disperse utility, and, second, they can compensate for differences in pay and compress utility.

To clarify what compensating differentials mean in this setting and what can be identified, I write down a simple partial equilibrium model of a maximizing firm where there are two motives for variation in nonpay characteristics across firms: a “Rosen” motive corresponding to variation in the marginal cost of providing amenities, and a “Mortensen” motive corresponding to variation in the desired level of utility provided by the firm. I show that knowledge of the pay and value perceived by workers identifies the variation in amenities due to the “pure” Rosen motive, i.e., the piece of the Rosen motive that is orthogonal to the Mortensen motive. I label nonpay characteristics arising from this source of variation as compensating differentials. In Appendix F I present an alternative motivation for this decomposition where I show that the compensating differentials component

\footnote{For example, Hwang et al. (1998), Lang and Majumdar (2004) and Mortensen (2003) construct models where utility is not equalized across jobs and, in equilibrium, pay and nonpay characteristics are positively correlated. Similarly, Pierce (2001) presents evidence that benefits are positively correlated with earnings.}
corresponds to the minimum variance of the nonpay components that rationalizes the firm-level measures of value and pay assuming additive separability of pay and nonpay components.

Suppose that workers’ utility functions are given by the following equation

\[ V^e_j = \omega (\Psi_j + \ln a_j), \]

(20)

where \( V^e_j \) is the utility of being employed at firm \( j \), \( \Psi_j \) is the (log) earnings at the firm, and \( a_j \) is the level of nonpay characteristics (amenities) provided by the firm. \(^{28}\) (I include an \( \omega \) to allow for there to be an arbitrary unit conversion from (log) dollars to utility. In what follows, for notational simplicity I set \( \omega = 1 \), but never use the fact that I know the units on \( V^e_j \).) Firm \( j \)'s objective function is:

\[ \max_{\Psi_j, a_j} \pi_j - c_j a_j - \exp(\Psi_j) \]

(21)

such that \( V^e_j = \bar{V}^e_j \). Here, \( \pi_j \) is the profit from hiring the marginal (and average) worker, \( c_j \) is the marginal cost of amenity provision, and \( \bar{V}^e_j \) is the desired level of utility provision. I allow for an arbitrary covariance between the marginal cost of amenity provision and the desired level of utility provision. Before considering the general case, it is helpful to consider two special cases of this model where only the “Rosen” or “Mortensen” motives are operative.

**Heterogeneity in cost of amenity provision:** We can interpret \( \text{Rosen (1986)} \) as considering the case where utility is equalized across firms so that \( V^e_j = \bar{V}^e \), a constant. Then the only stochastic element in the amenity provision expressions is the marginal cost of amenity provision, \( c_j \), and the following statements are true. First, wages and amenities are perfectly negatively correlated:

\[ \text{Corr}(\Psi_j, \ln a_j) = -1. \]

Second, and abusing notation, wages and utility are uncorrelated, that is,

\[ \text{Corr}(\Psi_j, V^e_j) = 0. \]

Finally, the only source of variation in nonpay characteristics is the variation in the cost of amenity provision across firms:

\[ \text{Var}(\ln a_j) = \frac{1}{4} \text{Var}(\ln c_j). \]

**Heterogeneity in utility:** We can interpret \( \text{Lang and Majumdar (2004) and Mortensen (2003, pg. 26-28)} \) as highlighting the strikingly different implications of the case where there is no variation in the marginal cost of amenity provision, \( c_j = \bar{c} \), and there is variation in utility. For simplicity, set \( \bar{c} = 1 \). Then the only stochastic element in amenity provision is \( \bar{V}^e_j \) and the following statements are true. First, wages and amenities are perfectly positively correlated:

\[ \text{Corr}(\Psi_j, \ln a_j) = 1. \]

Second, \(^{28}\) In Appendix \[ I \] show how to use the \( \text{Hotz and Miller (1993)} \) inversion to recover the flow-payoff, \( v_j \). I then estimate \( v_j = \omega' (\Psi_j + a_j') \), and find nearly the same relationship.
wages and utility are perfectly positively correlated, $\text{Corr}(\Psi_j, \bar{v}_j) = 1$. Finally, the only source of variation in nonpay characteristics is the variation in desired utility across firms: $\text{Var}(\ln a_j) = \frac{1}{4} \text{Var}(\bar{V}_j^e)$.

Both sources of heterogeneity: Now consider the case with both sources of heterogeneity. The previous stark results become ambiguous and depend on the relative strength of the two motives and their covariance. First, the sign of the correlation (or the covariance) between wages and amenities is ambiguous, depending on the relative variances of the two motives: $\text{Cov}(\Psi_j, \ln a_j) = \frac{1}{2} \text{Var}(\bar{V}_j^e) - \frac{1}{2} \text{Var}(\ln c_j)$. Second, the magnitude of the covariance between wages and utility ranges from 0 to 1 depending on both the variance in desired utility provision and its covariance with the cost of amenity provision, $\text{Cov}(\Psi_j, \bar{V}_j^e) = \frac{1}{2} \text{Var}(\bar{V}_j^e) + \frac{1}{2} \text{Cov}(\ln c_j, \bar{V}_j^e)$. Finally, the variance of nonpay characteristics depends on the variance of the two motives as well as their covariance, $\text{Var}(\ln a_j) = \frac{1}{4} \left[ \text{Var}(\bar{V}_j^e) + \text{Var}(\ln c_j) - 2 \text{Cov}(\bar{V}_j^e, \ln c_j) \right]$.

In this general case, what can be identified in terms of primitives of the firm’s problem given observables ($V_j^e$ and $\Psi_j$)? I now show that I can identify the variation in amenities that comes from the “pure” Rosen motive, or the variation in the cost of amenity provision that is orthogonal to variation in desired utility level. But I cannot identify the variation in amenities that contribute to utility dispersion, i.e., those that come from the Mortensen motive. The fundamental reason is that I do not know the variance of utility in log dollar units.

To see this result, consider the hypothetical regression of the cost of amenity provision on the variation in desired level of utility (and assume that variables are mean zero), $c_j = \beta \bar{V}_j^e + \tilde{c}_j$, where $\tilde{c}_j$ is the residual and thus is orthogonal to $\bar{V}_j^e$ by construction. This orthogonal component in the cost of amenity provision has the following variance: $\text{Var}(\tilde{c}_j) = \text{Var}(\ln c_j) - \frac{\text{Cov}(\ln c_j, \bar{V}_j^e)^2}{\text{Var}(\bar{V}_j^e)}$. Similarly, consider the variance in pay that is orthogonal to variation in utility in terms of primitives of the firm problem, or $(1 - R^2) \text{Var}(\Psi_j) = \frac{1}{4} \left[ \text{Var}(\ln c_j) - \frac{\text{Cov}(\ln c_j, \bar{V}_j^e)^2}{\text{Var}(\bar{V}_j^e)} \right]$, where $R^2 = \text{Corr}(V^e, \Psi)^2$. Up to rescaling, these terms are identical. Hence, I decompose the variation in amenities into two pieces:

$$\text{(22)} \quad \text{Var}(\ln a_j) = \frac{1}{4} \left[ \text{Var}(\ln c_j) - \frac{\text{Cov}(\ln c_j, \bar{v}_j)^2}{\text{Var}(\bar{v}_j)} + \text{Var}(\bar{v}_j) - 2 \text{Cov}(\bar{v}_j, \ln c_j) + \frac{\text{Cov}(\ln c_j, \bar{v}_j)^2}{\text{Var}(\bar{v}_j)} \right].$$

Rosen

Mortensen
where the Rosen piece can be measured by \((1 - R^2) \text{Var}(\Psi_j)\). This logic also leads to a decomposition of the variance of earnings into a rents and compensating differentials component:

\[
\text{Var}(\Psi) = \frac{1}{4} \left[ \text{Var}(\tilde{V}_j^e) + 2 \text{Cov}(\tilde{V}_j^e, \ln c_j) + \frac{\text{Cov}(\tilde{V}_j^e, \ln c_j)^2}{\text{Var}(\tilde{V}_j^e)} \right] \text{ rents}
\]

\[
+ \text{Var}(\ln c_j) - \frac{\text{Cov}(\tilde{V}_j^e, \ln c_j)^2}{\text{Var}(\tilde{V}_j^e)} \text{ compensating differentials}
\]

\[= R^2 \text{Var}(\Psi) + (1 - R^2) \text{Var}(\Psi) \text{ rents}
\]

\[- \text{ compensating differentials}\]

(23)

Figure Va illustrates this result. It shows a binned scatterplot of the firm-level values and earnings as well as the line of best fit. Rents can be seen in the upward slope in the line of best fit, which shows the variation in firm-level earnings that is reflected in variation in values. Compensating differentials—or Rosen amenities—are variation in firm-level earnings holding utility constant. This variation is depicted in the red dashed lines that show plus and minus one standard deviation bands of firm-level earnings holding utility constant. In contrast, because the x-axis is in units of the standard deviation of the idiosyncratic utility draw, Mortensen amenities are not identified. Mortensen amenities stretch out the x-axis in log dollar units, but the conversion from log dollars to the standard deviation of the idiosyncratic draw is not known.

V.B Decomposing firm-level pay into compensating differentials and rents

Table IV shows that about 70% of the firm component of pay is compensating differentials, while about 30% is rents. Aggregating this finding implies that compensating differentials account for at least 15% of the variance of earnings, since firms account for about 21% of the variance of earnings.

The three approaches to addressing measurement error deliver quantitatively similar results.

While this estimate of the role of compensating differentials might strike some readers as large, it relies on a conservative interpretation of the EE moves to lower-paying firms. Recall from Panel B of Table II that 43% of EE moves are to lower-paying firms. Panel D shows, however, that the model only interprets 40% of these moves as being value-increasing moves—i.e., moves where the increase in the nonpay characteristics fully offsets the decline in pay. More generally, the lower panel of Figure IIIb shows the share of moves that the model interprets as being to higher-value jobs.

\[\text{Table II shows that the Pearson correlation in the microdata underlying this plot is } 0.53 \text{ and the Spearman (rank) correlation is } 0.51. \text{ So given a monotone relationship, the deviations from linearity are not large.}\]

\[\text{The unadjusted } R^2 \text{ is } 0.28. \text{ The bootstrap-adjusted share is } 0.30. \text{ Using the correlations between samples 1 and 2 implies an } R^2 \text{ of } 0.306 = \frac{0.281}{0.970 \times 0.948}. \text{ And in the large firm sample the } R^2 \text{ is } 0.36. \text{ The Monte Carlo evidence in Appendix } I \text{ suggests that these adjustments might be biased down by one or two percentage points.}\]
firms as a function of the change in pay. For the largest earnings cuts, the model only interprets 20% of them as value-increasing moves.\footnote{Moving from $\tilde{V}^{EE}$ to $V^e$ addresses two issues highlighted in section \ref{sec:lagoffs} layoffs, and differences in offer intensity. The table shows three ways in which addressing these issues generates more “reasonable” answers. First, I find a tighter alignment between pay and $V^e$ than between pay and $\tilde{V}^{EE}$. Second, I find a tighter relationship between size and $V^e$ than size and $\tilde{V}^{EE}$, and size is often taken as an external marker of quality (e.g., Brown and Medoff (1989) or Moscarini and Postel-Vinay (2016).) Third, adjusting for offer intensity—which uses information in the NE flows—shifts down the probability of making a value increasing EE move, as we would expect if some of the systematic pattern was due to differences in offer intensity.}

The model explains the large number of non-value increasing moves to lower-paying firms through a large variance in the idiosyncratic shocks. Figure \ref{fig:distributions} plots the distribution of the values of the firms, and the idiosyncratic shocks. The variance of the idiosyncratic shocks is larger than the variance of the values of firms.\footnote{The variance of the idiosyncratic shocks is $\pi^2/6 \approx 1.6$ by normalization. The raw variance of the values is 0.847, while taking into account measurement error gives an estimate of 0.814.} Intuitively, the relative variance of the values and the idiosyncratic shocks is identified by the extent to which gross flows exceeds net flows.\footnote{To see this, return to equation \ref{eq:balanced}, and suppose that flows are exactly balanced ($M_{jk} = M_{kj}$). Then all flows are gross flows and the values of the two firms are identical. Alternatively, suppose that there are only flows in one direction and all flows are net flows. Then one firm is infinitely better than the other firm.} Given the importance of the idiosyncratic shocks in explaining mobility, it is logically possible that all the moves to lower-paying firms would have been best explained by idiosyncratic shocks rather than the pursuit of common features of firms valued by all workers.

While I find an important role for compensating differentials, I also provide two sources of evidence of rents in the labor market. First, I find that pay and values are positively correlated and this explains a non-trivial share of the variance of firm-level pay. As I discuss in the conclusion, this finding is consistent with a long tradition in labor economics. Second, I find that there is significant dispersion in the underlying values of the firms relative to the idiosyncratic utility draws. Thus, this paper provides way of quantifying the importance of a common job ladder in explaining worker mobility, which is at the heart of many frictional models of the labor market.

\subsection*{V.C Relating compensating differentials and rents to observables}

One potential concern with using revealed preference is that this approach amounts to labelling a residual and so compensating differentials are observationally equivalent to measurement error. To show that these differences do not look like measurement error, I show that these are related to industry and locations in plausible ways.
The first characteristic is sector. Table V shows that sector accounts for about 30% of the variance of nonpay characteristics, while going to 4 digit industry increases the variance share to 45%.

The estimates of sectoral-level nonpay characteristics are plausible and match intuitions about differences in working conditions. Figure Vb plots the sector-level values and earnings (and implicitly nonpay characteristics). The x-axis shows the values, which are in units of a standard deviation of a type I extreme value distribution. The y-axis shows the pay, which are in log-dollar units. The vertical slices in the graph show the Rosen amenities, since these slices reflect variation in pay while holding overall value constant. As evidence of plausibility, some sectors that are relatively high-paying relative to the value they offer workers are mining, construction and transportation/warehousing, which I interpret as evidence of compensating differentials. Similarly, education, public administration and arts/entertainment/recreation are low-paying relative to the value they offer workers, which is evidence of desirable nonpay characteristics.

The second characteristic is location. Table V shows that state captures 8% of the variance of nonpay characteristics. Similarly, county explains 15%. Consistent with the intuition from the Rosen-Roback model, the location-level variation in pay is weakly related to variation in value.

Jointly, industry and location account for about 60% of nonpay characteristics. Hence, much of the variation in nonpay characteristics is related to observable characteristics, and does so in intuitively plausible ways.

V.D Are the compensating differentials just variation in hours?

One observable that might be important is variation in hours. In my data (as in the U.S. tax data used by Piketty and Saez (2003) and Chetty et al. (2014b) among others), I observe earnings and not hours, and so a low-earnings job might just be a low-hours job. To the extent that variation in hours explains the compensating differentials, this finding would add credibility to the methods of the paper by mapping the compensating differentials to an observable. But it would detract from the novelty as there is a long tradition in economics that people value leisure.

---

34 Holzer et al. (1991, Figure 1) present a similar sector-level scatterplot, in their case of industry wage differentials and application differentials and interpret it in a similar way.

35 Without shrinkage, this statement would be exact because the count and 4-digit industry means are estimated in a common regression. With shrinkage, this statement is no longer exact.

36 This interpretation views variation in hours as coming from labor supply factors. Of historical interest, Smith (1776/2003, pg. 143-144) discusses variation in hours coming from labor demand factors as a source of compensating differentials in hourly wages and hence implicitly argues that it is annual earnings that should be equalized across jobs: “In the greater part of manufactures, a journeyman may be pretty sure of employment almost every day of
To attempt to quantify the role of hours in explaining compensating differentials, I consider the extent to which variation in hours can explain aspects of Figure Vb. The reason to focus on the sectoral-level picture is that the data I have do not contain hours, but I can construct measures of sectoral weekly hours using the March CPS. Hours variation explains about 15% of the variation in sectoral compensating differentials. To the extent that the sectoral analysis maps into the firm analysis, I conclude that while some of the compensating differential is variation in hours, it is unlikely to be the main compensating differential.

VI Other implications

VI.A Inequality

This section considers the consequences for earnings inequality of equalizing the $a_{Rosen}$ portion of nonpay characteristics across jobs and compensating workers (this leaves the $a_{Mortensen}$ portion unpriced). Mechanically, this amounts to replacing the $\Psi$ in the earnings equations with $\beta V^e$, where the $\beta$ is estimated from a regression of $\Psi$ on $V^e$. I find that equalizing nonpay characteristics and compensating workers reduces earnings inequality.

The effect on inequality of pricing the $a_{Rosen}$ component is ambiguous and depends on the correlation between nonpay characteristics and overall earnings. To see how earnings inequality might understate well-being inequality, suppose economists dominate the data. Economists are well-paid workers in jobs with good nonpay characteristics. Were we to equalize nonpay characteristics and compensate workers for it, then a highly-paid worker would be even more highly paid, and so earnings inequality would understate well-being inequality. In contrast, suppose artists dominate the data. Artists are low-paid workers in jobs with good nonpay characteristics. Were we to equalize nonpay characteristics and then compensate workers, then a low-paid worker would make more. Hence, the variance of earnings would go down, and earnings inequality would overstate well-being inequality.

---

37Panel A of Table A9 shows the sectoral measures of hours, and the footnote provides more detail on sample construction. Panel B reports the $R^2$. The year that he is able to work. A mason or bricklayer, on the contrary, can work neither in hard frost nor in foul weather, and his employment at all other times depends upon the occasional calls of his customers. He is liable, in consequence, to be frequently without any...The high wages of those workmen, therefore, are not so much the recompence of their skill, as the compensation for the inconstancy of their employment."
Table VI shows that equalizing nonpay characteristics and compensating workers would reduce inequality, that is, the artist example dominates the data. About half of the effect occurs at the industry level.

Equalizing nonpay characteristics and compensating workers has surprising impacts on the structure of earnings. Figure VII plots the actual distribution of earnings and the distribution in this counterfactual. Inequality is reduced primarily by shifting in the lower tail of the distribution. In contrast, Figure VIII shows a naive counterfactual. To compute this counterfactual, I multiply the firm effects by $1 - R^2$ and recompute the variance of earnings. Relative to the data, the naive counterfactual shifts in both the lower and the upper tail of the income distribution.

VI.B Ability of search models to match earnings dispersion

The message of this paper is consistent with Hornstein et al. (2011). They argue that benchmark search models cannot rationalize the extent of earnings dispersion—measured as the residual in a Mincerian regression—in the labor market. Their observation is that unemployed workers find jobs quickly—which suggests that workers do not face a large amount of dispersion in job value in the offer distribution since otherwise they would wait for a better offer.

By focusing on the behavior of employed workers, rather than unemployed workers, this paper provides a complementary source of evidence to Hornstein et al. (2011) that rents do not explain all earnings dispersion. The key evidence is systematic patterns of employed workers making employer-to-employer transitions to lower-paying firms. This finding indicates that a large portion of measured earnings dispersion is not treated by workers as reflecting dispersion in value. I interpret this finding as evidence for compensating differentials.

Hornstein et al. (2011) argue that search models that explain earnings dispersion typically imply implausibly low values of unemployment. Because utility is only measured up to an additive constant, I cannot compare the value of nonemployment to the average value of a job. A statistic about the value of nonemployment that I can compare to other estimates is the share of offers accepted among the unemployed. Hall and Mueller (2017, pg. 22) report that in their sample of job seekers in New Jersey collecting unemployment insurance, unemployed workers had accepted 71.9% of offers. I estimate that the nonemployed accept 78.1% of offers, which provides evidence that my estimate of the value of nonemployment is plausible.

I use the word counterfactual simply to summarize this mechanical measurement. The model is partial equilibrium and so I implicitly assume that this is costless for firms.
VII Robustness

I now present model results estimated on a variety of subgroups, which addresses concerns that my results are driven by learning, preference heterogeneity, or inappropriately pooling labor markets.\footnote{Limitations on computational resources meant that I only bootstrapped the main results in the paper.}

VII.A Learning: age subsamples

One explanation for patterns of wage increases and decreases at job changes is that this reflects learning in the labor market. For example, high-paying firms might be firms that hire workers when the market learns good things about them, while low-paying firms are firms that hire workers when the market learns bad things about them.

While a complete analysis of learning is well-beyond the scope of this paper, it is worth pointing out the empirical implications of two extreme cases. In the symmetric updating and public learning case the market can learn both good and bad things about a worker, this learning generates offsetting flows that are captured in the idiosyncratic shocks and it is not obvious how it biases either the estimates of pay or overall values. In contrast, some high-skill labor markets (like academia or high-powered corporate law) are characterized by asymmetric updating where learning only contributes to downward mobility (e.g., assistant professors do not get tenure, law firm associates do not make partner). In this case, such learning could lead to substantial bias in that I might recover the opposite of the correct ranking and hence find a substantial role for compensating differentials.

I consider two ways of detecting the presence of learning. The first is to split the sample on the basis of age and estimate the model among young (18-34) and old (35-61) year old workers.\footnote{This divides the sample at approximately half of the EE transitions.} The reason age is a useful way to divide the data is that Lange (2007) emphasizes that learning is quite quick, and so it is plausible that the effects of learning on mobility and earnings dynamics are concentrated in the younger sample. Table IV shows that the correlation in the pay between older and younger worker is 0.87, the values is 0.83, and the estimated non-pay characteristics is 0.78. Similarly, while there are differences in the decomposition between the two subsamples, in both subsamples the firm-level compensating differentials account for over 15% of the variance of earnings. The second way is to follow Card et al. (2013, pg. 991) and note that in the event studies in Figure A4 there are no differential pre-trends between the workers who move to lower- and higher-paying firms. In contrast, under a spot market assumption with public learning where
ability is more rewarded in some sectors (or firms) than others (as in Gibbons et al. (2005)), we would expect to see larger increases in earnings among those transitioning to higher-paying firms in the pre-period. The reason we would expect to see pre-trends is because the market is updating positively about the workers, and the spot market assumption implies that this learning would be reflected in pay. Combining these pieces of evidence, I conclude that learning is unlikely to drive the results.

VII.B Preference heterogeneity: gender subsamples

One concern is that the model assumes that all workers have common preferences (up to the idiosyncratic utility shocks) and that there is no heterogeneity in the treatment effect of firms; that is, all workers agree on the ranking. As with learning, this form of model misspecification could lead me to find a weaker relationship between values and earnings than I would find if I indeed took this heterogeneity into account.

It is not enough to observe preference heterogeneity for the model to be misspecified in a way that generates evidence of compensating differentials. In terms of the earnings, heterogeneity in treatment effects of firms will tend to find that the firm effects are less important, and so the overall role of compensating differentials in the earnings structure would be smaller. In terms of the values, heterogeneity tends to weaken the systematic patterns of mobility in the labor market; that is, it will increase the variance of idiosyncratic draws relative to the values. Combined, the simplest forms of heterogeneity do not push me towards finding a larger role for compensating differentials.

Nevertheless, one simple way to assess the role of heterogeneity is to consider gender as an observable form of heterogeneity. Estimating the model separately by men and women, in Table IV shows that the decomposition into rents and compensating differentials is quite similar within each group as overall, and the weighted average says that the firm-level compensating differentials explain about 15% of the variance of earnings. Moreover, the values and earnings of men and women are quite correlated.

VII.C Worker heterogeneity: subsamples by worker fixed effects

The assumption in this paper is that all workers share a common ranking of firms. One of the stark predictions of models of sorting based on comparative advantage is that workers of different types will have different rankings of firms. In addition, one of the ways in which AKM relies on weaker
assumptions than the search model is that it allows for different mobility patterns as a function of
the worker type.

I split the data on the basis of the estimated worker effects ($\alpha_w$ in equation (1)) and then re-
estimate the model in both samples. Table IV shows that while there are differences between the
groups, the weighted average estimate is that firm-level compensating differentials account for 21%
of the variance of earnings.

VII.D Pooling labor markets: state and sector subsamples

The assumption in this paper is that I can treat all data as coming from one labor market. While
this aggregate perspective follows the broader Abowd et al. (1999) literature, it is helpful to consider
whether aggregation drives results.

One way to disaggregate the labor market is to estimate the model sector-by-sector. Table IV
shows the results for the five sectors where the model finds the smallest role for nonpay character-
istics in explaining within-sector variation in pay. The key take away is that restricting to within
sector moves does not generate a tighter relationship between pay and value.

An alternative way to disaggregate the labor market is to go state-by-state. As such, I re-estimate
the model using only within state moves. Table IV shows that the overall weighted average of the
decomposition is that about 14% of the variance of earnings is firm-level compensating differentials.

VIII Discussion

This paper develops a framework to measure compensating differentials that takes into account
the difficulty of measuring nonpay characteristics and the possibility of utility dispersion in the
labor market. Implementing the framework, I find some evidence that compensating differentials
play a role in explaining the variance of earnings. I reach this conclusion by focusing on the firm
component of the variance of earnings, and separating this component into rents and compensating
differentials.

My approach takes into account several explanations for why the firm component of pay does
not perfectly line up with values. I started from the observation that 43% of EE moves are to lower-
paying firms. I showed that these moves could not be explained by trading off a lower intercept in
pay for a steeper slope. I then showed that these moves reflected a systematic pattern of mobility

Recall from Table V that 40 percent of the variation in pay is between sector.
and so could not be explained by idiosyncratic utility differences. Then I estimated a model which took into account the following explanations for these moves: idiosyncratic shocks, layoffs, and differences in the offer intensity. Even taking these factors into account, I still found a large role for compensating differentials. Nevertheless, my results still require many caveats. There might still be many layoffs in the data. Or the simple model of the labor market I use might be a poor approximation; for example, the model does not generate sorting between firms and workers.

This paper echoes some themes in the interindustry wage differentials literature. That literature found differences in pay across sectors that were not explained by worker observables. Similar to the limited success of the compensating differentials literature, Krueger and Summers (1988) found that observable industry characteristics explained little of interindustry wage differentials, supporting the interpretation of these differentials as rents. The exercise in this paper is related to another exercise that supported the interpretation of interindustry differentials as rents. A long line of work in the interindustry wage differential literature and elsewhere relates some measure of the value of the firm derived from worker behavior to some measure of the pay “premium” and finds a positive correlation. Specifically, Ulman (1965), Dickens and Katz (1987), Krueger and Summers (1988), and Katz and Summers (1989) found a negative relationship between the quit rate and the industry pay premium, which they interpreted as suggesting industry wage differentials reflect ex post rents. Similarly, Holzer et al. (1991) found a positive relationship between the number of applications per job opening and the industry pay premium, which they interpreted as suggestive of ex-ante rents. Consistent with this line of work, I interpret the positive correlation between values and earnings as evidence of rents in the labor market. Following Holzer et al. (1991, pg. 759-760), I depart from the line of work and interpret the correlation coefficient being less than one as evidence of compensating differentials. With a small survey dataset, Holzer et al. (1991) emphasize that the imperfect correlation might reflect sampling variability. With a large administrative dataset, I have shown some evidence that sampling variability is unlikely to drive my results.

This paper demonstrates that there is much to be learned about the labor market from studying the “quantity” dimension—how people do and do not flow across firms—of matched employer-employee data. The paper has shown that focusing on this dimension of the data allows a reassessment of the role of compensating differentials in the earnings structure. This paper suggests numerous promising avenues for research. One promising substantive direction is to use the framework

42Similarly, Tjaden and Wellschmied (2014) argue that earnings cuts on EE transitions do not provide evidence of nonpay characteristics because on average workers are more likely to leave jobs after earnings cuts on EE transitions than earnings increases, and Card et al. (2012) show that jobs at higher-paying firms last longer on average.
in this paper to study other aspects of the earnings structure—for example, the gender earnings gap. One promising methodological direction is to extend the framework to incorporate worker preference heterogeneity.

Stanford University and NBER
References


Table I: Summary statistics and the variance of earnings

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>≥ 15 people-years (per year)</th>
<th>Strongly Connected by EE</th>
<th>Strongly Connected by EE (restrictions)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People-years</td>
<td>504,945,000</td>
<td>411,088,000</td>
<td>409,550,000</td>
<td>408,961,000</td>
</tr>
<tr>
<td>People</td>
<td>105,921,000</td>
<td>91,142,000</td>
<td>90,895,000</td>
<td>90,803,000</td>
</tr>
<tr>
<td>Employers</td>
<td>6,155,000</td>
<td>484,000</td>
<td>476,000</td>
<td>472,000</td>
</tr>
<tr>
<td><strong>Summary statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean log earnings</td>
<td>10.43</td>
<td>10.48</td>
<td>10.48</td>
<td>10.48</td>
</tr>
<tr>
<td>Variance of log earnings</td>
<td>0.70</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Ensemble decomposition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employers</td>
<td></td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td></td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xb</td>
<td></td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variance components</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of employer effect</td>
<td></td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of person effect</td>
<td></td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of Xb</td>
<td></td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2cov(person, employer)</td>
<td></td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2cov(Xb, person + employer)</td>
<td></td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corr(person, employer)</td>
<td></td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall fit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Match effects model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Sample counts are rounded to the nearest thousand. The data is at an annual frequency. There is one observation per person per year. The observation is the job from which a person made the most money, but only if she made at least $3,250 (in $2011, using the CPI-U). Earnings are annualized. The table includes person-years in which on December 31 the person was aged 18-61 (inclusive). The restriction in column (2) is that the employer had on average 15 or more non-singleton observations each year (a total of 90 or more over the sample). The restriction in column (3) is that the employers lie in the set of firms strongly connected by employer-to-employer mobility, where strongly connected is discussed in section III. The extra restrictions in column (4) are that an employer hire a worker from nonemployment, and appears in 20 or more of the 50 bootstrap repetitions. The ensemble decomposition reports the covariance of earnings and the relevant characteristic relative to the overall variance of earnings. The variance components reports the variance of each component relative to the overall variance of earnings. Both decompositions sum to the overall fit (with a degrees of freedom correction). The match effects model adds interactions between employers and persons. Xb are the covariates. EE is employer-to-employer and ENE and is employer-to-nonemployment-to-employer.
Table II: Earnings declines, value changes, and firm-level pay

<table>
<thead>
<tr>
<th>Panel A. Pr($y \downarrow$)</th>
<th>All</th>
<th>EE</th>
<th>ENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>0.429</td>
<td>0.374</td>
<td>0.469</td>
</tr>
<tr>
<td>Unconditional (nominal)</td>
<td>0.402</td>
<td>0.343</td>
<td>0.445</td>
</tr>
<tr>
<td>When moving to a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...higher-paying firm</td>
<td>0.297</td>
<td>0.268</td>
<td>0.321</td>
</tr>
<tr>
<td>...lower-paying firm</td>
<td>0.578</td>
<td>0.515</td>
<td>0.618</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Pr($\Psi \uparrow$)</th>
<th>All</th>
<th>EE</th>
<th>ENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>0.530</td>
<td>0.570</td>
<td>0.501</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Pr($V^{EE} \uparrow$)</th>
<th>All</th>
<th>EE</th>
<th>ENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>0.621</td>
<td>0.664</td>
<td>0.589</td>
</tr>
<tr>
<td>When moving to a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...higher-paying firm</td>
<td>0.753</td>
<td>0.774</td>
<td>0.735</td>
</tr>
<tr>
<td>...lower-paying firm</td>
<td>0.472</td>
<td>0.519</td>
<td>0.442</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D. Pr($V^{e} \uparrow$)</th>
<th>All</th>
<th>EE</th>
<th>ENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>0.521</td>
<td>0.566</td>
<td>0.488</td>
</tr>
<tr>
<td>When moving to a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...higher-paying firm</td>
<td>0.672</td>
<td>0.696</td>
<td>0.652</td>
</tr>
<tr>
<td>...lower-paying firm</td>
<td>0.351</td>
<td>0.394</td>
<td>0.324</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel E. Correlations</th>
<th>Spearman Pearson Pearson</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V^{EE}$ and $\Psi$</td>
<td>0.427</td>
</tr>
<tr>
<td>$V^{EE}[-10%, +20%]$ and $\Psi$</td>
<td>0.475</td>
</tr>
<tr>
<td>$V^{EE}$ (adjust for size and offers) and $\Psi$</td>
<td>0.542</td>
</tr>
<tr>
<td>$V^{e}$ and $\Psi$</td>
<td>0.514</td>
</tr>
<tr>
<td>$V^{EE}$ and log(size)</td>
<td>0.042</td>
</tr>
<tr>
<td>$V^{e}$ and log(size)</td>
<td>0.177</td>
</tr>
<tr>
<td>$\Psi$ and log(size)</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Notes: The pay of a firm is defined by its firm effect ($\Psi$). $y$ is the log of annualized individual earnings. $\tilde{V}^{EE}$ uses information only in EE transitions. $V^{e}$ comes from estimating the full model. Size is defined as the number of person-years at the firm over the entire sample. The $\tilde{V}^{EE}[-10\%, +20\%]$ row uses only annual observations where the firm growth rates is between $-10\%$ and $+20\%$ to compute $\tilde{V}^{EE}$. The adjustment for size and offers is to multiply by the log size, and divide by the log share of workers hired from nonemployment who are hired by firm $j$. The correlations are adjusted for noise in ways described in Appendix H. The correlations are weighted by person-years at the employer-level in column (4) of Table I. EE is employer-to-employer and ENE is employer-to-nonemployment-to-employer.
Table III: Transition probabilities and model parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall EE Transition Probability</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>Overall ENE Transition Probability</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>EE Share of Transitions</td>
<td>0.413</td>
</tr>
<tr>
<td>$\bar{\delta}$</td>
<td>Exogenous ENE</td>
<td>0.039(0.000)</td>
</tr>
<tr>
<td>$\bar{\rho}$</td>
<td>Exogenous EE probability</td>
<td>0.029(0.000)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>Probability of offer on-the-job</td>
<td>0.168(0.000)</td>
</tr>
<tr>
<td></td>
<td>Pr(accept offer</td>
<td>nonemployed)</td>
</tr>
</tbody>
</table>

Notes: All probabilities and parameters are annual. A worker only counts as separating if she appears again in the dataset. The sample is column (4) of Table I. The $\bar{\delta}$ and $\bar{\rho}$ are averaged over the entire sample. Standard deviations across bootstrap samples in parentheses. EE is employer-to-employer and ENE is employer-to-nonemployment-to-employer.
Table IV: Why do some firms pay so much and some so little?

<table>
<thead>
<tr>
<th>Panel A. Decompositions</th>
<th>$\frac{\text{Cov}(\Psi,y)}{\text{Var}(y)}$</th>
<th>$R^2(\Psi,V^e)$</th>
<th>Comp. diff. share of all</th>
<th>Person-year share of all</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (adjust for noise)</td>
<td>0.209</td>
<td>0.295</td>
<td>0.148</td>
<td>1.000</td>
</tr>
<tr>
<td>All</td>
<td>0.211</td>
<td>0.281</td>
<td>0.152</td>
<td></td>
</tr>
<tr>
<td>Large ($\geq 1000$ per year)</td>
<td>0.198</td>
<td>0.356</td>
<td>0.128</td>
<td>0.372</td>
</tr>
<tr>
<td>Sample 1 (use sample 2 $\Psi$ in (1))</td>
<td>0.211</td>
<td>0.273</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>Sample 2 (use sample 1 $\Psi$ in (1))</td>
<td>0.211</td>
<td>0.274</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>0.211</td>
<td>0.273</td>
<td>0.153</td>
<td>0.982</td>
</tr>
<tr>
<td>Young (18-34)</td>
<td>0.254</td>
<td>0.390</td>
<td>0.155</td>
<td></td>
</tr>
<tr>
<td>Old (35-61)</td>
<td>0.192</td>
<td>0.126</td>
<td>0.167</td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>0.216</td>
<td>0.227</td>
<td>0.163</td>
<td>0.974</td>
</tr>
<tr>
<td>Men</td>
<td>0.216</td>
<td>0.334</td>
<td>0.144</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>0.210</td>
<td>0.242</td>
<td>0.159</td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>0.213</td>
<td>0.288</td>
<td>0.152</td>
<td>0.937</td>
</tr>
<tr>
<td>Low $\hat{\alpha}_w$ workers</td>
<td>0.343</td>
<td>0.302</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>High $\hat{\alpha}_w$ workers</td>
<td>0.234</td>
<td>0.181</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>0.282</td>
<td>0.235</td>
<td>0.213</td>
<td>0.964</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.180</td>
<td>0.343</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>Admin/Support/Waste</td>
<td>0.160</td>
<td>0.279</td>
<td>0.115</td>
<td></td>
</tr>
<tr>
<td>Hotel/Food</td>
<td>0.137</td>
<td>0.244</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>Retail Trade</td>
<td>0.120</td>
<td>0.206</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.169</td>
<td>0.202</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>State-by-state average</td>
<td>0.196</td>
<td>0.314</td>
<td>0.135</td>
<td>0.982</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Correlations</th>
<th>$\text{Corr}(\Psi)$</th>
<th>$\text{Corr}(V^e)$</th>
<th>$\text{Corr}(a_{\text{Rosen}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1 and 2</td>
<td>0.970</td>
<td>0.948</td>
<td>0.942</td>
</tr>
<tr>
<td>Young and Old</td>
<td>0.867</td>
<td>0.828</td>
<td>0.775</td>
</tr>
<tr>
<td>Men and Women</td>
<td>0.920</td>
<td>0.898</td>
<td>0.867</td>
</tr>
<tr>
<td>Low and high $\hat{\alpha}_w$</td>
<td>0.801</td>
<td>0.836</td>
<td>0.734</td>
</tr>
</tbody>
</table>

Notes: This table reports the main decomposition in this paper, and also shows the relationship between quantities estimated in different subsamples. In all cases, I start with the sample in column (3) of Table I and then re-estimate in the relevant subsample. In Panel A, column (1) reports the share of the variance of earnings in the sample that can be explained by firms as in equation (2). Column (2) computes the $R^2$ between $\Psi$ and $V^e$. Column (3) follows equation (23) and multiplies column (1) and (2) to show the share of the variance of earnings in the subsample that is attributable to compensating differentials. Column (4) shows the share of person-year observations that are represented in the subsample relative to the top row, where the sample is column (4) of Table I. The only row that adjusts for noise is the first row. In Panel B, I report correlations in the two subsamples described in the relevant rows.
Table V: Relationship to observables

<table>
<thead>
<tr>
<th></th>
<th>Variance shares</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Earnings (Ψ)</td>
<td>Nonpay (a_{Rosen})</td>
<td>R^2(Ψ, V^e)</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>0.052</td>
<td>0.077</td>
<td>0.000</td>
</tr>
<tr>
<td>County</td>
<td>0.102</td>
<td>0.150</td>
<td>0.049</td>
</tr>
<tr>
<td>Sector</td>
<td>0.381</td>
<td>0.297</td>
<td>0.497</td>
</tr>
<tr>
<td>4-digit industry</td>
<td>0.547</td>
<td>0.445</td>
<td>0.438</td>
</tr>
</tbody>
</table>

Notes: Columns (1) and (2) report variance shares at the level of the group reported in the row. The \( a_{Rosen} \) term is \( -\hat{\epsilon}_j \) from the regression \( \Psi_j = \beta_0 + \beta_1 V^e_j + \epsilon_j \). In all cases, these variance shares reflect the variance shares of shrunken version of the characteristic-level mean, as described in Appendix H. The \( R^2 \) term is the relationship between \( \Psi \) and \( V^e \), which is adjusted for noise.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Variance of earnings</th>
<th>Change relative to data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>0.672</td>
<td>N/A</td>
</tr>
<tr>
<td>Equalize location</td>
<td>0.658</td>
<td>-2.0%</td>
</tr>
<tr>
<td>Equalize industry</td>
<td>0.617</td>
<td>-8.2%</td>
</tr>
<tr>
<td>Equalize location and industry</td>
<td>0.604</td>
<td>-10.1%</td>
</tr>
<tr>
<td>Equalize all</td>
<td>0.570</td>
<td>-15.2%</td>
</tr>
<tr>
<td>“Naive”</td>
<td>0.518</td>
<td>-22.9%</td>
</tr>
<tr>
<td>Remove firm effects</td>
<td>0.481</td>
<td>-28.5%</td>
</tr>
</tbody>
</table>

Notes: This table shows how equalizing nonpay characteristics and compensating workers affects the variance of earnings. In the top panel, each row reports $\text{Var}(y_{it} - \tilde{a}_{Rosen,J(i,t)})$, where the $J(i,t)$ refers respectively to location, industry, location and industry, and firm. Location refers to county and industry to 4-digit NAICS industry. The bottom panel considers the effects of deflating the firm effects ($\text{Var}(y_{it} - (1 - R^2)\tilde{\Psi}_J(i,t))$) and subtracting off the firm effects and recomputing the variance of earnings ($\text{Var}(y_{it} - \tilde{\Psi}_J(i,t))$). Column (1) reports the variance of log annualized earnings. Column (2) reports the percent change relative to the data. The sample is the person-years in column (4) of Table [I]
Notes: These figures show the probability of an earnings cut as a function of the change in firm level pay for workers who switch annual dominant jobs. The earnings are the annualized earnings in the last year at the previous job and in the first year at the new job. The top panel looks at all transitions and the bottom panel looks at employer-to-employer (EE) transitions. In both panels, the transitions are sorted on the basis of the change in firm effects into 20 bins with an equal number of transitions in each bin.
Figure II: Change in firm pay is not related to change in slope of earnings

(a) All

Notes: These figures show how the firm-level slope of pay relates to the change in firm-level pay for workers who switch annual dominant jobs. The slope of firm-level pay is estimated using the earnings changes of the stayers. The top panel looks at all transitions and the bottom panel looks at employer-to-employer (EE) transitions. I sort the job changers into 20 bins on the basis of the change in the firm effects in the level of earnings. The circles plot the bin means. The solid line plots the best-fitting line estimated based on the micro-data. The dashed red line plots the 45 degree line. The coefficient in the upper panel is 0.000, and in the bottom panel it is 0.002.
Figure III: Moves to lower-paying firms are systematic

(a) Probability of a higher $\tilde{V}^{EE}$ firm

(b) Probability of a higher-value firm ($V^e$)

Notes: This figure considers the sample of workers who switch annual dominant jobs on EE transitions. I sort the job changers into 20 bins on the basis of the change in the firm effects. The upper panel reports the probability of an increase in $\tilde{V}^{EE}$, which is the systematic direction of mobility uses information only on EE transitions. The bottom panel reports the probability of an increase in $V^e$, which comes from estimating the full model. An increase in $V^e$ means that the decline in pay is outweighed by an increase in nonpay characteristics.
Notes: These figures show the data used to construct the exogenous weights for employer-to-employer (EE) and employer-to-nonemployer (EN) transitions. The probabilities and growth rates are quarterly. The probabilities are computed in one percentage point wide bins of employer growth rates. The figure plots a five-bin moving average. The exogenous weight is $\frac{\text{excess}}{\text{excess} + \text{expected}}$. At an expanding employer, the exogenous weight is zero by construction.
Figure V: Relationship between values and earnings

(a) Overall

Notes: The top panel sorts firms on the basis of firm-level values. The circles plot 20 bins with the same number of person-years, while the solid line plots the regression line estimated on the firm-level data. The thin-dashed line shows plus and minus one standard deviation of the firm-level earnings within each value bin. The bottom panel of this figure plots the sector-level means of the earnings and values. The thick-solid line plots the regression line run at the sector level and weighting by the number of person-years represented by each sector.
Figure VI: Dispersion in the labor market

Notes: This figure plots the dispersion in the value of the firms relative to the dispersion in the idiosyncratic utility draw. The idiosyncratic utility draw distribution is type I extreme value with scale parameter 1 and variance $\frac{\pi^2}{6} \approx 1.6$, and is a normalization. Both distributions are normalized so that the median is 0. The figure plots the shrunken values of $V^e$. The estimated variance of $V^e$ is 0.817.
Figure VII: Counterfactual inequality

(a) Counterfactual

(b) “Naive” counterfactual

Notes: The top panel of this figure plots the distribution of earnings in the data and in a counterfactual where I equalize nonpay characteristics. The bottom panel considers a “naive” counterfactual where I deflate all the firm components of earnings by a constant fraction and then recompute earnings.