The U-Shapes of Occupational Mobility*

Fane Groes
Copenhagen Business School

Philipp Kircher
London School of Economics and
University of Pennsylvania

Iourii Manovskii
University of Pennsylvania

Abstract

Using administrative panel data on the entire Danish population we document a new set of facts characterizing occupational mobility. For most occupations, mobility is U-shaped and directional: both low and high wage earners within an occupation have a particularly large probability of leaving this occupation, and the low (high) earners tend to switch to new occupations with lower (higher) average wages. Exceptions are occupations with steeply rising (declining) productivity, where mainly the lower (higher) paid workers within this occupation tend to leave. The facts conflict with several existing theories that are used to account for endogeneity in occupational choice, but it is shown analytically that the patterns are explained consistently within a theory of sorting under absolute advantage that includes learning about workers' abilities.

JEL Classification: E24, E25, J24, J31

Keywords: Occupational mobility, Learning, Labor markets, Sorting, Comparative and Absolute Advantage

---

*First version: June 2008. This version: August 2010. We would like to thank Martin Browning, Mette Ejrnæs, Pieter Gautier, Jonathan Heathcote, Hans Christian Konsted, Dale Mortensen, Giuseppe Moscarini, Kjetil Storesletten, and seminar participants at the University of Copenhagen, Copenhagen Business School, University of Essex, University of Maryland, the Minneapolis Fed, University of Oxford, University of Southampton, the 2008 Copenhagen Microeconometrics Summer Workshop, the Search & Matching workshop at the University of Pennsylvania, 2009 NBER Summer Institute, 2009 Society for Economic Dynamics annual meeting, 2009 Macroeconomics across Time and Space conference, 2009 Econometric Society European meeting, and 2010 Lincoln College Applied Microeconometrics conference for their comments. This research has been supported by the National Science Foundation Grants No. SES-0922406 and No. SES-0752076.

1 Introduction

Danish employers report that every year close to a fifth of the their workers change occupations (e.g., technician, engineer, manager). Similar levels of occupational mobility are reported for the US.[1] Moreover, these gross flows are much larger than the net flows that are needed to account for the changing sizes of occupations. What induces workers to undertake these occupational changes? The answer to this question seems especially interesting because occupational choices and wages are closely related. First, the differences in average occupational wages are substantial and persistent. Second, it has recently been argued that the returns to occupational tenure are nearly as large as the returns to labor market experience and much larger than the returns to firm or industry tenure.[2] Thus, understanding workers' occupational choices is important for understanding the allocation of the labor force across productive activities, for interpreting earning patterns, for measuring the returns to human capital accumulation, and for assessing the effects of various policies affecting sorting of workers across occupations. Since workers choose occupations endogenously, the outcome of such analysis will depend on the theory used to account for selection of workers across occupations. While there exist a number of theories of occupational choice, it remains an open empirical question which selection process is consistent with the data.

This paper contributes to our understanding of selection in occupational choices by looking at occupational mobility data in a novel way. Using administrative data on 100% of the Danish workforce we provide new direct evidence on patterns of worker mobility across occupations. This evidence conflicts with several existing theories that are often used to account for the endogeneity in occupational choice, but we can show analytically that the patterns are explained consistently within a theory of vertical occupational mobility combined with learning about worker ability.

We document that for most occupations, mobility is U-shaped and directional: it is both the low wage and the high wage workers within an occupation who have a particularly large probability of leaving that occupation, while the lowest probability of leaving is associated with the medium wage workers within the occupation. More than three-quarters of the labor force are employed in occupations exhibiting this pattern. While switching probabilities are particularly high at both ends of the wage spectrum within an occupation, the direction of sorting is very different for high and low wage earners. Those earning low wages relative to other workers in the same occupation tend to leave for new occupations that on average pay less to their workforce than the old occupation, while those with high relative wages in their occupation tend to leave

[1] Kambourov and Manovskii (2008), Moscarini and Thomsson (2007).

[2] Shaw (1984, 1987), Kambourov and Manovskii (2009b), Sullivan (2009), Groes (2010), Zangelidis (2008).

for occupations that on average pay more to their workforce. These patterns remain whether we focus on workers who stay with the same firm or on those who switch firms and across various ways of defining occupations. The U-shaped pattern where low and high wage earners in an occupation are most likely to leave it is predominant except for occupations with steeply rising (declining) productivity, where mainly the lower (higher) paid workers within this occupation tend to leave.

We are able to document these patterns because the data allows us to compare the behavior of workers relative to that of other workers with similar characteristics in the same occupation. Such analysis has been missing in the literature partly because most longitudinal datasets that have traditionally been analyzed feature only panels of a few thousand workers, and with around three hundred occupations, an analysis on a per occupation basis was not feasible. This new look at the data has at least two important direct implications: First, selection is not just one-sided. In particular, the well documented wage growth with tenure in an occupation is not just due to low wage earners leaving and high wage earners staying (wage growth based on composition). In fact, a large number of high wage earners are leaving their occupations as well, and models generating the wage implications based on worker selection need to take this into account. Second, occupations with strong productivity growth nevertheless shed a large fraction of their workforce, a stark feature of the data not featured by the commonly used models.

A number of prominent models of occupational choice feature counter-factual one-sided selection, typically with relatively low wage earners leaving the occupation while high wage earners stay. One popular class of such models is based on horizontal sorting due to match-specific shocks. Originating from Jovanovic (1979) and extended to occupational mobility by, e.g., McCall (1990) and Neal (1999), this work is based on the idea that occupations are identical (e.g., not different with respect to skill requirements), but workers find out the quality of their idiosyncratic match with an occupation over time. Horizontal re-sorting occurs when workers realize that their match-specific productivity is low and abandon the match in favor of (the search for) a better one. Thus, the model predicts that workers with low wages (low quality matches) leave the occupation, and their next occupational choice is a random draw. Both predictions do not match up with our findings in the data. Similarly, island economy models based on human capital extensions of Lucas and Prescott (1974), such as Kambourov and Manovskii (2005) and Alvarez and Shimer (2009), typically predict that it is the the low human capital and hence, low wage, workers who are the first to switch if occupational demand declines since high human capital workers have more incentive to wait for the conditions to improve. If occupational demand rises, no one leaves the occupation. The wage a switcher obtains in the new occupation is independent of her relative wage in the previous occupation. Once again, these implications do not match up with the patterns we find in the data.

The main shortcoming of these models relative to our findings in the data is that high wage earners have no-where better to go to, and therefore stay in their occupation. This is different in the classic Roy (1951) model with absolute advantage: High wage earners in a low productivity occupation leave if the productivity of their current occupation drops further relative to the productivity in the next better occupation. Yet, standard versions of the Roy (1951) model have the drawback that they only feature net occupational mobility due to shocks to occupations, and an occupation with rising productivity or demand does not shed any of its workers. Moreover, in most specifications only workers in one side of the occupational wage spectrum (either high or low earners) switch occupations, which holds even in models that introduce gross mobility through search frictions (e.g. Moscarini (2001)).[3]

Nevertheless, we show in the theoretical part of this paper that a tractable combination of the previous models − learning combined with a notion of absolute advantage − does account well for all of the qualitative patterns that we find in the data. The theoretical work is based on an equilibrium extension of the decision-theoretic model in Gibbons and Waldman (1999) that was successfully used to understand promotion dynamics within firms. Our model features the combination of sorting similar to the standard Roy (1951) model with absolute advantage, and learning about one's own abilities. As workers learn that the current occupation is not a good fit, it is not optimal to sample randomly from the remaining occupations, but to use the prior information in their new choice. If they did particularly well, workers find it optimal to move up to a more demanding occupation, if they did particularly poorly, they are more suitable for a less demanding occupation. They will be better-off moving there rather than taking large wage cuts in their current occupation. Our theory abstracts from an explicit notion of firms since in the data the pattern of occupational switching is similar for workers who stay with the same firm as for those who switch firms.[4]

[3]The most popular specifications consider absolute and relative advantage across two occupations. For this setting we discuss the predictions in more detail in Section 5.1.

[4]Our analysis suggests that firm affiliation might not be of first importance to understand the basic patterns of occupational mobility. Nevertheless, an extension of our model in which workers maximize the quality of their occupational fit but also receive idiosyncratic shocks that affect the performance at each individual firm, is likely to generate similar features as those outlined in Neal (1999) in his horizontal learning model despite the different predictions for wage patterns: Workers would first choose the appropriate occupation and subsequently look for the appropriate firm with an employment opportunity in such an occupation. Also, the presence of any labor market frictions makes firms with employment in more occupations more attractive because the internal labor market can facilitate transitions between these occupations (see Papageorgiou (2010) for such a model, yet with horizontal sorting which does not generate the U-shaped connection between wage and occupational choice in the internal labor market that we document).

In addition to accounting for the qualitative U-shaped mobility patterns and the direction of switching, this theory also has secondary implications that conform well with the data. Considering occupations with roughly constant productivities, the theory predicts that workers who switch to occupations with higher average wages see faster wage growth than workers who stay, who, in turn, see faster wage growth than workers who move to occupations with lower average wages. In terms of wage levels, those who switch to an occupation with higher average wages do better than those who remain in the old occupation, but worse than those who already work in the new occupation. The opposite holds for workers that move occupations with lower average wages. Finally, the equilibrium nature of the model implies that occupations with sharply increasing productivity will retain their high earners but shed their low earners (who are driven out by an inflow of higher qualified workers), and the opposite holds for occupations with a substantial decline in productivity. These predictions are confirmed in the data. The last prediction, in particular, requires an equilibrium model and is not a feature of decision-theoretic models like the Roy model where an occupation with rising productivity attracts more workers but does not shed any.

The idea to link sorting in vertical hierarchies to learning about one's own ability has received less attention than horizontal sorting due to match-specific learning, but nevertheless has a long tradition (see, e.g., Johnson (1978), Miller (1984), Gibbons and Katz (1992), Biddle and Roberts (1994), Jovanovic and Nyarko (1997), Papageorgiou (2007), and Eeckhout and Weng (2009)). However, most existing work has restricted the exposition to only two occupations. This is partly due to the focus on different speed of learning in different occupations, which tends to significantly increase the difficulty of the analysis. For our purposes it is crucial to extend the analysis to more than two occupations because U-shapes arise only in intermediate occupations: in the highest-ranked occupation top earners have no-where else to go to, and in the lowest-ranked occupation low-wage workers have no-where else to go to. To consider several occupations in a parsimonious model we abstract from differences in the speed of learning.

In the next section we present in detail the empirical findings that switching is U-shaped and directional, the main indicators that suggest the use of the comparative advantage model that we outline in Section 3. In the initial theoretical part we stay deliberately simple in order to present a theory on the same level of tractability as existing work on learning under horizontal sorting. We show that the model conforms well with the basic facts and its additional predictions also match up with the data. We should mention that, as in all models of learning mentioned here, the patterns of mobility induced by learning about workers' ability can also be generated by a particular exogenous process for ability itself and no learning. The main natural feature of a learning model is that workers become more convinced of their ability over time (the variance of "ability realizations" decreases over time), and we show in Section 3.3 that this naturally explains some of the prominent additional patterns in the data. While this is suggestive of a learning model, a more agnostic specification based on ability changes with decreasing variance would have the same explanatory power (as is the case in all of the related literature).

While the initial model is stationary and occupational productivities are constant over time, in Section 4.1 we extend the model to allow for occupations with time-varying productivities. We show that occupations with strong productivity growth retain their high wage (high ability) workers but shed their low wage (low ability) workers whose skills are no longer commensurate. Occupations experiencing a substantial decline in productivity loose their high ability workers who find it more worthwhile to work elsewhere and retain their low ability (low wage) workers. We also document such patterns in the data.

In Section 4.2 we extend the stationary environment to allow for human capital accumulation. In the basic model workers switch to better occupations only when they learn that they have higher ability than expected. The introduction of general human capital implies that workers move up the occupational career ladder also because they become more skilled over time. In its extreme, if we shut down learning, this channel would still generate occupational switching, but only to better occupations (for a related model with a similar career-ladder feature, see Sichernam and Galor (1990)). While such human capital accumulation is important to understand why we find somewhat more upward mobility in the data than downward mobility, the fact that a substantial fraction of occupational switches (even within a firm) is downward suggests that in a substantial number of cases human capital accumulation is outweighed by some other force, such as gradual learning about the capabilities of the worker, which is the main feature of our baseline model. We also extend the model to take into account that even within vertical occupational hierarchies a switch may require a new set of skills which induces costs to occupational switching. For example, engineers that move up to manage small groups need to adjust their human resource skills, while those that move down to become technicians need to adjust their applied skills.[5] We extend our analysis to allow for general and occupation-specific human capital accumulation as well as retraining costs, and show that the predictions of the model remain robust.

After laying out our model, in Section 5 we return to a more in-depth discussion of some related literature. In particular, we discuss in more detail similarities and differences to the

[5]We will show that U-shaped mobility patterns and directional sorting remain when we omit the occupational categories of managers, and even when we combine all occupations with relatively similar descriptions into one. This suggests that retraining and skill costs are indeed present, since the descriptions of the remaining occupations differ substantially.

standard Roy (1951) model. We also discuss the connection to the work by Gibbons, Katz, Lemieux, and Parent (2005), who propose an empirical strategy to estimate parameters of a closely related model based on using lagged occupational choices as instruments. In their decision-theoretic work they specify a somewhat different functional form for the workers' choice than the one that comes out of our equilibrium specification, yet we highlight the close connection which suggests that their instruments might be useful in future empirical work with our model.

Of course, we do not think that the simple vertical sorting mechanism that we propose accounts for the full extent of occupational mobility. In the Conclusion we discuss the broader research agenda, and the challenges to empirically assess the exact quantitative implications of the patterns presented in this paper. Both vertical and horizontal moves likely arise in the labor market, i.e., some occupations are considered better than others while some are just different and people switch along both of these dimensions. And among those occupations that can be ranked as better or worse, the ranking might change over time. Therefore, it is likely that match-specific components and the volatility of productivities of occupations or of the demands for their services are responsible for a nontrivial share of mobility. An important part of a future agenda is to identify which occupations form vertical hierarchies in order to identify the costs of switching within and across hierarchies. Our analysis suggests that many of the occupational switches do arise within hierarchies. Therefore, we do think that the mechanism we emphasize should be an important part of any comprehensive theory of occupational mobility.

2 The U-shapes of Occupational Mobility: Evidence

2.1 Data

We use the administrative Danish register data covering 100% of the population in the years 1980 to 2002. The first part of the data is from the Integrated Database for Labor Market Research (IDA), which contains annual information on socioeconomic variables (e.g., age, gender, education, etc.) and characteristics of employment (e.g., private sector or government, occupations, industries, etc.) of the population. Information on wages is extracted from the Income Registers and consists of the hourly wage in the job held in the last week in November of each year. Wage information is not available for workers who are not employed in the last week of November. The wages are deflated to the 1995 wage level using Statistics Denmark's consumer price index and trimmed from above and below at the 0.99 and 0.01 percentile for each year of the selected samples described below.

We use the Danish rather than the U.S. data for two reasons. First, the sample size is much larger. Our objective is to document the patterns of occupational mobility depending on the position of the individual in the wage distribution within her occupation. A sample sufficiently large to be representative *in each occupation* is essential for this purpose. Second, the administrative data minimizes the amount of measurement error in occupational coding that plagues the available U.S. data (see Kambourov and Manovskii (2009b)). Nevertheless, we find that the features of occupational mobility that can be compared between the U.S. and Denmark are quite similar.[6] This leads us to expect that the patterns of occupational mobility that we describe using Danish data generalize to, e.g., the U.S.

As is standard in the literature, the hourly wage variable is calculated as the sum of total labor market income and mandatory pension fund payments of the job held in the last week in November of a given year divided by the total number of hours worked in the job held in November of that year. The labor income and the pension contributions are from the tax authorities and are considered to be highly reliable. Wage structure is potentially affected by the presence of centralized wage bargaining in Denmark (see Dahl, le Maire, and Munch (2009) for a detailed description of the system). However, only around 13% of workers are covered by industry-wide bargaining where wages cannot be modified at the firm level. In other cases wages are bargained at the firm level, potentially subject to the lower bound on wages of the very inexperienced workers set at the industry level.

Occupational affiliation is defined by the so-called DISCO code, which is the Danish version of the ISCO-88 classification (International Standard Classification of Occupations).[7] The validity of the codes is considered to be high, in particular, because they are monitored by employers and unions and form the basis of wage bargaining at the national level. We use the most disaggregated definition of the occupational classification available, i.e., the 4-digit code. This classification corresponds fairly closely to the 3-digit Standard Occupational Classification used by the U.S. Census. We perform our analysis at this level of aggregation because it appears to better match the characteristics of the tasks performed by the workers than more aggregated classifications. For example, the following pairs of occupations have distinct 4-digit codes but the same 3-digit ones: economists and foreign language translators, hair-dressers and undertakers, radio-announcers and circus clowns, plumbers and electricians, etc. Moreover, the main variable used in our analysis is the position of the worker in the wage distribution of his occupation. This is affected by the coarseness of the classification used. For example, only 28% of economists

[6]For example, we find that the level of occupational mobility in Denmark is similar to the estimates of mobility in the U.S. that account for the coding error. Groes (2010) documents that the relationship between occupational tenure and wages in Denmark is similar to that found in the U.S. She also reports that the hazard rates of leaving an occupation in Denmark are similar to those estimated for the U.S.

[7]ISCO-88 codes are described at $http://www.ilo.org/public/english/bureau/stat/isco/isco88/major.htm.$

in the lowest decile of the economists' wage distribution are in the lowest decile of their 3-digit occupational group. Similarly, some workers in the lowest decile of the wage distribution of chemical engineers are in the 7^{th} decile of the wage distribution of their 3-digit occupation. These arguments notwithstanding, however, we will show that all of the results reported below are qualitatively similar when the analysis is performed at the 1-, 2-, and 3-digit levels.

2.1.1 Sample Selection

While the Danish register data dates back to 1980, because information on firm tenure is available only after 1995 and because of a change in the occupational classification in 1995, we study the data spanning the 1995-2002 period (the latter cut-off was dictated by the data availability at the time we performed the analysis). We use the pre-1995 data in constructing some of the variables. For example, in 1995 the two occupational classifications used in the Danish register data are linked to the worker's job which allows us to construct measures of occupational tenure. For example, a worker will be considered to have 5 years of occupational experience in 1996 if he is observed in the same occupation in 1995 and 1996 according to the new occupational classification and at the same time has the same occupation from 1992 to 1995 according to the old occupational classification.

We only select male workers in order to minimize the impact of fertility decision on labor market transitions. The sample is restricted to employees because we do not observe earnings for the self-employed. Since we study occupational mobility between consecutive years, the sample only includes workers with valid occupation data in the year after we use them in the analysis (e.g., we use information from 2002 for this purpose). To construct experience and tenure variables we need to observe each individual's entire labor market history. Thus, our sample includes all individuals completing their education in or after 1980 if they remain in the sample at least until 1995. The sample includes graduates from all types of education from 7th grade to a graduate degree conditional on observing the individual not going back to school for at least three years after graduation. Thus, a worker who completed high school, worked for three years, then obtained a college degree and went back to full-time work will have two spells in our sample: first, the three years between high school and college, and second, after graduating from college. If he worked for less than three years between high school and college, he joins our sample only after graduating from college.

We conduct our analysis using two samples that differ in additional restrictions that we impose. We label these samples a *Small Sample* and a *Large Sample*. Their construction is as follows.

Our overriding concern in constructing the Small Sample is the reliability and consistency of the data. This sample is restricted to full time workers in the private sector. The restriction to private-sector workers is due to the concern that wage setting and mobility patterns in the government sector may be partially affected by non-market considerations. Moreover, in the period 1995 to 1998 we do not observe the workplace of public employees, which makes it difficult to condition on employer tenure if these workers are included in the sample. Part-time workers are excluded because they do not have as dependable wage information and do not have any occupational codes. We truncate workers' labor market histories the first time we observe them in part-time employment, public employment, self-employment, or at the first observation with missing wage data or missing firm or occupational codes.[8] In order to have the same distribution of experience in the period 1995 to 2002 we truncate worker histories 15 years after graduation.

Our main objective in constructing the Large Sample is to maximize the size of the sample. Consequently, it is much less restrictive. It includes public-sector workers and includes workers who have spells of part-time work and non-employment.[9] It also includes workers who re-enter the sample after having a missing firm, industry, or occupational spell.[10]

Descriptive statistics of the main samples used in the analysis are provided in Appendix Table A-1. The results reported in the body of the paper are mainly based on the Small Sample that contains approximately $400,000$ observations. The results based on the Large Sample that includes approximately 1.3 million observations are reported in Appendix A5.3. We have also verified that all the results hold for the "intermediate" samples that impose some but not all of the restrictions of the Small Sample.

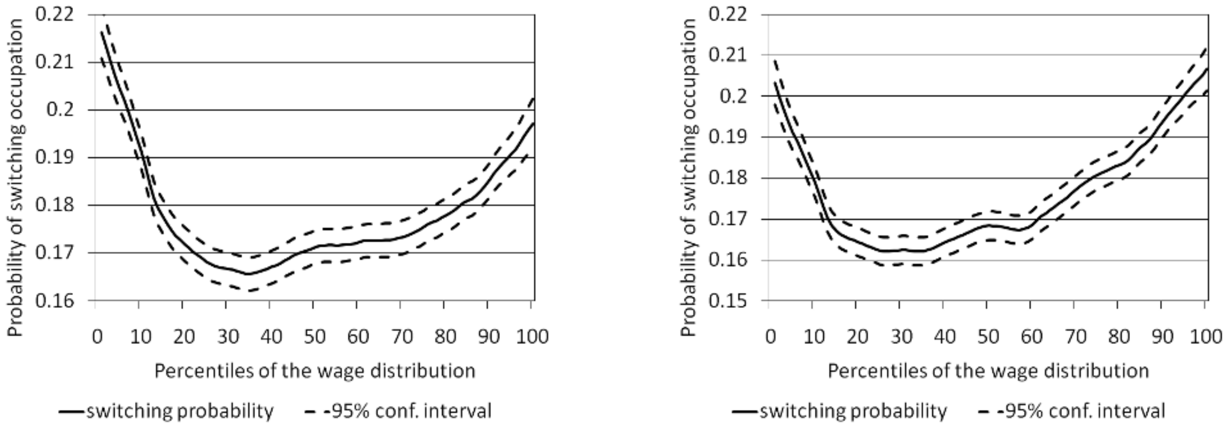
2.2 U-shapes in the Probability of Occupational Switching

In this section we present evidence of U-shapes in the probability of occupational switching. For each worker that we observe in a given year of our sample, we compare his wage to the wages of the other workers *in the same occupation* in the same year. This gives us this worker's rank in the wage distribution of his occupation. That is, it gives the fraction of other workers in the

[8]Workers are allowed to be either unemployed or out of the labor force up to two years after graduation without being dropped from the sample.

[9]We treat part time work as non-employment.

[10]We exclude observations with missing occupational or firm affiliation data. After an observation with missing occupation (firm) affiliation we cannot reliably calculate occupational (firm) tenure until the worker is observed switching occupations (firms). Upon an occupational (firm) switch the corresponding tenure is set to zero and from that point on the observations are included into the sample. For example, a worker who is a cook in period t, has missing occupation in period $t+1$, is a cook in period $t+2$, and a truck driver in period $t+3$, will be included in the sample in period t and again in period $t+3$ − the two observations with reliable occupational tenure information.



(a) Distribution of raw wages within occupation and year.

(b) Distribution of wage residuals.

Figure 1: Non-parametric plot of probability of switching occupation by worker's percentile in the relevant wage distribution.

same occupation that earn lower wages than him this year. We plot the probability of switching to a new occupation in the following year against this rank. Figure 1(a) is a non-parametric plot (from a kernel smoothed local linear regression with bandwidth of 5 percentiles) of the probability of switching out of an occupation as a function of a worker's position in the wage distribution in that occupation in a given year.[11] The probability of switching occupation is clearly U-shaped in wages. It is the workers with the highest or lowest wages in their occupations who have the highest probability of leaving the occupation. The workers in the middle of the wage distribution of their occupation have the lowest probability of switching occupations.

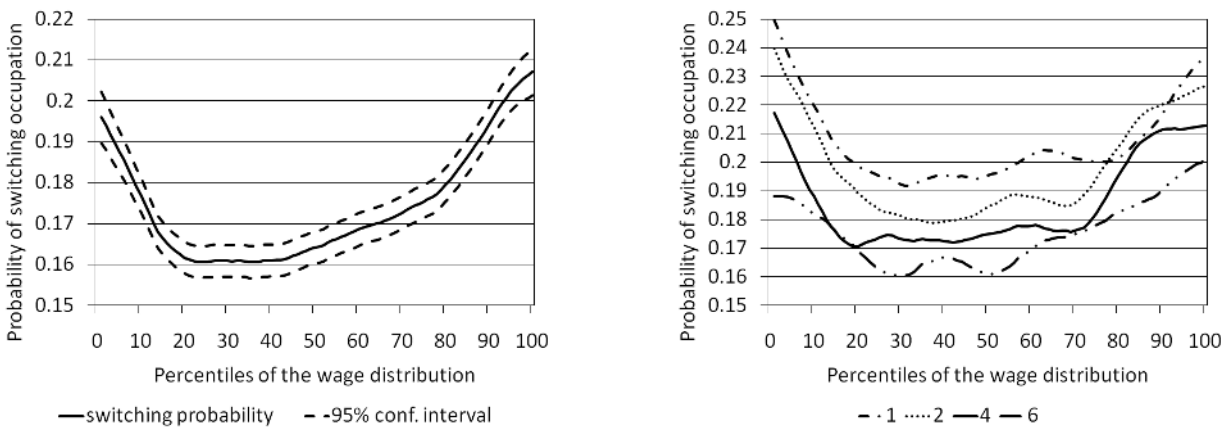
Figure 1(a) is based on raw wage data. Figure 1(b) indicates that we also observe a U-shaped pattern of occupational mobility in the position of the worker in the distribution of residual wages in his occupation in a given year. We generate residual wages by estimating a standard wage regression

$$ln\, w_{ijt} = X_{ijt}\beta + \epsilon_{ijt}, \tag{1}$$

where w_{ijt} is real hourly wage of an individual i working in occupation j in period t. The explanatory variables in X include calendar year dummies, third degree polynomials in general experience, occupational tenure, industry tenure, a second degree polynomial in firm tenure, the sequence number of occupational spell, education, marital status, union membership, and regional dummies. These wage regressions are estimated separately for each occupation.[12]

[11]The occupations are restricted to include a minimum of ten workers per year in order to find the percentiles of the wage distribution within an occupation. The confidence intervals are pointwise estimates from 60% of the sample.

[12]Figure A-1 in Appendix A5 shows that excluding firm and industry tenure or dummies for the sequence



(a) Overall.

(b) For different years after graduation.

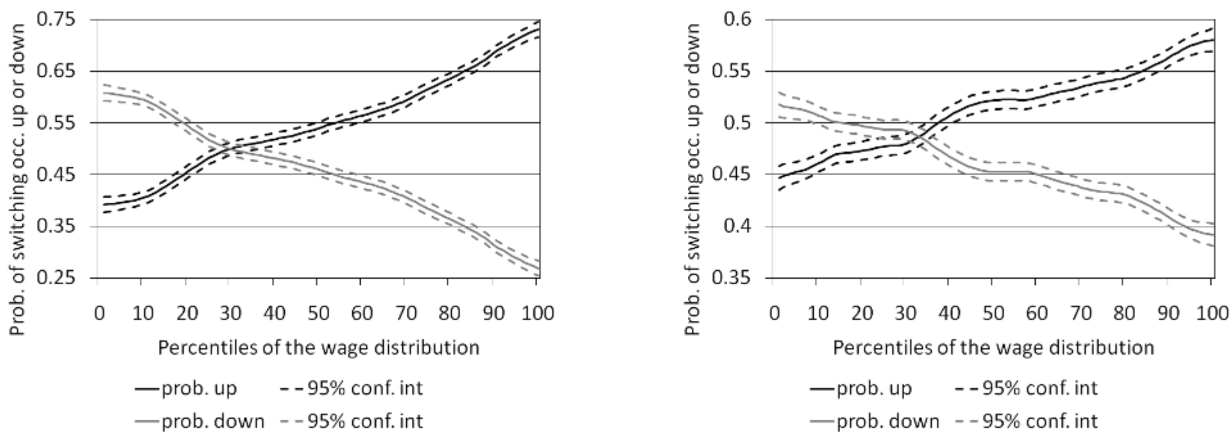
Figure 2: Non-parametric plot of probability of switching occupation by worker's percentile in the distribution of raw wages within occupation, year, and years after graduation.

The U-shaped pattern of mobility is also evident in Figure 2(a) where we plot the probability of switching out of an occupation against worker's rank in the distribution of wages within occupation, year, and among workers with the same number of years after graduation. That is, we compute the rank of the individual in the distribution of wages of workers who completed their education in the same year and work in the same occupation in a given year. Figure 2(b) separately graphs occupational mobility for workers who graduated 1, 2, 4, and 6 years ago. While the rate of occupational mobility generally declines with labor market experience, the U-shaped pattern of occupational mobility is pronounced for all years after graduation.

In Appendix Figure A-5 we show that all the findings reported above are similar when computed on the Large Sample that includes workers in the public sector and allows for spells of non-employment and part-time work.

To assess the prevalence of U-shaped pattern of occupational mobility we compute the fraction of occupations featuring U-shapes and the fraction of workers employed in these occupations. Computing these statistics requires enough workers in each occupation in each year to accurately predict the probability of changing occupation in different parts of the wage distribution of that occupation. Thus, we restrict the sample to occupations that include at least 100 workers in a given year. Separately for each occupation, we estimate the probit regression of the probability of switching occupation on a 2^{nd} degree polynomial in worker's percentile in the wage distribution within occupation and year, i.e. $Pr(switch) = \Phi\left[\alpha + \beta \cdot perc + \gamma \cdot perc^2\right]$. The partial effect of the wage percentile on the probability of switching occupation is $\frac{\partial Pr(switch)}{\partial perc} =$

number of the occupational spell from the wage regression does not qualitatively affect the results. Appendix Figures A-2 to A-4 illustrate that the U-shaped pattern of mobility is robust to alternative bandwidths choices.



(a) Distribution of raw wages within occupation and year. Average wage in occupation from population.

(b) Distribution of wage residuals. Average wage in occupation from time constants in wage regression.

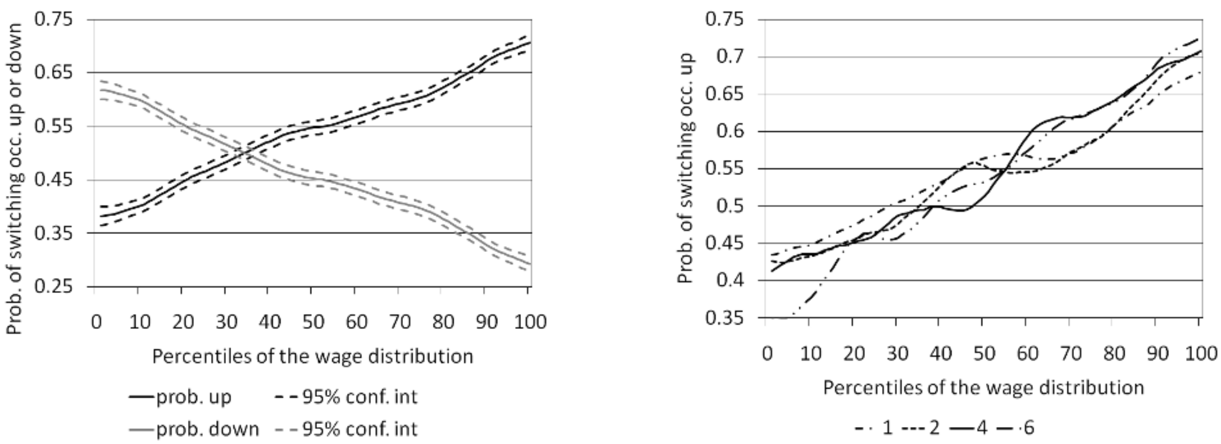
Figure 3: Non-parametric plot of direction of occupational mobility, conditional on switching occupation, by worker's percentile in the relevant wage distribution before the switch.

$\phi\left(\alpha + \beta \cdot perc + \gamma \cdot perc^2\right)(\beta + 2\gamma \cdot perc)$. The U-shaped pattern implies that this derivative evaluated at $perc = 0$ must be negative, that is $\phi(\alpha)(\beta) < 0$. Similarly, the U-shaped pattern also implies that the derivative evaluated at $perc = 1$ must be positive, i.e., $\phi(\alpha + \beta + \gamma)(\beta + 2\gamma) > 0$. In our Small Sample, 65% of occupations (employing 83% of workers) satisfy both of these criteria when percentiles are defined in raw wages. If percentiles are defined in wage residuals, 75% of occupations (employing 86% of workers) satisfy these criteria. In our Large Sample, 71% of occupations (employing 86% of workers) satisfy both of these criteria when percentiles are defined in raw wages. If percentiles are defined in wage residuals, 83% of occupations (employing 92% of workers) satisfy these criteria.

2.3 U-shapes in the Direction of Occupational Switching

In this section we document another prominent feature of the data: conditional on changing occupation, workers with higher (lower) relative wages *within* their occupation tend to switch to occupations with higher (lower) *average* wages than the average wage in their current occupation. We first find the average wage of all occupations in a given year in order to determine the ranking between occupations. Similarly to our analysis of probability of occupational switching, we rank occupations based on their raw wages or residual wages adjusted for worker characteristics. To obtain the ranking based on raw wages, we find the average real wage of all full-time private-sector workers in a given occupation in a given year.[13] To obtain the ranking based on residual wages,

[13]Note that this is a bigger sample than our selected sample, which only consists of workers who graduated after 1980 and who never worked in the public sector, worked part-time, etc. The results are, however, robust to



(a) Overall.

(b) For different years after graduation.

Figure 4: Non-parametric plot of direction of occupational mobility, conditional on switching occupation, by worker's percentile in the distribution of raw wages within occupation, year, and years after graduation before the switch.

we use our selected sample to run a similar wage regression as in Equation 1 for each occupation where we include time dummies in the regression (without the intercept). We interpret the coefficients on these time dummies as the average occupational wage in a given year, adjusted for human capital accumulation of workers in the occupation as well as other worker characteristics such as education, regional dummies, and marital status.

Figure 3(a) plots the probability of switching to an occupation with a higher or lower average wage as a function of the worker's position in the wage distribution of the occupation he or she is leaving. The sample on which the figure is based consists of all workers who switched occupation in a given year and occupations are ranked based on the raw average wages. Figure 3(b) presents corresponding evidence when occupations are ranked based on residual wages and the direction of occupational mobility is plotted against the percentile in the distribution of residual wages within an occupation the worker is switching from. The evidence contained in these figures suggest that, conditional on switching occupations, the higher relative wage a person has in his occupation before the switch, the higher is the probability that she will switch to an occupation with a higher average wage. Similarly, the lower relative wage a worker has in his occupation before the switch, the higher is the probability that he will switch to an occupation with a lower average wage than in the occupation he switches from.

Figure 4(a) illustrates that similar results hold if we further condition on workers position in the distribution of wages in his occupation in a given year and among people with the same

only looking at the average wages in our selected sample.

number of years since graduation. This figure is comparable to Figure 3(a) in that occupational average wages are calculated from raw wages of the population in the occupation in a given year. Finally, Figure 4(b) shows that the direction of occupational mobility is similar for individuals who graduated 1, 2, 4, or 6 years prior.

Appendix Figure A-6 illustrates that all these patterns are also observed on the less restrictive Large Sample.

2.4 Discussion of Empirical Evidence

2.4.1 U-shapes of Occupational Mobility within and between Firms

Our findings remain robust if we separately consider occupational switches who stay with their firms and occupational switchers who change firms as well. In both samples the probability of switching remains U-shaped in the position of the worker in the wage distribution of her occupation. Moreover, in both samples mobility is directional so that the relatively high (low) wage workers in their occupation tend to switch to occupations that pay on average higher (lower) wages. While the average probability of switching occupations is higher among those who switch firms than among those who stay with the same firm, possibly because occupational switching often necessitates switching firm if the new occupation is not represented in the old firm, the directional switching probabilities are virtually indistinguishable between the two samples. Figures 5 and 6 summarize this evidence when worker relative position in the wage distribution is determined based on raw wages. Appendix Figures A-9 and A-10 summarize this evidence when worker relative position in the wage distribution is determined based on wages residuals.[14]

These results suggest that unemployment is not the main driver of the occupational switching. There is sizable rate of occupational mobility within firms in Denmark, and this mobility exhibits similar patterns as those for the entire population of workers. High occupational mobility within firm has also been documented for the U.S. by Kambourov and Manovskii (2008). Moreover, a non-trivial fraction of workers who stay with the same firm switch to occupations that on average pay less to their workers.

This raises the question whether switches to lower-ranked occupations within a firm are indeed associated with lower wage growth for the individual worker, or whether they are just labels that are inconsequential for the actual wage and position that the individual workers has within the firm. Appendix Table A-2 contains evidence that the consequences for workers' wage

[14]We use worker's position in the overall wage distribution to plot these figures (i.e., the same distribution on which our unconditional on firm switching findings were based). An alternative is to define worker position in the wage distribution of the subsample he belongs to (i.e., firm switchers or firm stayers) and plot the probability of switching and the direction of switching against this rank. Qualitatively, this does not affect our findings.

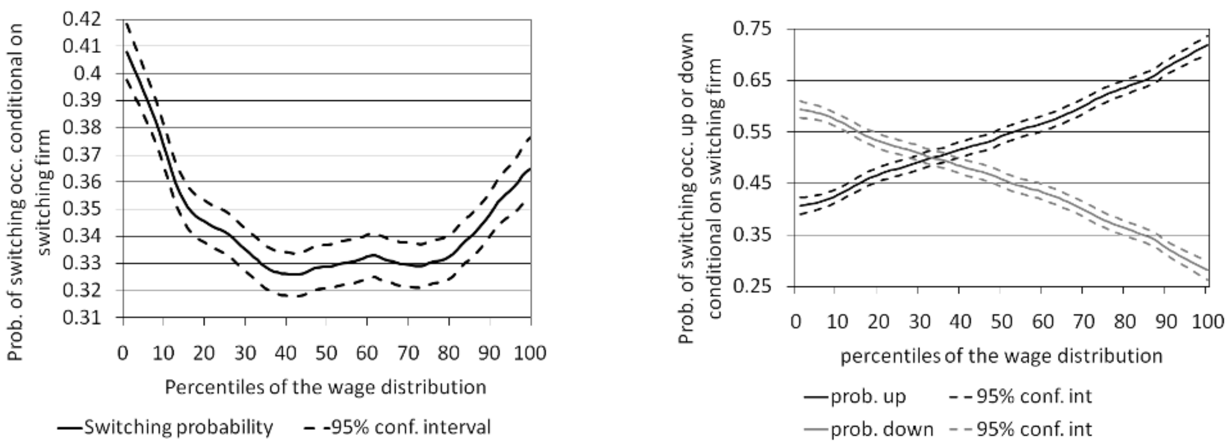


Figure 5: Non-parametric plots of probability of switching occupation and of direction of occupational mobility *conditional on switching firms* by worker's percentile in the distribution or raw wages.

growth are substantial both for workers that stay within the same firm as well as for those that switch firms. Among workers who stay with their firm, those who move to higher-ranked occupations see significantly higher wage growth than those who stay in the same occupation. Those workers who move to lower-ranked occupations see a significantly lower wage growth than those who stay in the same occupation. These patterns are similar for workers switching firms although the wage changes for this group of workers are somewhat larger.

2.4.2 Alternative Occupational Classifications

In this section we explore robustness of our findings to a number of alternative ways to define occupations. We begin by considering 1-, 2-, and 3-digit occupational classifications and compare the results to the 4-digit classification used in our main analysis. Figure A-11 in Appendix A5.5 illustrates that our results are robust to using alternative occupational classifications. While the level of mobility falls as occupational classifications become coarser, the U-shaped pattern of mobility remains unaffected. This provides further indication that a considerable part of mobility is driven by movements across occupations that can be vertically ranked which is clearly the case at the 1-digit level.

A potential concern is that some 4-digit occupations may not be sufficiently clearly differentiated (e.g., "Primary education teaching professionals" and "Primary education teaching associate professionals"). This may result in some spurious re-classification of workers' occupations because of reporting errors or when a worker continues to perform essentially the same task but gets re-classified because of a change in an institutional setting (such as teaching a different

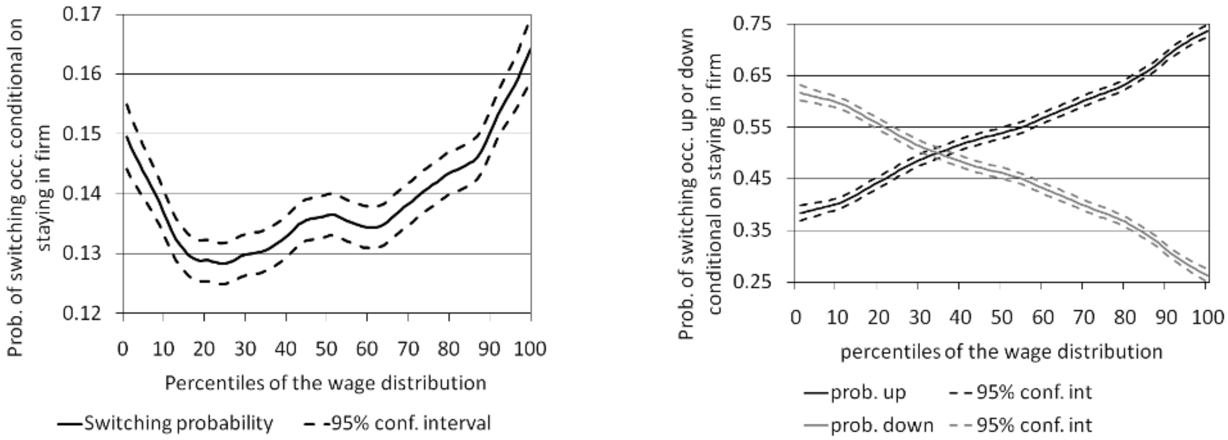


Figure 6: Non-parametric plots of probability of switching occupation and of direction of occupational mobility *conditional on staying with the firm* by worker's percentile in the distribution of raw wages.

class level). To address this concern we perform the following experiment. We access the Statistics Denmark's web page that firms can use to search for the correct occupational category of their employees. Typing in a description of the tasks performed by an employee into a search engine provided on this web page, returns one or more 4-digit occupational codes related to the query. For example, if we search for the word "painter," four distinct 4-digit occupations are returned. These are "Painter and related work," "Varnisher and related painters," "Glass, ceramics, and related decorative painters," and "Sculpture, painters and related artists." Similarly the search for the word "accountant" or "accounting" returns three 4-digit occupations, which are "Accountants", "Bookkeepers," and "Accounting and bookkeeping clerks." We go through all 4-digit occupations, excluding managers, and search for the word that describes the given occupation. We then group together all occupations returned by the search engine. This means that a switch from "Accountant" to "Bookkeeper" or to "Accounting and bookkeeping clerks" will not be registered as an occupational switch. In Figure 12(a) in Appendix A5.6 we plot the probability of switching across these occupational groups as a function of the worker's position in the wage distribution of their occupation.[15] We find that the U-shaped mobility patterns are robust to this re-classification of related occupations, while the level of occupational mobility is naturally somewhat lower.

[15]We keep the wage percentiles from the 4-digit occupations rather than the new defined occupational groups because the groups are not in a "closed relation." As an example, an "Accountant" is grouped with "Accounting and bookkeeping clerks" who, in turn, are grouped with "Administrative secretaries and related associate professionals." However, "Accountants" are not grouped with "Administrative secretaries and related associate professionals."

To assess whether our finding that workers with relatively high wages are more likely to leave their occupations is predominantly driven by promotions to managerial occupations we perform the following two experiments. First, we reclassify all managers as one occupation. Second, we exclude all managers from the sample. The results, plotted in Figures 12(b) and 12(c), respectively, indicate that U-shaped pattern of mobility is not mainly driven by movements in and out of managerial occupations.

Finally, in Figure 12(d) we plot the mobility patterns on the sample that excludes "... not elsewhere classified" occupations (their codes end with the number "9"). The U-shaped mobility patterns are not affected by this change in the sample.

2.4.3 The Effects of Measurement Error

While the occupational affiliation data we use is generally regarded to be highly reliable, some coding error might be present. Since the occupational code is provided to Statistics Denmark by the firm it is more likely for a worker's occupational affiliation to be miscoded when the worker switches firms. However, we have just seen that the U-shapes are robust to workers switching occupation *conditional on switching* firms as well as workers switching occupation *conditional on staying* with the same firm. Similarly, the direction of occupational mobility is also unchanged when conditioning on occupation *and* firm switchers or conditioning on occupation but *not* firm switchers. If measurement error were sizable, we would expect switches across firms to be more random and have a flatter curve than switches within firms. We do not find any evidence of this. These results suggest that measurement error is unlikely to substantially affect our findings. Moreover, we have also seen that grouping occupations together based on the similarity of their descriptions also did not affect our findings, again suggesting only limited possibility for measurement error to play an important role.

To further investigate the potential role of the measurement error we document the patterns of mobility for workers whose occupational affiliation is stable over several time periods and therefore less prone to possible temporary coding errors. When considering whether a worker switches occupation between periods t and $t+1$ we now only consider workers who have been in the same occupation for at least the two years $t-1$ and t and then stay in the same occupation for at least the two years $t+1$ and $t+2$. We also consider workers with at least three years of occupational affiliation before and after the switching point. The shape and direction of mobility for these workers is reported in Figures A-13 through A-16 in Appendix A5.7. We find that our results remain robust. Similar results are obtained on the samples of occupational switchers within and across firms.

2.4.4 Focus on Occupational Mobility

Our primary focus is on worker mobility across occupations which were shown in prior work to be major predictors of individual earnings. We have repeated the analysis on industries and found that mobility across industries does not exhibit U-shapes. Instead, it is the poor matches in the bottom part of the wage distribution in an industry that are more likely to be destroyed. This difference between occupation and industry switching might be due to the fact that industries have less of a natural ladder. Workers that find out that they are very talented may not change industries, but are likely to switch to more demanding occupations. It is also very interesting to assess the extent and patterns of sorting across firms. Unfortunately we cannot apply our methodology at the firm level because firms in Denmark are generally too small for this purpose, especially since one would presumably need to condition on workers' occupation.

2.4.5 Summary

To summarize the evidence presented so far, the probability of switching out of most occupations is U-shaped in the position of the worker in the wage distribution of that occupation. Workers with high wages relative to their occupational average switch to occupations with higher average wages. Workers with low wages relative to their occupational average switch to occupations with lower average wages.

The fact that high paid workers tend to switch to occupations that on average pay more suggests a model in which absolute advantage (high pay) goes hand in hand with comparative advantage in the more productive occupations (switching to better occupations). This is called positive sorting in traditional Roy models, and will be a central element of the following theory. We confront additional implications of the theory with the data as we derive them.

3 The U-shapes of Occupational Mobility: Theory

In this section we present a model of vertical sorting, where gross mobility arises since workers initially have only limited information about their ability and learn about it over time. In the model it is efficient that workers of higher ability work in the occupations where ability is most valued. If a worker learns that he is much better (or worse) than expected, he adjusts (has to adjust) to an occupation commensurate of his ability.

We show that the combination of learning and sorting is sufficient to generate the qualitative patterns that we find in the data. To highlight the basic impact of these two features, we abstract from other factors such as human capital accumulation and costs of occupational switching. We discuss in the following section how these features can be integrated. They do not offset the qualitative implications of sorting, but we discuss them because we expect them to be important for any quantitative assessment of the theory. Finally, we consider the validity of some secondary implications of the theory.

3.1 The Model

Workers: Workers choose employment in different occupations over time. Time is discrete and runs forever. Each period a unit measure of workers enters the labor market. The index for an individual worker will be i throughout. Each worker is in the labor force for T periods. Workers are risk-neutral and discount the future with a common discount factor β. Each worker has an innate ability level a_i that is drawn at the beginning of his life from a normal distribution with mean μ_a and variance σ_a^2. For the baseline model without human capital accumulation we assume that this ability remains constant throughout worker's life (we relax this in Section 4). The amount of output that a worker can produce depends on his ability. In particular, he produces

$$X_{i,t} = a_i + \varepsilon_{i,t} \tag{2}$$

in a given period t of his life, where $\varepsilon_{i,t}$ is a normally distributed noise term with mean zero and variance σ_ε^2. Workers do not know their ability (and neither do firms), but workers observe the output they produce. We assume that the worker observes an initial draw after finishing school, i.e., before entering the labor market.[16]

Over time, workers learn about their true ability. Let $\phi_a = 1/\sigma_a^2$ and $\phi_\varepsilon = 1/\sigma_\varepsilon^2$ denote the precision for each distribution, which is defined as the inverse of the variance. Define the cumulative precision of a worker at the beginning of his t^{th} year in the labor market as $\phi_t := \phi_a + t\phi_\varepsilon$. Initially every worker only knows that his ability is distributed with mean $A_0 = \mu_0$ and precision $\phi_0 = \phi_a$. Standard results on updating of normal distributions establish that his belief at the beginning of every period $t > 0$ of his life is normally distributed with mean $A_{i,t}$ and precision ϕ_t, where the mean is determined successively by the output realizations that he observes. After observing some output realization $X_{i,t}$ the new mean is given by the precision-weighted average of the prior mean and the output observation:

$$A_{i,t+1} = \frac{\phi_t}{\phi_{t+1}} A_{i,t} + \frac{\phi_\varepsilon}{\phi_{t+1}} X_{i,t}. \tag{3}$$

[16] In general we allow this error term to be distributed with a different variance σ_0^2 that might not coincide with the variance of the labor market error term σ_ε^2. While our exposition is presented for $\sigma_0^2 = \sigma_\varepsilon^2$, the more general expressions can be obtained by adjusting the cumulative precision defined below as $\phi_t = \phi_a + \phi_0 + (t-1)\phi_\varepsilon$.

From the point of view of the individual, this evolution of the posterior is a martingale with decreasing variance: The weight on the prior increases the more observations have already been observed in the past, i.e., the higher is t (see, e.g., Chamley (2004)). Correspondingly, the weight on the most recent observation decreases with years in the labor market. For all practical purposes, (3) can be interpreted as some exogenous change in workers ability, even though the learning interpretation appears to be particularly natural. In the following we will refer to $A_{i,t+1}$ as the expected ability or simply as the belief, and drop the person-identifier i and/or the time identifier t when there is no danger of confusion.

For completeness, we note two points. First, let $G_t(A_{t+1}|A_t)$ denote the distribution of next period's belief for a worker with current belief A_t. It is normal with mean A_t.[17] In particular, its density g_t is single-peaked and symmetric around its peak at A_t, and shifting the prior mean A_t simply shifts the entire distribution about the posterior horizontally in the sense that $g_t(A_{t+1}|A_t) = g_t(A_{t+1} + \delta|A_t + \delta)$ for any δ. This is all we need for most proofs.[18] We call the latter property *lateral adjustment*. Second, the cross-sectional distribution $F_t(A)$ of beliefs among workers that start the t'th period of their working life can be computed from (3) and is independent of any choices that agents make.[19] Therefore, the measure of agents with belief below A accross all cohorts at any point in time, $F(A) = \sum_{t=1}^{T} F_t(A)$, can be computed prior to any analysis of occupational choice. This simplifies the specification of an equilibrium.

Occupations: There are a finite number of occupations, indexed by $k \in \{0, 1, ..., K\}$, each with some fixed measure γ_k of available jobs. We treat the number of jobs as exogenous in this exposition, yet Appendix A2 discusses how endogenous entry can be accommodated (limited entry and associated competition among workers for scarce jobs will be most important in Section 4.1 to explain the mobility patterns when occupational productivities change).

Each unit of the good (or service) that is produced sells in the market at some exogenously

[17] Conditional on knowing the true ability a of a worker, the output X_t is distributed normally with mean a and precision ϕ_ϵ, i.e. $X_t \sim N(a, \phi_\epsilon)$. Yet the ability is not known. Rather, the individual only knows his expected ability A_t while his true ability is a draw $a \sim N(A_t, \phi_t)$. Integrating out the uncertainty over his ability implies that output is distributed $X \sim N(A_t, \phi_\epsilon\phi_t/\phi_{t+1})$. We are not interested in the output per se, but in the update $A_{t+1} = (\phi_\varepsilon X_t + \phi_t A_t)/\phi_{t+1}$ as a function of output. This linear combination implies that that the posterior distribution $G_t(A_{t+1}|A_t)$ is a normal with mean A_t and precision $\phi_t\phi_{t+1}/\phi_\varepsilon$, i.e. $A_t \sim N(A_t, \phi_t\phi_{t+1}/\phi_\varepsilon)$.

[18] For one proof (Proposition 4) we also need concavity of $g_t(A_{t+1}|A_t)$ in A_{t+1} locally for A_{t+1} near A_t, which holds for the normal distribution.

[19] At the beginning of period t the workers have observed t output observations (one in school and $t-1$ in the labor force). The only relevant information for the worker is the average $\bar{X}$ of these output realizations. Conditional on a this is distributed normally with mean a and precision $t\phi_\varepsilon$. Since a is not known, an agent with prior μ_a faces realizations of $\bar{X}$ that are normal with mean μ_a and precision $t\phi_\varepsilon\phi_a/\phi_t$. Since the update is $A^t = (t\phi_\varepsilon\bar{X} + \phi_a\mu_a)/\phi_t$, F^t is normal mean μ_a and precision $\phi_t\phi_a/(t\phi_\varepsilon)$.

given price P_k. Therefore, worker i employed in a job of type k generates revenue

$$R_{ki} = P_k X_i. \tag{4}$$

Equivalently, we can interpret P_k as the productivity in terms of efficiency units of the labor (at a common sales price of unity). We rank occupations in order of increasing productivity such that $P_K > ... > P_k > ... > P_0 = 0$. Therefore, any given worker produces more in a higher ranked occupation. One can view the lowest ranked occupation as home production. An output signal is observable even in home production, and home production is available to everybody (more jobs than population size: $\gamma_0 \geq T$). All other jobs are assumed to be scarce (less jobs than workers with positive ability: $\sum_{k=1}^{K} \gamma_k < T - F(0)$).[20]

Wages: We consider a competitive economy without matching frictions. The only frictions are information frictions in the sense that workers' actual abilities are not known. There are (at least) two ways to think about wage-setting in our economy. Wages might be output-contingent contracts $w(X)$ that specify different wages based on the particular output that is realized. If a firm wants to obtain profits Π_k it can simply offer the wage contract

$$w_k(X) = P_k X - \Pi_k \tag{5}$$

to any worker who is willing to take this contract. Since workers are risk-neutral, they choose the occupation with the highest expected wage. Therefore, it is not necessary the firm has as much information as the workers, since workers would self-select. Since workers are risk-neutral, the relevant sorting criterion for them is their expected wage given their belief A about their mean ability:

$$W_k(A) = P_k A - \Pi_k. \tag{6}$$

Alternatively, if the firm has the same information as the worker it can directly pay expected wages according to (6). In this case the firm absorbs all the risk. It would need to have the same information as the worker because otherwise it might attract workers with low expected abilities who try to get a high pay. Given risk-neutrality, whether firms or workers face the output risk does not affect the occupational choices by workers because in either case workers only care about expected wages (given by (6)), but observed wages differ according to the specification and could potentially lead to different assessments of observed patterns. We will show our main qualitative results under both wage setting regimes. In fact, firms might pay workers according to some weighted average of (5) and (6) to provide both incentives for self-selection as well as insurance

[20]Otherwise the lowest occupations would not attract any workers and would simply not be observed in the data.

to workers, and our arguments can easily be extended to show that our main propositions hold for any such convex combination.

Equilibrium: We are considering a standard stationary competitive equilibrium in this matching market between occupations and workers. As market prices one can use either profits or wages, as one determines the other via (5) [or (6)]. It is notationally more convenient to focus on the profits. Stationary means that the entrepreneurs' profits $\Pi = (\Pi_1, \Pi_2, ..., \Pi_K)$ and the associated wage offers are constant over time. The tractability of the baseline model arises from the fact that every period workers can costlessly re-optimize and therefore the sequence of decisions that maximize their life-time income coincides to the sequence of decisions that maximizes their payoff in each period. Since the cross-sectional distribution of beliefs $F(A)$ remains constant, we can use standard tools for the analysis of static matching models. In particular, a worker will work in occupation k rather than $k-1$ if the expected wage is higher: $P_k A - \Pi_k \geq P_{k-1} A - \Pi_{k-1}$. There is exactly one level of expected ability, call it B_k, at which this holds at equality:

$$B_k \equiv \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}}, \text{ for } k \in \{1, .., K\}. \tag{7}$$

Therefore, workers optimally choose to work in occupation k if their expected ability is within the interval $[B_k, B_{k+1})$, where we define $B_0 \equiv -\infty$ and $B_{K+1} \equiv \infty$.[21] Market clearing then means that the number of workers $F(B_{k+1}) - F(B_k)$ that would like to work in occupation k coincides with the number of jobs γ_k available in this occupation:

$$\gamma_k = F(B_{k+1}) - F(B_k), \text{ for } k \in \{1, ..., K\}. \tag{8}$$

The system (7) and (8) can easily be solved recursively: Summing (8) accross all k and noting that $F(B_{K+1})$ equals the total popluation T, we get $\sum_{k \in \{1,...,K\}} \gamma_k = T - F(B_1)$ which determines B_1. Then successive application of (8) yields the remaining cutoff levels $(B_2, ..., B_K)$. Since zero productivity in the lowest occupation implies zero profit, (7) then delivers the profits of the firms in the various occupations $(\Pi_1, ..., \Pi_K)$. To sum up:

Definition 1 *An equilibrium is a vector of profits* $\Pi = (\Pi_0, ..., \Pi_K)$ *with* $\Pi_0 = 0$ *and a vector of optimal worker cutoff level* $(B_1, B_2, ..., B_K)$ *such that equations (7) and (8) hold.*

[21] This sorting property is driven by the fact that our expected revenue function is supermodular, as highlighted in the seminal contribution by Becker (1973). Nevertheless, there are revenue functions different from the one that we assume that give rise to exactly the same wage patterns despite the fact that more able workers work in less productive occupations (this arises for example when revenues are $R_{ki} = P_k + (1 - P_k)X_i$, see Eeckhout and Kircher (2009) for details). In general, what is important for our results is that there is some benefit from sorting into the appropriate occupation given one's skills, so that workers adjust their occupation as they learn their type more precisely.

3.2 Analysis: Shape and Direction of Occupational Mobility

Consider a worker who chooses occupation k in his t^{th} year of labor market experience, and earns wage W as in (6). Let $S_{k,t}(W)$ be the probability that this worker switches, i.e., that he chooses a different occupation in $t+1$. We will use the superscript "+" to indicate the probability of switching to a higher occupation, and "-" to indicate the probability of switching to a lower occupation. Clearly $S_{k,t}(W) = S^{+}_{k,t}(W) + S^{-}_{k,t}(W)$. Similarly, if wages are set by (5), then denote the corresponding switching probabilities by lower-case letters $s_{k,t}(w)$, $s^{+}_{k,t}(w)$ and $s^{-}_{k,t}(w)$.

To analyze the shape of the switching probability formally, we adopt the following definition. Fix the cohort indicator t and the occupation k, and consider how the switching probability changes in the wage.

Definition 2 (U-shapes) *A function is U-shaped if it has local maxima at the boundaries of its domain and one of these is a global maximum. A function is strictly U-shaped if it is U-shaped and quasi-convex.*

U-shapes capture the qualitative feature that switching probabilities increase towards each of the ends of the domain, i.e., in the context of $S_{k,t}(\cdot)$ switching becomes more likely for workers with low and high expected wages. Strict U-shapes additionally ensure that the switching probability increases monotonically from its interior minimum toward the extremes of the domain. A particular property of a U-shaped function S is that $g \circ S$ is also U-shaped whenever g is strictly monotone. This has the practical relevance that it will not matter whether we refer to the actual wage of a worker or to the rank of the worker in the wage distribution, since the rank is just a monotone transformation of the actual wage.

It is easiest to highlight why the model generates U-shapes by looking at the case where workers get paid according to expected ability (6). The wage directly reflects the worker's expected ability, as $A = (W + \Pi_k)/P_k$. If a worker chose occupation k, it has to be the case that his prior about his expected ability was within the relevant cutoffs, i.e., $A \in [B_k, B_{k+1})$. Next period he will switch down only if his posterior falls below the cutoff B_k, he will switch up only if his posterior falls above B_{k+1}, and overall he will switch if either of these two happens. Therefore

$$S^{-}_{k,t}(W) = G_t(B_k|A), \tag{9}$$

$$S^{+}_{k,t}(W) = 1 - G_t(B_{k+1}|A) \text{ and} \tag{10}$$

$$S_{k,t}(W) = G_t(B_k|A) + 1 - G_t(B_{k+1}|A). \tag{11}$$

Consider interior occupations, i.e., occupations $k \in \{1, ..., K-1\}$ that are not at the extreme end of the spectrum. Since the distribution $g_t(A'|A)$ is symmetric and quasi-concave, the

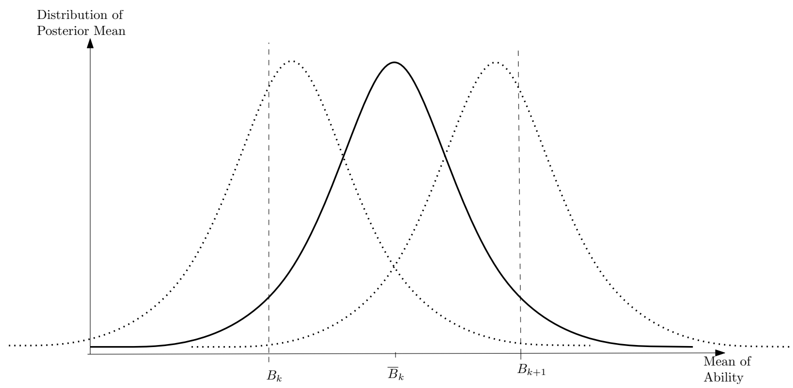


Figure 7: Illustration of the proof of Propositions 1 and 2.

switching probability is lowest when the prior A is at the midpoint between B_k and B_{k+1} and increases the more the prior moves toward either side of the interval. Figure 7 illustrates this. The solid curve is the density $g_t(A'|A)$ of the posterior belief A' for an agent with a prior belief at the midpoint $A = \overline{B}_k := (B_k + B_{k+1})/2$. For this worker it is least likely that his posterior lies outside the boundaries B_k and B_{k+1}. The dotted curve to the right is the density of the posterior for a worker starting with a prior above $\overline{B}_k$. He is more likely to switch because his posterior has more mass outside the "stay" interval $[B_k, B_{k+1})$. This is particularly clear for large intervals: agents with prior in the middle need very large shocks to induce them to leave, while agents on the boundaries only need small shocks to induce them to switch to the adjacent occupation. The following proposition establishes this for occupations of all sizes:

Proposition 1 (U-Shapes in Mobility) *Consider some interior occupation k and cohort t. The switching probability $s_{k,t}(w)$ is U-shaped; the switching probability $S_{k,t}(W)$ is strictly U-shaped.*

Proof. For the formal proof see Appendix A1. ■

For interior occupations, U-shapes are likely to persist even when we do not condition on cohort t.[22] For the extreme occupations of home production $k = 0$ and of $k = K$ the switching probability $S(\cdot, t)$ is also quasi-convex, but the minimum is at the extreme of the domain: In the

[22] This can be easily proved for the wage setting process (5), because extreme wages reflect extreme updates no matter what cohort the worker is in, and so switching probabilities approach one for extreme wages accross cohorts. Under (6), there is still a tendency for U-shapes unconditional on cohort, yet it is possible to construct examples where U-shapes do not hold for all occupations. The reason is that at the same expected ability older workers have more precision and switch less. If young workers are mainly in the middle of the interval $[B_k, B_{k+1})$ while old workers are more at one side, this composition effect between cohorts can lead workers with interior abilities (and wages) to switch more than those with abilities that are a bit more to the side.

case of home production workers at the bottom are least likely to switch since there is no lower occupation to switch down to, while in the case of the highest occupation workers at the top are least likely to switch because there is nothing better to move to.[23] Therefore, U-shapes cannot be derived in models that focus on two occupations only.

Next, we describe the direction of switching. Consider some interior occupation k. Intuitively, workers with high ability within this occupation and associated high average wages are the ones that are most likely to have output realization that indicate that they are appropriate for more productive occupations. This is visible in Figure 7 because the tail of the distribution that exceeds the upper bound increases as the distribution is shifted to the right. Workers with low belief about their mean ability are more likely to find out that they are not good enough and should move to a less productive occupation. Such a switch might manifest itself through firing if the employer has the same information as the worker, or as a quit due to the fact that the wage in absence of high performance is not good enough in the current occupation. The following proposition formalizes this intuition about switching behavior. It characterizes the probability for upward and downward switches conditional on switching. If the switching probability $S_{k,t}(W) > 0$, then the conditional probability of switching up is $S^{+}_{k,t}(W)/S_{k,t}(W)$, and similar for downward switches.[24]

Proposition 2 (Direction of Sorting) *Consider workers of cohort t in interior occupation k that switch. The higher ability workers are more likely to switch up and the lower ability workers are more likely to switch down: $s^{+}_{k,t}(w)/s_{k,t}(w)$ is increasing and $s^{-}_{k,t}(w)/s_{k,t}(w)$ is decreasing; $S^{+}_{k,t}(W)/S_{k,t}(W)$ is increasing and $S^{-}_{k,t}(W)/S_{k,t}(W)$ is decreasing.*

Proof. We focus on the wage setting process (6); see Appendix A1 for (5). Recall that a worker earns wage W only if he has belief $A = (W + \Pi_k)/P_k$. By (10) we can write $S^{+}_{k,t}(W) = 1 - G_t(B_{k+1}|A) = 1 - G_t(B_{k+1} - A|0)$, where the second equality follows from lateral adjustment. Since $G_t(\cdot|0)$ is a CDF it is increasing, and so $-G_t(B_{k+1} - A|0)$ is increasing in A, and thus in W. By a similar argument $S^{-}_{k,t}(W)$ is decreasing in W. This immediately implies that $S^{+}_{k,t}(W)/(S^{-}_{k,t}(W) + S^{+}_{k,t}(W))$ is increasing, while $S^{-}_{k,t}(W)/(S^{-}_{k,t}(W) + S^{+}_{k,t}(W))$ is decreasing. ■

Therefore, this simple model about learning one's absolute advantage generates the main predictions about sorting that we documented in the data.

[23] Proposition 2 provides a more general formal proof for this.

[24] If $s_{k,t}(w) = 0$, then $s^{+}_{k,t}(w) = s^{-}_{k,t}(w) = 0$. In this case, notational consistency in the following proposition requires a convention about conditional probabilities. In this case it is a convenient to define the conditional probability of switching up or down as $s^{+}_{k,t}(w)/s_{k,t}(w) = s^{+}_{k,t}(w)/s_{k,t}(w) = 1/2$.

3.3 Other Implications of the Model

In this section we derive and contrast with the data several additional implications of the model. Before turning to some implications that are specific to our equilibrium theory, we review two standard results from the learning literature. These have important implications for understanding the earnings process. First, we point out that our learning model is obviously able to account for the fact that switching probabilities decline with age, which is visible in, e.g., Figure 2(b). Second, the model can reproduce the important empirical pattern that cross-sectional variance in wages increases with labor market experience. This pattern has received much attention in the literature, going back to, e.g., Mincer (1974). The fact that the learning model captures these features so naturally makes it a strong candidate for modeling the process (3) by which A_t evolves.

Younger workers of cohort t switch less often than older workers of cohort $t' > t$, as long as t' is sufficiently large. This well-known result follows immediately from the fact that over time worker's information becomes more precise. Therefore, for any given belief A about one's mean ability and associated wage W, the likelihood that this prior will change substantially given the new output realization is lower for older workers. That is, $S_t(W)$ is decreasing in t and older workers switch less conditional on the same ability (same expected wage).[25]

If wages are paid according to expected ability (6)*, the cross-sectional variance in wages for young workers (cohort t) is smaller than for older workers (cohort $t' > t$).* Exactly because the information about each individual becomes more precise, the wages in (6) diverge for older workers. Since the ability of young workers is not very precise, they get similar wages. (Think about the extreme case in which initially there is no information about workers ability, and so all workers obtain exactly the same wage in their first period of their working life.) As information gets revealed in the output process, it becomes clearer which workers have high ability and which have low ability, and the former get paid more while the latter get paid less. Thus, their remuneration naturally fans out. Process (5) is analytically less tractable since the variance in the distribution of mean ability is confounded with the variance in the output process, but the result holds in the simulations we have conducted.[26]

[25] Since the distribution of abilities is different for older workers, it is theoretically possible that a particular older generation t' has abilities that are more concentrated around some switching cutoff B_k and therefore they switch more than a younger generation t. This is not possible as t' becomes large because information becomes nearly perfect while concentration does not go up substantially around any cutoff given our normal distribution assumptions, and we did not find any such effect in any of our simulations.

[26] Under (5), if there exists only one occupation the variance in wages would be unchanged as workers simple obtain a wage equal to their innate ability plus shock, multiplied by P_1. With multiple occupations, if workers start mainly in occupation k initially most output realizations are multiplied by P_k, but later generations sort better and low abilities get multiplied by smaller productivities $P_{k'} < P_k$ while higher abilities get multiplied by

3.3.1 Other Implications of the Model: Theory

We also obtain additional predictions specific to our model by considering wages of workers of the same cohort who switch occupations relative to those who stay in an occupation. It is clear that our model can generate the pattern regarding the wage changes that we reported in Section 2.4: upward switchers tend to see higher wage increases than workers who remain in the occupation, who in turn see higher wage increases than workers who switch down. This immediately arises if wages are set according to (6) but depends on parameters when wages are set according to (5).[27] More robust predictions of the model that are independent of whether the wage setting process is assumed to be given by (5) or (6) involve the comparison of wage levels between occupational switchers and stayers.

When we compare wages of workers who start in the same occupation, but some switch and some stay, we obtain the following prediction:

Proposition 3 *Consider workers of the same cohort, and compare the wages in period $t+1$ for those who stayed in occupation k with the wages of those workers who switched from k to k' between periods t and $t+1$: The average wage of the stayers is higher than the average wage of downward switchers ($k' < k$), but is lower than the average wage of upward switchers ($k' > k$).*

Proof. The result follows immediately because expected wages are increasing in the belief about one's mean ability, and these beliefs are strictly ranked: workers who switch up do so because their belief went up above B_{k+1}, while workers who stayed have a belief in $[B_k, B_{k+1}]$, and workers who switched down have a belief below B_k. ■

A less immediate implication arises when we compare switchers and stayers, but consider those that end up in the same occupation. Here, the predictions are exactly reversed:

Proposition 4 *Consider workers of the same cohort, and compare the wages in period $t+1$ for those who stayed in occupation k' with the wages of those workers who switched from adjacent occupation k to k' between periods t and $t+1$. Assume occupation k' and its adjacent occupations are not too large. Then the average wage of the stayers is lower than the average wage of downward switchers ($k' < k' + 1 = k$), but is higher than the average wage of upward switchers ($k' > k' - 1 = k$).*

higher productivities $P_{k''} > P_k$, which tends to increase the variance.

[27] Wages set according to (6) strictly increase in ability. Since a worker only switches up if his ability improved more than the ability for those people who stay, the wage of a worker who switches up improves more than for those how stay. A similar argument applies to workers who switch down. This pattern is less obvious if wages are set according to (5) because of reversion to the mean: a worker who had a particularly good shock will switch up to get his output multiplied by a higher productivity but will likely not have such a good shock again (and therefore not such high output) next period.

Proof. See Appendix A1. ■

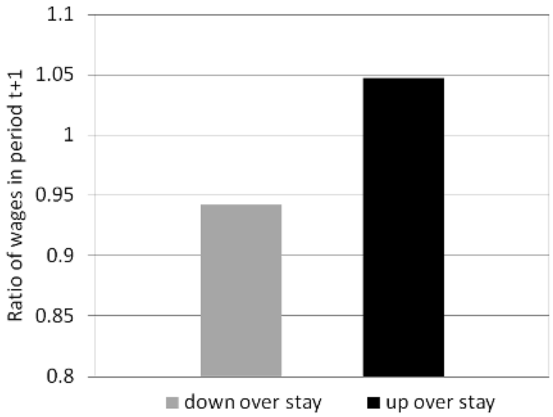
The logic behind this result is the following. Consider downward switchers. They enter the new occupation from above, and their expected ability is more concentrated at the upper end of interval $[B_{k'}, B_{k'+1})$ relative to the expected abilities of the workers who were in this occupation all along.

Thus, the predictions about the relative wages of stayers vs. switchers depend in an interesting way on the definition of stayers. This provides us with observable implications that can be verified in the data.

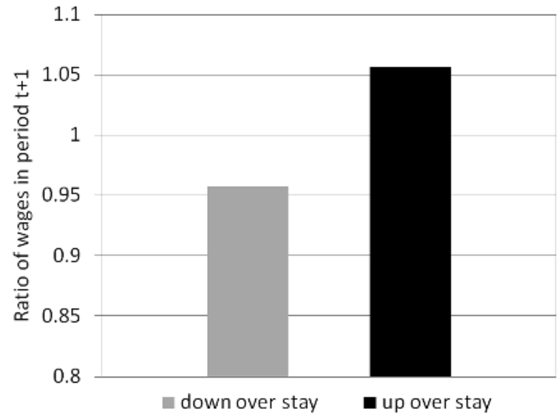
3.3.2 Other Implications of the Model: Evidence

To assess the empirical validity of the implication in Propositions 3, we compare the wage levels of stayers to those of switchers in the period after the switch. Consider all workers in occupation k in period t. In period $t+1$ we compute the ratio of wages of workers who left occupation k for a higher ranking occupation to wages of those who remained in occupation k between t and $t+1$. Similarly, we compute the ratio of wages of workers who left occupation k for a lower ranking occupation in period $t+1$ to wages of those who remained in occupation k between t and $t+1$. Next, we compute the average of these ratios across all occupations weighted by the number of workers in each occupation who switched either up or down. Figure 8(a) presents the results. The wage ratio of up-switchers over stayers is above 1, which indicates that the wages in period $t+1$ of workers who switch up from t to $t+1$ is higher than the wage of workers who stayed from period t to $t+1$. The wage ratio of down-switchers over stayers is below 1 indicating that workers who switched to lower ranking occupations have lower wages after the switch than workers who stayed in the same original occupation. This is consistent with the predictions of Proposition 3. Figure 8(b) shows that the ranking implied by Proposition 3 remains consistent with the data when we also condition on number of years after graduation in addition to being in occupation k in period t.

To assess empirically the predictions of Proposition 4, we compare wages of switchers into occupation k' to wages of those who stayed in occupation k' between t and $t+1$. In Figure 9(a) we construct the weighted average of period $t+1$ wage ratios of switchers into occupation k' over stayers in occupation k'. Figure 9(a) illustrates that workers who switched to higher ranking occupations have lower wages after the switch than the stayers in the occupation into which the up-switchers moved and that the opposite is true for workers who switched to lower ranking occupations. Figure 9(b) illustrates that these patterns are robust to also conditioning on the number of years after graduation.



(a) Unconditional.



(b) Conditional on the number of years after graduation.

Figure 8: Weighted average of year $t+1$ ratios of real wages of workers who switch occupations between years t and $t+1$ over workers who stay in the same *original* occupation in years t and $t+1$ by direction of the switch (i.e., whether the switch involves moving to an occupation that pays more or less on average than the source occupation.)

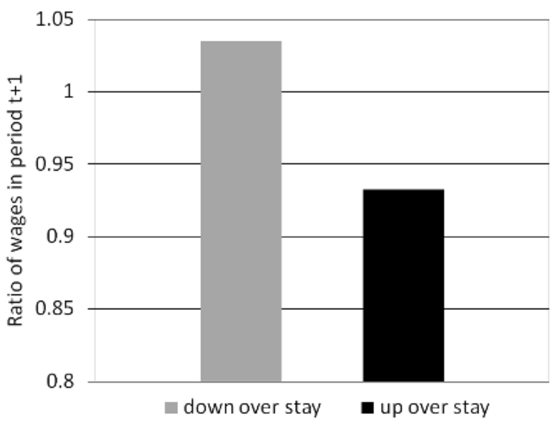
Appendix Figure A-7 presents the results of these experiments on the Large Sample. The predictions of Propositions 3 and 4 remain consistent with the data.

4 Extensions

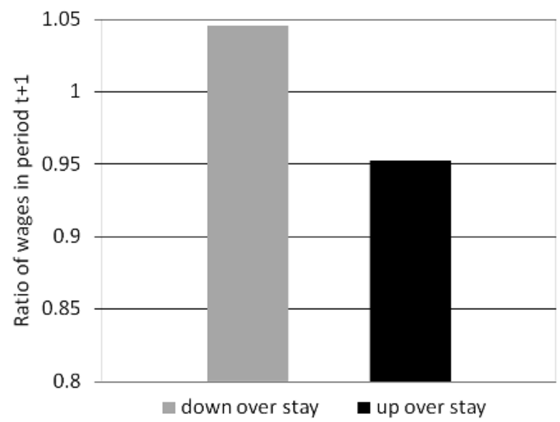
In this section we discuss two important ways in which the model can be extended. For each extension we depart from the baseline model laid out so far, even though it is easy to see how the extensions can be combined.

First, we introduce changes to the productivity of occupations – and see that the competition among workers for jobs implies that occupations with rising productivity attract and retain high ability (high wage) workers that drive out low ability (low wage) workers. We show that this pattern is indeed present in the data for the occupations with fastest wage growth. Opposite predictions arise for declining occupations, and these are also confirmed in the data. This captures the behavior for the bulk of the occupations for which we did not find U-shapes.

Second, we allow for human capital and switching costs. We show that general human capital introduces a career-ladder motive: as human capital grows workers switch into better occupations. In the absence of learning this would only induce upward mobility, the presence of learning then still induces downward mobility. Occupation specific capital acts as a particular form of switching cost as it is lost when leaving the occupation. We show how this can be accommodated, and then highlight numerically that the empirical switching patterns documented



(a) Unconditional.



(b) Conditional on the number of years after graduation.

Figure 9: Weighted average of year $t+1$ ratios of real wages of workers who switch occupations between years t and $t+1$ over workers who stay in the same *destination* occupation in years t and $t+1$ by direction of the switch (i.e., whether the switch involves moving to an occupation that pays more or less on average than the source occupation.)

above are robust under plausible values of human capital accumulation processes.

4.1 Changing Occupational Productivities

4.1.1 Mobility in Response to Changing Occupational Productivity: Theory

So far we have considered a stationary environment where the only reason for a change of occupation is learning about one's own ability. This generates mobility even though the productivity of all occupations remains constant over time. Nevertheless, some occupations such as computer programmers in the 1990s have seen substantial wage increases for their workforce (conditional on staying in the occupation, so excluding obvious selection effects), while other occupations such as textile machine operators have seen substantial wage declines.[28]

In this section we will allow occupational productivities to change over time. As mentioned, occupations with high productivity growth retain only their high ability workers while their low ability workers are driven out because their skill is no longer appropriate, and vise versa for declining occupations. In the subsequent subsection we confirm empirically that occupations with large productivity changes exhibit this mobility pattern. This accounts for most of the occupations that do not feature U-shapes in the probability of switching.

We note here that these insights arise because we consider an equilibrium model where workers compete with each other for jobs. This is very different from the standard analysis in

[28]Kambourov and Manovskii (2009a) measure the magnitude of changes in occupational productivities.

the Roy-type models, where each worker has the decision-theoretic task of choosing the most appropriate occupation without concern for the choices of the other workers. In such a setting, a more productive occupation attracts more workers but does not loose any of its existing workers (we discuss this further in Section 5.1).

To make these points theoretically, we need to slightly expand the notation. Denote calendar time by τ and index occupations by a name $r \in \{0, 1, ..., K\}$ instead of their rank in terms of productivity (since the rank is now changing), with $r = 0$ still being home production with constant productivity of zero. We retain the same notation as in the previous section, with the adjustment for the name of the occupation and an additional superscript indicating calendar time. For example, $P_r^\tau > 0$ denotes the productivity of occupation r at calendar time τ. Productivities can be deterministic functions of calendar time, but are also allowed to be realizations of some stochastic process. Stochasticity does not affect the analysis since workers can still costlessly change occupations each period. Importantly, the cross-sectional distribution F of beliefs remains unchanged because it does not rely on occupational choice. Therefore, the model can be solved period by period as in the previous section. We assume that each period productivities can be strictly ordered.

We also continue to assume that the measure γ_r of entrepreneurs in each occupation r remains constant, although our results are robust as long as entry is sufficiently inelastic to induce competition among workers for scarce jobs.[29] Inelastic labor demand might arise, for example, because a job in an occupation needs a particular type of physical capital that is not easily adjusted when the demand for the services of the occupation changes. See the further discussion in Appendix A2.

Given the productivities that prevail in period τ, let $\underline{B}_r^\tau$ and $\overline{B}_r^\tau$ be the lower and upper bounds on the ability (analogous to bounds B_k and B_{k+1} in the preceding section). That means that workers with beliefs in $[\underline{B}_r^\tau, \overline{B}_r^\tau)$ choose to work in occupation r. Equation (8) readily reveals that these beliefs depend exclusively on the number of jobs that offer lower wages, not on the level of productivity per se. It will therefore be convenient to define Γ_r^τ as the measure of all jobs that have weakly lower productivity than the jobs in occupation r in period τ. We call Γ_r^τ the *position* of occupation r in the distribution of productivities. When the position of a specific occupation r stays constant for two periods, i.e. $\Gamma_r^\tau = \Gamma_r^{\tau+1}$, it follows immediately that the cutoffs that determine who stays in the occupation remain constant, i.e. $\underline{B}_r^\tau = \underline{B}_r^{\tau+1}$ and $\overline{B}_r^\tau = \overline{B}_r^{\tau+1}$, and the switching behavior of workers in occupation r remains essentially unchanged compared to

[29]We discuss entry in the Appendix A2. When entry is completely elastic, the model resembles the Roy (1951) model, since each worker essentially decided by himself whether to "buy" a job in occupation k, independent of the choices of all other workers.

the baseline model analyzed in the previous section.[30] Switching is maximal at both ends of the earnings (and ability) spectrum, and is lowest at intermediate income levels.

When an occupation improves in rank between period τ and $\tau+1$ in the sense that $\Gamma_r^{\tau+1} > \Gamma_r^{\tau}$, the bounds on ability improve in the sense that $\underline{B}_{r_\tau(k)}^{\tau+1} > \underline{B}_{r_\tau(k)}^{\tau}$ and $\overline{B}_{r_\tau(k)}^{\tau+1} > \overline{B}_{r_\tau(k)}^{\tau}$. An immediate implication of the increased bounds is that workers who stay in the occupation between the two periods are a positive selection of the initial workforce. Therefore, these stayers improve their position in the aggregate cross-occupational wage distribution. This allows us to use the wage increase of stayers relative to the increase of other workers as a measure of the productivity gain of an occupation relative to the productivities of other occupations.

Another implication of the increased bounds of an improving occupation is that high ability workers join while low ability workers are driven out. This has direct consequence of the patterns of switching that we observe. In particular, in rising occupations high wage workers tend to stay while low wage workers tend to leave. The following proposition is proved for the case where firms absorb the uncertainty of the production process, but a proposition with similar content can be proved when workers are residual claimants.[31]

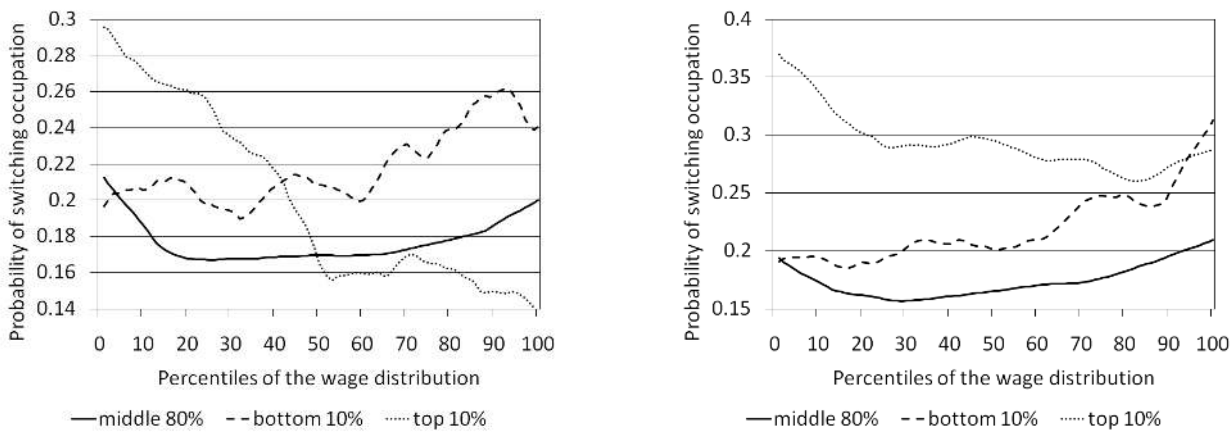
Proposition 5 *Consider an occupation r with a sufficient relative rise in productivity such that $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$. If wages are set according to (6), the probability of switching out of occupation r between and τ and $\tau+1$ decreases with higher wages for workers in the same cohort. The reverse holds for a sufficient decline in relative productivity such that $\Gamma_r^{\tau+1} \leq \Gamma_r^{\tau} - \gamma_r$.*

Proof. We prove the result for a rising occupation; analogous steps establish the result for a declining occupation. Wages in (6) rise in the prior A, and the distance $|\underline{B}_r^{\tau+1} - A|$ decreases in A for all workers that choose occupation r in period τ [since $A \leq \overline{B}_r^{\tau}$ and $\overline{B}_r^{\tau} \leq \underline{B}_r^{\tau+1}$ when $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$]. Thus, workers with a higher prior are closer to the region $[\underline{B}_r^{\tau+1}, \underline{B}_r^{\tau+1})$ where they stay in r, and therefore it is more likely that their posterior falls into this region (which follows formally from single-peakedness and lateral adjustment of the update G_t). ■

This implies that rising occupations with particularly fast growth in wages of stayers (i.e., occupations whose rank is sufficiently increasing) exhibit switching patterns where mainly the low

[30] For the baseline model in Section 3 where productivities do not change, define $\hat{s}_{k,t}(X) = s_{k,t}(P_k X - \Pi_k)$ and $\hat{S}_{k,t}(A) = S_{k,t}(P_k A - \Pi_k)$. This gives the switching probabilities based on output/ability rather than on the wages. It can be shown that $\hat{s}_{k,t}(X)$ and $\hat{S}_{k,t}(A)$ are invariant to the exact productivity level of occupation k, as long as it retains the same position among the occupations.

[31] For wages set according to (5) we can prove the following (proof available upon request). Consider an occupation r that rises sufficiently in position, $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$, and consider the probability of staying in r between τ and $\tau + 1$. Then only workers who had wages above the occupational mean in τ stay, while all lower wage workers leave. The reverse holds for a sufficient decline in position, $\Gamma_r^{\tau+1} \leq \Gamma_r^{\tau} - \gamma_r$.



(a) Distribution of raw wages within occupation and year. Growth rates of average wage in occupation from population.

(b) Distribution of wages residuals. Growth rates of average wage in occupation from time constants in wage regression.

Figure 10: Non-parametric plot of probability of switching occupation by worker's percentile in the relevant wage distribution. For the fastest growing 10% of occupations, the slowest growing 10% of occupations, and the remaining 80% of occupations.

wage workers switch out and the higher wage workers stay, and the opposite holds for declining occupations.

4.1.2 Mobility in Response to Changing Occupational Productivity: Evidence

Consistent with the theory, in the data we find that lower paid workers in a given occupation tend to leave it when occupational productivity rises, while higher paid workers in a given occupation are more likely to leave it when productivity of the occupation declines. We examine this in the data by studying occupations with different growth rates of the average wage of workers who stay in the occupation between years t and $t+1$ (to avoid obvious selection effects on average wages). Similar to Section 2.3 we compute the average wage based on either the raw wages or wage residuals. For each of these two notions of the average wage, we calculate the percent increase between each two consecutive years between 1995 and 2002.

Figure 10 plots three groups of occupations, separated by the growth rates of average wages of stayers between years t and $t+1$. The first group consists of the 10 percent of occupations with the lowest growth rates, the second group is the 10 percent of occupations with the highest growth rates, and the third group is the occupations with growth rates in average occupational wages in the middle 80 percent. For the three different occupational groups we plot the probabilities of switching occupation as a function of the workers' position in wage distribution in their occupation in year t. Figures 10(a) and 10(b) show that workers in occupations with the lowest growth rate of wages between t and $t+1$ have a higher probability of leaving their occupation between t and $t+1$ if they are from the upper end of the occupational wage distribution in year t. Workers in the fast growing occupations have a higher probability of leaving their occupation if they are in the lower part of the wage distribution in their occupation. Workers in occupations that grow faster than the slowest growing 10 percent but slower than the fastest growing 10 percent, have a probability of changing occupation that is U-shaped in their wage percentile.

The results of the corresponding analysis on the Large Sample are reported in Appendix Figures 8(a) and 8(b). They exhibit the same qualitative patterns.

4.2 Human Capital and Switching Costs

Incorporating human capital and switching costs into our theory is important. It is likely that not all wage growth is due to sorting, but human capital accumulation is also an important element. On the theoretical side we highlight the role of general human capital for upward mobility, and we show how to accommodate specific human capital that is tied to the occupation and acts as a switching cost that make transitions between occupations less attractive. We introduce these elements into the basic environment of Section 3, and then show in simulations that the basic patterns for mobility still arise for reasonable parameter values.

For general human capital, assume that a worker at the beginning of his t^{th} year in the labor market has human capital $H(t)$. Moreover, a worker who starts his ι^{th} consecutive year in occupation k has human capital $h_k(\iota)$. We normalize both forms of human capital to be zero in the first year, and assume that the human capital functions are weakly increasing. If a worker switches occupation, he loses his occupation-specific human capital and has tenure $\iota = 1$ in his new occupation. This introduces switching costs, and thus the optimal decisions have to be calculated from a dynamic program that trades off the future gains from switching with the immediate costs. For completeness, we also allow other switching costs κ_k that may arise when a worker switches from occupation k to a different occupation, which might capture application effort, retraining costs, etc.

Consider a worker with t years of general labor market experience and ι years of occupational experience in occupation k. There are various ways in which human capital can influence the output process. Our preferred specification is in analogy to (2)

$$X_k = a_i + H(t) + h_k(\iota) + \varepsilon_i. \tag{12}$$

leading to expected wage

$$W_k(A) = P_k(A_t + H(t) + h_k(\iota)) - \Pi_k. \tag{13}$$

Since human capital accumulation is deterministic, a worker who observes his output can back out $a_i + \varepsilon_i$, and therefore learning is not affected by human capital accumulation and the distribution F of beliefs in the population remains unchanged.[32] For this adjusted output process (12) the wages are still determined by (5) given the profit Π_k that firms want to obtain. The main difference to the previous sections is that workers solve a dynamic programming problem when deciding on the optimal occupation decision. We again consider a stationary equilibrium where firms' equilibrium profits Π_k remain constant over time. We define the precise notion of an equilibrium for this setup in Appendix A3.

Consider first the implication of general human capital accumulation ($H(t)$ strictly increasing) for occupational switching, abstracting from switching costs ($h_k(\iota) = 0$, $\kappa = 0$). Compared to a world without human capital the distribution of worker productivity now shifts by $H(t)$ for workers with t years of experience, since the relevant measure of a worker's ability in producing output is $a_i + H(t)$. Even though the new labor market entrants have the same distribution of ability as in the setting without human captial, with general human capital older workers become more productive and induce tougher competition for jobs in more productive occupations. Therefore, young workers start lower and in expectation move up to better occupations over the lifetime. *Human capital induces a drift toward more productive occupations,* creating another force for the upward movement through the occupation ladder beyond learning. This leads naturally to a somewhat higher aggregate probability of switching to higher than to lower occupations, as is visible in Figure 3.[33] If the variance in output is reduced to zero, all movements are upward and are entirely driven by improvements in skills over the workers lifetime, yet this would be missing the substantial part of downward movements (even when staying with the same firm) and associated wage implications that we observe in the data.

Our insights on U-shapes carry over to the setting with switching costs ($h_k(\iota)$ increasing, $\kappa > 0$). U-shapes still obtain for any wage setting that is weighted average (5) and (6) with

[32] Alternatively, we could e.g. exponentiate the right hand side of (12), which would still leave beliefs in the cross-section unchanged.

[33] Hall and Kasten (1976) and a number of later papers (e.g., Miller (1984), Sichernam and Galor (1990)) have also found that there is a systematic tendency for workers to move up to higher paying occupations with age. Wilk and Sackett (1996) have noted the tendency of workers to move to occupations requiring higher cognitive skills with age. Note that human capital accumulation is not necessary to induce an upward bias in switching: Depending on the precise values of the γ_k's and P_k's the workers might enter mostly in low occupation when young and then move up (or the reverse, depending on parameters). The main effect of general human capital is that it adds an *additional* element that unambiguously shifts young workers to less productive occupations since on average they cannot compete with older (more productive) workers in more productive occupations.

As an aside, note that we can add some additional terms $\alpha H(t)$ with $\alpha \geq 0$ to (13) to account for general human capital that increases the productivity in all occupations but does not interact with productivity of the occupation. This makes it possible to fit a wider range of wage growth patterns. In particular, this type of human capital does not affect sorting and does not induce a drift toward the more productive occupations.

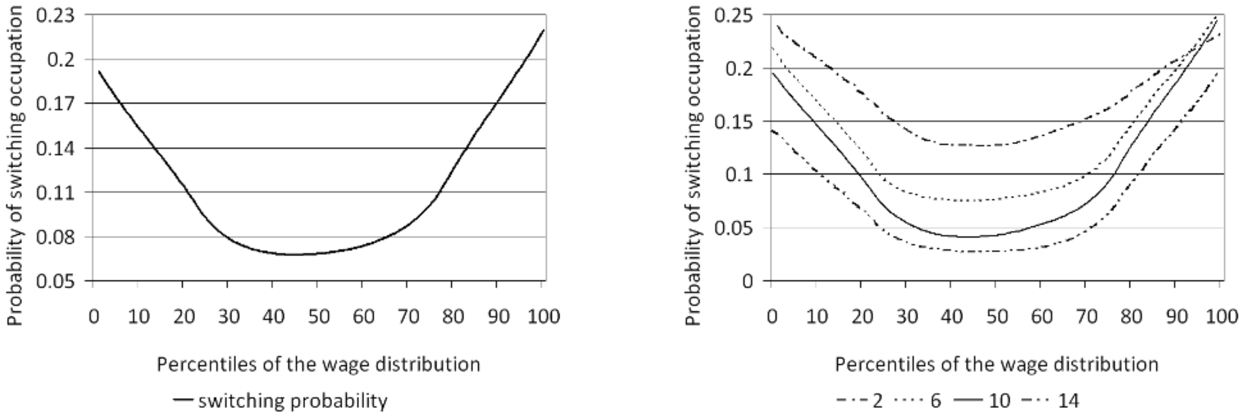
positive weight on (5). In this case wages partially reflect the new information obtained through the realized output, and very high (low) outlier wages can only arise because of very high (low) output realizations, in which case the agent learned that he is much better (worse) than he expected and it can be shown that at the extreme wages the update must be so large that the gains from switching outweigh any finite switching costs. In contrast, when workers are fully insured against the output risk by receiving the expected wage according to (6), the current period wage does not reveal any information about what the worker learned in the current period and the logic of the preceding argument does not apply. In this case, it could be that U-shapes do not arise. This could happen, for example, if older workers are more productive and therefore earn higher wages, but face higher switching costs and therefore have low probability of leaving the occupation.

However, in numerical simulations we always found U-shapes for reasonable parameter values. For instance, consider the following numerical example. We set the model period to be one year and assume that workers are in the labor market for 40 years. We assume that there are 25 occupations (plus home production) of approximately equal size with prices given by $P_k = 1 + 0.05k$ for $k \geq 1$. We set $H(t) = 0.008t$ and $h_k(\iota) = 0.008\iota$ for $\iota \leq 5$ and $h_k(\iota) = 0.04$ otherwise. These choices imply that during the first 5 years in an occupation wages grow by 10% and half of this wage growth is due to accumulation of occupation-specific human capital and half due to accumulation of general human capital. To ensure that (nearly) all workers have positive ability we normalize average ability to a sufficiently high value $\mu_a = 50$. Finally, we set the precision $\phi_a = 0.667$ and $\phi_\varepsilon = 0.052$. At these parameter values the model generates the occupational mobility rate of approximately 10% and the variance of log wages of 0.15. Taken together, sorting and human capital accumulation account for a life-time wage growth of 60%.

Figure 11 describes the patterns of occupational switching estimated in the model-generated data. The probability of switching is clearly U-shaped in the position of a worker in wage distribution in his occupation. Moreover, this pattern is also apparent when we condition on years of labor market experience. We emphasize that this is just a numerical example and not an attempt to calibrate the model. However, it is representative of the patterns we observe in simulations for various parameterizations under wage setting given by by (6).

5 Connection to Existing Models

Our work connects to several strands of the literature discussed in the introduction. As we mentioned, our empirical findings on the shape and the direction of sorting conflict with predictions of match-specific learning models (Jovanovic (1979), McCall (1990), Neal (1999)) and of island-



(a) Overall.

(b) For different years after graduation.

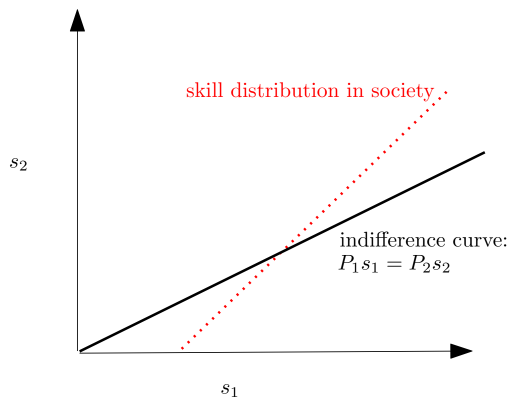
Figure 11: Non-parametric plot of probability of switching occupation by worker's percentile in the wage distribution within occupation, year, and years after graduation. Model Simulations.

economy models with human capital (Kambourov and Manovskii (2005, 2009a)). In both types of models low wage earners tend to switch, and since they did not receive any additional information about their fit to other occupations they take a random draw for their next occupation. In contrast, the crucial part of our model is that the experience of workers in their current occupation determines their choice of the next occupation, and that the occupations can be ranked. In such a world a bad fit can be characterized by underqualification *or overqualification* of a worker for a particular job. This means that it is not only low wage workers who leave an occupation, but also very qualified workers with high wages. This sorting property is very similar to sorting in the classic Roy (1951) model, and we briefly highlight the similarities and differences to this model in the following. The main difference relates to the equilibrium interaction among workers and to the net mobility induced by learning. The latter is linked to recent work by Gibbons, Katz, Lemieux, and Parent (2005), and we make this connection precise in the second part of this section.

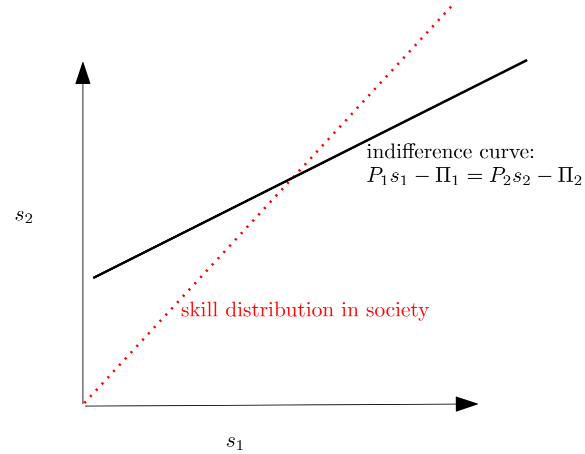
5.1 Roy Model

The idea that occupations might be vertically ordered goes back at least to the Roy (1951) model with absolute advantage. In the basic version of the Roy model according to the formalization in Heckman and Honore (1990) and Bils and McLaughlin (2001) there are two occupations 1 and 2 with output prices P_1 and P_2, respectively.[34] Each worker is endowed with a two-dimensional

[34]Bils and McLaughlin (2001) are concerned with industries rather than occupations, but in terms of exposition these can be used interchangeably. The later sections in Heckman and Honore (1990) also discuss in detail identification with multiple occupations.



(a) Illustration of the Roy model. s_i and P_i: skill level and price of output in occupation $i \in \{1, 2\}$.



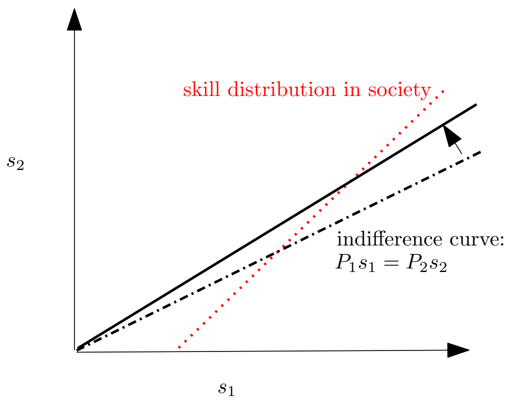
(b) Illustration of our model. $s_i = a$: skill level in occupation $i \in \{1, 2\}$. P_i and Π_i: price of output and profit in occupation $i \in \{1, 2\}$.

Figure 12: Comparison between Roy model and our model for fixed productivities P_1 and P_2.

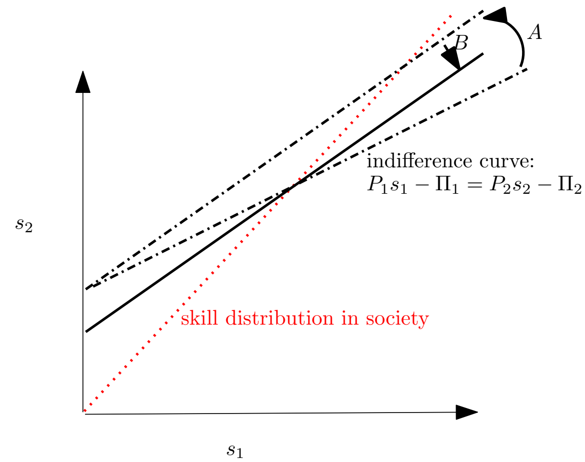
skill set (s_1, s_2) that describes his output in each occupation. Each worker chooses the occupation where he obtains the highest profit, i.e., he chooses Occupation 1 if $P_1 s_1 > P_2 s_2$. Figure 12(a) illustrates this. The solid line through the origin depicts all skill combinations (s_1, s_2) where the workers are exactly indifferent between each occupation, and workers with skill combinations to the right obtain a higher return in Occupation 1 and so choose it, while to the left they choose Occupation 2.

Mobility arises only when prices change. For example, if the low productivity occupation P_1 improves, then the solid line becomes steeper and more agents choose Occupation 1. Which agents change depends on the skill distribution in society. If the skills that are prevalent in society are given by the dotted line in Figure 12(a), then the worst workers in Occupation 2 will leave and become the highest wage workers in Occupation 1. This is depicted in Figure 13(a). This is a setting with absolute advantage: A worker who is better in one occupation is also more skilled in the other. The alternative picture with relative advantage would have a decreasing dotted skill line where workers that are better in one occupation are not as skilled in the other, in which case the lowest wage workers would leave occupation 2 when P_1 increases and become the lowest wage workers in occupation 1. U-shapes are not likely to arise unless the skill distribution has a particular curvature.

Our model resembles the Roy model with absolute advantage. In our specification workers choose Occupation 1 if $P_1 s_1 - \Pi_1 > P_2 s_2 - \Pi_2$ as indicated in Figure 12(b). All skills are on the diagonal line since $(s_1, s_2) = (a, a)$. The fact that Figure 12(b) is essentially a rotation of 12(a) shows the close resembles when switching is driven by absolute advantage, yet our setup does



(a) Illustration of the Roy model: when productivity P_1 in Occupation 1 increases, more workers choose Occupation 1.



(b) Illustration of our model: When productivity P_1 in Occupation 1 goes up, then the indifference line becomes steeper (step A, but also the firms' profit and therefore the intercept changes when the number of jobs is fixed (step B).

Figure 13: Comparison between Roy model and our model when productivity P_1 increases.

exhibits two major departures from the standard Roy model.

First and most importantly, workers learn about their skill over time, and therefore their position on the dotted line in Figure 12(b) changes over time, which induces them to switch occupation even if the aggregate productivity of all occupations remains unchanged.

Second, the interaction is not just decision-theoretic. If P_1 increases, the line that divides who selects into which occupation becomes steeper, as indicated in step A in Figure 13(b). When the number of jobs in this occupation remains limited, then also the profits that the firms make change and therefore the intercept changes as depicted by step B in Figure 13(b). If the number of jobs is fixed, exactly the same skills select into each occupations as before, unless the solid curve becomes steeper than the skill distribution in which case the matching pattern reverse. The latter case generates the effects described in Section 4.1.1, where an increase (and reversal) in occupational productivity changes the workforce composition because only workers that learned that they are much better than they thought stay in Occupation 1, whereas those workers that did not see their skills improve will be driven out by incoming high-skill workers out of Occupation 2. This second effect is not present in the standard Roy model where there is no interaction between the decisions of workers.

5.2 Relation to Gibbons, Katz, Lemieux, and Parent (2005)

Our model of learning is related to work by Gibbons, Katz, Lemieux, and Parent (2005). They also extend the Roy (1951) model to allow for learning about workers' abilities. They do not use an equilibrium model, and do not explicitly analyze the switching behavior of workers as a function of their earnings. Rather, their focus is on the decision-theoretic problem of an individual worker, for which they propose a instrumental variables method based on lagged occupational choices in order to estimate his choice parameters consistently. Since adaptations of their model allow to back out underlying parameters such as productivities even in our model, it is important to review the connection.

Consider the expected wages in our model, and assume that productivities are constant over time. Therefore, the profit vector $(\Pi_0, \Pi_1, ..., \Pi_K)$ remains constant over time. This vector implies that a worker at the beginning of his t^{th} period in the labor market who observed output realizations $(X_0, X_1, ..., X_{t-1})$ obtains an expected wage according to (6) of

$$E[P_k(a_i + \varepsilon_{it}) - \Pi_k | X_0, X_1, ..., X_{t-1}] = P_k A_{it} - \Pi_k,$$

where we left out the additive human capital terms for notational convenience. For the decision-theoretic problem of individual worker, profits Π_k can be interpreted as parameters.

Now consider the following transformation where we raise the wage of workers into the exponent:

$$E[e^{\{P_k(a_i+\varepsilon_{it})-\Pi_k\}} | X_0, X_1, ..., X_{t-1}]. \tag{14}$$

In this alternative process output can be viewed as $e^{P_k(a_i+\varepsilon_{it})}$, and profits are a fraction of output. The latter part is harder to interpret in a standard equilibrium setting, but nevertheless this specification gives rise to similar switching patterns, as we will see now. It corresponds to the specification in Gibbons, Katz, Lemieux, and Parent (2005), (who also have additional additive terms in the exponent capturing occupational and overall tenure and other observed characteristics of the worker). Expression (14) is equal to

$$e^{\{P_k A_{it} + (1/2) P_k^2 \phi_t^{-2} - \Pi_k\}}.$$

Workers sort themselves to the occupation with the highest expected wage. Since the ranking of wages is preserved under monotone transformations, we can take logarithms and obtain the sorting criterium:

$$P_k A_{it} \ - \Omega_{kt},$$

where $\Omega_{kt} := \Pi_k + (1/2)P_k^2\phi_t^{-2}$ now reflects the opportunity cost of obtaining the revenue P_kA_{it} in occupation k, in contrast to only Π_k in our model. This is due to the fact that the upside potential of uncertainty is larger than the downside potential after exponentiating. This makes young employees especially attractive, as their uncertainty is higher. To see this formally, note that a worker will choose occupation k if his belief satisfies $A_{i,t} \in [B_{k,t}, B_{k+1,t})$ where the cutoffs $B_{kt} = \Omega_{kt} - \Omega_{k-1,t}/(P_k - P_{k-1})$. This still has the potential to generate U-shapes, but since B_{kt} is increasing in labor market experience t, older agents with the same belief as younger agents sort themselves into a lower occupation, yielding a downward drift. If that drift is too strong, then there will be no U-shapes if workers are paid their expected wage. This downward drift can be offset once accumulation of general human capital is introduced, since it induces an upward drift, yielding overall the potential for a balanced U-shape.

Based on wages according to (14), Gibbons, Katz, Lemieux, and Parent (2005) propose a method of quasi-differencing of the wages and using lagged occupational choices as instruments to estimate the underlying parameters. In this paper we provide evidence on mobility patterns and show that it is consistent with the type of selection that Gibbons, Katz, Lemieux, and Parent (2005) provide a method to control for. Since their method can be adapted to the setting in this paper, we view the two papers as complementary to each other.

6 Conclusion

Using administrative panel data on 100% of Danish population we document a new set of facts characterizing the patterns of occupational mobility. We find that a worker's probability of switching occupation is U-shaped in his position in the wage distribution in his occupation. It is the workers with the highest or lowest wages in their occupations who have the highest probability of leaving the occupation. Workers with higher (lower) relative wage within their occupation tend to switch to occupations with higher (lower) average wages. Higher (lower) paid workers within their occupation tend to leave it when relative productivity of that occupation declines (rises) steeply.

To account for these patterns we suggest that it might be productive to think of occupations as forming vertical hierarchies. Complementarities between the productivity of an occupation and the ability of the workers induces workers to sort themselves into occupations based on their absolute advantage. Since their absolute advantage is not fully known initially they update after observing their output and re-sort themselves according to their update on their ability. Employment opportunities in each occupation are scarce, inducing competition among workers for them. We present an equilibrium model of occupational choice with these features and show analytically that it is consistent with patterns of mobility described above.

This theory captures the patterns of occupational mobility in a very parsimonious model. In particular, it generates the patterns of "promotions" and "demotions" observed in the data. An investigation of the occupational classification suggests that both of these switches up or down the occupational hierarchy represent real occupational changes in the sense that the required skill set changes substantially. Moreover, it is essential to take the pattern of selection implied by the model into account to estimate the returns to occupational tenure, interpret earnings dynamics, and to assess the effects of economic policies. While neither models of learning in the absence of occupational differentiation (horizontal learning) nor models solely of comparative advantage generate the data patterns that we find, a tractable combination of the two acocunts well for the observed pattern. We also show that standard upward career progression due to human captial accumulation can easily be intergrated into the framework, yet by itsself fails to account for the downward movements observed in the data.

The analysis in this paper shows the qualitative ability of the model to account for the new data patterns that we find (and for a number of patterns documented in prior work). Our simulations suggests that the model might also generate the right quantitative magnitudes. In terms of the future agenda, the main objective is to explore more fully its quantitative implications. In particular, while we think that the vertical sorting mechanism we described is an important part of any comprehensive theory of occupational mobility, it appears unlikely that it accounts for the full extent of occupational mobility. The main goal in this agenda will be to embed and distinguish different economic forces − such as learning, fluctuations in occupational productivities or demands, and search frictions in locating jobs in various occupations[35] − in a dynamic general equilibrium model of occupational choice and to quantitatively evaluate their contribution to the amount of occupational mobility observed in the data. The key challenge in this regard is to identify the sets of occupations forming distinct hierarchies in the data and the extent of transferability of skills across occupations within and across these hierarchies.[36] For

[35]If agents cannot instantaneously change jobs, but have to go through a search phase before they find a new job, adjustment based on new information is not instantaneous. Nevertheless, if search frictions are sufficiently small the allocation is close to the competitive outcome that we outline in this work and we expect the basic properties to carry over (for convergence when the periods between search activities becomes small see for example Atakan (2006a,b) and for convergence when the short side of a market gets matched with near certainty see Eeckhout and Kircher (2010)).

[36]Note that the parameters of the stationary environments in Section 3 and Section A2 such as occupational productivity P_k, profits Π_k and human capital accumulation functions $H(t)$ and $h_k(\tau)$ can be consistently estimated using the methodology proposed by Gibbons, Katz, Lemieux, and Parent (2005) even if the econometrician does not know exactly which occupation belongs to which hierarchy. It suffices that the workers know this. If they stay within distinct hierarchy, their past choices serve as instruments. However, if the environment is not stationary (as in Section 4.1) or if switching occurs also across hierarchies, further investigation is necessary.

example, one hierarchy could be electrical equipment assembler, electrician, electrical engineer, and manager. Another could be truck driver, taxi driver, motor vehicle mechanic, and sales representative. Yet another could be an economics consultant, economics professor, and dean. It is likely that switches within and across these hierarchies are present in the data. It is also likely that human capital is not equally transferable within and between hierarchies. Developing a way to identify such hierarchies in the data and the transferability of human capital between them seems essential to enable future quantitative analysis.

References

Alvarez, F., and R. Shimer (2009): "Unemployment and Human Capital," mimeo, University of Chicago.

Atakan, A. E. (2006a): "Assortative Matching with Explicit Search Costs," *Econometrica*, 74(3), 667–681.

——— (2006b): "Assortative Matching with Explicit Search Costs: Existence and Asymptotic Analysis," Working Paper, Northwestern University.

Becker, G. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Biddle, J., and K. Roberts (1994): "Private Sector Scientists and Engineers and the Transition to Management," *Journal of Human Resources*, 29(1), 82–107.

Bils, M., and K. J. McLaughlin (2001): "Interindustry Mobility and the Cyclical Upgrading of Labor," *Journal of Labor Economics*, 19(1), 94–135.

Chamley, C. P. (2004): *Rational Herds: Economic Models of Social Learning*. Cambridge University Press.

Dahl, C., D. le Maire, and J. Munch (2009): "Wage Dispersion and Decentralization of Wage Bargaining," Discussion Paper No. 09-15, University of Copenhagen.

Eeckhout, J., and P. Kircher (2009): "Identifying Sorting in Theory," Working Paper, University of Pennsylvania.

——— (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78(2), 539–574.

Eeckhout, J., and X. Weng (2009): "Assortative Learning I: Frictionless Case," mimeo, University of Pennsylvania.

Gibbons, R., and L. F. Katz (1992): "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *Review of Economic Studies*, 59(3), 515–535.

Gibbons, R., L. F. Katz, T. Lemieux, and D. Parent (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 23(4), 681–724.

Gibbons, R., and M. Waldman (1999): "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, 114(4), 1321–1358.

Groes, F. (2010): "Occupational Experience, Mobility, and Wages: Patterns in the Danish Data," Manuscript, Copenhagen Business School.

Hall, R. E., and R. Kasten (1976): "Occupational Mobility and the Distribution of Occupational Success among Young Men," *American Economic Review, Papers and Proceedings*, 66, 309–315.

Heckman, J. J., and B. E. Honore (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–1149.

Johnson, W. R. (1978): "A Theory of Job Shopping," *Quarterly Journal of Economics*, 92(2), 261–278.

Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

Jovanovic, B., and Y. Nyarko (1997): "Stepping Stone Mobility," *Carnegie-Rochester Conference Series on Public Policy*, 46(1), 289–326.

Kambourov, G., and I. Manovskii (2005): "Accounting for the Changing Life-Cycle Profiles of Earnings," mimeo, The University of Pennsylvania.

——— (2008): "Rising Occupational and Industry Mobility in the United States: 1968-1997," *International Economic Review*, 49(1), 41–79.

——— (2009a): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76(2), 731–759.

——— (2009b): "Occupational Specificity of Human Capital," *International Economic Review*, 50(1), 63–115.

Lucas, R. J., and E. Prescott (1974): "Equilibrium Search and Unemployment," *Journal of Economic Theory*, 7, 188–209.

McCall, B. P. (1990): "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, 98(1), 45–69.

Miller, R. A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92(6), 1086–1120.

Mincer, J. (1974): *Schooling, Experience, and Earnings*. Columbia University Press, New York, NY.

Moscarini, G. (2001): "Excess Worker Reallocation," *Review of Economic Studies*, 68, 593–612.

Moscarini, G., and K. Thomsson (2007): "Occupational and Job Mobility in the US," *Scandinavian Journal of Economics*, 109(4), 807–836.

Neal, D. (1999): "The Complexity of Job Mobility Among Young Men," *Journal of Labor Economics*, 17(2), 237–261.

Papageorgiou, T. (2007): "Learning Your Comparative Advantages," mimeo, Pennsylvania State University.

——— (2010): "Large Firms and Internal Labor Markets," mimeo, Pennsylvania State University.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Shaw, K. (1984): "A Formulation of the Earnings Function Using the Concept of Occupational Investment," *Journal of Human Resources*, 14, 319–40.

——— (1987): "Occupational Change, Employer Change, and the Transferability of Skills," *Southern Economic Journal*, 53, 702–19.

Sichernam, N., and O. Galor (1990): "A Theory of Career Mobility," *Journal of Political Economy*, 98(1), 169–192.

Sullivan, P. J. (2009): "Empirical Evidence on Occupation and Industry Specific Human Capital," *Labour Economics*, forthcoming.

Wilk, S. L., and P. R. Sackett (1996): "Longitudinal Analysis of Ability - Job Complexity Fit and Job Change," *Personnel Psychology*, 49(4), 937–967.

Zangelidis, A. (2008): "Occupational and Industry Specificity of Human Capital in the British Labour Market," *Scottish Journal of Political Economy*, 55(4), 420–443.

APPENDICES

A1 Omitted Proofs and Derivations

Remainder of Proof of Proposition 1:

Proof. Consider wage setting process (6) and associated switching probability $S_{k,t}$ first. Define $\delta_k = (B_{k+1} - B_k)/2$ to be half of the distance of interval $[B_k, B_{k+1})$, and recall that $\overline{B}_k = B_k + \delta_k$. Any other belief A can be written in terms of the distance δ from $\overline{B}_k$. Then

$$\begin{aligned} & S_{k,t}(P_k\overline{B}_k - \Pi_k) - S_{k,t}(P_k(\overline{B}_k + \delta) - \Pi_k) \\ = \;& G_t(B_k|\overline{B}_k) - G(B_k|\overline{B}_k + \delta) + G_t(B_{k+1}|\overline{B}_k + \delta) - G_t(B_{k+1}|\overline{B}_k) \qquad \text{(A1)} \\ = \;& G_t(-\delta_k|0) - G(-\delta_k - \delta|0) + G_t(\delta_k - \delta|0) - G_t(\delta_k|0) \\ = \;& \int_0^{\delta} \left[g_t(-\delta_k - \varepsilon|0) - g_t(\delta_k - \varepsilon|0)\right] d\varepsilon, \qquad \text{(A2)} \end{aligned}$$

where the second equality follows from lateral adjustment. Clearly this distance is zero when $\delta = 0$. Symmetry around zero and single-peakedness imply that the integrand in (A2) is strictly negative for any $\varepsilon > 0$. Therefore, this integral is strictly negative for $\delta > 0$. When $\delta < 0$ the integrand of (A2) is positive for all relevant ε but the integral is negative because of integration from *zero* to a negative number. The proposition obtains because integral (A2) decreases in the absolute value $|\delta|$.

Now consider instead the wage setting process (5). If we observe the worker of cohort t in occupation k at wage w, (5) implies that his output must have been $X(w) = (w + \Pi_k)/P_k$. If we knew the prior A that this person had, we could by (3) calculate his posterior as $A' = \alpha A + (1-\alpha)X(w)$, where $\alpha = \phi_t/\phi_{t+1}$.

Since we only know the wage but do not know his prior A, we can only determine the range of priors for which the worker would switch. He switches up if $A' \geq B_{k+1}$,which we can rewrite as $\alpha A + (1-\alpha)X(w) \geq B_{k+1}$. He switches down if $A' \leq B_k$,which we can rewrite as $\alpha A + (1-\alpha)X(w) \leq B_k$. Since the worker chose occupation k in period t, we know that $A \in [B_k, B_{k+1})$. Therefore, neither of the two inequalities can be satisfied if $X(w) \in [B_k, B_{k+1})$ or equivalently $w \in [B_kP_k - \Pi_k, B_{k+1}P_k - \Pi_k)$. Therefore, for such intermediate wages the switching probability $s_{k,t}(w) = 0$, which constitutes a local minimum.

We complete the proof by showing that very high and low wages constitutes a local maximum for the switching probability. This is the case since at such wages the switching probability is one. Since $A \in [B_k, B_{k+1})$, the condition for upward switching is satisfied for any of the priors if it holds for the lowest possible prior, i.e., $\alpha B_k + (1-\alpha)X(w) \geq B_{k+1}$ which yields equivalently $X(w) \geq \frac{B_{k+1} - \alpha B_k}{1-\alpha}$ or $w \geq \frac{B_{k+1} - \alpha B_k}{1-\alpha}P_k - \Pi_k$. So for wages above this threshold $s_{k,t}(w) = 1$. Similarly, the condition for downward switches is satisfied for all priors if it holds for the highest prior, which means $\alpha B_{k+1} + (1-\alpha)X(w) \leq B_k$ or equivalently $X(w) \leq \frac{B_k - \alpha B_{k-1}}{1-\alpha}$ or $w \leq \frac{B_k - \alpha B_{k-1}}{1-\alpha} P_k - \Pi_k$. For such low wages again the switching probability is $s_{k,t}(w) = 1$.

This establishes the U-shape property. It does not establish strict U-shapes even though the range of priors at which the workers will switch expands with the distance of the wage from the "no-switching" region $[B_k P_k - \Pi_k, B_{k+1} P_k - \Pi_k)$. Consider a wage realization $w > B_{k+1} P_k - \Pi_k$ and a different wage realization $w' > w$. After the first, agents with priors in (A, B_{k+1}) switch for some A, while after the latter agents with priors in (A', B_{k+1}) switch, and $A' < A$ because at the higher wage updating is stronger. While this might suggest that more agents switch after w', this need not be true. The probability that the prior is in (A, B_{k+1}) *conditional* on observing w may in fact be higher than the probability that the prior was in (A', B_{k+1}) *conditional* on realizing wage w'. One can construct examples where this happens, and in such a case more agents switch after w than after w'. This arises because the conditional probability does not have to be monotone. ■

Remainder of Proof of Proposition 2:

Proof. Consider workers that chose interior occupation k in their t^{th} year in the labor market. We will use the notation as in the proof of Proposition 1, and exploit the following result shown there: Workers switch only if $X(w)$ is either below B_k or above B_{k+1}; if they switch and $X(w) \leq B_k$, they switch down; if they switch and $X(w) \geq B_{k+1}$, they switch up. Note that $X(w) \leq B_k$ is equivalent to $w \leq B_k P_k - \Pi_k$, while $X(w) \geq B_{k+1}$ is equivalent to $w \geq B_{k+1} P_k - \Pi_k$. Conditional on switching, the switch will be downward with probability 1 if $w \leq B_k P_k - \Pi_k$ and will be upward with probability 1 if $w \geq B_{k+1} P_k - \Pi_k$, leading to an increasing schedule. ■

Proof of Proposition 4:

Proof. Consider workers with t years of labor market experience that chose occupation k and those that chose occupation k'. In year $t+1$ we compare their wages, conditional on choosing k'. All workers that we compare have some belief in $[B_{k'-1}, B_{k'+2})$ in period t, and a belief in $[B_{k'}, B_{k'+1})$ in period $t+1$ of their work-life. The distribution of the update is concave in the relevant region if $B_{k'+2} - B_{k'-1}$ is not too large since normal distributions are concave around their mean (in particular we require $B_{k'+2} - B_{k'-1} < \sqrt{\phi_{t+1}/(\phi_\epsilon + \phi_t)}$). Since by market clearing $F(B_{k'+2}) - F(B_{k'-1}) = \gamma_{k'+1} + \gamma_{k'} + \gamma_{k'-1}$, this holds if the measure of firms in these occupations is not too large.

The workers' update A_{t+1} is distributed symmetrically around A_t. If $k' > k$, the density $g_t(A_{t+1}|A_t)$ of the update evaluated at any point A_{t+1} in $[B_{k'}, B_{k'+1})$ is higher (because of symmetry and single-peakedness) and has a larger derivative (because of concavity) for any stayer (person with $A_t \in [B_{k'}, B_{k'+1})$ than for any switcher (person with $A_t \in [B_{k'-1}, B_{k'})$). It then follows directly that the conditional distribution of the update, conditional on $A_{t+1} \in [B_{k'}, B_{k'+1})$, for stayers first order stochastically dominates the distribution for switchers. The implication for expected wages follows immediately. For $k' < k$ the density is still higher but the derivative is lower, which directly implies that the distribution for switchers first order stochastically dominates the distribution for stayers. ■

A2 Free Entry into Occupations

In the main body of the paper we have taken the number of jobs per occupation as fixed. Here we briefly outline that the model extends to an economy in which jobs can be created at some opportunity cost. Clearly entry costs have to differ between occupations to sustain several occupations with different productivities (since otherwise only the most productive occupations will operate). Assume that the per-period cost to create and maintain a job in occupation k (or r, if we adopt the notation from section 4.1) is given by $C_k(\gamma_k) = \bar{c}_k + c(\gamma_k)$, except for home production sector $k = 0$ where entry costs are $C_0(\gamma_0) = 0$. That is, there is a fixed cost $\bar{c}_k$ independent of the number of other entrepreneurs who create jobs, and a component $c(\gamma_k)$ that depends on the overall number of entrants into the occupation.

If we assume that $c(\gamma_k) = 0$, then we have perfectly elastic supply of jobs. This corresponds to a model in which workers can simply rent jobs at cost $\bar{c}_k$. Occupations with lower productivity have to have lower costs as otherwise no worker would rent the job. The model is particularly simple to solve because firms profits are exogenously tied to the entry costs:

$$\Pi_k = \bar{c}_k. \tag{A3}$$

This entry assumption corresponds to the standard Roy models which are essentially decision-theoretic: any worker that wants to enter occupation k can do so by "buying" a machine at cost c_k, there are no further congestion effects, and competition between workers is essentially absent.

The drawback of having only fixed costs $\bar{c}_k$ is the response of the market when productivities change over time, as we analyzed for the basic model in Section 4.1. In a model with absolute advantage, if an occupation becomes more productive than another one but retains its lower entry cost, then the other occupation completely disappears. There are various reasons why we don't expect this to occur: Prices might change in response to output changes or costs might change in response to the number of jobs in the occupation. Costs change for example if there is heterogeneity among entrepreneurs and $c(\gamma_k)$ reflects the costs of the marginal entrant: the more entrepreneurs enter the less able the marginal one is.[37] We integrate this idea into the model by assuming that $c(.)$ is increasing and convex. If prices are always high enough to cover the fixed cost, then Inada conditions on the second component ensure that even with changing

[37] In this interpretation all infra-marginal entrants will generate profits larger than their costs. Only the marginal entrant will be exactly indifferent to entering.

productivities no occupation completely vanishes, but the level of operation might substantially vary.[38,39]

An equilibrium is now a tuple $\Pi = (\Pi_0, ..., \Pi_K)$ of profits and a tuple $\gamma = (\gamma_0, ...\gamma_K)$ of entry levels such that all conditions in Equilibrium Definition 1 are satisfied and additionally it holds that $\Pi_k = C(\gamma_k)$ for all $k > 0$. All results regarding switching behavior from Section 3 apply, only that now the cutoffs B_k are determined in a way that incorporates optimal entry. It is easy to solve for these cutoffs by considering the following set of equations in analogy to (7) and (8)

$$\frac{C(\gamma_k) - C(\gamma_{k-1})}{P_k - P_{k-1}} = B_k, \tag{A4}$$

$$F(B_k) - F(B_{k-1}) = \gamma_k, \tag{A5}$$

for all $k > 0$.

Equation system (A4) and (A5) allows us to determine the size of each occupation in each period even in the case when productivities are changing as in Section 4.1. We can now define an improving occupation in the sense of Proposition 5 as one that improves its position at both the high and the low end, i.e. $\Gamma_r^{\tau+1} > \Gamma_r^{\tau}$ and $\Gamma_r^{\tau+1} - \gamma_r^{\tau+1} > \Gamma_r^{\tau} - \gamma_r^{\tau}$, where again superscripts indicate the time period. A sufficient increase additionally means $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r^{\tau}$. With these extended definitions the proposition remains valid. If on the other hand an occupation with increasing productivity expands so much in size that the measure of jobs with strictly lower productivities $\Gamma_r - \gamma_r$ actually decreases, it starts to employ not only more high ability but also more low ability workers. When we consider a smooth increase in the productivity of occupation r and hold the other productivities fixed, it is easy to see that the expansion of the workforce is continuous but the position switches upward when it overtakes another occupation, at which point indeed both upper and lower position Γ_r and $\Gamma_r - \gamma_r$ increase jointly and the ability of the work force improves substantially in the sense of first order stochastic dominance.

[38]In particular, it is easy to verify that the following conditions ensure employment in all occupations $k > 0$ in all periods. Assume that $c'(0) = 0$ and there is some constant $\psi > 0$ and employment level $e = [\alpha T - F(\psi)]/K$ such that $\lim_{\gamma \to e} c'(\gamma) = \infty$, which ensures that no occupation employs more than e workers. Moreover, assume that prices evolve according to some (possibly stochastic) process with the feature that there exists a lowest price $P > 0$. That is, no occupation $k > 0$ ever draws a price below P. Then $\psi P > \max_k \bar{c}_k$ ensures that it is optimal to have at least some employment in each occupation at each point in time because the worker with ability ψ never gets employed and therefore could be hired for free.

[39]Another alternative formulation that ensures the operation of all occupations is that prices are changing while entry costs remain constant, i.e. $P_k(\gamma_k)$ is dependent on the level of employment and C_k is fixed. Together with some Inada conditions still all occupation remain active, but the requirement that $\Pi_k = C_k$ implies that the equilibrium ordering of the productivities $P_k(\gamma_k)$ of occupations cannot change.

A3 Equilibrium Definition with Human Capital and Switching Costs

The output-contingent wages of workers are still given by (5), where output is now determined by (12). The expected wage for a worker in occupation k with prior mean A at the beginning of his t^{th} period in the labor force and his ι^{th} consecutive period in this occupation is in analogy to (6)

$$W_k(A, t, \iota) = P_k[A + H(t) + h_k(\iota)] - \Pi_k.$$

For any given profit vector $\Pi = (\Pi_0, ..., \Pi_K)$ the worker can forecast his expected wage in all occupations for given prior and given experience. He can then evaluate his optimal choice of occupation by simple backward induction. His state vector at the beginning of each period is (t, k, ι, A) : his year in the labor market t, the occupation k he was last employed in, his consecutive years of experience in this occupation ι, and his belief about his mean ability A. Newborns start with home production as their previous occupation. In the last year of his life the worker optimizes

$$V(T, k, \iota, A) = \max \left\{ W_k(A, T, \iota), \max_{m \neq k} \{W_m(A, T, 1) - \kappa_m\} \right\},$$

i.e. he chooses whether to stay in his previous occupation or to switch to a new occupation where this would be his first year of experience and pay the switching costs. This gives a decision rule $d(T, k, \iota, A|\Pi) \in \{0, ..., K\}$ regarding the occupation that the worker chooses given the profits that firms make. Similarly, a worker with $t < T$ years of experience maximizes his expected payoff including the continuation value

$$V(t, k, \iota, A) = \max \left\{ \begin{array}{c} W_k(A, t, \iota) + \beta E_{A'} V(t+1, k, \iota+1, A'), \\ \max_{m \neq k} \{W_m(A, t, 1) - \kappa_m + \beta E_{A'} V(t+1, m, 2, A')\} \end{array} \right\},$$

where $\beta \in (0, 1]$ is the discount factor and A' is the update about the worker's mean ability. The solution to this problem gives again a decision rule $d(t, k, \iota, A|\Pi) \in \{0, ..., K\}$. It is straightforward to show that for given profit vector Π these decision rules are unique for almost all ability levels A. Given the distribution $F_t(A)$ of priors of each cohort and these decision rules, one can derive for given Π the steady-state number of agents that choose occupation k, call it $v_k(\Pi)$. Similar to Equilibrium Definition 1 we can now define:

Definition 6 *An equilibrium is a vector of profits* $(\Pi_0, ...\Pi_K)$ *such that* Π_0 *and* $v_k(\Pi) = \gamma_k$ *for all* $k > 0$.

A4 Appendix Tables

Table A-1: Summary statistics for the Large and Small samples and subsamples.

	Small Sample		Large Sample	
	Sample A	Sample B	Sample A	Sample B
Number of observations	402,136	368,520	1,292,932	1,229,339
Number of occupations	229	143	324	242
Age	29.66	29.49	33.25	33.20
Occupational tenure	4.41	4.41	4.54	4.57
Occupational spell number	1.69	1.67	2.30	2.30
Occupational switchers	0.18	0.17	0.19	0.18
Employer tenure	2.36	2.33	2.78	2.77
Employer switchers	0.18	0.19	0.15	0.15
Industry tenure	3.38	3.35	3.78	3.79
Years after graduation	6.49	6.40	9.56	9.54
12 years of school or less	0.73	0.74	0.65	0.66
13 years of school or more	0.27	0.26	0.35	0.34
Hourly wage in DKK in 1995	170.13	168.75	172.66	172.29
Married	0.30	0.30	0.42	0.42
Number of children	0.71	0.70	0.94	0.94

Note − The table contains the descriptive summary statistics of the Large and Small samples defined in the main text. For each of the two main samples two subsamples A and B are defined. Sample A imposes a restriction that there are at least 10 workers in an occupation in a given year. Sample B imposes a restriction that there are at least 10 workers from the same cohort (defined by the year of completing education) in an occupation in a given year.

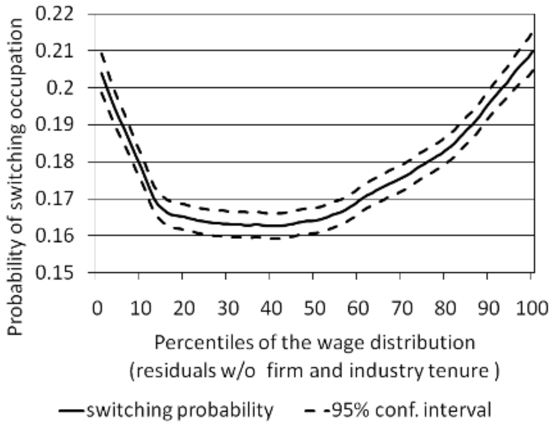
Table A-2: Net wage changes for various occupational transitions.

	Type of occupational transition	
	Switch up	Switch down
All workers	2.22	-0.67
	(0.17)	(0.19)
Firm stayers	1.16	-0.75
	(0.19)	(0.21)
Firm switchers	2.07	-2.73
	(0.45)	(0.48)

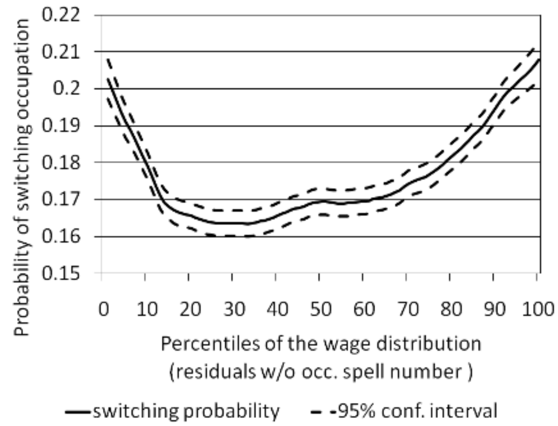
Note − The table contains wages changes for workers experiencing various types of occupational transitions net of the wage change of the corresponding group of occupational stayers. Standard errors in parenthesis.

A5 Appendix Figures

A5.1 Alternative Wage Regression Specifications



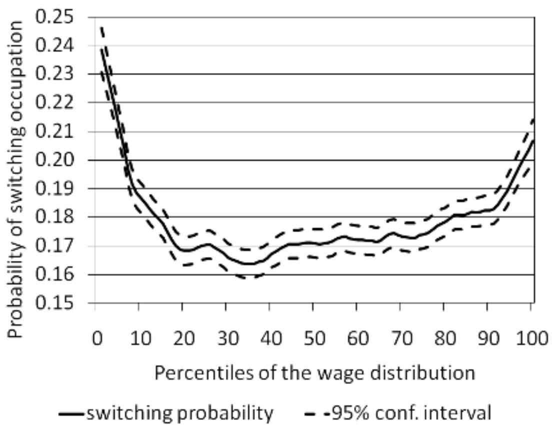
(a) Wage regression excluding firm and industry tenure.



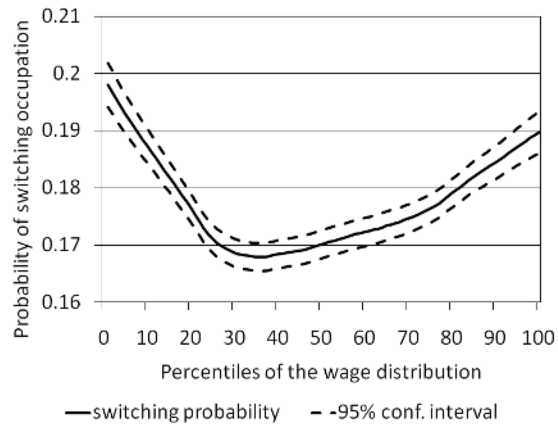
(b) Wage regression excluding occupational spell number.

Figure A-1: Non-parametric plot of probability of switching occupation by worker's percentile in residual distributions from alternative wage regression specifications.

A5.2 Sensitivity to Bandwidth Choice

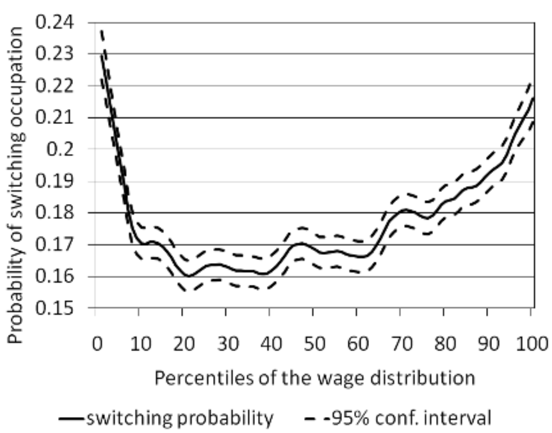


(a) Half bandwidth.

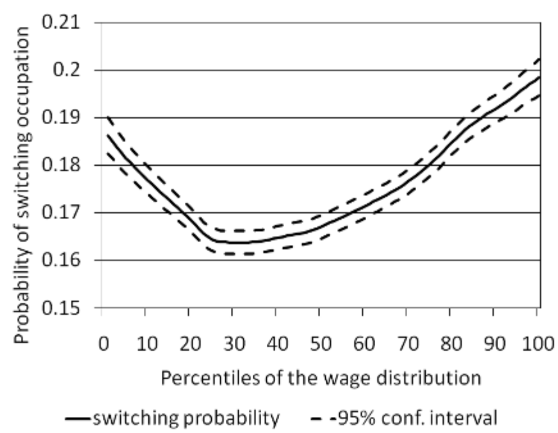


(b) Double bandwidth.

Figure A-2: Non-parametric plot of probability of switching occupation by worker's percentile in the wage distribution within occupation and year for half and double bandwidth.

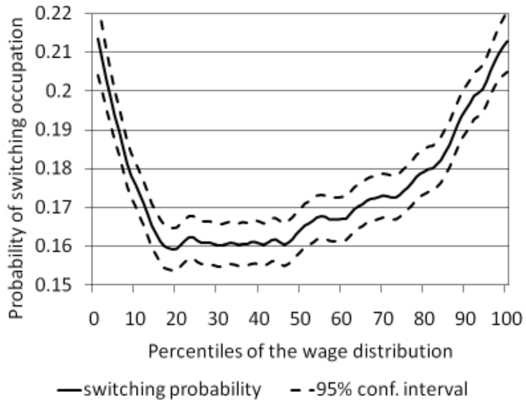


(a) Half bandwidth.

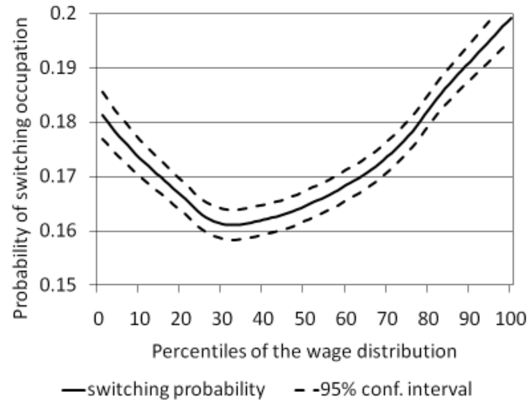


(b) Double bandwidth.

Figure A-3: Non-parametric plot of probability of switching occupation by worker's percentile in the distribution of wage residuals for half and double bandwidth.



(a) Overall. Half bandwidth.



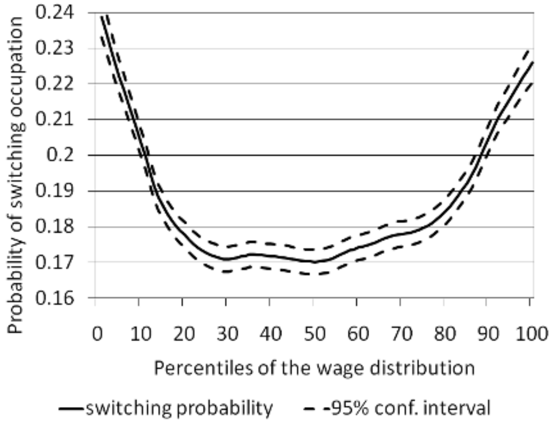
(b) Overall. Double bandwidth.

(c) For various years after graduation. Half bandwidth.

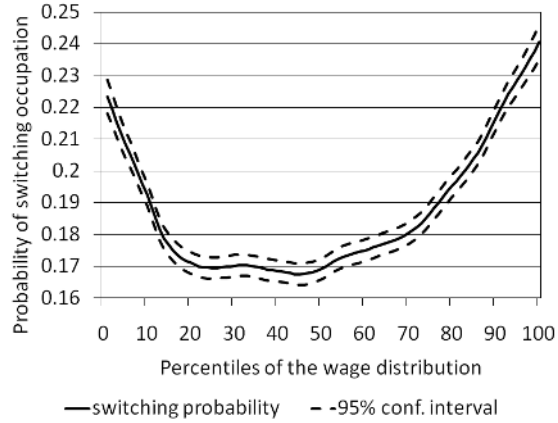
(d) For various years after graduation. Double bandwidth.

Figure A-4: Non-parametric plot of probability of switching occupation by worker's percentile in the wage distribution within occupation, year, and years after graduation for half and double bandwidth.

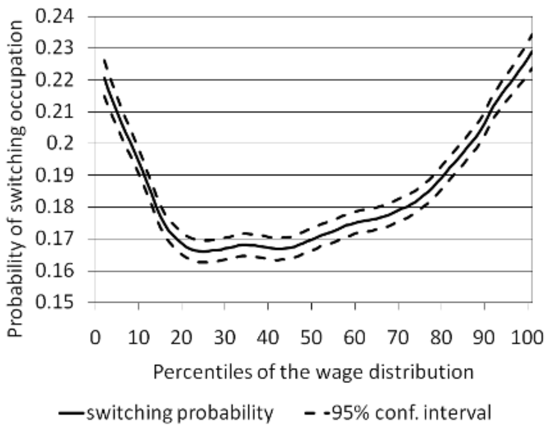
A5.3 Results on the Large Sample



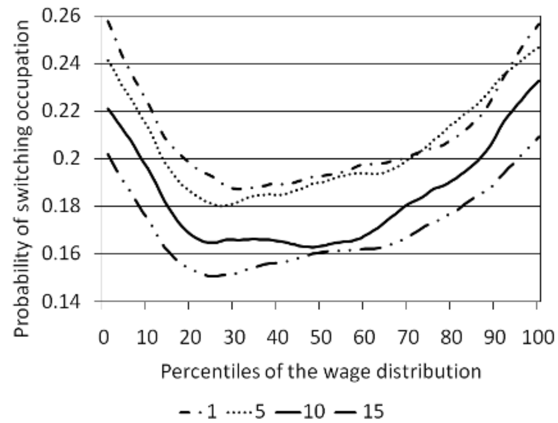
(a) Distribution of raw wages within occupation and year.



(b) Distribution of wages residuals.

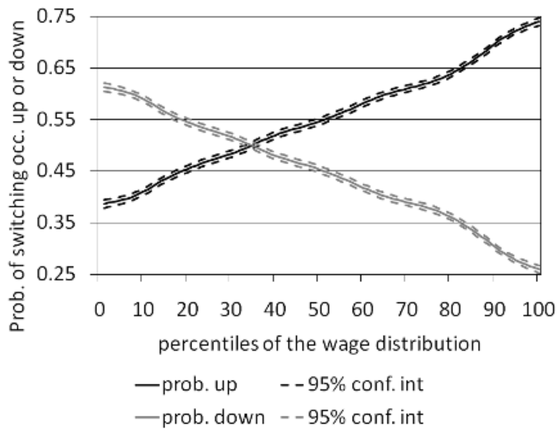


(c) Distribution of raw wages within occupation, year, and year after graduation.

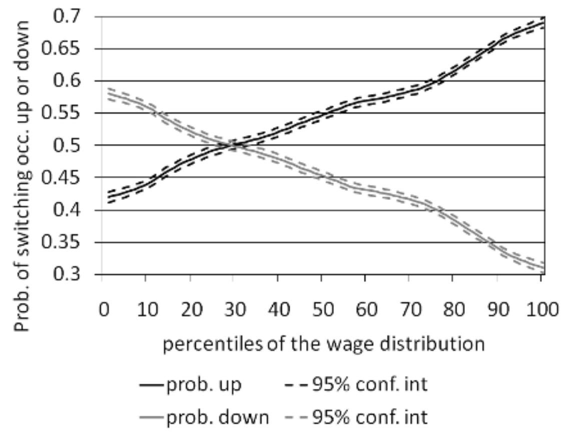


(d) Distribution of raw wages within occupation, year, and year after graduation for various years after graduation.

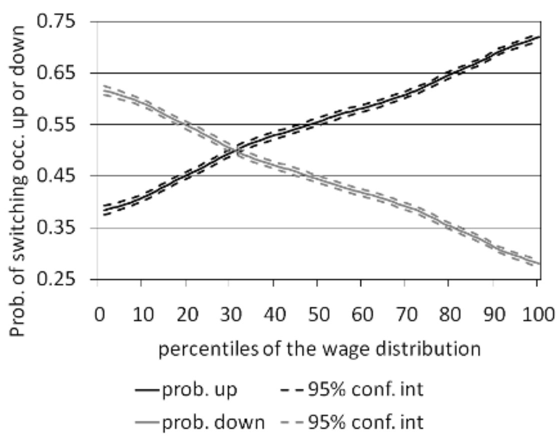
Figure A-5: Non-parametric plot of probability of switching occupation by worker's percentile in the relevant wage distribution. Large Sample.



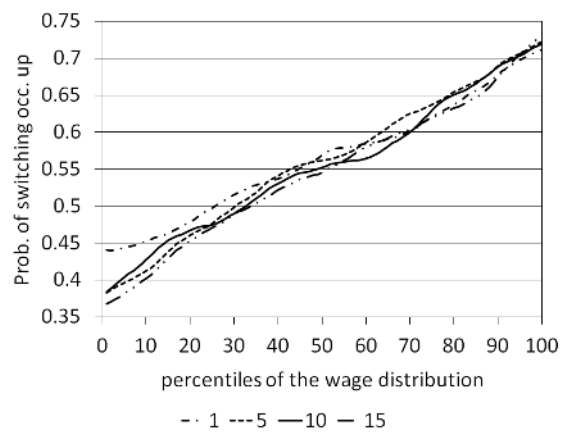
(a) Distribution of raw wages within occupation and year. Average wage in occupation from population.



(b) Distribution of wages residuals. Average wage in occupation from time constants in wage regression.

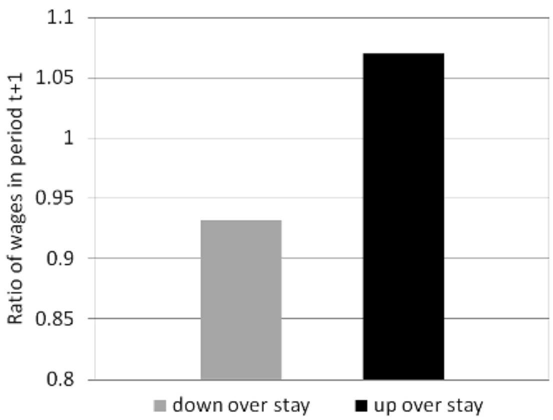


(c) Distribution of raw wages within occupation, year, and year after graduation. Average wage in occupation from population.

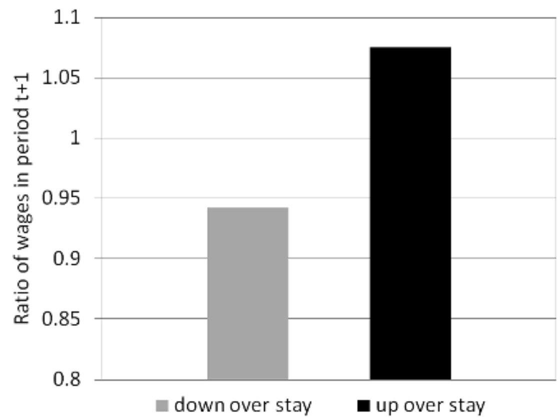


(d) Distribution of raw wages within occupation, year, and year after graduation for different years after graduation. Average wage in occupation from population.

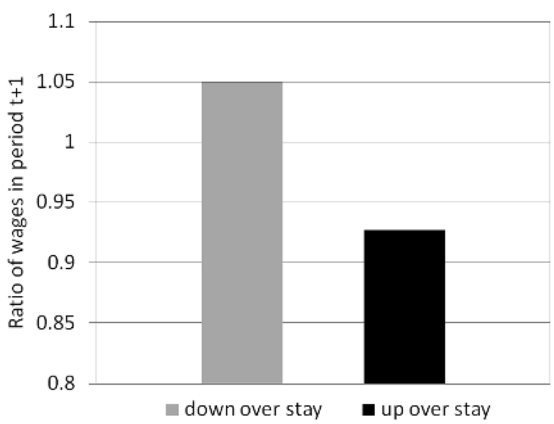
Figure A-6: Non-parametric plot of direction of occupational mobility, conditional on switching occupation, by worker's percentile in the relevant wage distribution before the switch. Large Sample.



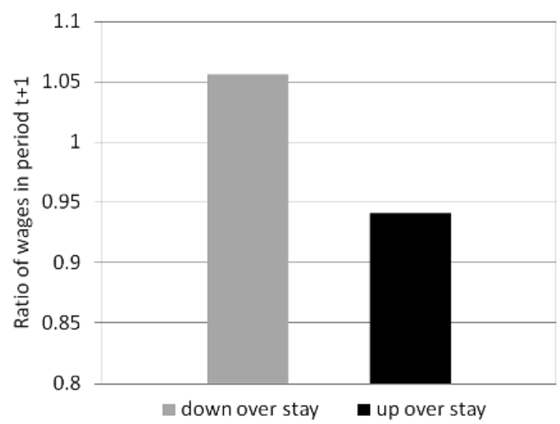
(a) Unconditional.



(b) Conditional on the number of years after graduation.

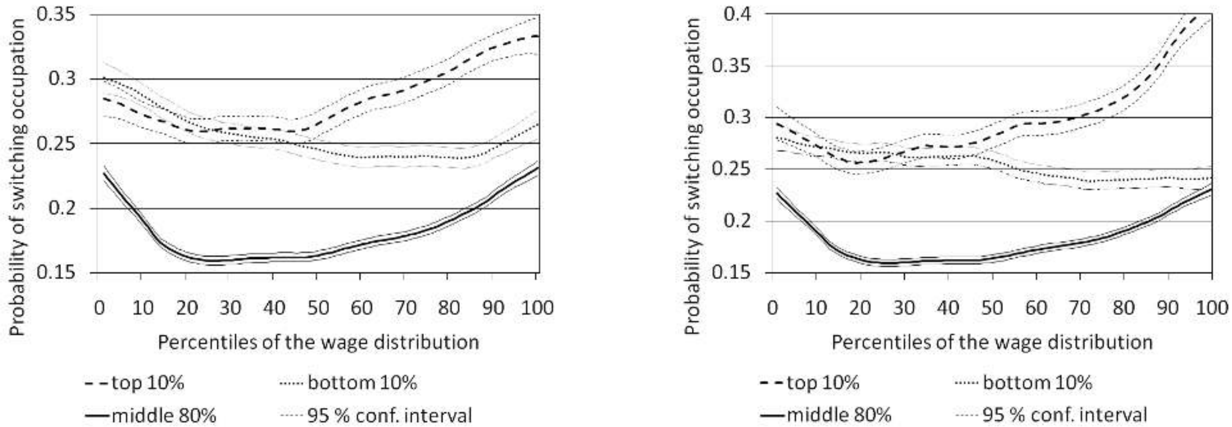


(c) Unconditional.



(d) Conditional on the number of years after graduation.

Figure A-7: Weighted average of year $t+1$ ratios of real wages of workers who switch occupations between years t and $t+1$ over (1) workers who stay in the same *original* occupation in years t and $t+1$ (Panels 7(a) and 7(b)) or (2) workers who stay in the same *destination* occupation in years t and $t+1$ (Panels 7(c) and 7(d)) by direction of the switch (i.e., whether the switch involves moving to an occupation that pays more or less on average than the source occupation). Large Sample.



(a) Distribution of raw wages within occupation and year. Growth rates of average wage in occupation from population.

(b) Distribution of wage residuals. Growth rates of average wage in occupation from time constants in wage regression.

Figure A-8: Non-parametric plot of probability of switching occupation by worker's percentile in the relevant wage distribution. For the fastest growing 10% of occupations, the slowest growing 10% of occupations, and the remaining 80% of occupations. Large Sample.

A5.4 Patterns of Occupational Mobility Within and Across Firms Conditional on Worker's Position in the Distribution of Wage Residuals

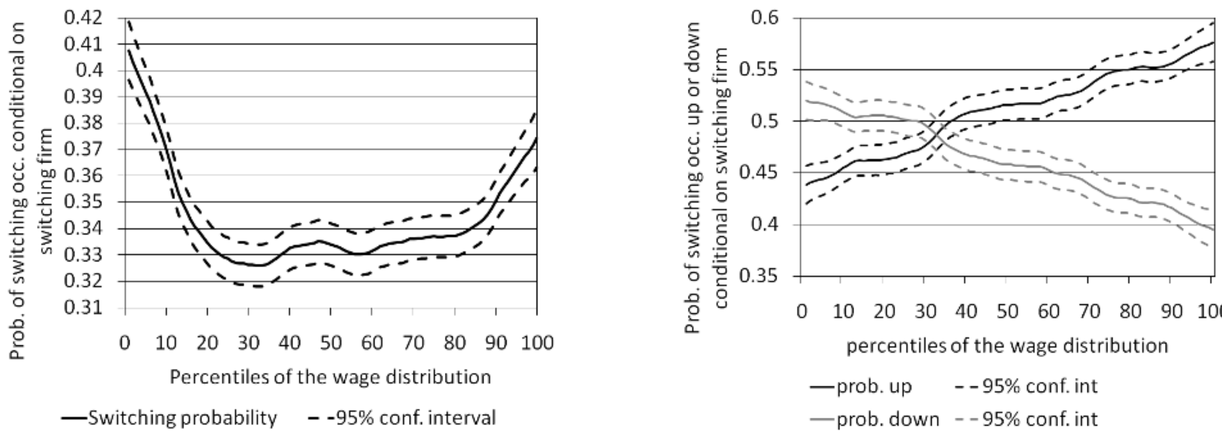


Figure A-9: Non-parametric plots of probability of switching occupation and of direction of occupational mobility *conditional on switching firms* by worker's percentile in the distribution residual wages.

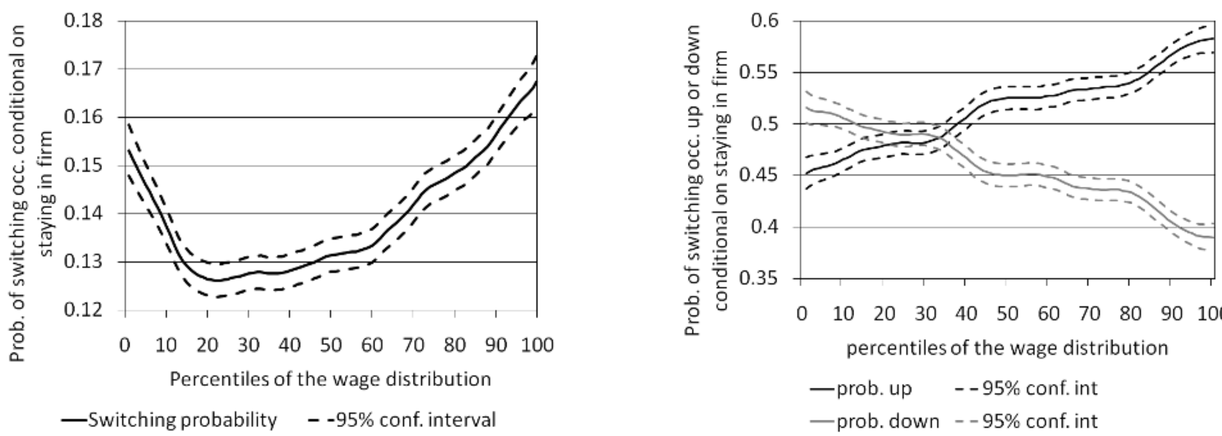
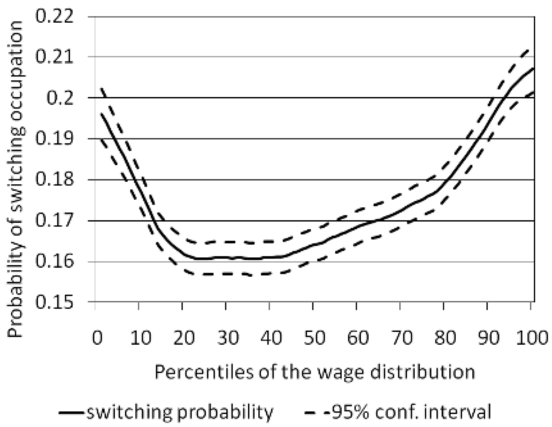
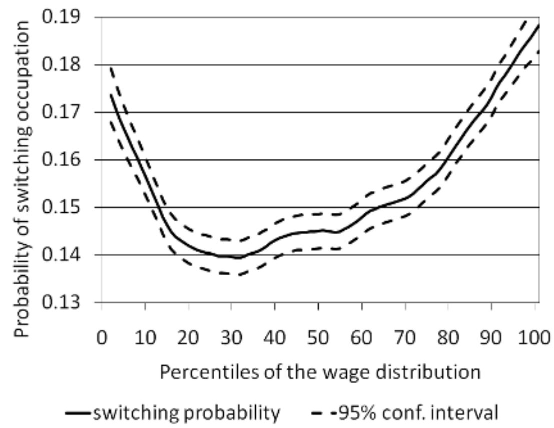


Figure A-10: Non-parametric plots of probability of switching occupation and of direction of occupational mobility *conditional on staying with the firm* by worker's percentile in the distribution of residual wages.

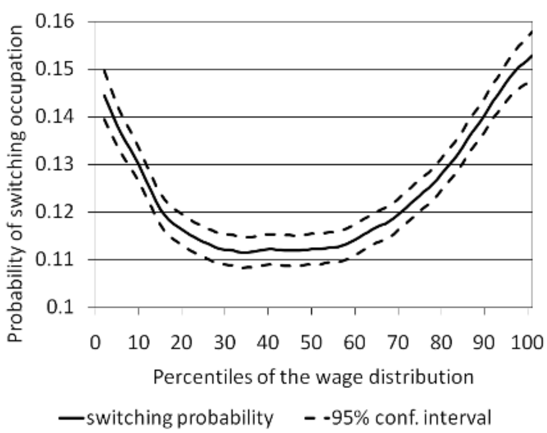
A5.5 Results on Various Levels of Occupational Aggregation



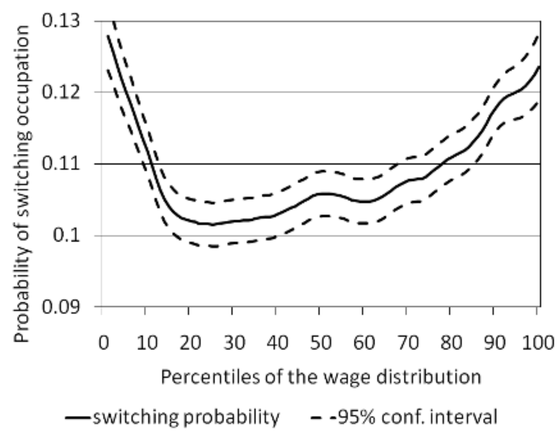
(a) Four-digit classification.



(b) Three-digit classification.



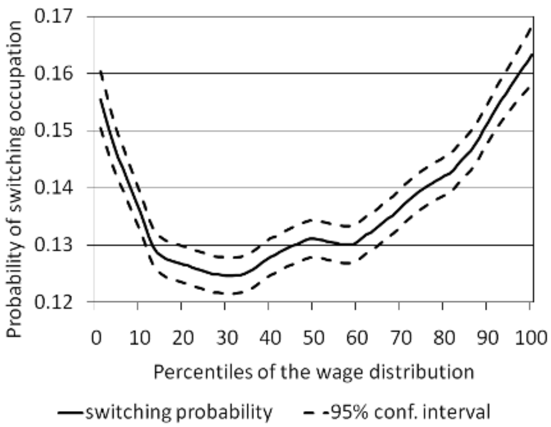
(c) Two-digit classification.



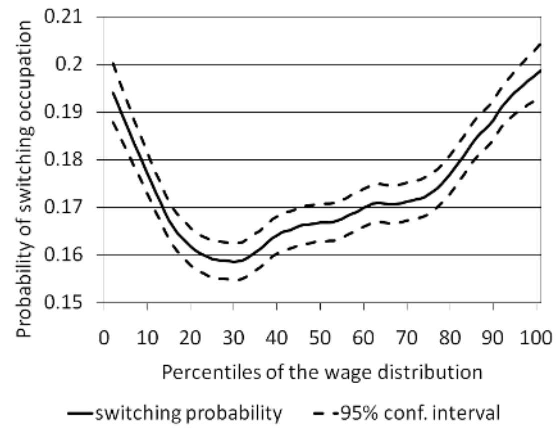
(d) One-digit classification.

Figure A-11: Non-parametric plot of probability of switching occupation by worker's percentile in the distribution of raw wages within occupation, year, and number of years after graduation. Various occupational classifications.

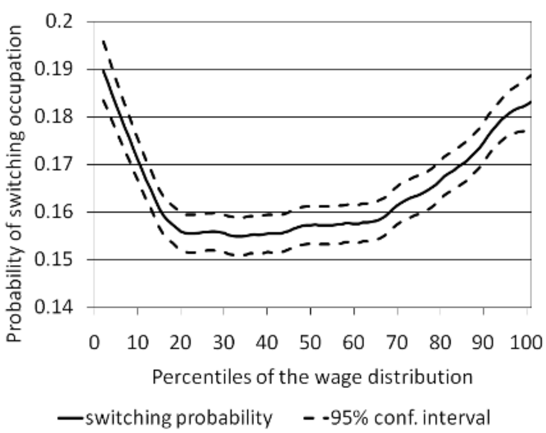
A5.6 Results on Various Occupational Grouppings



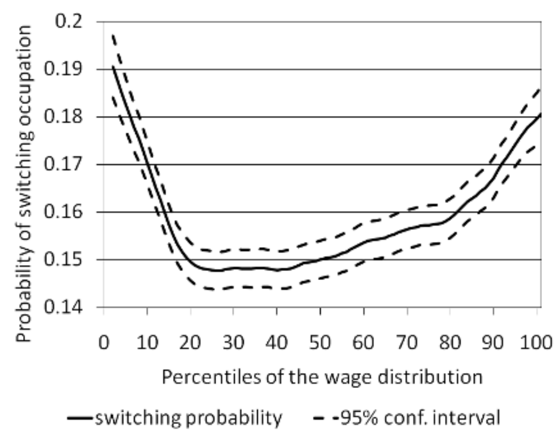
(a) Consructed occupational groups.



(b) All managers in one occupation.



(c) No managers in sample.



(d) No “Not elswhere classified” occupations.

Figure A-12: Non-parametric plot of probability of switching occupation by worker’s percentile in the distribution of raw wages within occupation, year, and number of years after graduation. Various occupational groupings.

A5.7 Assessing the Role of Measurement Error

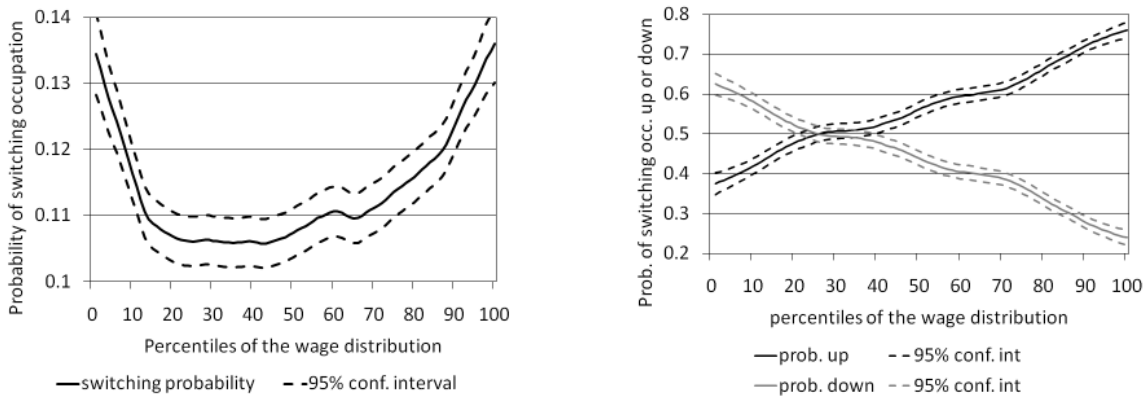


Figure A-13: Non-parametric plots of probability of switching occupation between years t and $t+1$ and of direction of occupational mobility conditional on staying in the same occupation in years $t-1$ and t and staying the same occupation in years $t+1$ and $t+2$ by worker's percentile in the distribution of raw wages.

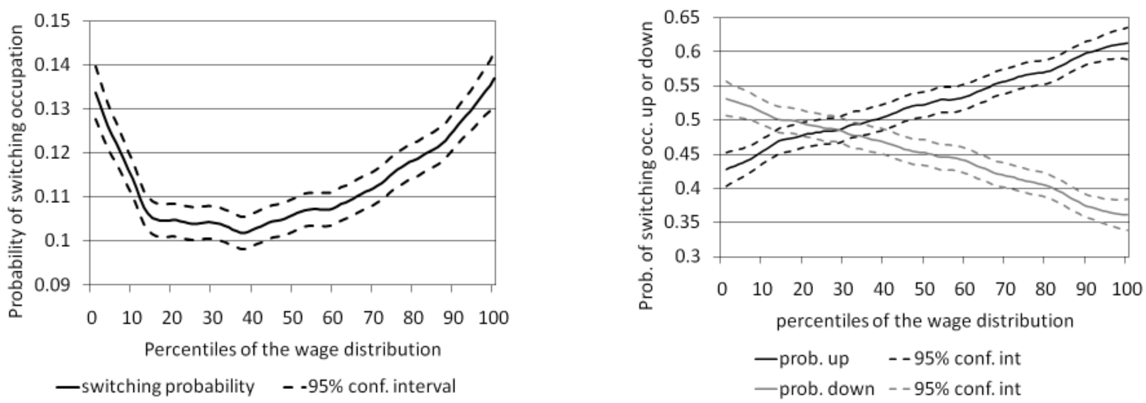


Figure A-14: Non-parametric plots of probability of switching occupation between years t and $t+1$ and of direction of occupational mobility conditional on staying in the same occupation in years $t-1$ and t and staying the same occupation in years $t+1$ and $t+2$ by worker's percentile in the distribution of residual wages.

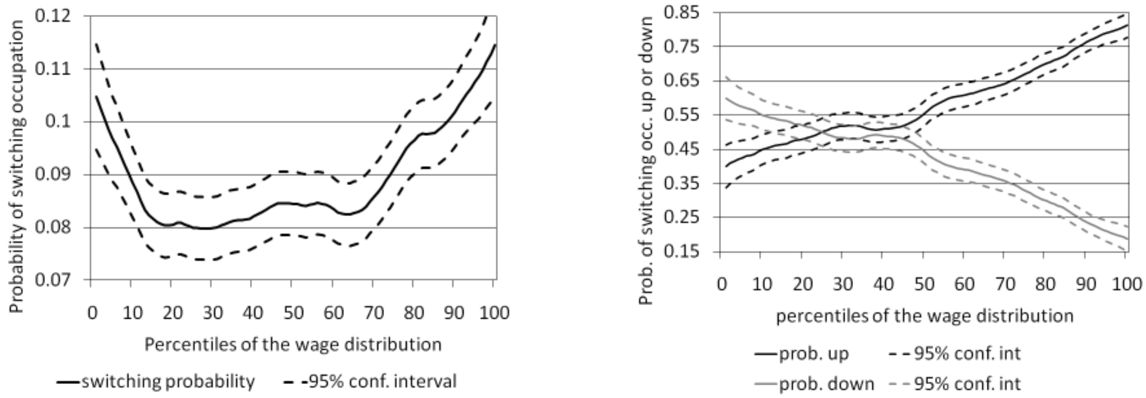


Figure A-15: Non-parametric plots of probability of switching occupation between years t and $t+1$ and of direction of occupational mobility conditional on staying in the same occupation in years $t-2$, $t-1$, and t and staying the same occupation in years $t+1$, $t+2$, and $t+3$ by worker's percentile in the distribution of raw wages.

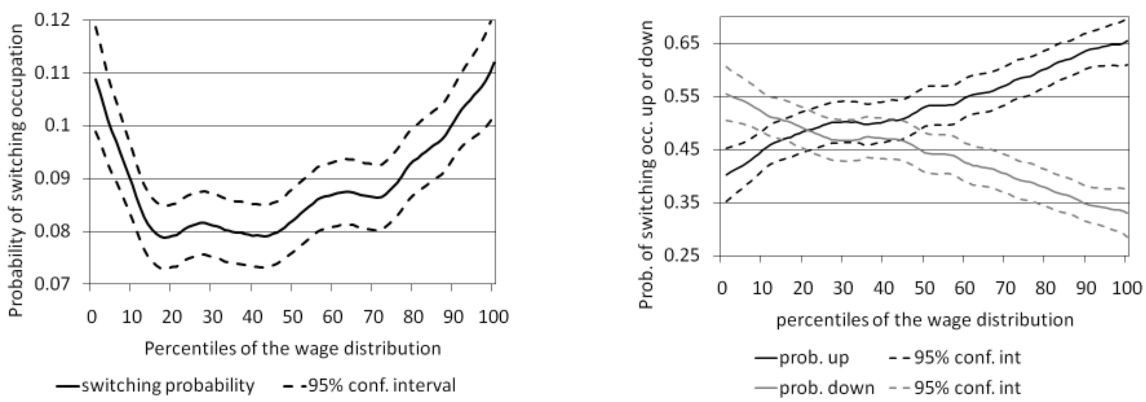


Figure A-16: Non-parametric plots of probability of switching occupation between years t and $t+1$ and of direction of occupational mobility conditional on staying in the same occupation in years $t-2$, $t-1$, and t and staying the same occupation in years $t+1$, $t+2$, and $t+3$ by worker's percentile in the distribution of residual wages.