NBER WORKING PAPER SERIES

RANKING FIRMS USING REVEALED PREFERENCE

Isaac Sorkin

Working Paper 23938
http://www.nber.org/papers/w23938

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017

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 differential 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.

Isaac Sorkin
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
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.[1] But these findings are typically viewed as somewhat special leading to the conclusion that compensating differentials are not relevant for understanding the structure of earnings.[2]

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. 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 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.[3] This

---

[1] For recent work on fatality risk see Lavetti and Schmutte (2016) and Lavetti (2017). For PhD scientists and their research agenda, see Stern (2004). See Mas and Pallais (2017) for an interesting recent study of alternative work arrangements.

[2] For example, Hornstein, Krusell, and Violante (2011, pg. 2883) survey some literature on the hedonic approach and write that compensating differentials "does not show too much promise" in explaining earnings dispersion.

[3] 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.

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, LaLonde, and Sullivan (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.

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, Kramarz, and Margolis (1999) (AKM), Andersson et al. (2012), Card, Heining, and Kline (2013), Song et al. (2016), Barth et al. (2016), Card, Cardoso, and Kline (2016), Goldschmidt and Schmieder (2017), Engbom and Moser (2017), and Abowd, Mckinney, and Zhao (2017)).[4] 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

---

[4]In this paper, I use the word firm and employer interchangeably.

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), 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 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, Krusell, and Violante (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 screening, systematic forms of preference heterogeneity, mobility costs, and counteroffers. Combined, the first two mechanisms rule out any sorting between workers and firms. 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.[5] 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.[6]

---

[5]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 propogates through estimation.

[6]There is also literature (e.g., Gronberg and Reed (1994), Dey and Flinn (2005), Bonhomme and Jolivet (2009),

This paper uses how workers move across firms in a new way relative to existing literature. Bagger and Lentz (2016) 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 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, Chaudhuri, and McLaren (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 1 introduces the data. Section 2 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. I also show that these EE moves to lower-paying firms reflect systematic patterns of mobility. Section 3 takes up the task of how to rank firms using mobility data. Section 4 presents the main results of the paper. Section 5 presents the earnings inequality counterfactual and discusses the relationship to the Hornstein, Krusell, and Violante (2011) puzzle. Section 6 shows that the main results are robust across a variety of subgroups. Finally, section 7 relates this paper to the interindustry wage differential literature, and discusses some caveats and promising avenues for future work.

1 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.[7] The LEHD is matched employer-employee data and so allows me to follow workers across firms.

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 calender year.[8] I also restrict attention to workers aged 18-61 (inclusive) and, following Card, Heining, and Kline (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

Aizawa and Fang (2015) and Jarosch (2015)) which estimates the value of specific amenities in a search environment.

[7]See Abowd et al. (2009) for details.

[8]Reduction to one observation per person per year is common. See Abowd, Kramarz, and Margolis (1999) (France), Abowd, Lengermann, and McKinney (2003) (US), Card, Heining, and Kline (2013) (Germany), and Card, Cardoso, and Kline (2016) (Portugal). Even outside of estimating statistical wage decompositions, Bagger et al. (2014) also reduce to one such observation per year.

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. See Appendix A for details on dataset construction (including how earnings are annualized).

Three features of the LEHD should be kept in mind when interpreting the results. First, the unemployment insurance system measures earnings, but not hours.[9] 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 4.4. Second, the notion of an employer is a state-level unemployment insurance account, though the dataset follows workers across states.[10] Third, only employers that are covered by the unemployment insurance system appear in the dataset.[11] 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.[12] 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.[13] The second restriction is that I look at the strongly connected subset of those firms (I define strongly connected in section 3.2). 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.[14]

---

[9]The 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)).

[10]This can understate firm size for two reasons. First, for employers that operate in multiple states, this understates true employer size. Second, it is also possible for a given employer to have multiple unemployment insurance accounts within a state, which would also lead to an understatement of true employer size, though this is quantitatively unimportant (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.") That said, 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.

[11]This restriction results in the exclusion of certain sectors of the economy. In particular, small nonprofit (organizations 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/uilawcompar/2012/coverage.pdf`.

[12]I 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 K for a map.

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

[14]An additional motivation to impose a minimum size threshold is to minimize variation in the identified set of

---

6

Table 1 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.

2 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. I then develop a method to find the systematic pattern of mobility. I show that this systematic pattern also includes moves to lower paying firms, and so these moves to lower-paying firms are unlikely to be explained by idiosyncratic shocks.

2.1 Firms play an important role in earnings determination

To measure firm-level earnings, I use the following equation for log earnings (known as the Abowd, Kramarz, and Margolis (1999) decomposition):

$$\underbrace{y_{it}}_{\text{log earnings}} = \underbrace{\alpha_i}_{\text{person effect}} + \underbrace{\Psi_{\boldsymbol{J}(i,t)}}_{\text{firm effect}} + \underbrace{x'_{it}\beta}_{\text{covariates}} + \underbrace{r_{it}}_{\text{residual/error term}}, \tag{1}$$

where y_{it} is log earnings of person i at time t, α_i is a person fixed effect, $\Psi_{\boldsymbol{J}(i,t)}$ is the firm fixed effect at the employer j where worker i is employed at time t (denoted by $\boldsymbol{J}(i,t)$), r is an error term, and x is a set of covariates including higher-order polynomial terms in age.[15]

I quantify the role of firms in earnings using the following decomposition of the variance of earnings:[16]

$$\underbrace{\text{Var}(y_{it})}_{\text{variance of earnings}} = \underbrace{\text{Cov}(\alpha_i, y_{it})}_{\text{person effect}} + \underbrace{\text{Cov}(\Psi_{\boldsymbol{J}(i,t)}, y_{it})}_{\text{firm effect}} + \underbrace{\text{Cov}(x'_{it}\beta, y_{it})}_{\text{covariates}} + \underbrace{\text{Cov}(r_{it}, y_{it})}_{\text{residual}}. \tag{2}$$

The share of the variance in earnings accounted for by firms is:

$$\frac{\text{Cov}(\Psi_{\boldsymbol{J}(i,t)}, y_{it})}{\text{Var}(y_{it})}. \tag{3}$$

Firms play an important role in earnings determination. The third portion of Table 1 shows that firms account for about 21% of the variance of earnings.

The fourth portion of Table 1 reports an alternative decomposition. It shows results that are quite similar to results for an identical time period in the US using a different dataset. In particular,

---

firms across the bootstrap resamples.

[15]Because 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, Heining, and Kline (2013), are omitted. Following Card, Heining, and Kline (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).

[16]Card et al. (2017, pg. 10) call this the "ensemble" decomposition.

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 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).

While for finite-sample reasons it is likely that this 21% estimate is biased upwards, three alternative approaches presented in column (1) of Table 4 suggest that this bias is negligible. First, I shrink the firm effects using bootstrapped estimates of standard errors in a way that I describe in more detail in section 3.3. Using the shrunken estimates, I find that firms explain 21% of the variance of earnings. Second, I consider a sample of very large firms (1000 or more non-singleton person-year observations per year). Using the large firm sample, I find that firms explain 20%. Third, I split workers randomly into two mutually exclusive subsamples and estimate the AKM decomposition in each of these subsamples. I then use the firm effects from one sample in the AKM decomposition in the other sample. Using this approach, I find that firms explain 21% of the variance of earnings. In combination, these three approaches suggest that the magnitude of the bias is small and firms account for about 20% of the variance of earnings. Finally, Appendix H discusses Monte Carlo evidence on the baseline measurement and shows that the bias is negligible.

2.2 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 2 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%).[17] 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 and then discuss why this finding is not mechanical. To begin to show that the earnings declines are related to firm-level characteristics that we expect to be priced in the labor market, I show that the firm

---

[17]The share of earnings declines is quantitatively consistent with evidence from survey datasets where researchers are able to calculate changes in hourly earnings. Jolivet, Postel-Vinay, and Robin (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 Wellschmied (2014, Table 2) find 34%. And for the National Longitudinal Survey of Youth 1997, Sullivan and To (2014, Table 1) find 36%.

effects capture the probability of an earnings decline. Panel B of Table 2 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 1 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.

The firm effects also capture the magnitude of the earnings declines. As emphasized by Chetty, Friedman, and Rockoff (2014), a measure of bias in firm effects (or, in their case, teacher value-added) is to consider the β_1 coefficient in the following regression:

$$y_{i,t}^r - y_{i,t-1}^r = \beta_0 + \beta_1 \left[\hat{\tilde{\Psi}}_{J(i,t)} - \hat{\tilde{\Psi}}_{J(i,t-1)} \right] + \epsilon_{i,t}, \forall \text{ i,t s.t. } J(i,t) \neq J(i,t-1), \tag{4}$$

where $y_{i,t}^r = y_{i,t} - x_{it}'\hat{\beta}$ is the residualized earnings, and $\hat{\tilde{\Psi}}_{J(i,t)}$ is the shrunken firm effect. If the firm effects are unbiased, then we expect $\hat{\beta}_1 = 1$. The top panel of Figure 2 shows that this is the case. The figure plots 20 bins of changes in firm effects at all transitions between annual dominant employers against the average individual-level change in earnings on these transitions. The solid line plots the best-fitting line from a regression run on the individual-level data. The thin-dashed line shows the line that would be expected if the firm effects were unbiased. The lines are identical and the coefficient is 1.005. The bottom panel shows the analogous figure for the EE transitions, and the slope is 0.813. Formally this finding could be interpreted as indicating misspecification, though it is not clear whether the departure is quantitatively important. Figure A3 reports the results of a conceptually similar exercise where, following Card, Heining, and Kline (2013, pg. 997), I plot event studies around transitions from lower- to higher-paying firms and vice-versa and show that earnings change in opposite directions with equal magnitudes.

While it may seem mechanical that the firm effects would predict the individual-level changes, this finding does not hold if the AKM decomposition is seriously misspecified. To show this, I simulate data from a model where mobility is on the basis of the comparative advantage (e.g., Eeckhout and Kircher (2011), Lopes de Melo (2016) and Hagedorn, Law, and Manovskii (2017)). This type of model implies that the residual plays a large role in determining mobility and generates at least two implications which are at odds with the data. First, there are no earnings cuts on EE transitions, and the individual-level earnings changes always lie above the x-axis. Second, there is not the approximate symmetry in earnings changes from moving to a better or a worse firm (Card, Heining, and Kline (2013, pg. 990) emphasize this symmetry property). Figure A2 plots the analogous figure to Figure 2b with data simulated from the example production function in Eeckhout and Kircher (2011). The estimate of β_1 is about 0.4, and unlike in the data, the earnings changes display a v-shape in the firm effects changes.[18]

---

[18]The v-shape comes from earnings increases accruing to workers whose comparative advantage is working at the lowest productivity firms. In a model of circular heterogeneity without absolute advantage like Marimon and Zilibotti (1999), there is no variance in the estimated firm effects because all firms are equally high- and low-paying for the same number of workers and in the same way.

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, Kramarz, and Margolis (1999), I estimate firm-specific earnings slopes using the wage growth of the stayers. When workers move to lower-paying firms, Figure 3 shows that they do not move to firms offering steeper slopes in earnings (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). These results are quantitatively different than a model that explains earnings cuts as a function of an option value of a future increase. I simulate Papp (2013)'s calibration of Cahuc, Postel-Vinay, and Robin (2006), which matches the share of earnings cuts in the data. In the simulated data, the correlation between the firm effects in the intercept and the slope is -0.90. This result implies that earnings cuts are not explained by the possibility of future earnings increases at the same firm.

2.3 Moves to lower-paying firms are systematic

While many employer-to-employer transitions are to lower-paying firms, Panel B of Table 2 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.

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 3, 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. 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, ι, 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.[19] This distributional assumption implies that the probability of choosing firm k over j is $\frac{exp(\tilde{V}_k^{EE})}{exp(\tilde{V}_k^{EE})+exp(\tilde{V}_j^{EE})}$.

---

[19]Because 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.

A simple estimator of the relative common values is the ratio of the empirical choice probabilities:

$$\frac{M_{kj}^o/N}{M_{ki}^o/N} = \frac{M_{kj}^o}{M_{jk}^o} = \frac{exp(\tilde{V}_k^{EE})}{exp(\tilde{V}_j^{EE})}. \tag{5}$$

While it would be possible to take equation (5) 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 C 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 (5) gives

$$M_{kj}^o exp(\tilde{V}_j^{EE}) = M_{jk}^o exp(\tilde{V}_k^{EE}), \ \forall j \in \mathcal{E},$$

where the "for all" holds because (5) holds for all pairs of employers. Summing across all employers on both sides gives

$$\underbrace{\sum_{j \in \mathcal{E}} M_{kj}^o}_{\text{\# entering k}} \underbrace{exp(\tilde{V}_j^{EE})}_{\text{value}} = \underbrace{\sum_{j \in \mathcal{E}} M_{jk}^o}_{\text{\# exiting k}} \underbrace{exp(\tilde{V}_k^{EE})}_{\text{value}}. \tag{6}$$

Dividing by the sum on the right-hand side gives

$$\frac{\overbrace{\sum_{j \in \mathcal{E}} M_{kj}^o exp(\tilde{V}_j^{EE})}^{\text{value weighted entry}}}{\underbrace{\sum_{j \in \mathcal{E}} M_{jk}^o}_{\text{exits}}} = \underbrace{exp(\tilde{V}_k^{EE})}_{\text{value}}. \tag{7}$$

Equation (7) implies one linear restriction per firm. The equation generates a recursive definition

11

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 (7). Specifically, define a diagonal matrix S with the ith diagonal entry being $S_{ii}^o = \sum_{j \in \mathcal{E}} M_{ji}^o$. Then letting $exp(\tilde{V}^{EE})$ be the $|\mathcal{E}| \times 1$ vector that contains the firm-level $exp(\tilde{V}_k^{EE})$ yields the following:

$$\underbrace{S^{o,-1}M^o}_{\text{normalized flows}} \underbrace{exp(\tilde{V}^{EE})}_{\text{values}} = \underbrace{exp(\tilde{V}^{EE})}_{\text{values}}. \tag{8}$$

This equation allows me to solve for $exp(\tilde{V}^{EE})$. Intuitively, $exp(\tilde{V}^{EE})$ is the fixed point of the function $S^{o,-1}M^o : \mathbb{R}^{|\mathcal{E}|} \to \mathbb{R}^{|\mathcal{E}|}$. 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 $exp(\tilde{V}^{EE})$ exists. To show when the $exp(\tilde{V}^{EE})$ 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 $S^{o,-1}M^o$ 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:*

$$S^{o,-1}M^o exp(\tilde{V}^{EE}) = exp(\tilde{V}^{EE}).$$

Proof. See Appendix D (also for graph theory definitions). $\square$

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 (7). 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, Creecy, and Kramarz (2002) who show that the employer fixed effect in Abowd, Kramarz, and Margolis (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, Creecy, and Kramarz (2002) is helpful in understanding the data requirements to estimate $\tilde{V}^{EE}$. As in Abowd, Creecy, and Kramarz (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.[20]

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 Perron-Frobenius theorem is used to show that the top eigenvector is unique). S^o divides the ith row of M^o by the ith 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 ith *column* of M^o by the ith 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 $S^{o,-1}M^o$ is invariant to the value of the diagonal entries in M^o.

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

Proof. See Appendix D. □

Remark on Result 2: A natural statistic to compute to assess the noise in the estimation of $\tilde{V}^{EE}$ 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.[21]

Panel C of Table 2 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 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 4 shows this pattern more generally.

---

[20]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.

[21]To 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 $\{1, \frac{x-y}{x+y}\}$. 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 $(-\infty, 1)$.

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 2 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 Ψ 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 Ψ is only 0.07.

Moves to lower-paying firms do not reflect (mass) layoffs: A natural concern is that moves to lower-paying firms might not reflect workers' preferences. Instead, workers might have been laidoff. 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, LaLonde, and Sullivan (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 F of Table 2 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 $exp(\tilde{V}^{EE})\frac{g}{f^o}$, 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 F of Table 2 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, Ψ, rises to 0.57, but is still far from 1.

The next section develops an economic model that nests equation (7) 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.

3 Ranking firms using revealed preference

3.1 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 2 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.[22] One example of such a paper is Stern (2004) 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, Hall and Mueller (2017) 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

---

[22]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.

economics has discussed whether the quit-layoff distinction is meaningful in the presence of efficient turnover.[23] 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, Flaaen, Shapiro, and Sorkin (2016) 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, LaLonde, and Sullivan (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 Vilhuber (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.

---

[23]See, for example, Becker, Landes, and Michael (1977) and McLaughlin (1991).

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 δ_j and ρ_j, where δ_j is the probability of an exogenous job destruction shock that sends a worker to nonemployment, and ρ_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_j^e. This value includes both pay and nonpay components, $V_j^e = \omega(\Psi_j + a_j)$, where a_j is the nonpay characteristic at firm j and ω 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:[24]

$$
\underbrace{V^e(v_j, \delta_j, \rho_j)}_{\text{value of being at j}} = \underbrace{v_j}_{\text{flow payoff}} + \underbrace{\beta}_{\text{discounter}} \mathbb{E}\Bigg\{ \underbrace{\delta_j \int_{\iota_1} \{V^n + \iota_1\}dI}_{\text{exogenous job destruction}}
$$

$$
+ \underbrace{\rho_i(1-\delta_j)\sum_k \int_{\iota_2} \{V_k^e + \iota_2\}dI\tilde{f}_k}_{\text{exogenous employer-to-employer (reallocation)}}
$$

$$
+ \underbrace{(1-\rho_j)(1-\delta_j)}_{\text{no exogenous shocks}} \times \Bigg[\underbrace{\underbrace{\lambda_1}_{\text{offer}}\sum_k \int_{\iota_3}\int_{\iota_4} \max\{\underbrace{V_k^e + \iota_3}_{\text{accept}}, \underbrace{V_j^e + \iota_4}_{\text{reject}}\}dIdIf_k}_{\text{endogenous employer-to-employer}}
$$

$$
+ \underbrace{\underbrace{(1-\lambda_1)}_{\text{no offer}}\int_{\iota_5}\int_{\iota_6} \max\{\underbrace{V^n + \iota_5}_{\text{accept}}, \underbrace{V_j^e + \iota_6}_{\text{reject}}\}dIdI}_{\text{endogenous job destruction}} \Bigg]\Bigg\}. \tag{9}
$$

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

The continuation value weights the expected value of four mutually exclusive possibilities. Two

---

[24]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).

possibilities generate exogenous separations. With probability δ_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 ι_1 is an idiosyncratic draw from a type I distribution. With probability ρ_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 employers. If this happens, the worker receives value $V_j^e + \iota_2$, where V_j^e is the common component of the value, and ι_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 λ_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 $V_k^e + \iota_3$ and $V_i^e + \iota_4$ and makes a maximizing decision. Note that the ιs 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 (9), 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_j^e, as well as the idiosyncratic utility draw, ι. By taking into account the ι, the model allows workers to make off-setting moves and, more generally, for gross flows to exceed net flows. Second, the ρ_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:

$$
\underbrace{V^n}_{\text{value of being nonemployed}} = \underbrace{b}_{\text{flow payoff}}
$$
$$
+ \underbrace{\beta}_{\text{discounter}} \mathbb{E}\{\underbrace{\lambda_0}_{\text{offer}} \sum_j \int_{\iota_7} \int_{\iota_8} \underbrace{\max\{V_j^e + \iota_7, \underbrace{V^n + \iota_8}_{\text{reject}}\} dI dI f_j}_{\text{accept}}
$$
$$
\underbrace{\phantom{+ \beta \mathbb{E}\{\lambda_0 \sum_j \int_{\iota_7} \int_{\iota_8} \max\{V_j^e + \iota_7, V^n + \iota_8\} dI dI f_j}}_{\text{endogenous nonemployment-to-employment}}
$$
$$
+ \underbrace{(1 - \lambda_0) \int_{\iota_9} \{V^n + \iota_9\} dI\}}_{\text{no offer}}. \tag{10}
$$

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 (10) is the same f in the third row of equation (9)— which firms hire

18

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 Ψ, 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, Kramarz, and Margolis (1999) (AKM). In the search model, mobility depends on V_j^e and ι. 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 (α_i), the covariates ($X_i\beta$) and the history (past and future) of firms (Ψ_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.[25] 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. 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 (ι). I can selection correct the earnings equation and allow mobility to depend on the error term. See Appendix B for details.

3.2 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.

Summarizing the systematic pattern of worker flows

This section rewrites the model so that it gives rise to the same expression used in section 2 to measure the systematic pattern of worker flows. Relative to the derivation in section 2, 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

---

[25]It is possible to include a persistent worker-firm match effect in AKM, which is not possible in the search model.

I adopt the standard continuum assumption in discrete choice models, such flows from employer k to employer j are given by

$$\underbrace{M_{jk}}_{\text{k to j flows}} = \underbrace{g_k W}_{\text{\# of workers at k}} \underbrace{(1 - \delta_k)(1 - \rho_k)}_{\text{no shocks}} \underbrace{\lambda_1 f_j \Pr(j \succ k)}_{\text{get offer from j and accept}}, \tag{11}$$

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 $\Pr(j \succ k)$.

To connect this model to the expressions in section 2, 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:

$$\frac{M_{jk}}{M_{kj}} = \frac{f_j g_k (1 - \delta_k)(1 - \rho_k) \Pr(j \succ k)}{f_k g_j (1 - \delta_j)(1 - \rho_j) \Pr(k \succ j)}. \tag{12}$$

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

$$\underbrace{\frac{M_{jk}}{M_{kj}}}_{\text{relative flows}} = \underbrace{\frac{f_j}{f_k}}_{\text{relative offers}} \times \underbrace{\frac{g_k (1 - \delta_k)(1 - \rho_k)}{g_j (1 - \delta_j)(1 - \rho_j)}}_{\text{effective relative size}} \times \underbrace{\frac{exp(V^e(v_j))}{exp(V^e(v_k))}}_{\text{relative values}}. \tag{13}$$

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:

$$\underbrace{exp(\tilde{V}_j)}_{\text{flow-relevant value}} \equiv \underbrace{\frac{f_j exp(V_j^e)}{g_j (1 - \delta_j)(1 - \rho_j)}}_{\text{offer} \times \text{value / effective size}}. \tag{14}$$

$exp(\tilde{V}_j)$ 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.[26]

---

[26] For flows between employers and the nonemployed state, an analogous derivation implies:

$$\frac{M_{nj}}{M_{jn}} = \frac{(1 - \lambda_1) g_j W (1 - \delta_j)(1 - \rho_j) \Pr(n \succ j)}{\lambda_0 f_j U \Pr(j \succ n)} = \frac{exp(\tilde{V}_n)}{exp(\tilde{V}_j)},$$

Combining (13) and (14) gives relative flows between employers in terms of $exp(\tilde{V}_i)$:

$$\underbrace{\frac{M_{jk}}{M_{kj}}}_{\text{relative flows}} = \underbrace{\frac{exp(\tilde{V}_j)}{exp(\tilde{V}_k)}}_{\text{relative flow-relevant values}}. \tag{15}$$

This expression is the same as that in section 2. 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 2, the main feature of the data that I use to pin down the offer distribution is where nonemployed workers are hired. Relative to the intuitive approach in section 2, 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 2 and is that nonemployed and employed workers search from the same distribution. In the Bellman equations, the employed and nonemployed draw offers from a common F. The second assumption is new relative to section 2 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 (9)), 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_i^o, 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:

$$f_j^o \equiv \frac{M_{jn}}{\sum_{k \in \mathcal{E}} M_{kn}}. \tag{16}$$

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

where

$$\underbrace{exp(\tilde{V}_n)}_{\text{flow-relevant value}} = \underbrace{\frac{(1 - \lambda_1) W exp(V^n)}{\lambda_0 U}}_{\text{offers} \times \text{ values / size}},$$

U is the number of nonemployed workers, λ_0 is the probability of an offer when nonemployed, and n is the nonemployed state.

21

ment to employment flows:

$$\sum_{j \in \mathcal{E}} M_{jn} exp(\tilde{V}_n) = \sum_{j \in \mathcal{E}} \frac{\lambda_0 U f_j exp(V_j^e)}{exp(V^n) + exp(V_j^e)} \frac{(1 - \lambda_1) W exp(V^n)}{\lambda_0 U}. \tag{17}$$

Finally, the third equation captures the level of employer-to-employer flows and pins down λ_1:

$$\frac{\sum_{k \in \mathcal{E} \setminus \{j\}} \sum_{j \in \mathcal{E}} M_{kj}}{W \sum_{j \in \mathcal{E}} g_j(1 - \delta_j)(1 - \rho_j)} = \lambda_1 \sum_j g_j(1 - \delta_j)(1 - \rho_j) \sum_k f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)}. \tag{18}$$

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 2 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 5.[27] 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.[28]

I now write down the equations that correspond to this idea for the EE separations and analogous equations apply for the EN separations. 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

---

[27]This figure is inspired by Davis, Faberman, and Haltiwanger (2012, Figure 6).
[28]This approach is conceptually related to that in Flaaen, Shapiro, and Sorkin (2016).

right hand side of Figure 5—is given by

$$\Pr(EE(j))^{endog} = (1 - \delta_j)(1 - \rho_j) \left[\lambda_1 \sum_{k \in \mathcal{E}} f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)} \right], \quad (19)$$

where this is the probability that a worker is not laid off, gets an offer, and accepts it.

Now develop notation to measure the separation rate when the firm is contracting. Let $\mathcal{C}(t,j)$ be an indicator variable for firm j being on the right hand side of Figure 5 and contracting in terms of the number of workers in period t. The following expression records this probability (where the o stands for observed):

$$\Pr(EE|\mathcal{C}(t,j) = 1)^o. \quad (20)$$

Finally, combine the previous two expressions to generate an expression for the probability that a separation at a contracting firm was "expected anyway" and so counts as an endogenous separation:

$$\Pr(Endog = 1|EE = 1, \mathcal{C}(t,j) = 1) = \min\left\{ \frac{\Pr(\Pr(EE(j))^{endog}}{\Pr(EE|\mathcal{C}(t,j) = 1)^o}, 1 \right\}, \quad (21)$$

while the exogenous (exog) probability is the complementary probability: $\Pr(Exog = 1|EE = 1, \mathcal{C}(t,j) = 1) = 1 - \Pr(Endog = 1|EE = 1, \mathcal{C}(t,j) = 1)$. In terms of the figure (and the EN measure), the endogenous probability is the black solid line relative to the total length of the line, while the exogenous probability is the red 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 *endogenous 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, $k \in \mathcal{E}$ $M_{kj} = \sum_t M_{kj,t}^o \Pr(Endog = 1|EE = 1, t, j)$, and, for flows to nonemployment (n), and for all $k \in \mathcal{E}$ $M_{nj} = \sum_t M_{nj,t}^o \Pr(Endog = 1|EN = 1, t, j)$. 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: $\rho_j = \frac{\sum_t \sum_{k \in \mathcal{E}} \sum_{i \in \mathcal{E}} M_{kj,t}^o \Pr(Exog=1|EE=1,t,j)}{\sum_t \sum_{k \in \mathcal{E} \cup n} \sum_{i \in \mathcal{E}} M_{kj,t}^o}$, and $\delta_j = \frac{\sum_t \sum_n \sum_{j \in \mathcal{E}} M_{kj,t}^o \Pr(Exog=1|EN=1,t,j)}{\sum_t \sum_{k \in \mathcal{E} \cup n} \sum_{i \in \mathcal{E}} M_{kj,t}^o}$.

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 δ and ρ, to limit the number of parameters to estimate I aggregate to the sectoral level in two places. Specifically, in equation (19) 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 (14) to unravel for the firm-specific value, I use the sectoral average of the δ and ρ.

Summary

Appendix F 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_j^e). 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 λ_1 which generates the extent of EE transitions. A byproduct of an estimate of λ_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 δ and ρ) and then repeat the three steps.

3.3 Measurement error

As detailed in section 4, 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 G provides more details and Appendix H provides Monte Carlo evidence.

3.4 Other model-consistent approaches to ranking firms

To provide more intuition about the model, Appendix I 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 (2016) 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

24

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.

3.5 Model parameters

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 3 shows the estimated model parameters, as well as the split of separations into EE and ENE. The top rows of Table 3 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 (δ) of 0.04. Combining the exogenous weights and the separation probabilities, gives an annual exogenous employer-to-employer, or reallocation, rate (ρ) of 0.03. The feature of the data that drives the estimates of δ and ρ 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 5.

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.

4 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.

4.1 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.[29] 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 E, I present an alternative motivation for this decomposition where I show that the compensating differentials component 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

$$V_j^e = \omega(\Psi_j + \ln a_j), \tag{22}$$

where V_j^e is the utility of being employed at firm j, Ψ_j is the (log) earnings at the firm, and a_j is the level of nonpay characteristics (amenities) provided by the firm.[30] (I include an ω 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_j^e.) Firm j's objective function is:

$$\max_{\Psi_j, a_j} \pi_j - c_j a_j - exp(\Psi_j) \tag{23}$$

such that $V_j^e = \bar{V}_j^e$. Here, π_j is the profit from hiring the marginal (and average) worker, c_j is the marginal cost of amenity provision, and $\bar{V}_j^e$ 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 Rosen (1986) as considering the case where utility is equalized across firms so that $\bar{V}_j^e = \bar{V}^e$, 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: $Corr(\Psi_j, \ln a_j) = -1$. Second, and abusing notation, wages and utility are uncorrelated, that is, $Corr(\Psi_j, V_j^e) = 0$. Finally, the only source of variation in nonpay characteristics is the variation in

---

[29]For example, Hwang, Mortensen, and Reed (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.

[30]In appendix J I show how to use the 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.

the cost of amenity provision across firms: $Var(\ln a_j) = \frac{1}{4} Var(\ln c_j)$.

Heterogeneity in utility: We can interpret 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 $\bar{V}_j^e$ and the following statements are true. First, wages and amenities are perfectly positively correlated: $Corr(\Psi_j, \ln a_j) = 1$. Second, wages and utility are perfectly positively correlated, $Corr(\Psi_j, \bar{v}_j) = 1$. Finally, the only source of variation in nonpay characteristics is the variation in desired utility across firms: $Var(\ln a_j) = \frac{1}{4} Var(\bar{V}_j^e)$.

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: $Cov(\Psi_j, \ln a_j) = \frac{1}{4} \left[Var(\bar{V}_j^e) - Var(\ln c_j) \right]$. 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, $Cov(\Psi_j, \bar{V}_j^e) = \frac{1}{2} Var(\bar{V}_j^e) + \frac{1}{2} Cov(\ln c_j, \bar{V}_j^e)$. Finally, the variance of nonpay characteristics depends on the variance of the two motives as well as their covariance, $Var(\ln a_j) = \frac{1}{4} \left[Var(\bar{V}_j^e) + Var(\ln c_j) - 2Cov(\bar{V}_j^e, \ln c_j) \right]$.

In this general case, what can be identified in terms of primitives of the firm's problem given observables (V_j^e and Ψ_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: $Var(\tilde{c}_j) = Var(\ln c_j) - \frac{Cov(\ln c_j, \bar{V}_j^e)^2}{Var(\bar{V}_j^e)}$. Similarly, consider the variance in pay that is orthogonal to variation in utility in terms of primitives of the firm problem, or $(1 - R^2)Var(\Psi_j) = \frac{1}{4} \left[Var(\ln c_j) - \frac{Cov(\ln c_j, \bar{V}_j^e)^2}{Var(\bar{V}_j^e)} \right]$, where $R^2 = Corr(V^e, \Psi)^2$. Up to rescaling, these terms are identical. Hence, I decompose the variation in amenities into two pieces:

$$Var(\ln a_j) = \frac{1}{4} \left[\underbrace{Var(\ln c_j) - \frac{Cov(\ln c_j, \bar{v}_j)^2}{Var(\bar{v}_j)}}_{\text{Rosen}} + \underbrace{Var(\bar{v}_j) - 2Cov(\bar{v}_j, \ln c_j) + \frac{Cov(\ln c_j, \bar{v}_j)^2}{Var(\bar{v}_j)}}_{\text{Mortensen}} \right], \quad (24)$$

where the Rosen piece can be measured by $(1 - R^2)Var(\Psi_j)$. This logic also leads to a decomposition

27

of the variance of earnings into a rents and compensating differentials component:

$$\underbrace{Var(\Psi)}_{\text{variance of firm effects}} = \frac{1}{4}\left[\underbrace{Var(\bar{V}_j^e) + 2Cov(\bar{V}_j^e, \ln c_j) + \frac{Cov(\tilde{V}_j^e, \ln c_j)^2}{Var(\bar{V}_j^e)}}_{\text{rents}} + \underbrace{Var(\ln c_j) - \frac{Cov(\tilde{V}_j^e, \ln c_j)^2}{Var(\bar{V}_j^e)}}_{\text{compensating differentials}}\right]$$

$$= \underbrace{R^2 Var(\Psi)}_{\text{rents}} + \underbrace{(1 - R^2)Var(\Psi)}_{\text{compensating differentials}} \quad . \tag{25}$$

Figure 6a illustrates this result. It shows a binned scatterplot of the firm-level values and earnings as well as the line of best fit.[31] 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.

4.2 Decomposing firm-level pay into compensating differentials and rents

Table 4 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.[32]

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 2 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 4b shows the share of moves that the model interprets as being to higher-value 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.[33]

---

[31]Table 2 shows that the Pearson correlation in the microdata underlying this plot is 0.53 and the Spearman (rank) correlation is 0.51. So given a monotone relationship, the deviations from linearity are not large.

[32]The unadjusted R^2 is 0.28. The bootstrap-adjusted share is 0.30. Using the correlations between samples 1 and 2 implies an R^2 of $0.306 = \frac{0.281}{0.970 \times 0.948}$. And in the large firm sample the R^2 is 0.36. The Monte Carlo evidence in Appendix H suggests that these adjustments might be biased down by one or two percentage points.

[33]Moving from $\tilde{V}^{EE}$ to V^e addresses two issues highlighted in section 2: 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 7 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.[34] Intuitively, the relative variance of the values and the idiosyncratic shocks is identified by the extent to which gross flows exceeds net flows.[35] 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 l 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.

4.3 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 5 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 6b 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.[36] 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.

---

[34]The variance of the idiosyncratic shocks is $\frac{\pi^2}{6} \approx 1.6$ by normalization. The raw variance of the values is 0.847, while taking into account measurement error gives an estimate of 0.814.

[35]To see this, return to equation (5), 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.

[36]Holzer, Katz, and Krueger (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.

The second characteristic is location. Table 5 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.[37] Hence, much of the variation in nonpay characteristics is related to observable characteristics, and does so in intuitively plausible ways.

4.4 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. (2014) 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.[38]

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 6b. 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.[39] 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.

5 Other implications

5.1 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 $a_{\text{Mortensen}}$ portion unpriced). Mechanically, this amounts to replacing the Ψ in the earnings equations with βV^e, where the β is estimated from a regression of Ψ on V^e. I find that equalizing nonpay characteristics and compensating workers reduces earnings inequality.

---

[37]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.

[38]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 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."

[39] Panel A of Table A10 shows the sectoral measures of hours, and the footnote provides more detail on sample construction. Panel B reports the R^2.

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.

Table 6 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 8 plots the actual distribution of earnings and the distribution in this counterfactual.[40] Inequality is reduced primarily by shifting in the lower tail of the distribution. In contrast, Figure 8b 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.

5.2 Ability of search models to match earnings dispersion

The message of this paper is consistent with Hornstein, Krusell, and Violante (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, Krusell, and Violante (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, Krusell, and Violante (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

---

[40]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.

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.

6   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.[41]

6.1   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.[42] 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 4 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, Heining, and Kline (2013, pg. 991) and note that in the event studies in Figure A3 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

---

[41]Limitations on computational resources meant that I only bootstrapped the main results in the paper.
[42]This divides the sample at approximately half of the EE transitions.

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.

6.2 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 4 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. As such, this evidence suggests that heterogeneity is unlikely to be driving my results.

6.3 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 (α_w in equation (1)) and then re-estimate the model in both samples. Table 4 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. Taking this heterogeneity into account generates a larger role for compensating differentials than in the benchmark results, which suggests that ignoring this form of heterogeneity is not driving the results.

6.4 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, Kramarz, and Margolis (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 4 shows the results for the five sectors where the model finds the *smallest* role for nonpay characteristics in explaining within-sector variation in pay.[43] 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 4 shows that the overall weighted average of the decomposition is that about 14% of the variance of earnings is firm-level compensating differentials.

7 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.

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 (e.g., Ulman (1965), Dickens and Katz (1987), Krueger and Summers (1988), Holzer, Katz, and Krueger (1991), and Card, Heining, and Kline (2012)).[44] 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, Katz, and Krueger (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, Katz, and Krueger (1991) emphasize that the

---

[43]Recall from Table 5 that 40 percent of the variation in pay is between sector.

[44]Similarly, 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.

imperfect correlation might reflect sampling variability. With a large administrative dataset, I have shown in multiple ways that sampling variability does not drive my results.

Besides sampling variability, I have taken into account several other 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. 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. Although I have tried to provide evidence on these concerns, given the nature of the data and the exercise there are necessarily limitations to how reassuring I can be. Besides investigating these issues more fully, future work could try to relate the nonpay characteristics to observables, or find richer datasets that require fewer assumptions.

References

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Fixed Effects Using Linked Longitudinal Employer-Employee Data." Working Paper TP-2002-06, US Census Bureau: Longitudinal Employer Household Dynamics.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2):251–333.

Abowd, John M., Paul Lengermann, and Kevin L. McKinney. 2003. "The Measurement of Human Capital in the U.S. Economy." Working paper.

Abowd, John M., Kevin L. Mckinney, and Nellie L. Zhao. 2017. "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data." Working paper, NBER.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. University of Chicago Press, 149–230.

Aizawa, Naoki and Hanming Fang. 2015. "Equilibrium Labor Market Search and Health Insurance Reform." Working paper.

Andersson, Fredrik, Elizabeth E. Davis, Matthew L. Freedman, Julia I. Lane, Brian P. McCall, and Kristin Sandusky. 2012. "Decomposing the Sources of Earnings Inequality: Assessing the Role of Reallocation." *Industrial Relations* 51 (4):779–810.

Arcidiacono, Peter and Paul B. Ellickson. 2011. "Practical Methods for the Estimation of Dynamic Discrete Choice Models." *Annual Review of Economics* 3:363–394.

Artuc, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review* 100 (3):1008–1045.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90 (2):300–323.

Bagger, Jesper, Francois Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin. 2014. "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics." *American Economic Review* 104 (6):1551–1596.

Bagger, Jesper and Rasmus Lentz. 2016. "An Equilibrium Model of Wage Dispersion and Sorting." Working paper.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman. 2016. "It's Where You Work: Increases in Earnings Dispersion Across Establishments and Individuals in the U.S." *Journal of Labor Economics* 34 (S2):S67–S97.

Becker, Dan. 2011. "Non-Wage Job Characteristics and the Case of the Missing Margin." Working paper.

Becker, Gary S., Elisabeth M. Landes, and Robert T. Michael. 1977. "An Economic Analysis of Marital Instability." *Journal of Political Economy* 85 (6):1141–1187.

Benedetto, Gary, John Haltiwanger, Julia Lane, and Kevin Mckinney. 2007. "Using Worker Flows to Measure Firm Dynamics." *Journal of Business and Economic Statistics* 25 (3):299–313.

Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer. 2011. "Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data." *Journal of Business and Economic Statistics* 29 (4):493–505.

BLS. 1997. *U.S. Bureau of Labor Statistics Handbook of Methods*.

Bonhomme, Stephane and Gregory Jolivet. 2009. "The Pervasive Absence of Compensating Differentials." *Journal of Applied Econometrics* 24:763–795.

Bowlus, Audra and Lars Vilhuber. 2002. "Displaced workers, early leavers, and re-employment wages." Work in progress.

Brown, Charles and James Medoff. 1989. "The Employer Size-Wage Effect." *Journal of Political Economy* 97 (5):1027–1059.

Burdett, Kenneth and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2):257–273.

Burgess, Simon, Julia Lane, and David Stevens. 2000. "Job Flows, Worker Flows, and Churning." *Journal of Labor Economics* 18 (3):473–502.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence." *Econometrica* 74 (2):323–364.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2017. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* .

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131 (3):633–686.

Card, David, Joerg Heining, and Patrick Kline. 2012. "Workplace Heterogeneity and the rise of West German wage inequality." Working Paper 18522, NBER.

———. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3):967–1015.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the U.S. Health Care Sector." *American Economic Review* 106 (8):2110–2144.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9):2593–2632.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129 (4):1553–1623.

Davis, Steven J., R. Jason Faberman, and John Haltiwanger. 2012. "Labor Market Flows in the Cross Section and Over Time." *Journal of Monetary Economics* 59 (2012):1–18.

Dey, Matthew S. and Christopher J. Flinn. 2005. "An Equilibrium Model of Health Insurance Provision and Wage Determination." *Econometrica* 73 (2):571–627.

Dickens, William T. and Lawrence F. Katz. 1987. "Inter-Industry Wage Differences and Theories of Wage Determination." Working paper.

Eeckhout, Jan and Philipp Kircher. 2011. "Identifying Sorting—In Theory." *Review of Economic Studies* 78 (3):872–906.

Engbom, Niklas and Christian Moser. 2017. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." Working paper.

Flaaen, Aaron, Matthew D. Shapiro, and Isaac Sorkin. 2016. "Reconsidering the Consequences of Worker Displacements: Firm versus Worker Perspective." Work in progress.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. 2005. "Comparative Advantage, Learning, and Sectoral Wage Determination." *Journal of Labor Economics* 23 (4):681–724.

Goldschmidt, Deborah and Johannes F. Schmieder. 2017. "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure." *Quarterly Journal of Economics* 132 (3):1165–1217.

Gronberg, Timothy J. and W. Robert Reed. 1994. "Estimating Workers' Marginal Willingness to Pay for Job Attributes Using Duration Data." *Journal of Human Resources* 29 (3):911–931.

Hagedorn, Marcus, Tzuo Hann Law, and Iourri Manovskii. 2017. "Identifying Equilibrium Models of Labor Market Sorting." *Econometrica* 85 (1):29–65.

Hall, Robert E. and Andreas I. Mueller. 2017. "Wage Dispersion and Search Behavior: The Importance of Non-Wage Job Values." *Journal of Political Economy* .

Haltiwanger, John, Henry R. Hyatt, Lisa B. Kahn, and Erika M. McEntarfer. 2017. "Cyclical Job Ladders by Firm Size and Firm Wage." *American Economic Journal: Macroeconomics* .

Holzer, Harry J., Lawrence F. Katz, and Alan B. Krueger. 1991. "Job Queues and Wages." *Quarterly Journal of Economics* 106 (3):739–768.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *American Economic Review* 101:2873–2898.

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

Hwang, Hae-Shin, Dale T. Mortensen, and W. Robert Reed. 1998. "Hedonic Wages and Labor Market Search." *Journal of Labor Economics* 16 (4):815–847.

Hyatt, Henry R. and Erika McEntarfer. 2012. "Job-to-Job Flows and the Business Cycle." Working paper.

Hyatt, Henry R., Erika McEntarfer, Kevin McKinney, Stephen Tibbets, and Doug Walton. 2014. "Job-to-Job Flows (JTJ) Flows: New Labor Market Statistics From Linked Employer-Employee Data ." Working Paper CES 14-34.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83 (4):685–709.

Jarosch, Gregor. 2015. "Searching for Job Security and the Consequences of Job Loss ." Working paper.

Jolivet, Gregory, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US." *European Economic Review* 50:877–907.

Kline, Patrick. 2008. "Understanding sectoral labor market dynamics: An equilibrium analysis of the oil and gas field services industry." Working paper.

Kornfeld, Robert and Howard S. Bloom. 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals." *Journal of Labor Economics* 17 (1):168–197.

Krueger, Alan B. and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (2):259–293.

Lang, Kevin and Sumon Majumdar. 2004. "The Pricing of Job Characteristics When Markets Do Not Clear: Theory and Policy Implications." *International Economic Review* 45 (4):1111–1128.

Lange, Fabian. 2007. "The Speed of Employer Learning." *Journal of Labor Economics* 25 (1):1–35.

Lavetti, Kurt. 2017. "The Estimation of Compensating Wage Differentials: Lessons from the Deadliest Catch." Working paper.

Lavetti, Kurt and Ian M. Schmutte. 2016. "Estimating Compensating Wage Differentials with Endogenous Job Mobility." Working paper.

Lopes de Melo, Rafael. 2016. "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence." *Journal of Political Economy* .

Marimon, Ramon and Fabrizio Zilibotti. 1999. "Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits." *Economic Journal* 455:266–291.

Mas, Alexandre and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements."

McLaughlin, Kenneth J. 1991. "A Theory of Quits and Layoffs with Efficient Turnover." *Journal of Political Economy* 99 (1):1–29.

Minc, Henryk. 1988. *Nonnegative matrics*. Wiley.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78 (381):47–55.

Mortensen, Dale T. 2003. *Wage Dispersion: Why are similar workers paid differently?* MIT Press.

Moscarini, Giuseppe and Fabien Postel-Vinay. 2016. "Did the Job Ladder Fail After the Great Recession?" *Journal of Labor Economics* 34 (S1):S55–S93.

Nunn, Ryan. 2013. "Match Quality with Unpriced Amenities." Working paper.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1998. "The PageRank Citation Ranking: Bringing Order to the Web." Working paper, Stanford. URL `http://ilpubs.stanford.edu:8090/422/`.

Palacios-Huerta, Ignacio and Oscar Volij. 2004. "The Measurement of Intellectual Influence." *Econometrica* 72 (3):963–977.

Papp, Tamas K. 2013. "Frictional Wage Dispersion with Bertrand Competition: An Assessment." *Review of Economic Dynamics* 16:540–552.

Pierce, Brooks. 2001. "Compensation Inequality." *Quarterly Journal of Economics* 116 (4):1493–1525.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118 (1):1–39.

Pinski, Gabriel and Francis Narin. 1976. "Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics." *Information Processing and Management* 12:297–312.

Rosen, Sherwin. 1986. "The Theory of Equalizing Differences." In *Handbook of Labor Economics*, vol. 1, edited by Orley Ashenfelter and Richard Layard. Elsevier, 641–692.

Smith, Adam. 1776/2003. *The Wealth of Nations.* New York: Bantam Dell.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2016. "Firming Up Inequality." Working paper.

Stern, Scott. 2004. "Do Scientists pay to Be Scientists?" *Management Science* 50 (6):835–853.

Sullivan, Paul and Ted To. 2014. "Search and Non-Wage Job Characteristics." *Journal of Human Resources* 49 (2):472–507.

Taber, Christopher and Rune Vejlin. 2016. "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market." Working paper.

Tjaden, Volker and Felix Wellschmied. 2014. "Quantifying the Contribution of Search to Wage Inequality." *American Economic Journal: Macroeconomics* 6 (1):134–161.

Ulman, Lloyd. 1965. "Labor Mobility and the Industrial Wage Structure in the Postwar United States." *Quarterly Journal of Economics* 79 (1):73–97.

Table 1: Summary statistics and the variance of earnings

	All	$\geq$ 15 people-years (per year)	S. Connected by EE	S. Connected by EE (restrictions)
	(1)	(2)	(3)	(4)
Sample size				
People-years	$504{,}945{,}000$	$411{,}088{,}000$	$409{,}550{,}000$	$408{,}961{,}000$
People	$105{,}921{,}000$	$91{,}142{,}000$	$90{,}895{,}000$	$90{,}803{,}000$
Employers	$6{,}155{,}000$	$484{,}000$	$476{,}000$	$472{,}000$
Summary statistics				
Mean log earnings	10.43	10.48	10.48	10.48
Variance of log earnings	0.70	0.67	0.67	0.67
Share of variance of earnings explained by each parameter set				
Employers			0.21	
People			0.57	
Xb			0.11	
Variance components				
Variance of emp. effect			0.14	
Variance of person effect			0.51	
Variance of Xb			0.07	
2cov(person, emp.)			0.10	
2cov(Xb, person + emp.)			0.08	
Corr(person, emp.)			0.19	
Overall fit of AKM decomposition				
Adj. R^2			0.86	
Match effects model				
Adj. R^2			0.92	

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 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. AKM is the Abowd, Kramarz, and Margolis (1999) decomposition (see equation (1)). EE is employer-to-employer.

Table 2: Earnings declines, value changes, and firm-level pay

Panel A. $\Pr(y \downarrow)$	All	EE	ENE
Unconditional	0.429	0.374	0.469
Unconditional (nominal)	0.402	0.343	0.445
When moving to a			
...higher-paying firm	0.297	0.268	0.321
...lower-paying firm	0.578	0.515	0.618
Panel B. $\Pr(\Psi \uparrow)$	All	EE	ENE
Unconditional	0.530	0.570	0.501
Panel C. $\Pr(\tilde{V}^{EE} \uparrow)$	All	EE	ENE
Unconditional	0.621	0.664	0.589
When moving to a			
...higher-paying firm	0.753	0.774	0.735
...lower-paying firm	0.472	0.519	0.442
Panel D. $\Pr(V^e \uparrow)$	All	EE	ENE
Unconditional	0.521	0.566	0.488
When moving to a			
...higher-paying firm	0.672	0.696	0.652
...lower-paying firm	0.351	0.394	0.324
Panel E. Correlations	Spearman	Pearson	Pearson
	(rank)		(adjusted for noise)
$\tilde{V}^{EE}$ and Ψ	0.427	0.400	0.435
$\tilde{V}^{EE}[-10\%, +20\%]$ and Ψ	0.475	0.219	0.573
$\tilde{V}^{EE}$ (adjust for size and offers) and Ψ	0.542	0.551	0.571
V^e and Ψ	0.514	0.530	0.543
$\tilde{V}^{EE}$ and log(size)	0.042	0.045	0.049
V^e and log(size)	0.177	0.151	0.154
Ψ and log(size)	0.069	0.093	0.094

Notes: The pay of a firm is defined by its firm effect (Ψ). 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 correlations are adjusted for noise in ways described in Appendix G. The correlations are weighted by person-years at the employer-level in column (4) of Table 1. EE is employer-to-employer and ENE is employer-to-nonemployment-to-employer.

Table 3: Transition probabilities and model parameters

Symbol	Description	Value
	Overall EE Transition Probability	0.096
	Overall ENE Transition Probability	0.138
	EE Share of Transitions	0.413
$\bar{\delta}$	Exogenous ENE	0.039(0.000)
$\bar{\rho}$	Exogenous EE probability	0.029(0.000)
λ_1	Probability of offer on-the-job	0.168(0.000)
	Pr(accept offer — nonemployed)	0.781(0.002)

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 1. The $\bar{\delta}$ and $\bar{\rho}$ are averaged over the entire sample. Standard deviations across bootstrap samples in parentheses.

Table 4: Why do some firms pay so much and some so little?

Panel A. Decompositions

	$\frac{Cov(\Psi,y)}{Var(y)}$	$R^2(\Psi,V^e)$	Comp. diff. share	Person-year share of all
	(1)	(2)	(3)	(4)
All (adjust for noise)	0.209	0.295	0.148	1.000
All	0.211	0.281	0.152	
Large ($\geq$ 1000 per year)	0.198	0.356	0.128	0.372
Sample 1 (use sample 2 Ψ in (1))	0.211	0.273	0.153	
Sample 2 (use sample 1 Ψ in (1))	0.211	0.274	0.153	
Weighted average	0.211	0.273	0.153	0.982
Young (18-34)	0.254	0.390	0.155	
Old (35-61)	0.192	0.126	0.167	
Weighted average	0.216	0.227	0.163	0.974
Men	0.216	0.334	0.144	
Women	0.210	0.242	0.159	
Weighted average	0.213	0.288	0.152	0.937
Low $\hat{\alpha}_w$ workers	0.343	0.302	0.239	
High $\hat{\alpha}_w$ workers	0.234	0.181	0.192	
Weighted average	0.282	0.235	0.213	0.964
Agriculture	0.180	0.343	0.119	
Admin/Support/Waste	0.160	0.279	0.115	
Hotel/Food	0.137	0.244	0.103	
Retail Trade	0.120	0.206	0.095	
Manufacturing	0.169	0.202	0.135	
State-by-state average	0.196	0.314	0.135	0.982

Panel B. Correlations

	Corr(Ψ)	Corr(V^e)	Corr(a_{Rosen})
Sample 1 and 2	0.970	0.948	0.942
Young and Old	0.867	0.828	0.775
Men and Women	0.920	0.898	0.867
Low and high $\hat{\alpha}_w$	0.801	0.836	0.734

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 1 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 Ψ and V^e. Column (3) follows equation (25) 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 1. 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 5: Relationship to observables

	Variance shares		
	Earnings (Ψ)	Nonpay (a_{Rosen})	$R^2(\Psi, V^e)$
	(1)	(2)	(3)
State	0.052	0.077	0.000
County	0.102	0.150	0.049
Sector	0.381	0.297	0.497
4-digit industry	0.547	0.445	0.438

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 $\Psi_j = \beta_0 + \beta_1 V_j^e + \epsilon_j$. In all cases, these variance shares reflect the variance shares of shrunken version of the characteristic-level mean, as described in Appendix G. The R^2 term is the relationship between Ψ and V^e, which is adjusted for noise.

Table 6: Implications for inequality

Scenario	Variance of earnings	Change relative to data
	(1)	(2)
Data	0.672	N/A
Equalize location	0.658	-2.0%
Equalize industry	0.617	-8.2%
Equalize location and industry	0.604	-10.1%
Equalize all	0.570	-15.2%
"Naive"	0.518	-22.9%
Remove firm effects	0.481	-28.5%

Notes: This table shows how equalizing nonpay characteristics and compensating workers affects the variance of earnings. In the top panel, each row reports $Var(y_{it} - \tilde{\hat{a}}_{Rosen,\boldsymbol{J}(i,t)})$, where the $\boldsymbol{J}(i,t)$ 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 ($Var(y_{it} - (1-R^2)\tilde{\hat{\Psi}}_{\boldsymbol{J}(i,t)})$) and subtracting off the firm effects and recomputing the variance of earnings ($Var(y_{it} - \tilde{\hat{\Psi}}_{\boldsymbol{J}(i,t)})$). 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 1.

Figure 1: Change in firm pay related to probability of an earnings cuts
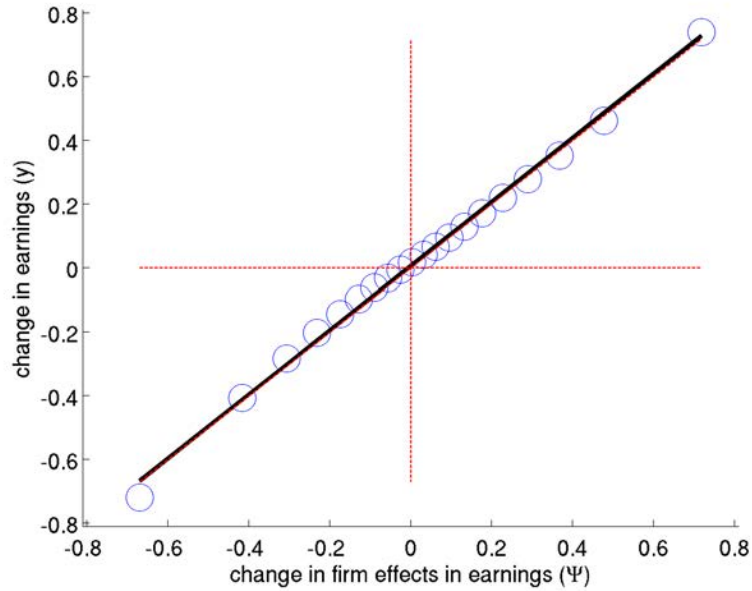
(a) All

(b) EE
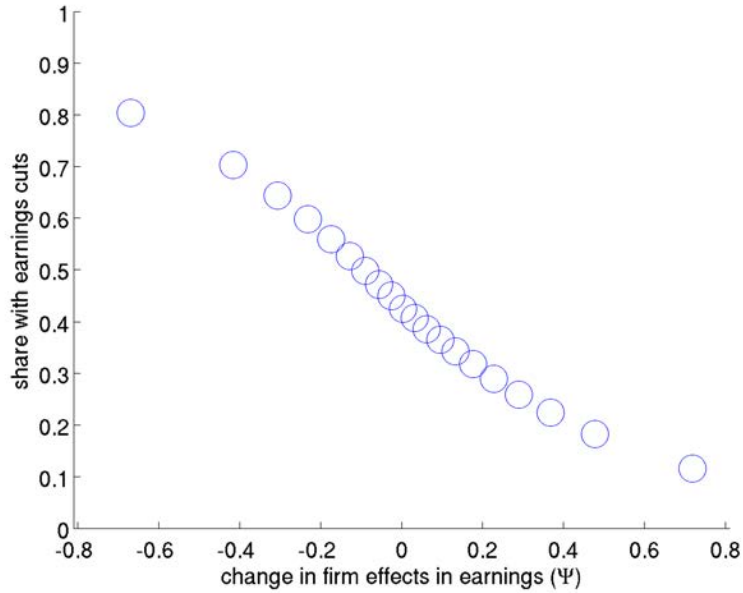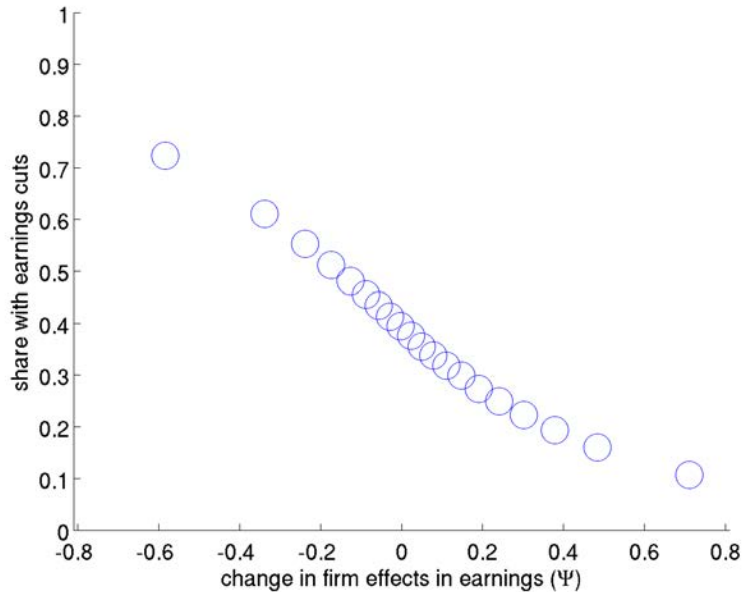
Notes: These figures show the probability of an earnings cut as a function of the change in firm level pay for workers who switch annual dominant jobs. The earnings are the annualized earnings in the last year at the previous job and in the first year at the new job. The top panel looks at all transitions 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 2: Change in firm pay related to magnitude of earnings change

(a) All

(b) EE
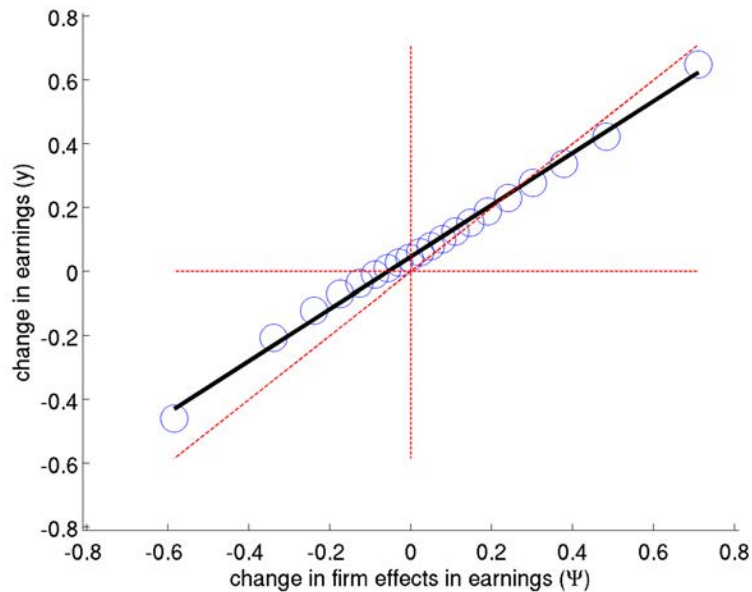
Notes: These figures show how the magnitude of earnings changes relate to the change in firm-level pay for workers who switch annual dominant jobs. The earnings are the residualized annualized earnings in the last year at the previous job and in the first year at the new job. The top panel looks at all transitions and the bottom panel looks at employer-to-employer (EE) transitions. I sort the job changers into 20 bins on the basis of the change in the firm effects. 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 1.005 (standard error: 0.0003), and in the bottom panel is 0.813 (standard error: 0.0003).

Figure 3: Change in firm pay is not related to change in slope of earnings

(a) All

(b) EE
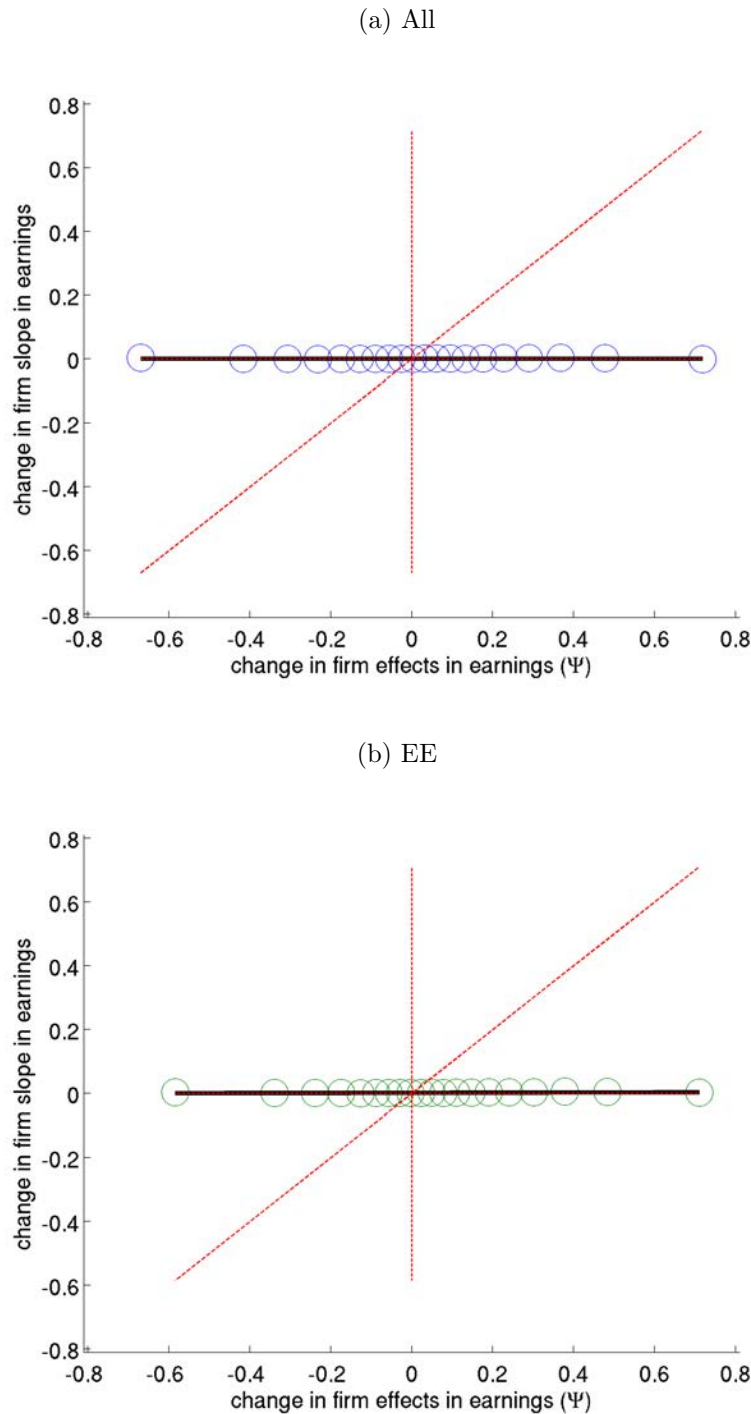
Notes: These figures show how the firm-level slope of pay relates to the change in firm-level pay for workers who switch annual dominant jobs. The slope of firm-level pay is estimated using the earnings changes of the stayers. The top panel looks at all transitions and the bottom panel looks at employer-to-employer (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 circles plot the bin means. The solid line plots the best-fitting line estimated based on the micro-data. The coefficient in the upper panel is 0.000, and in the bottom panel it is 0.002.

Figure 4: Moves to lower-paying firms are systematic

(a) Probability of a higher $\tilde{V}^{EE}$ firm
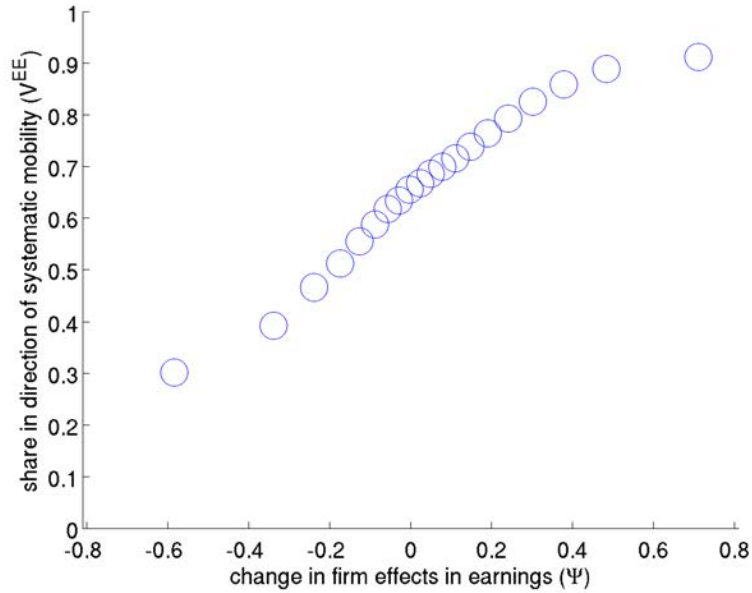
(b) Probability of a higher-value firm (V^e)
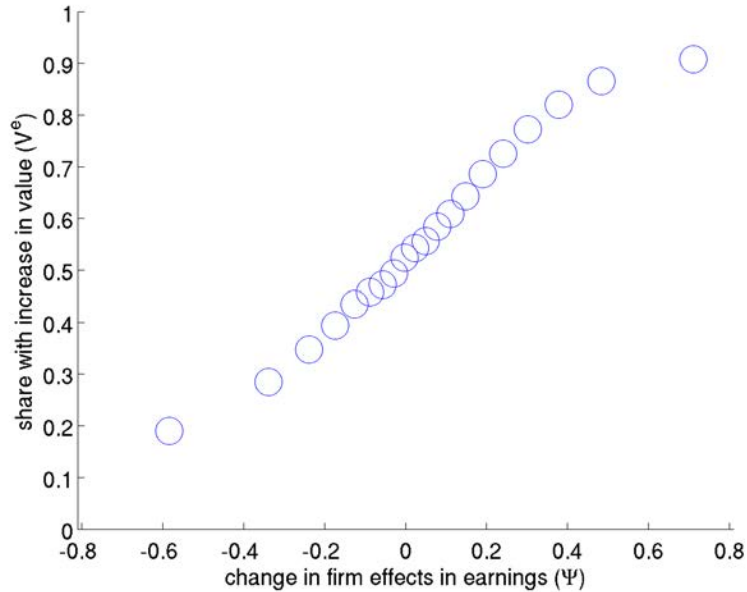
Notes: This figure considers the sample of workers who switch annual dominant jobs on EE transitions. I sort the job changers into 20 bins on the basis of the change in the firm effects. The upper panel reports the probability of an increase in $\tilde{V}^{EE}$, which is the systematic direction of mobility 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.

49

Figure 5: Constructing exogenous rates
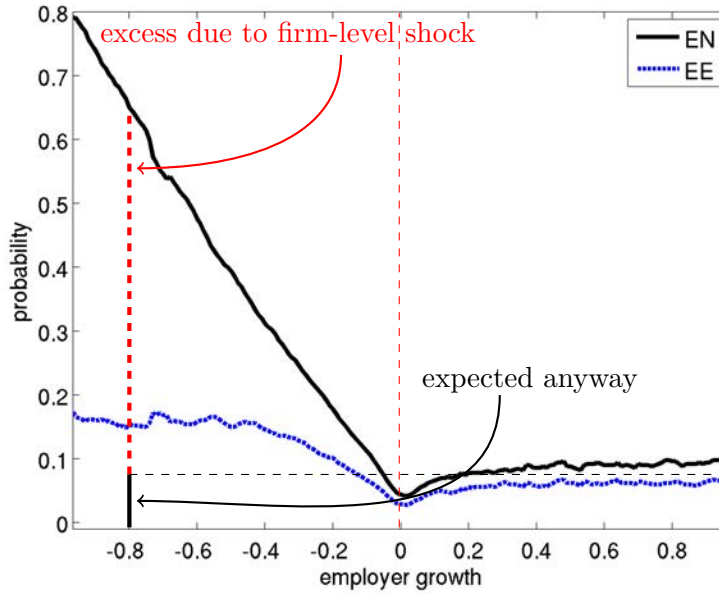
Notes: These figures shows the data used to construct the exogenous weights for employer-to-employer (EE) and employer-to-nonemployer (EN) transitions. The probabilities and growth rates are quarterly. The probabilities are computed in one percentage point wide bins of employer growth rates. The figure 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 6: Relationship between values and earnings

(a) Overall
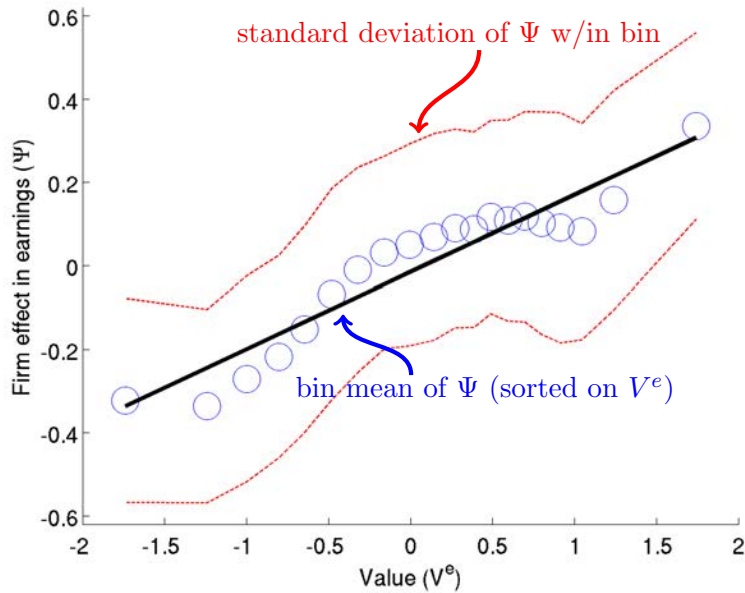
(b) Sector
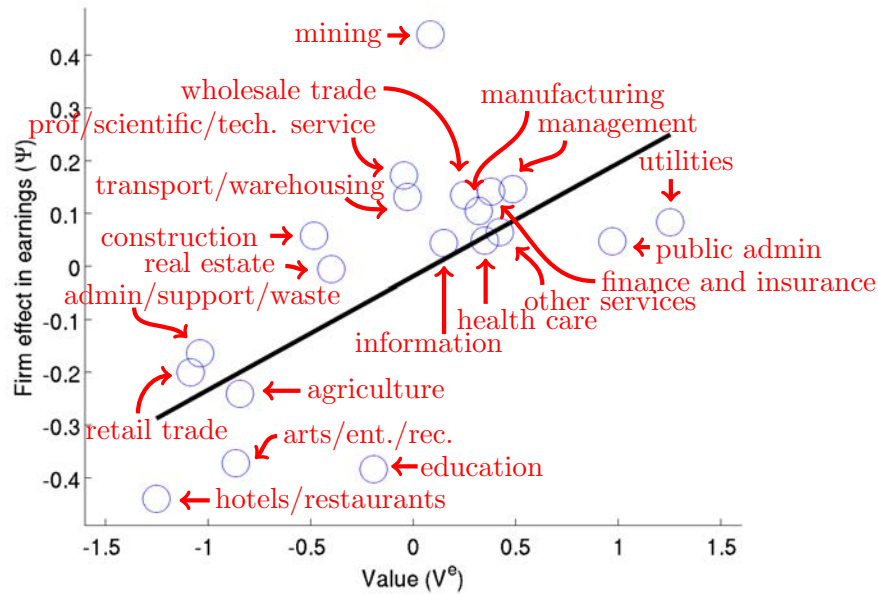
Notes: The top panel sorts firms on the basis of firm-level values. The circles plot 20 bins with the same number of person-years, while the solid line plots the regression line estimated on the firm-level data. The red line shows plus and minus one standard deviation of the firm-level earnings within each value bin. The bottom panel of this figure plots the sector-level means of the earnings and values. The solid black line plots the regression line run at the sector level and weighting by the number of person-years represented by each sector.

Figure 7: Dispersion in the labor market
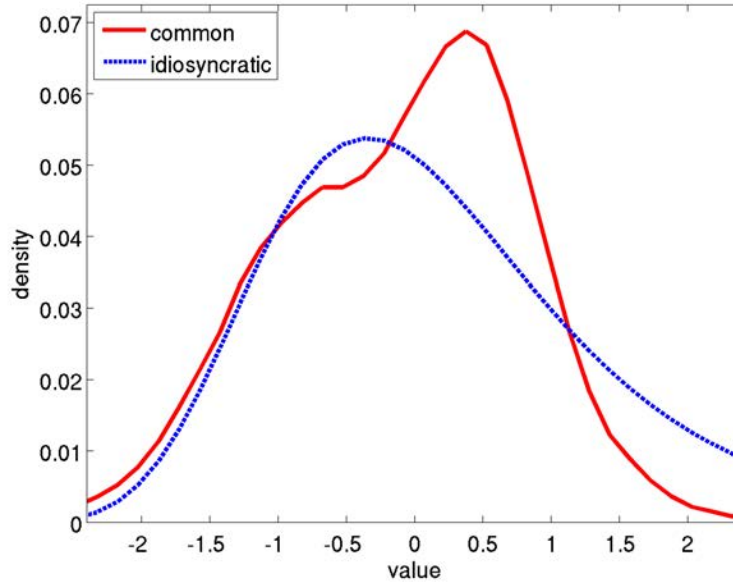
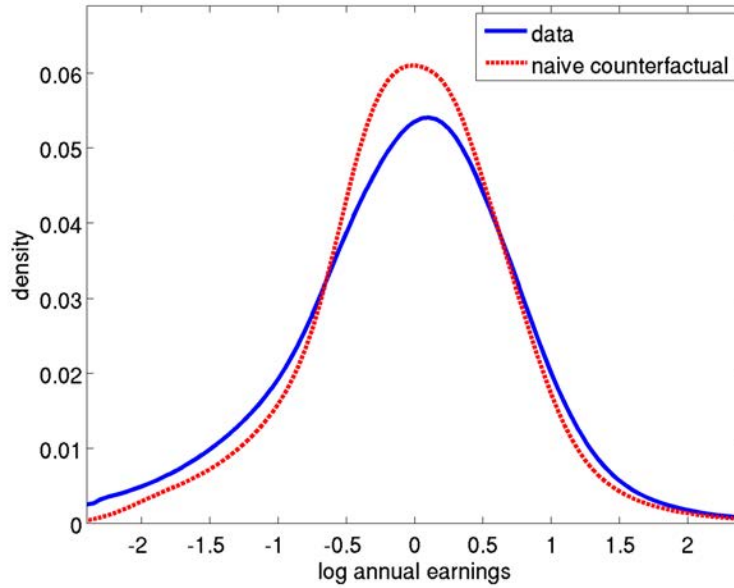*Notes:* This figure plots the dispersion in the value of the firms relative to the dispersion in the idiosyncratic utility draw. The idiosyncratic utility draw distribution is type I extreme value with scale parameter 1 and variance $\frac{\pi^2}{6} \approx 1.6$, and is a normalization. Both distributions 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.
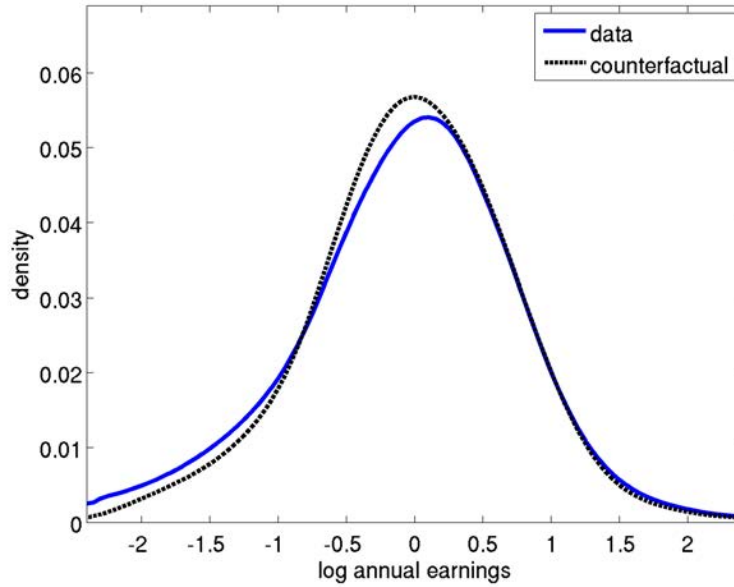
(a) Counterfactual

(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.

A   Appendix: Constructing datasets

Being able to track employers over time is central to measuring employer-to-employer flows and administrative errors in the employer identifiers would lead to an overstatement of flows. Following Benedetto et al. (2007), I assume that large groups of workers moving from employer A to employer B in consecutive periods—especially if employer B did not previously exist—likely reflects errors in the administrative data rather than a genuine set of flows. As such, I correct the employer identifiers using worker flows. I use the Successor-Predecessor File and assume that if 70% or more of employer A's workers moved to employer B, then either 1) employer B is a relabelling of employer A or 2) employer B acquired employer A. Therefore, I do not count such "moves" as employer-to-employer transitions.

A.1   Annual dataset

I follow Abowd, Lengermann, and McKinney (2003) to construct the dataset to estimate the earnings decomposition. I depart from them to define employment in a way that is consistent with how employment is defined to construct employer-to-employer flows, to follow more recent literature in imposing age restrictions, and to follow more recent literature in dropping jobs with very low earnings.

For the purposes of estimating the earnings decomposition, the *annual dominant employer* is the employer from which the worker had the highest earnings in the calendar year. This employer is chosen from the employers from which the worker had received earnings for two or more consecutive quarters within the calendar year; the reason to make this restriction is to allow me to code transitions between employers as employer-to-employer or employer-to-nonemployment-to-employer.[45] In this set of jobs, the annual dominant employer is the one with the highest total earnings in the calendar year.

To construct annualized earnings, for each quarter within a year I first identify the nature of the worker's attachment to the employer. Specifically, code quarter t of earnings into one of the following two mutually exclusive categories: full-quarter (if earnings from the employer are in quarters $t-1$, t and $t+1$) or continuous (if earnings are in quarters $t-1$ and t or in t and $t+1$). Annualize these earnings as follows. First, if the worker had any quarters of full-quarter earnings, take the average of these quarters and multiply by 4 to get an annualized salary. Second, if the worker did not have full-quarter earnings and has any quarters of continuous earnings, take the average of these and multiply by 8 to get an annualized salary. The justification for this procedure is that if a worker is present in only two consecutive quarters and if employment duration is uniformly distributed then on average the earnings represent $\frac{1}{2}$ a quarter's work, while if a worker is present

---

[45]This eliminates quarters of employment that Abowd, Lengermann, and McKinney (2003, pg. 15-16) term "discontinuous," that is, where a worker is observed in neither adjacent quarter. Abowd, Lengermann, and McKinney (2003, pg. 15-16) report that such discontinuous quarters of employment accounted for 5 percent of person-year observations in their final dataset. Second, it eliminates "continuous" quarters of employment where the first quarter of the match is quarter IV within the year, and the second quarter is quarter I of the following year. Under the assumption that continuous quarters are uniformly distributed within the year, this eliminates $\frac{1}{8}$ of continuous workers. Abowd, Lengermann, and McKinney (2003, pg. 15-16) report that continuous quarters account for 11 percent of observations in their final dataset, so this eliminates about 1.4 percent of observations.

in both adjacent quarters then the earnings reflect a full quarter's work.[46] Then take the log of these earnings.

I then make two additional sample restrictions. First, I keep workers aged 18-61 (on December 31st of the year), inclusive. This is an attempt to avoid issues with retirement. This age restriction is similar to, e.g., Card, Heining, and Kline (2013) (20-60 in Germany) and Taber and Vejlin (2016) (19-55 in Denmark), though Abowd, Lengermann, and McKinney (2003) do not report imposing any age restriction. Second, following Card, Heining, and Kline (2013) I drop observations with *annualized* earnings of less than $3,250 in 2011:IV dollars.[47]

I now summarize how the various sample restrictions affect the same size. Table A2 in Appendix K shows that there are about 650 million person-employer-years before imposing an earnings test, 614 million after imposing an earnings test, and 505 million after going down to one observation per person per year. This means that after dropping the low-earnings jobs, there are an average of 1.2 employers per person per year.[48] Table A3 shows the distribution of the number of jobs per year in row 2 of Table A2.

Table A4 shows that on the full annual dataset, 91% share of person-year observations are full-quarter and 9% are continuous.[49] Table A5 shows the distribution of the number of years per person. About 40% of the people are in the dataset for all 7 years, and only 13% are in the dataset for only a single year. Table A6 shows that there is a substantial amount of mobility in this sample: half of the workers have two or more employers. Table A7 shows that about 10% of person-employer matches (or 30% of person-years) last for the entire span of my data. However, almost half of matches (20% of person-years) only last for a single year.

A.2 Quarterly dataset

I build on ideas developed in Bjelland et al. (2011) and Hyatt et al. (2014). Specifically, the procedure of restricting to jobs with two quarters of earnings and using overlapping quarters of earnings to label an employer-to-employer transition comes from Bjelland et al. (2011, pg. 496, equation 2). The idea of using earnings in the two quarters to select the dominant job is found in Hyatt et al. (2014, pg. 3).

For the purposes of measuring flows, the *quarterly dominant employer* in quarter t is the employer from which the worker had the highest earnings summing over quarter t and quarter $t-1$. This job is chosen from among the employers where the worker had positive earnings in *both* quarter t *and* quarter $t-1$. To count as employment, the earnings must pass the the same earnings test as for the annual dataset.[50]

For the person-quarters that remain after the earnings test, the goal is to select a single employer—the quarterly dominant employer. The quarterly dominant employer is the employer from which the worker has the most total earnings summing across $t-1$ and t. There is one

---

[46]In the small number of cases where a worker had forward-looking continuous employment in quarter IV as well as another quarter of continuous employment at the same employer, I included this quarter in the earnings calculation.

[47]Card, Heining, and Kline (2013) drop daily wages of less than 10 euros. 10 euros $\times \approx 1.3$ euros per dollar $\times 250$ days per year= 3,250.

[48]For Germany, Card, Heining, and Kline (2013, Appendix Table 1a, row 5) find 1.10 employers per person per year, and this number is stable from 1985 through 2009.

[49]Abowd, Lengermann, and McKinney (2003, pg. 15-16) find 84% are full-quarter, 11% are continuous, and 5% are discontinuous.

[50]Sum together the two quarters of earnings and multiply by 4. If the earnings are below $3,250, then drop the person-employer match. Multiplying by 4 is justified if one assumes that each quarter is a continuous quarter of employment. The assumption that this is a continuous quarter of employment does lead to more jobs being included than the annual dataset; specifically, if a job is actually full-quarter, then the annualized earnings treating it as full quarter can be lower than the annualized earnings assuming it is a continuous quarter.

exception to this selection rule. If a worker has earnings from her *annual* dominant employer in quarters $t-1$ and t, then this employer is the quarterly dominant employer regardless of whether it is the employer with the most total earnings summing across $t-1$ and t. The reason for prioritizing the annual dominant job is that I want to use this quarterly dataset to code transitions between annual dominant jobs so it is important that they appear in the quarterly dataset.

If a worker has different quarterly dominant employers in quarter t and quarter $t+1$, then this worker had earnings from both employers in quarter t and I label the worker as having undergone an employer-to-employer transition in quarter t. If a worker has no dominant employer in quarter $t+1$, then, with one exception highlighted a little later in this section, I consider that worker to have been nonemployed in quarter $t = 1$, so I label the transition from the quarter t dominant employer as a transition into nonemployment.[51][52]

I depart from prior work to address the possibility that workers move on the seam between two quarters (Hyatt and McEntarfer (2012) emphasize that on some outcomes these transitions look like employer-to-employer moves). To make this concrete, suppose that I observe a worker at firm A in quarter $t-2$ and $t-1$ and at firm B in t and $t+1$. Then the definitions developed previously say that in quarter $t-1$ firm A is the dominant employer and in quarter $t+1$ firm B is the dominant employer. But in quarter t the worker had no dominant employer because it was not the second consecutive quarter of any employment relationship. So the transition from A to B was an employer-to-nonemployment-to-employer transition. It might be, however, that the worker's last day at A was the last day of quarter $t-1$ and her first day at B was the first day of quarter t, so this was actually an employer-to-employer transition. The way I attempt to capture these transitions is to use the stability of earnings across quarters to suggest that a worker was probably employed for the full quarter in both quarters. So, if the earnings from firm A in quarters $t-2$ and $t-1$ are within 5% of each other (using quarter $t-1$ earnings as the denominator), then this employer is the dominant employer in quarter t. This then allows me code the transition from A to B as employer-to-employer. Table A8 shows that this correction accounts for 3.5% of the employer-to-employer transitions in my dataset.

The final result is a dataset that at the quarterly level says where the person was employed and, if this is a new job, says whether the worker came to this job directly from another job, or had an intervening spell of nonemployment.

A.3 Combining the quarterly and annual datasets

The goal of combining the datasets is to use the detail of the quarterly dataset to label each transition between annual dominant employers as an employer-to-employer or an employer-to-nonemployment-to-employer transition.

To label the transition as employer-to-employer or employer-to-nonemployment-to-employer, I proceed as follows. First, identify consecutive observations where a worker has a different annual dominant employer; to be concrete, suppose that the worker's annual dominant employer is A in 2002 and B in 2003.[53] Second, look at the quarterly dataset and find the last quarter that the worker is employed at A (this might be in 2002 or 2003). Third, look at the quarterly dataset and find the first quarter that the worker is employed at B (this might be in 2002 or 2003). If

---

[51]Similarly, Burgess, Lane, and Stevens (2000) drop matches that only last a single quarter.

[52]This definition of a transition into nonemployment will pick up very few recalls as employer-to-nonemployment-to-employer transitions. The reason is that even if a worker is nonemployed awaiting recall for 13 weeks, the probability that I record a quarter with zero earnings from her employer is less than 10% ($\frac{1}{13}$).

[53]It is possible that a worker only appears in the annual dataset in nonconsecutive years—say, 2002 and 2004. In this case, the procedure ends up labelling the transition an employer-to-nonemployment-to-employer.

the last quarter at A and first quarter at B are adjacent, then there was an overlapping quarter of earnings and I label this an employer-to-employer transition. If not, then typically I label this an employer-to-nonemployment-to-employer transition. The exception to labelling the transition an employer-to-nonemployment-to-employer transition is if the worker made an employer-to-employer move through some third (and possibly fourth or fifth) employer en route to moving from A to B. Suppose, for example, that the worker makes the following transitions (where EE is employer-to-employer and ENE is employer-to-nonemployment-to-employer): $A \xrightarrow{EE} C \xrightarrow{EE} B$. Because the worker only made employer-to-employer transitions between A and B, I label this an employer-to-employer transition between annual dominant employers. Alternatively, suppose that I observe $A \xrightarrow{EE} C \xrightarrow{ENE} B$. Then I label the transition between annual dominant employers an employer-to-nonemployment-to-employer transition.

If a worker never has another employer, then I do not attempt to label this transition. For example, if a worker has a dominant employer in 2006 and no dominant employer in 2007, then I do not record a separation in 2006. The reason is that this could occur for any number of reasons: 1) a worker ages out of my age range, 2) a worker moves out of my states, or 3) a worker leaves the labor force.

B Appendix: Selection-correcting the earnings

I selection-correct the earnings equation by combining the proportionality assumption and the results of the search model. That is, I add the expectation of the error term from the search model to the earnings equation.

In the first period of a worker's employment relationship, this expectation depends on the identity of her prior firm in her first year at each firm. That is, suppose a worker moves from firm 2 to firm 1, then $\mathbb{E}[\iota_1 | V_1^e + \iota_1 > V_2^e + \iota_2] = \mathbb{E}[\iota_1 | \iota_1 - \iota_2 > V_2^e - V_1^e]$. In the second and subsequent years, this selection term for a worker at employer j is

$$\mathbb{E}[\iota | V_j^e, \text{not move}] = \frac{\sum_{\mathcal{E} \backslash j, n} Pr(\text{offer from k and not move}) \mathbb{E}[\iota | \text{offer from k and not move}]}{\sum_{\mathcal{E} \backslash j, n} Pr(\text{offer from k and not move})}. \quad \text{(A1)}$$

For a worker at j, these terms—when involving other firms—are:

$$Pr(\text{offer from k and not move}) = \lambda_1 f_k \frac{exp(V_j^e)}{exp(V_k^e) + exp(V_j^e)}, \quad \text{(A2)}$$

$$\mathbb{E}[\iota | \text{offer from k and not move}] = \gamma - \log\left(\frac{exp(V_j^e)}{exp(V_k^e) + exp(V_j^e)}\right). \quad \text{(A3)}$$

For a worker at j, these terms are (when involving nonemployment):

$$Pr(\text{offer from nonemp and not move}) = (1 - \lambda_1) \frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)}, \quad \text{(A4)}$$

$$\mathbb{E}[\iota | \text{offer from nonemp and not move}] = \gamma - \log\left(\frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)}\right). \quad \text{(A5)}$$

In implementation there are a couple issues. First, for the first year that a worker appears in the dataset I do not know which selection correction term to apply; that is, it might be that

the worker showed up from another firm, or it might be that the worker had already been there. To address this, I assume that all such observations are in the second or subsequent years of the employment relationship. Second, there are firms that I cannot estimate the revealed value of, even though I can estimate the value of the firm in the earnings equation (these are firms in the strongly connected set for which I cannot estimate either f or g). For the purposes of the selection correction, I assume that $\frac{g}{f} = 1$ and therefore use the mobility relevant value. Third, to speed up computation, I discretize the firms into $1,000$ equally-sized (in terms of person-years) bins and use the bin means to compute the selection correction.

I then use the expectation of the ι in the earnings equation. Let $\mathbb{E}[\iota|i,t]$ denote the expectation of the ι given the worker i's history. Then I estimate:

$$y_{it} = \alpha_i + \Psi_{\boldsymbol{J}(i,t)} + x'_{it}\beta + \eta\mathbb{E}[\iota|i,t] + \tilde{r}_{it}, \tag{A6}$$

where $\tilde{r}$ indicates that this is a different residual than in equation (1) because I include the expectation of ι. I find that the correlation of Ψ_j with and without the selection correlation is 0.99986. Table A9 presents the variance decomposition with the selection correction.

C   Appendix: Overidentification test

The model has a theory of every entry in the M^o matrix and thus it is possible to construct a wide variety of overidentifying tests by comparing the model predictions for entries to the empirical matrix. The overidentifying test I conduct focuses on the "dense" part of the matrix and asks how well the top eigenvector is able to predict binary comparisons. Specifically, I study the (j,k) pairs where $M^o_{jk} \neq 0$ and $M^o_{kj} \neq 0$. For each such pair, I label the winner of the binary comparison—or the "local" winner— the firm that has the most flows; i.e. j wins if $M^o_{jk} > M^o_{kj}$ and k wins if $M^o_{kj} > M^o_{jk}$. In contrast, the model says that firm k wins if $\tilde{V}^o_k > \tilde{V}^o_j$, or the "global" winner.[54]

The extent of disagreement between the global and local rankings is consistent with the model being the data-generating process. When I weight the comparisons by the number of accepted offers represented in each comparison,[55] the model and the binary comparisons agree on 71.0% of comparisons. Is 70% big or small? This number allows me to reject the null of the model being equivalent to all firms having the same value.[56] I find that the 90% confidence interval under the random null is [49.77%, 50.23%]. Under the null that the model is the data-generating process, the 90% confidence interval is [77.37%, 77.49%].[57] This means that the data are statistically inconsistent with the model being the data-generating process, but the economic magnitude of the rejection is not large. Thus, I conclude that the top eigenvector of the mobility matrix does a reasonable job of summarizing the structure of the employer-to-employer transitions.

---

[54] The reason to focus on the top eigenvector is that adjustments for the offer distribution will affect both comparisons in the same way.

[55] I.e., for a comparison of j and k, I weight by $M^o_{jk} + M^o_{kj}$.

[56] If in the data I observe M^o_{BA} workers flowing from A to B and M^o_{AB} workers flowing from B to A, then I take $M^o_{AB} + M^o_{BA}$ draws from a binomial distribution, where the probability of choosing A is 0.5. I ask what share of weighted comparisons the model and the binary comparisons agree on. I repeat this procedure 50 times to generate a null distribution under the hypothesis of all firms are equally appealing.

[57] I repeat the procedure described in footnote 56 except that the probability of choosing A is given by $\frac{exp(\tilde{V}_A)}{exp(\tilde{V}_A)+exp(\tilde{V}_B)}$, where $exp(\tilde{V}_A)$ is what I estimate in the model (and similarly for B) and the probability of choosing B is the remaining probability.

D Appendix: Omitted Proofs

Proof of Result 1

Notational/definitional preliminaries: This follows the presentation in Minc (1988) of standard graph theory definitions. Let M be a matrix, where entry M_{ij} measures flows from employer j to employer i. Note that all entries in M are by construction nonnegative: the entries are either zeros, or positive values. Let $\mathcal{E}$ be a set (of employers) labelled from $1...n$. Let A be a set of ordered pairs of elements of $\mathcal{E}$. The pair $D = (\mathcal{E}, A)$ is a *directed* graph. $\mathcal{E}$ is the set of *vertices*, and the elements of A are the *arcs* of D, which represent *directed* flows between employers. A sequence of arcs $(i, t_1)(t_1, t_2)...(t_{m-2}, t_{m-1})(t_{m-1}, j)$ is a *path connecting* j to i. The *adjacency* matrix of a directed graph is the $(0,1)$ matrix whose (i,j) entry is 1 if and only if (i,j) is an arc of D. An adjacency matrix is associated with a nonnegative matrix M if it has the same zero pattern as M. The directed graph is *strongly connected* if for any pair of distinct vertices i and j there is a path in D connecting i to j and j to i. The directed graph is *connected* if for any pair of distinct vertices i and j there is a path in D connecting i to j or a path connecting j to i.

Proof. Observe that if M is strongly connected, then every column sum is nonzero so that the adjacency matrix associated with M is the same as the adjacency matrix associated with $S^{-1}M$.

By Minc (1988), chapter 4, theorem 3.2, a nonnegative matrix is irreducible if and only if the associated directed graph is strongly connected. By Minc (1988), chapter 1, theorem 4.4, an irreducible matrix has exactly one eigenvector in E^n (the simplex). If M represents a set of strongly connected firms then these two theorems (often jointly called the Perron-Frobenius theorem) guarantee the existence of a unique solution of the form:

$$S^{-1}M exp(\tilde{V}) = \lambda exp(\tilde{V}),$$

where all the entries in $exp(\tilde{V})$ are of the same sign.

All that remains to show is that $\lambda = 1$. Consider the jth row of $S^{-1}M exp(\tilde{V}) = \lambda exp(\tilde{V})$. Let e_j be the basis vector; that is, it is a zero vector with 1 in the jth row.

$$[S^{-1}M exp(\tilde{V})]_j = [\lambda exp(\tilde{V})]_j, \tag{A7}$$

$$\frac{e_j^T M exp(\tilde{V})}{||Me_j||_1} = \lambda e_j^T exp(\tilde{V}), \tag{A8}$$

where $|| \cdot ||_1$ is the l_1 norm of a matrix so for an arbitrary matrix A we have $||A||_1 = \sum_k \sum_j |a_{kj}|$. Note that $||Me_j||_1$ is a scalar.

Because M is a nonnegative matrix, we can rewrite the l_1 norm as a dot product with a vector of ones. Specifically, let $\mathbf{1}$ be a column vector of 1s:

$$||Me_j||_1 = \mathbf{1}^T Me_j. \tag{A9}$$

Rearrange:

$$\frac{e_j^T M exp(\tilde{V})}{||Me_j||_1} = \lambda e_j^T exp(\tilde{V}), \tag{A10}$$

$$\frac{e_j^T M exp(\tilde{V})}{\mathbf{1}^T Me_j} = \lambda e_j^T exp(\tilde{V}), \tag{A11}$$

$$e_j^T M exp(\tilde{V}) = \lambda \mathbf{1}^T Me_j e_j^T exp(\tilde{V}). \tag{A12}$$

Now sum over the rows:

$$\sum_j e_j^T M exp(\tilde{V}) = \sum_j \lambda \mathbf{1}^T Me_j e_j^T exp(\tilde{V}), \tag{A13}$$

$$\sum_j e_j^T M exp(\tilde{V}) = \lambda \sum_j \mathbf{1}^T Me_j e_j^T exp(\tilde{V}), \tag{A14}$$

$$\mathbf{1}^T M exp(\tilde{V}) = \lambda \sum_j \mathbf{1}^T Me_j e_j^T exp(\tilde{V}), \tag{A15}$$

$$\mathbf{1}^T M exp(\tilde{V}) = \lambda \mathbf{1}^T M \sum_j e_j e_j^T exp(\tilde{V}), \tag{A16}$$

$$\mathbf{1}^T M exp(\tilde{V}) = \lambda \mathbf{1}^T M exp(\tilde{V}). \tag{A17}$$

Hence, $\lambda = 1$. □

Proof of Result 2

Proof. The proof shows that the diagonal elements cancel out. First, use the identity from (6):

$$exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n} M_{k'j} = \sum_{k \in \mathcal{E} \cup n} M_{jk} exp(\tilde{V}_k).$$

Expand to write the diagonal elements explicitly:

$$exp(\tilde{V}_j) M_{jj} + exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n \setminus \{j\}} M_{k'j} = \sum_{k \in \mathcal{E} \cup n \setminus \{j\}} M_{jk} exp(\tilde{V}_k) + exp(\tilde{V}_j) M_{jj}.$$

Then cancel the diagonal terms to show that (6) holds with arbitrary diagonal elements:

$$exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n \setminus \{j\}} M_{k'j} = \sum_{k \in \mathcal{E} \cup n \setminus \{j\}} M_{kj} exp(\tilde{V}_k).$$

□

E Appendix: Alternative derivation of the decomposition in section 4

Result 3. *Suppose that the utility function is given by equation (22) and that $\{V_j^e\}_{j \in \mathcal{E}}$ and $\{\Psi_j\}_{j \in \mathcal{E}}$ are known. Then*

$$Var(a_{Rosen}) = (1 - R^2)Var(\Psi),$$
$$Var(a_{Mortensen}) \in [0, \infty)$$

and combining them, we have

$$Var(a) \in [Var(\Psi)(1 - R^2), +\infty),$$

where $R^2 = Corr(V^e, \Psi)^2$. The willingness to pay for Rosen *and* Mortensen *amenities is one. The* Rosen *amenities are related to earnings as follows:*

$$Corr(\Psi, a_{Rosen}) = -\sqrt{1 - R^2}.$$

When $Var(a_{Mortensen}) > 0$,

$$Corr(\Psi, a_{Mortensen}) = \sqrt{R^2}.$$

Bounds on the variance of utility in log dollar units are:

$$Var(\Psi + a) \in [Var(\Psi)R^2, \infty).$$

Proof. It is helpful to first have explicit expressions for a number of quantities. Write the R^2 between Ψ and V in terms of the known variable V and the unknown variable a:

$$R^2 = \frac{Cov(\Psi, V)^2}{Var(\Psi)Var(V)} \tag{A18}$$

$$= \frac{Cov(\Psi, \alpha(\Psi + a))^2}{Var(\Psi)Var(\alpha(\Psi + a))} \tag{A19}$$

$$= \frac{\alpha^2 Cov(\Psi, (\Psi + a))^2}{\alpha^2 Var(\Psi)Var((\Psi + a))} \tag{A20}$$

$$= \frac{[Var(\Psi) + Cov(\Psi, a)]^2}{Var(\Psi)[Var(\Psi) + Var(a) + 2Cov(\Psi, a)]}. \tag{A21}$$

It is also helpful to write $Var(a)$ in terms of one unknown quantity by rearranging equation (A21):

$$R^2[Var(\Psi)^2 + Var(\Psi)Var(a) + 2Var(\Psi)Cov(\Psi, a)] = Var(\Psi)^2 + 2Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2, \tag{A22}$$

$$R^2 Var(\Psi)Var(a) = (1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2, \tag{A23}$$

$$Var(a) = \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}{R^2 Var(\Psi)}. \tag{A24}$$

The following is a useful expression for $Corr(\Psi, a)$:

$$Corr(\Psi, a) = \frac{Cov(\Psi, a)}{\sqrt{Var(a)Var(\Psi)}}, \tag{A25}$$

$$= \frac{Cov(\Psi, a)}{\sqrt{\frac{(1-R^2)Var(\Psi)^2 + 2(1-R^2)Var(\Psi)Cov(\Psi,a) + Cov(\Psi,a)^2}{R^2 Var(\Psi)}}Var(\Psi)}, \tag{A26}$$

$$= \sqrt{R^2} \frac{Cov(\Psi, a)}{\sqrt{(1-R^2)Var(\Psi)^2 + 2(1-R^2)Var(\Psi)Cov(\Psi,a) + Cov(\Psi,a)^2}}. \tag{A27}$$

A lower bound on $Var(a)$: To minimize $Var(a)$, start with the expression for $Var(a)$ (equation (A24)) in terms of $Cov(\Psi, a)$ and take the first order condition with respect to $Cov(\Psi, a)$:

$$\frac{\partial Var(a)}{\partial Cov(\Psi, a)} = \frac{2(1 - R^2)Var(\Psi) + 2Cov(\Psi, a)}{R^2 Var(\Psi)} \tag{A28}$$

$$0 = \frac{2(1 - R^2)Var(\Psi) + 2Cov(\Psi, a)}{R^2 Var(\Psi)} \tag{A29}$$

$$Cov(\Psi, a) = -(1 - R^2)Var(\Psi). \tag{A30}$$

The second order condition is $\frac{2}{R^2 Var(\Psi)}$, which is positive. Substitute this into the expression for $Var(a)$ (equation (A24)) to get that the minimum value is given by:

$$Var(a) = \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)(-(1 - R^2)Var(\Psi)) + (-(1 - R^2)Var(\Psi))^2}{R^2 Var(\Psi)} \tag{A31}$$

$$= Var(\Psi)(1 - R^2). \tag{A32}$$

Compute the correlation between Ψ and a at this lower bound:

$$Corr(\Psi, a) = \frac{Cov(\Psi, a)}{\sqrt{Var(a)Var(\Psi)}} \tag{A33}$$

$$= \frac{-(1 - R^2)Var(\Psi)}{\sqrt{Var(\Psi)(1 - R^2)Var(\Psi)}} \tag{A34}$$

$$= -\sqrt{1 - R^2}. \tag{A35}$$

And compute the variance of utility in log dollar units:

$$Var(\Psi + a) = Var(\Psi) + Var(a) + 2Cov(\Psi, a) \tag{A36}$$

$$= Var(\Psi) + Var(\Psi)(1 - R^2) - 2(1 - R^2)Var(\Psi) \tag{A37}$$

$$= R^2 Var(\Psi). \tag{A38}$$

An upper bound on $Var(a)$: Take the limit of the expression for $Var(a)$ (equation (A24)) while treating R^2 as a constant (because it is observable data):

$$\lim_{Cov(\Psi,a) \to \infty} \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}{R^2 Var(\Psi)} = \infty. \tag{A39}$$

Note that this implies that $Var(a)$ goes to infinity with the square of $Cov(\Psi, a)$, which is why the R^2 expression remains finite.

What is $Corr(\Psi, a)$ in this case?

$$\lim_{Cov(\Psi,a)\to\infty} Corr(\Psi, a) = \lim_{Cov(\Psi,a)\to\infty} \sqrt{R^2} \frac{Cov(\Psi, a)}{\sqrt{(1-R^2)Var(\Psi)^2 + 2(1-R^2)Var(\Psi)Cov(\Psi,a) + Cov(\Psi,a)^2}} \tag{A40}$$

$$= \sqrt{R^2} \tag{A41}$$

$$= Corr(\Psi, V). \tag{A42}$$

And:

$$Var(\Psi + a) \to \infty. \tag{A43}$$

Rosen vs. Mortensen amenities: To decompose the a term into *Rosen* and *Mortensen* amenities, note that the properties of the *Rosen* amenities correspond to the lower bounds in these results, while the properties of the *Mortensen* amenities correspond to the upper bounds.

To see that the a_{Rosen} term captures variation in pay while holding value constant, consider the following equation:

$$\Psi = \beta V^e + \epsilon. \tag{A44}$$

Treating this equation as a regression (where we have demeaned V^e and Ψ, which is without loss of generality because they are only identified up to location), we have:

$$\hat{\beta} = \frac{Cov(\Psi, V^e)}{Var(V^e)}. \tag{A45}$$

And:

$$\hat{\epsilon} = \Psi - \hat{\beta} V^e. \tag{A46}$$

Note that $\hat{\epsilon} = -a_{Rosen}$ because it generates variation in pay (Ψ) while holding value constant. Hence, a_{Rosen} and Ψ are in the same units and so workers are willing to trade them off one-for-one.

For the variance of a_{Rosen}:

$$Var(a_{Rosen}) = Var(\hat{\beta} V^e - \Psi) \tag{A47}$$

$$= Var(\Psi) + \hat{\beta}^2 Var(V^e) - 2\hat{\beta} Cov(\Psi, V^e) \tag{A48}$$

$$= Var(\Psi) + \left(\frac{Cov(\Psi, V^e)}{Var(V^e)}\right)^2 Var(V^e) - 2\left(\frac{Cov(\Psi, V^e)}{Var(V^e)}\right) Cov(\Psi, V^e) \tag{A49}$$

$$= Var(\Psi) - \frac{Cov(\Psi, V^e)^2}{Var(V^e)} \tag{A50}$$

$$= Var(\Psi) - \frac{Cov(\Psi, V^e)^2}{Var(V^e)Var(\Psi)} Var(\Psi) \tag{A51}$$

$$= Var(\Psi)(1 - R^2). \tag{A52}$$

For the covariance of a_{Rosen} and Ψ:

$$Cov(a_{Rosen}, \Psi) = Cov(\hat{\beta}V^e - \Psi, \Psi) \tag{A53}$$

$$= \hat{\beta}Cov(V^e, \Psi) - Var(\Psi) \tag{A54}$$

$$= \left(\frac{Cov(\Psi, V^e)}{Var(V^e)}\right)Cov(V^e, \Psi) - Var(\Psi) \tag{A55}$$

$$= \frac{Cov(\Psi, V^e)^2}{Var(V^e)Var(\Psi)}Var(\Psi) - Var(\Psi) \tag{A56}$$

$$= -Var(\Psi)(1 - R^2). \tag{A57}$$

Finally, for the correlation of a_{Rosen} and Ψ:

$$Corr(a_{Rosen}, \Psi) = \frac{Cov(a_{Rosen}, \Psi)}{\sqrt{Var(a_{Rosen})Var(\Psi)}} \tag{A58}$$

$$= \frac{Var(\Psi)(1 - R^2)}{\sqrt{Var(\Psi)(1 - R^2)Var(\Psi)}} \tag{A59}$$

$$= -\sqrt{1 - R^2}. \tag{A60}$$

These are exactly the properties of a at the lower bounds.

The properties of a at the upper bound correspond to the properties of $a_{Mortensen}$ (conditional on the variance being positive). In terms of interpretation, $Corr(\Psi, a_{Mortensen}) > 0$ means that a hedonic regression would find a wrong-signed coefficient on $a_{Mortensen}$ and hence corresponds to the explanation for the absence of evidence of compensating differentials that desirable nonpay characteristics are positively correlated with pay. Note that $a_{Mortensen}$ does not correspond to nonpay characteristics that are orthogonal to pay.

Willingness to pay for Rosen and Mortensen amenities: Note that in these derivations $V^e = \omega(\Psi + a)$ so that by construction the Rosen and Mortensen amenities are in the same units as Ψ (log dollars) and workers are willing to trade-off one-for-one between log dollars and the amenities.

□

F   Appendix: Description of estimating the model

Step 0 Initialize the model EE and EN probabilities using EE and EN probabilities at expanding firms. The relative size of employers (g_j) and the number of workers (W) are summary statistics of the data. Initalize $\{V_j^e\}_{j \in \mathcal{E}}$ to be a constant.

Step 1 *Exogenous separations*: Using the method described in section 3.2, build M and compute δ and ρ using these probabilities.

Step 2 *Central tendency of worker flows*: Using equation (8), compute $exp(\tilde{V})$.

Step 3 *Offer distribution*: Compute f by doing a grid-search on λ_1 to match the level of EE flows. As an output this gives a new value of $\{V_j^e, f_j\}_{j \in \mathcal{E}}$ as well as λ_1. See below for more detail.

Step 4 Given the new values of $\{V_j^e\}_{j \in \mathcal{E}}$ use equation (19) to compute the new counterfactual separation probabilities. If the size-weighted correlation between the old and new $\{V_j^e\}_{j \in \mathcal{E}}$ is less than 0.999, then return to step 1.

Details on step 3

Define C_1 to be the share of offers that are accepted from nonemployment, or:

$$C_1 \equiv \sum_{j' \in \mathcal{E}} f_{j'} \frac{exp(V_j^e)}{exp(V_{j'}^e) + exp(V^n)}, \tag{A61}$$

so that f_j^o can be written in terms of model parameters as

$$f_j^o = \frac{f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V^n)}}{C_1}. \tag{A62}$$

Take an initial guess of λ_1:

- Evaluate two equations, where I maintain the convention of data or variables whose values are known by a given step are on the left-hand side, while unknowns are on the right-hand side. In the following equation, g_j and f_j^o are from step 0, δ_j and ρ_j are from step 1, and $\tilde{V}_j$ is from step 2. The first equation is an identity, where the right hand side comes from substituting in equations (14) and (A62) to the left hand side:

$$\frac{g_j exp(\tilde{V}_j)}{f_j^o}(1 - \delta_j)(1 - \rho_j) = g_j \frac{f_j exp(V_j^e)}{g_j(1 - \delta_j)(1 - \rho_j)} C_1 \frac{1}{f_j} \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e)}(1 - \delta_j)(1 - \rho_j) \tag{A63}$$

$$= C_1 [exp(V_j^e) + exp(V^n)]. \tag{A64}$$

The second equation comes from rewriting equation (17) using the C_1 notation:

$$\frac{1}{1 - \lambda_1} \frac{1}{W} \sum_{j \in \mathcal{E}} M_{jn} exp(\tilde{V}_n) = \frac{1}{1 - \lambda_1} \frac{1}{W} \sum_{j \in \mathcal{E}} \lambda_0 U f_j \frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)} \frac{(1 - \lambda_1) W exp(V^n)}{\lambda_0 U} \tag{A65}$$

$$= \frac{1}{1 - \lambda_1}(1 - \lambda_1) exp(V^n) \sum_{j \in \mathcal{E}} f_j \frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)} \tag{A66}$$

$$= exp(V^n) C_1. \tag{A67}$$

- Combine equations (A64) and (A67), to give the following two terms: $C_1 exp(V_j^e)$ and $C_1 exp(V^n)$.

- Rewrite equation (A62) by multiplying by $\frac{C_1}{C_1}$ and rearranging:

$$f_j^o = f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V^n)} \frac{1}{C_1} \tag{A68}$$

$$= f_j \frac{C_1 exp(V_j^e)}{C_1 exp(V_j^e) + C_1 exp(V^n)} \frac{1}{C_1} \tag{A69}$$

$$f_j^o \frac{C_1 exp(V_j^e) + C_1 exp(V^n)}{C_1 exp(V_j^e)} = \frac{f_j}{C_1}. \tag{A70}$$

In this equation, the terms on the left-hand side are known from step 1, so this step gives $\frac{f_j}{C_1}$.

- Now that $\frac{f_j}{C_1}$ is known, solve for C_1 by using the normalization $\sum_{j \in \mathcal{E}} f_j = 1$ (note that C_1 contains f_j, so the scale transformation cancels out.[58]).

$$\sum_{j \in \mathcal{E}} \frac{f_j}{C_1} = \frac{\sum_{j \in \mathcal{E}} f_j}{C_1} = \frac{1}{C_1}. \tag{A71}$$

Now that C_1 is known and from equation (A70) $\frac{f_j}{C_1}$ is known, it is possible to solve for f_j.

- Knowledge of C_1 gives $exp(V^n)$ and $exp(V_j^e)$, via equations (A64) and (A67).

- Given the parameters of the model, compute the number of endogenous employer-to-employer transitions implied by the model:[59]

$$\lambda_1 \sum_i g_j(1 - \delta_j)(1 - \rho_j) \sum_k f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)}. \tag{A72}$$

I search over a grid of width 0.001 and select the λ_1 that minimizes the absolute gap between equation (A72) and the probability of EE transitions in the data, or

$$\frac{\sum_{j \in \mathcal{E} \setminus \{j\}} \sum_{k \in \mathcal{E}} M_{jk}}{W \sum_{j \in \mathcal{E}} g_j(1 - \delta_j)(1 - \rho_j)}. \tag{A73}$$

G   Appendix: Addressing measurement error

This appendix provides more details on two approaches to addressing and quantifying the role of measurement error in driving my results.

Approach 1: Shrinkage

The first approach uses an empirical Bayes approach to shrinkage. Specifically, I follow Morris (1983) and use estimates of the standard errors to downweight noisier observations. In my context, measurement error means that I overstate the variance of the underlying values and so understate the correlation.

Standard errors

To compute standard errors, I use the bootstrap.

To maintain the dependency structure in the data, I resample at the level of worker-year-pairs. That is, if I have three earnings observations for worker w, $\{y_{i,t-1}, y_{i,t}, y_{i,t+1}\}$, then I create a set of two observations, $\{(y_{i,t-1}, y_{i,t}), (y_{i,t}, y_{i,t+1})\}$, where I record how the worker moved (i.e., EE, ENE, or not at all) from the employer in the first period to the second period.[60] The asymptotic thought experiment that this relates to is allowing W in equation (11) to grow.

---

[58] Define $\hat{f}_j = \alpha f_j$ and let $\hat{C}_1$ be the C_1 constructed using $\hat{f}_j$. Then $\frac{\hat{f}_j}{\hat{C}_1} = \frac{f_j}{C_1} = \frac{\alpha f_j}{\sum_{j' \in \mathcal{E}} \alpha f_{j'} \frac{exp(V_{j'}^e)}{exp(V_{j'}^e) + exp(V^n)}}$

[59] To make this computationally feasible, group firms into $1,000$ categories on the basis of the firm values (V^e).

[60] While this procedure places equal weight on each transition between employers, it double-counts the interior earnings observations (in this case, the year t earnings observations). Hence, to compute the earnings decomposition in the bootstrap resamples I delete duplicate interior earnings observations. Formally, if an interior earnings observation appears n times in the bootstrap replicate, then I include the observation $\lceil \frac{n}{2} \rceil$ times, where this notation is the ceiling operator that rounds up to the nearest integer.

Two issues arise in the bootstrap: first, how to normalize estimates and second, the identified set. Because both V_j^e and Ψ_j are only identified up to location, I need to normalize the value of one j. I normalize the location of Ψ by assuming that the estimates are noiseless for a very large firm. I normalize the location of V_j^e by setting $V^n = 0$ in all repetitions.

The second issue is that the identified set of firms varies across bootstrap resamples. The reason is that the strongly connected set of firms differs across the bootstrap resamples. This means that there will be a different number of observations to estimate the variability of each of the parameters (that is, smaller firms will typically have fewer resamples). I address this in two steps. First, my initial sample selection eliminates the smallest firms where this issue is likely to come up the most (i.e., I eliminate firms that have fewer than 90 non-singleton observations). Second, I compute 50 bootstrap replications and only keep firms that show up in at least 20 of them.

Shrinkage

I use the standard errors to shrink the estimates of V_j^e and Ψ_j. Formally, I follow the empirical Bayes approach laid out by Morris (1983). My exposition follows Online Appendix C in Chandra et al. (2016).

Define some notation. Let j be a firm. Let n_J be the number of firms. Let $n_{p(j)}$ be the number of person years represented by firm j. Let q_j be a measure of the quality of the firm—i.e., either V_j^e or Ψ_j. Let $\hat{q}_j$ denote the estimate of q for firm j. Let Q be the $n_j \times 1$ vector of $\hat{q}_j$. Let $\hat{\pi}_j^2$ denote the variance of the estimate. Let $\hat{\sigma}^2$ denote the estimate of the true variance of q_j. Let $\mathbf{x}_j$ be an $n_x \times 1$ vector of characteristics of firm j. I use a set of dummies for 4-digit industry and county. Let X be the stacked vector of the $\mathbf{x}_j'$. Let $\boldsymbol{\lambda}$ be an $n_x \times 1$ vector of coefficients. Finally, let w_j be the weight of firm j and W be the $n_J \times n_J$ matrix with w_j on the diagonal.

The following equations show how these terms relate.

$$w_j = n_{p(j)} \frac{1}{\hat{\pi}_i^2 + \hat{\sigma}^2} \tag{A74}$$

$$\hat{\sigma}^2 = \max \left\{ 0, \frac{\sum_j w_j \left\{ \frac{n_J}{n_J - n_x} (\hat{q}_j - \mathbf{x}_j' \boldsymbol{\lambda})^2 - \hat{\pi}_j^2 \right\}}{\sum_j w_j} \right\} \tag{A75}$$

$$\hat{\boldsymbol{\lambda}} \equiv (X'WX)^{-1} X'WQ. \tag{A76}$$

The two unknowns are $\hat{\sigma}^2$ and $\hat{\boldsymbol{\lambda}}$. These are solved for in a loop. Initialize $w_j = n_{p(j)}$. Then iterate the following till convergence:

1. Compute $\hat{\boldsymbol{\lambda}}$, then a new estimate of $\hat{\sigma}^2$ (using the above equation).

2. Check if $\hat{\sigma}^2$ has converged. If not, update the weights, w_j, and return to step 1.

The feasible shrinkage estimator is:

$$\hat{b}_j = \left(\frac{n_J - n_x - 2}{n_J - n_x} \right) \left(\frac{\hat{\pi}_j^2}{\hat{\pi}_j^2 + \hat{\sigma}^2} \right) \tag{A77}$$

$$q_j^{EB(f)} = (1 - \hat{b}_j)\hat{q}_j + \hat{b}_j \mathbf{x}_j' \hat{\boldsymbol{\lambda}}. \tag{A78}$$

The variance of the distribution unconditional on covariates is given by:

$$\hat{\zeta}^2 = \max \left\{ 0, \frac{\sum_j w_j \{ \frac{n_J}{n_J - 1} (\hat{q}_j - \bar{q})^2 - \hat{\pi}_j^2 \}}{\sum_j w_j} \right\}, \tag{A79}$$

where

$$\bar{q} = \frac{\sum_j w_j \hat{q}_j}{\sum_j w_j}. \tag{A80}$$

Now suppose we have two measure of firm quality A and B and we want to know their correlation. Let a tilde'd variable represent a variable that is adjusted for measurement error. Then:

$$\widetilde{Corr}_j(q_A, q_B) = \frac{Cov(\hat{q}_A, \hat{q}_B)}{\sqrt{\hat{\zeta}_A^2 \hat{\zeta}_B^2}}, \tag{A81}$$

$$\widetilde{R}^2(q_A, q_B) = \left(\widetilde{Corr}_i(q_A, q_B) \right)^2, \tag{A82}$$

where this reflects the assumption that the measurement error in A and B is uncorrelated.

Recall that $\hat{\boldsymbol{\lambda}}$ is the vector of coefficients, which reflects industry and location means of the measure of quality. For some purposes I am interested in comparing these. Hence, I shrink $\hat{\boldsymbol{\lambda}}$ using the observation that $\hat{\boldsymbol{\lambda}}$ is computed from the following regression:

$$\sqrt{W}Q = \boldsymbol{\lambda}\sqrt{W}X, \tag{A83}$$

and so I can estimate the variance around the $\hat{\boldsymbol{\lambda}}$ using analytical formulas for the variance-covariance matrix, and then shrink these estimates using the formulas above.

Approach 2: Split samples

A second approach to quantifying the importance of measurement error is to split the sample in half. Specifically, I divide the sample based on *people* by randomly allocating each unique person into sample 1 or sample 2. By splitting on the basis of people, I get two independent estimates of the value and pay at each firm that shows up in strongly connected set defined by each subsample. With two independent estimates of the same quantity, I can estimate how much of the variance is due to noise.

Formally, let $\hat{V}_j^{e,1}$ be the estimate in subsample 1 and $\hat{V}_j^{e,2}$ be the estimate of the value of being employed at firm j in subsample 2. Assume that:

$$\hat{V}_j^{e,1} = V_j^e + \epsilon_{j,1} \tag{A84}$$

and

$$\hat{V}_j^{e,2} = V_j^e + \epsilon_{j,2}. \tag{A85}$$

Because the samples are mutually exclusive, the errors are uncorrelated and $Cov(\epsilon_{j,1}, \epsilon_{j,2}) =$

$0, Cov(\hat{V}_j^{e,2}, \epsilon_{j,1}) = 0$, and $Cov(\hat{V}_j^{e,1}, \epsilon_{j,2}) = 0 \ \forall \ j$. Hence,

$$Corr(\hat{V}_j^{e,1}, \hat{V}_j^{e,2}) = \frac{Cov(\hat{V}_j^{e,1}, \hat{V}_j^{e,2})}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}} \tag{A86}$$

$$= \frac{Cov(V_j^e, V_j^e) + 2Cov(\epsilon_{j,1}, \epsilon_{j,2})}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}} \tag{A87}$$

$$= \frac{Var(V_j^e)}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}}. \tag{A88}$$

The core exercise of this paper reduces to:

$$R^2(\hat{V}_j^e, \hat{\Psi}_j^e) = \frac{Cov(\hat{V}_j^e, \hat{\Psi}_j^e)}{Var(\hat{V}_j^e)Var(\hat{\Psi}_j^e)}. \tag{A89}$$

Under the assumption that $Cov(\hat{V}_j^e, \hat{\Psi}_j^e) = Cov(V_j^e, \Psi_j^e)$, we have:

$$R^2(V_j^e, \Psi_j^e) = \frac{Cov(V_j^e, \Psi_j^e)^2}{Var(V_j^e)Var(\Psi_j^e)} \tag{A90}$$

$$< \frac{R^2(\hat{V}_j^e, \hat{\Psi}_j^e)}{Corr(\hat{V}_j^{e,1}, \hat{V}_j^{e,2})Corr(\hat{\Psi}_j^{e,1}, \hat{\Psi}_j^{e,2})}. \tag{A91}$$

H Appendix: Monte Carlo evidence

This appendix describes Monte Carlo evidence on the properties of the estimators used in this paper.

Simulation details

There are a few high-level issues that inform the design of the Monte Carlos. First, the search model does not impose steady state so that firms can grow or shrink over time. Hence, if I take the set of parameter values and run the model for a large number of periods then, in the limit, the data would be dominated by a small number of firms. Second, the method to identify exogenous separations relies on variation in the growth rates of a given firm over time. Hence, in order to have a simulation that generates data where it is possible to include all steps of estimation, I need to have multiple time periods. Third, as is well-known, in order to run AKM it is necessary to have multiple observations per worker and to follow workers across firms (whereas in the search model the identify of workers is irrelevant). Fourth, it is computationally intensive to estimate the model.

To balance these various considerations, I proceed as follows. First, to address the fact that it is computationally intensive to estimate the model, I randomly sample of firms. To preserve the size distribution of firms, I stratify the firms by size and then sample from each bin, where the bins contain an equal number of person-years. The bootstrapping is especially time-intensive, so I report a one-in-200 sample for the bootstrapping results and a one-in-10 sample for all results.[61] Second, to generate multiple observations over time, I divide the number of non-singleton person-years at a

---

[61]To show the effect of varying network density, I also report one-in-20 samples.

firm in four $(\frac{g_j W}{4})$ and thus allow for five periods. Third, to allow for variation in firm growth rates, I set the realization of δ_j and ρ_j to zero in three periods, and equal to $4\{\delta_j, \rho_j\}$ in one period (I do not tell the model estimation code which time period this occurs). Fourth, to generate a panel of workers, in each of the four years I follow workers between the two employers. If the firm grows from one year to the next, then I cut-off worker histories randomly. In contrast, if the firm shrinks, then I add new workers. The net result of this procedure is that workers appear in the simulated data for at most five periods, but there is a distribution of the length of the worker's labor market histories.

The last piece of the simulation is constructing earnings by person-year. I draw the person effects from a normal distribution. I draw residuals in an independent way from a normal distribution. Because the model does not generate sorting between firms and workers, these are picked so that the AKM decomposition generates that firms account for $\frac{0.21}{0.9}$ of the variance of earnings, workers are $\frac{0.57}{0.9}$ and the residual is $\frac{0.11}{0.9}$, where I divide by 0.9 because I do not generate (meaningful) covariates in the simulations (i.e., so that the rest of the code will run I generate years, ages, and an indicator for the type of quarter, but then I do not impute any explanatory power to these variables).

I report the results of 100 simulation runs.

Results

Table A1 reports the results of these simulations. The left-hand side of the table shows results from one in 10 sampling, the middle reports results from one in 20 sampling, and the right-hand side of the table shows the results from one in 200 sampling. Panel A reports the (percentage point) gap between the true and estimated values that are relevant for the quantitative bottom-line of the paper, while the italicized rows report the levels of the true values.[62] The basic point to take from Panel A is that the estimation procedure is slightly biased down, but the bias is quantitatively small. Focusing on the one in 10 sampling, the bias down in the raw measure is two percentage points while the bias in the various corrected measures is one percentage point. The basic point to take from comparing the one in 10 sampling and the one in 20 sampling is that the different sampling rate does not have a big effect on the estimates. For example, the median gap between the true and estimated R^2 between V^e and Ψ are both one percentage point. Finally, the one in 200 sampling shows that the bootstrap performs similarly to the split sample approach.

Panel B reports the correlation between the true and estimated values of Ψ, V^e, and $\tilde{V}^{EE}$ across simulation runs. The table allows us to understand the result in Panel A that the procedure is not particularly biased. Specifically, the table shows that given the sample sizes, Ψ and V^e are estimated without that much noise. The correlation between the true and estimated Ψ is 1.00, and the correlation between the true and estimated V^e is also very high, 0.98. In contrast, the table shows that the estimation of the $\tilde{V}^{EE}$ is much noisier. Why is $\tilde{V}^{EE}$ so much noisier than V^e? The basic reason is that estimation of $\tilde{V}^{EE}$ uses much less information than the estimation of V^e: specifically, it only uses information in the EE transitions, whereas V^e also uses information in the ENE transitions (this information is necessary to estimate the value of nonemployment, which is necessary to estimate the offer distribution). Since Table 3 shows that EE transitions account for only 40% of all transitions, it is not surprising that when we move from $\tilde{V}^{EE}$ to V^e that the estimation becomes less noisy.

---

[62]The levels of the true values vary across simulations because the weights are stochastic. To be specific, even though the set of true V^e and Ψ are fixed across simulation runs, the randomness in which firms end up in the connected set generates variation in firm size because I only keep person-years in which the worker reappears in the dataset.

Table A1: Monte Carlo Results

	One in 10 sampling				One in 20 sampling				One in 200 sampling			
Panel A. True minus estimated												
	Mean	50th	10th	90th	Mean	50th	10th	90th	Mean	50th	10th	90th
Raw R^2 of V^e and Ψ	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
Split sample adjusted R^2 of V^e and Ψ	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	-0.01	0.00
Bootstrap adjusted R^2 of V^e and Ψ	NA	NA	NA	NA	NA	NA	NA	NA	0.00	0.00	0.00	0.01
Raw large firm R^2 of V^e and Ψ	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.01
Level of true R^2 of V^e and Ψ	0.26	0.26	0.26	0.26	0.32	0.32	0.32	0.32	0.32	0.32	0.31	0.32
Level of true R^2 of V^e and Ψ at large firms	0.33	0.33	0.33	0.33	0.43	0.43	0.43	0.43	0.55	0.55	0.55	0.55
Firm share of variance of earnings (raw)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firm share of variance of earnings (large firm)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Level of true firm share	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
Level of true firm share (large firm)	0.22	0.22	0.22	0.22	0.24	0.24	0.24	0.24	0.22	0.22	0.22	0.22
Panel B. Correlations between true and estimated values												
Ψ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
V^e	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99
$\tilde{V}^{EE}$	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.84

Notes: This table reports Monte Carlo simulations of the estimation procedure. The table reports statistics across 100 simulation runs. The left-hand side reports simulations where I have drawn one-in-10 random sample; the middle panel reports a one-in-20 random sample and the right-hand side reports one-in-200 random sample. Because the bootstrap is very computationally expensive, I only report the bootstrap results for the one-in-200 sample. Panel A reports statistics on the gap between the true minus the estimated value, while Panel B reports statistics on the correlation between the true and estimated values. The interpretation of a positive number in Panel A is that the estimation procedure underestimates the true relationship. The definitions and procedures are the same as in Table 4.

I Appendix: Other model-consistent approaches to ranking firms

The first approach looks at worker inflows and ranks firms based on the share of hires that are on employer-to-employer transitions. Bagger and Lentz (2016, pg. 21) term this ratio the "poaching index." Formally, the observed poaching index is:

$$PR_j^o = \frac{\sum_{k \in \mathcal{E} \setminus i} M_{kj}^o}{\sum_{k \in \mathcal{E} \cup n \setminus j} M_{jk}^o}. \tag{A92}$$

The model-consistent version focuses on the endogenous flows:

$$PR_j^m = \frac{\sum_{k \in \mathcal{E} \setminus j} M_{kj}}{\sum_{k \in \mathcal{E} \cup n \setminus j} M_{jk}}. \tag{A93}$$

Result 4. *If the value of nonemployment (V^n) is low enough relative to the distribution of the value of employment (V_j^e), then PR_j^m is monotonically increasing in firm value.*

Proof. It is easier to work with a monotone transformation of the poaching index and look at the ratio of hires on employer-to-employer transitions, and consider employer-to-employer flows from a firm to itself:

$$PR_j^m = \frac{\sum_{k \in \mathcal{E}} M_{jk}}{M_{jn}} \tag{A94}$$

$$= \frac{\sum_{k \in \mathcal{E}} g_k W (1-\delta)(1-\rho) \lambda_1 f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V_k^e)}}{\lambda_0 U f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V^n)}} \tag{A95}$$

$$= \frac{\sum_{k \in \mathcal{E}} g_k W (1-\delta)(1-\rho) \lambda_1 \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_k^e)}}{\lambda_0 U} \tag{A96}$$

$$\propto \sum_{k \in \mathcal{E}} g_k \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_j^e)}, \tag{A97}$$

where $\propto$ means "proportional to" and drops all the constant terms. Consider how each term in the sum depends on $exp(V_j^e)$:

$$\frac{\partial}{\partial exp(V_j^e)} \left[\frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_k^e)} \right] = \frac{(exp(V_j^e) + exp(V_k^e)) - (exp(V_j^e) + exp(V^n))}{(exp(V_j^e) + exp(V_k^e))^2} \tag{A98}$$

$$= \frac{exp(V_k^e) - exp(V^n)}{(exp(V_j^e) + exp(V_k^e))^2}. \tag{A99}$$

Aggregate over all k:

$$\frac{\partial}{\partial exp(V_j^e)} \left[\sum_{k \in \mathcal{E}} g_k \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_k^e)} \right] = \sum_{k \in \mathcal{E}} g_k \frac{exp(V_k^e) - exp(V^n)}{(exp(V_j^e) + exp(V_k^e))^2}. \tag{A100}$$

We want conditions on V^n such that this expression is positive for all values of $exp(V_j^e)$ so that PR_j^m is increasing in $exp(V_j^e)$. If $V^n < min_k V_k^e$, then this is always true. Fixing $exp(V_j^e)$, equation

(A100) is monotone decreasing in $exp(V^n)$:

$$\frac{\partial}{\partial exp(V^n)}\left[\sum_{k\in\mathcal{E}}g_k\frac{exp(V_k^e)-exp(V^n)}{(exp(V_j^e)+exp(V_k^e))^2}\right]=\sum_{k\in\mathcal{E}}\frac{-g_k}{(exp(V_j^e)+exp(V_k^e))^2}<0. \tag{A101}$$

Hence, for small enough V^n equation (A100) is positive for all $exp(V_j^e)$ so that PR_j^m is increasing in $exp(V_j^e)$. $\qquad\square$

The intuition of the result is that "better firms hire from better firms," where nonemployment is viewed as an exceptionally bad firm.

The second approach looks at worker outflows and ranks firms based on the separation rate. This approach follows a long tradition in the inter-industry wage differential literature of using a survey-based measure of the quit rate as a measure of desirability (e.g., Ulman (1965, Table III) and Krueger and Summers (1988, Table IX)). The model offers several ways of operationalizing this idea, which hinges on how to interpret the survey response of "quit." One possibility is to interpret this as all EE transitions, which gives rise to the following pair of definitions:

$$QR_j(EE)^o=\frac{\sum_{k\in\mathcal{E}\setminus j}M_{kj}^o}{g_jW};\ \ QR_j(EE)^m=\frac{\sum_{k\in\mathcal{E}\setminus j}M_{kj}}{(1-\delta)(1-\rho)g_jW}. \tag{A102}$$

Alternatively, the quit rate could be interpreted as all separations, which gives rise to the following pair of definitions:

$$QR_j(ALL)^o=\frac{\sum_{k\in\mathcal{E}\cup n\setminus j}M_{kj}^o}{g_jW};\ \ QR_j(ALL)^m=\frac{\sum_{k\in\mathcal{E}\cup n\setminus j}M_{kj}}{(1-\delta)(1-\rho)g_jW}. \tag{A103}$$

Result 5. $QR_j(ALL)^m$ and $QR_j(EE)^m$ are monotonically decreasing in V_j^e.

Proof. **EE quit rate:** The probability of an EE quit is given by (for simplicity, this includes the probability of a worker at firm j quitting to firm j):

$$QR_j(EE)^m=\lambda_1\sum_k f_k\frac{exp(V_k^e)}{exp(V_k^e)+exp(V_j^e)}. \tag{A104}$$

Taking the derivative with respect to $exp(V_j^e)$:

$$\frac{\partial QR_j(EE)^m}{\partial exp(V_j^e)}=\lambda_1\sum_{k\in\mathcal{E}}f_k\frac{-exp(V_k^e)}{(exp(V_k^e)+exp(V_j^e))^2}<0. \tag{A105}$$

Hence, $QR_j(EE)^m$ is decreasing in $exp(V_j^e)$.
EN quit rate: The probability of an EN quit is given by:

$$QR_j(EN)^m=(1-\lambda_1)\frac{exp(V^n)}{exp(V^n)+exp(V_j^e)}. \tag{A106}$$

Taking the derivative with respect to $exp(V_j^e)$:

$$\frac{\partial QR_j(EN)^m}{\partial exp(V_j^e)}=(1-\lambda_1)\frac{-exp(V^n)}{(exp(V^n)+exp(V_j^e))^2}<0. \tag{A107}$$

Hence, $QR_j(EN)^m$ is decreasing in $exp(V_j^e)$.

All quit rate: This result follows form combining the previous two results. $\qquad\square$

The intuition of the result is that because workers at all firms face the same offer distribution and, in expectation, value all firms the same way, the probability of choosing to leave is decreasing in the quality of the firm.

J   Appendix: Inverting the value function

Following Hotz and Miller (1993), take advantage of two properties of Type I extreme value errors,[63] and to keep notation compact, use $p_{jk} = \Pr(j \succ k) = \frac{exp(V_j^e)}{exp(V_j^e) + exp(V_k^e)}$. Rearranging equation (9):

$$
\begin{aligned}
v_j = {} & V_j^e - \beta \mathbb{E}\{\delta\{V^n + \gamma\} + \rho_j(1 - \delta_j) \int_k \{V_k^e + \gamma\}\tilde{f} \\
& + (1 - \rho_j)(1 - \delta_j) \times [(1 - \lambda_1)\{p_{nj}(V^n + \gamma - \ln p_{nj}) + p_{jn}(V_j^e + \gamma - \ln p_{jn})\} \\
& + \lambda_1 \int_k \{p_{kj}(V_k^e + \gamma - \ln p_{kj}) + p_{jk}(V_j^e + \gamma - \ln p_{jk})\}f]\}.
\end{aligned}
\tag{A108}
$$

Solving this equation requires two objects that are not required to solve for V_j^e: β and $\tilde{f}$. I set $\beta = 0.95$ (reflecting the annual frequency of the model), and I set $\tilde{f} = f$. If there were no variation in δ_j and ρ_j across j, then v_j is just a monotone transformation of V_j^e since the first two terms do not vary in V_j^e and the second two terms are monotone increasing in V_j^e. With variation in δ_j and ρ_j, then this equivalence breaks. Nonetheless, the correlation between V_j^e and v_j is 0.937. Similarly, the correlation between v_j and Ψ_j is 0.573 (because it is computationally expensive to compute v_j, I did not compute this quantity in each bootstrap repetition and so this should be compared to the "raw" correlation between V_j^e and Ψ_j, which is 0.530.)

K   Appendix: Additional tables and figures

Table A2: Constructing Sample of Dominant Jobs

	Number	Unique People	Unique Employers
	(1)	(2)	(3)
Person-employer-year pre-earnings test	$650,288,000$	$108,002,000$	$6,688,000$
Person-employer-year post-earnings test	$613,341,000$	$105,921,000$	$6,511,000$
Person-years	$504,945,000$	$105,921,000$	$6,155,000$

Notes: All counts are rounded to the nearest thousand. Row 2 divided by row 3 is 1.215. The first row shows the total number of person-year-employer observations that are continuous quarter or full-quarter among workers in the relevant age range. The second row shows the number of person-year-employer observations where the person's dominant job in the particular year passes an earnings test. The third row goes down to the unique employer that provides the worker's "dominant" job, or the employer from which the worker made the most in the calendar year.

---

[63]First, $\gamma = \mathbb{E}[\iota] \approx 0.577...$ is Euler's constant. Second, the conditional expectation of ι is related to the choice probability in the following way, where the two choices are a and b: $\mathbb{E}[\iota_a|a] = \gamma - \ln(\Pr(a \succ b))$.

Table A3: Distribution of jobs per person per year

	Number of person-years
1	$413,553,000$
2	$77,735,000$
3	$11,611,000$
4+	$2,047,000$

Notes: All counts are rounded to the nearest thousand. This table deconstructs the gap between row 2 and row 3 in Table A2. The column sum is row 3 in Table A2. This shows among workers in the sample of workers with dominant jobs the distribution of the number of continuous and full-quarter jobs in a year.

Table A4: Type of earnings in the annual dominant job dataset

Type of earnings	Number of person-years
Full quarter	$458,017,000$
Continuous quarter	$46,928,000$
Continuous quarter share	0.093

Notes: All counts are rounded to the nearest thousand. The column sum is the number of person-years in row 3 in Table A2. A worker is employed full-quarter in quarter t if she has earnings from her employer in quarter t and quarters $t-1$ and $t+1$. A worker is employed in a continuous quarter way in quarter t if she has earnings from her employer in quarter t and quarter $t-1$ or quarter $t+1$.

Table A5: Number of years per person

	Number of people	Share
1	14,041,000	0.133
2	11,422,000	0.108
3	9,873,000	0.093
4	9,111,000	0.086
5	8,963,000	0.085
6	10,396,000	0.098
7	42,115,000	0.398

Notes: All counts are rounded to the nearest thousand. The column sum is the number of unique people in row 3 in Table A2.

Table A6: Dominant employers per person

Number of dominant employers	Number of people	Share of people
1	52,938,000	0.500
2	27,228,000	0.257
3	14,945,000	0.141
4	7,157,000	0.068
5	2,764,000	0.026
6	771,000	0.007
7	118,000	0.001

Notes: All counts are rounded to the nearest thousand. The column sum is the number of unique people in row 3 in Table A2.

Table A7: Number of years per match

Years per match	Matches (person-employers)	Share of matches	Share of person-years
1	93,327,000	0.466	0.185
2	39,176,000	0.196	0.155
3	19,842,000	0.099	0.118
4	12,295,000	0.061	0.097
5	8,573,000	0.043	0.085
6	6,745,000	0.034	0.080
7	20,175,000	0.101	0.280

Notes: All counts are rounded to the nearest thousand. The column sum in the first column is the number of matches and is approximately 200,000,000, which is between the number of unique people and the number of person-years. The next column shows the distribution by share of matches. The last column shows the distribution of person-years.

Table A8: Composition of separations in the quarterly dataset

Type of transition	Definition	Number
employer-to-nonemployment	Standard	131,621,000
employer-to-employer	Standard	76,152,000
employer-to-employer	New	2,680,000
employer-to-employer transition share		0.375
New definition share		0.035
Total separations		210,453,000

Notes: All counts are rounded to the nearest thousand. The dataset is the quarterly dataset, so it includes some workers not in the annual dataset. The standard definition uses overlapping quarters to measure employer-to-employer transitions. The new definition uses stability of earnings to measure employer-to-employer transitions.

Table A9: Summary statistics and the variance of earnings with the selection-correction

	S. Connected by EE
Sample size	
People-years	$409,550,000$
People	$90,895,000$
Employers	$476,000$
Summary statistics	
Mean log earnings	10.48
Variance of log earnings	0.67
Share of variance of earnings explained by each parameter set	
Employers	0.21
People	0.57
Xb	0.11
Selection-correction	0.00
Variance components	
Variance of emp. effect	0.14
Variance of person effect	0.50
Variance of Xb	0.07
Variance of selection-correction	0.00
2cov(person, emp.)	0.10
2cov(Xb, person + emp.)	0.08
2cov(selection-correction,person+emp.)	0.00
2cov(selection-correction, Xb)	0.00
Corr(person, emp.)	0.19
Overall fit of AKM decomposition	
Adj. R^2	0.86
Match effects model	
Adj. R^2	0.92

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). EE is employer-to-employer. The sample is the same as column (3) of Table 1.

Table A10: Hours and compensating differentials

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Variation in hours across sectors							
Agriculture	N/A	N/A	44.7	44.20	45.60	46.40	46.70
Mining	49.04	49.20	49.20	41.00	41.40	41.50	41.60
Construction	40.70	40.90	40.90	41.10	41.40	41.50	41.70
Manufacturing	42.36	42.30	42.30	42.20	42.60	42.60	42.80
Wholesale	38.22	38.10	42.60	38.60	39.00	39.70	40.40
Retail	38.22	38.10	36.90	37.80	38.20	39.00	39.60
Transport and Warehousing	42.20	42.30	42.20	42.30	42.80	42.80	42.90
Utilities	42.20	42.30	42.30	40.70	41.00	41.40	41.70
Information	39.88	39.60	39.60	40.20	40.50	40.80	41.00
FinInsurance	40.24	40.00	40.40	40.10	40.50	40.60	40.70
Real Estate	40.24	40.00	39.10	40.80	41.40	41.50	41.80
ProfSciTech Services	40.16	40.00	41.20	40.20	40.70	40.80	41.20
Management	40.16	40.00	42.70	37.10	37.60	37.80	38.40
Admin and Waste	40.16	40.00	38.20	37.50	37.80	38.00	38.50
Education	37.38	37.30	36.90	37.50	37.70	38.00	38.40
Health and Social	37.38	37.30	37.50	34.60	34.40	36.20	37.30
Arts and Rec	34.36	34.20	34.80	34.00	33.80	35.60	36.60
Accom and Food Serv	34.36	34.20	34.00	36.60	37.20	37.70	38.40
Other Services	36.88	36.70	36.70	40.00	40.50	40.70	41.00
Public Administration	40.84	40.70	40.70	40.20	40.30	40.40	40.60
N	625741	625739	681264	654715	467121	459650	445962
Panel B. Relationship between hours and compensating differentials							
R^2	0.323	0.325	0.307	0.149	0.158	0.160	0.169

Notes: Panel A report average hours worked last week across sector pooling data from 2003-2007. Columns (1) through (3) uses the monthly CPS. Column (1) reproduces the weighted average of the published BLS Annual Average tables (Table 21). The BLS tables are more aggregated than the sector level, and so the numbers are copied where two sectors are joined (for example, "education" and "health and social" are aggregated). Column (2) replicates column (1) using the micro-data. Column (3) disaggregates the sectors that were combined in columns (1) and (2), and adds in agricultural workers. Columns (4) through (7) use the March CPS. Column (4) imposes the same sample restrictions as in columns (2) and (3). Column (5) adds sample restrictions following Autor, Katz, and Kearney (2008) in terms of dropping imputed observations and observations with extreme earnings. Column (6) restricts to men and women aged 18-61. Finally, column (7) imposes an earnings floor of $3250 a year (in $2011). Panel B reports the relationship between estimated sectoral compensating differentials reported in Figure 6b and measures of the variation in hours worked across sectors reported in each column in Panel A. The R^2 comes from a regression of the sectoral compensating differentials on the the hours reported in the relevant column, where the observations are weighted by the sectoral sum of the relevant person weights.
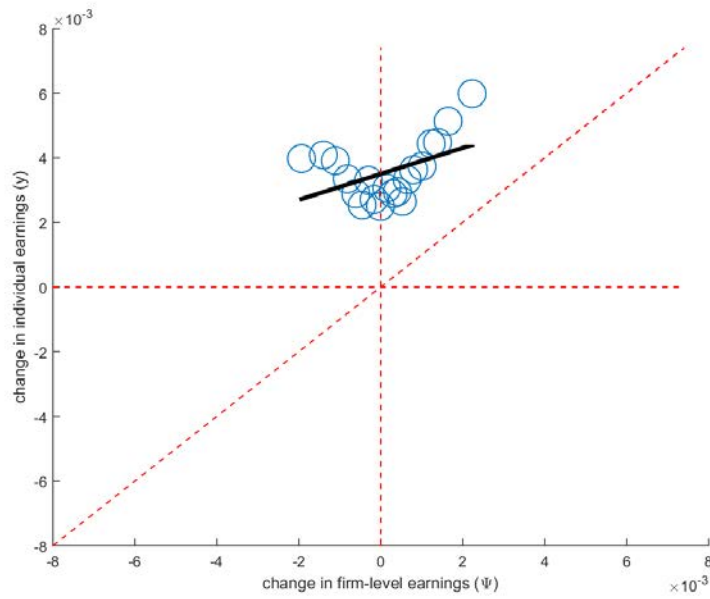
Figure A1: States used in analysis
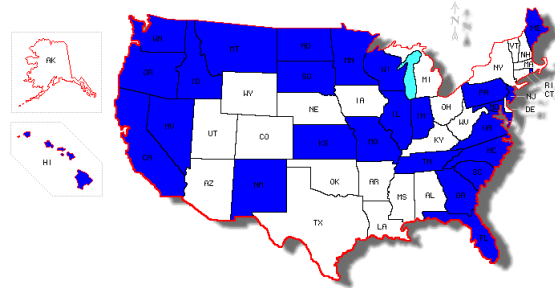
Notes: The states in blue are used in the analysis.

Figure A2: Change in firm effect does not predict magnitude of earnings change in a matching model

Notes: This figure is based on simulating the example production function in Eeckhout and Kircher (2011) and is constructed in a manner analogous to Figure 2.

Figure A3: Event studies of earnings changes
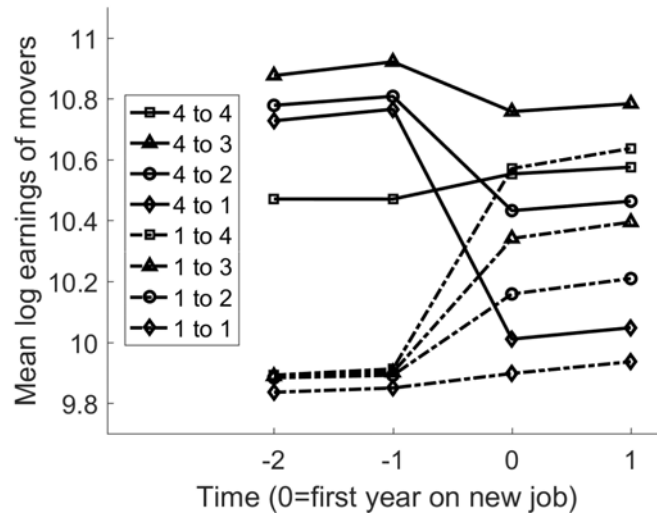
Notes: This figure shows the mean wages of workers who change jobs and held the preceeding job for two or more years, and the new job for two or more years. "Job" refers to the dominant annual job. Each job is classified into quartiles based on the estimated firm effects in Table 1 column (3).