The U-Shapes of Occupational Mobility

F. GROES

Copenhagen Business School

P. KIRCHER

University of Edinburgh

I. MANOVSKII

University of Pennsylvania

First version received September 2010; final version accepted August 2014 (Eds.)

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: not only low but also high wage earners within an occupation have a particularly large probability of leaving their occupation, and the low (high) earners tend to switch to new occupations with lower (higher) average wages. Exceptions to this pattern of two-sided selection 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 vertical 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

1. INTRODUCTION

Danish employers report that every year close to a fifth of their workers change occupations (e.g., technician, engineer, manager). Similar levels of occupational mobility are reported for the US. 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. 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 occupational
choices are endogenous, 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
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 mobility pattern is predominant except for
occupations with steeply rising (declining) productivity, from which mainly the lower (higher)
paid workers tend to leave.

We are able to document these patterns because the data allow us to compare the behavior
of different workers 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. 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 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.

Nevertheless, we show theoretically that selection based on vertical sorting where more able workers are matched with more productive occupations does account well for all of the qualitative empirical patterns: High-ability workers within an occupation tend to earn high wages, and changes to their perceived ability can lift them beyond the threshold where it is optimal to move to a more skill-intensive occupation, inducing (predominantly) upward mobility. Workers whose perceived ability is such that they obtain wages close to the middle of the occupational wage distribution are less likely to update their beliefs sufficiently to warrant a move to a new occupation, so their mobility is lower. Low-ability workers within an occupation tend to earn lower wages and changes to their perceived ability might induce them to move to an occupation with lower skill-requirements, inducing (predominantly) downward mobility. While there are many reasons why perceived ability may be changing (some are discussed below), one plausible reason is that workers’ ability gets revealed only slowly over time through observations of labor market performance, as formalized in the decision-theoretic work of Gibbons and Waldman (1999) that was successfully used to understand mobility and promotion dynamics within firms. Our model is an extension of this framework to a general equilibrium setting where wages are set in competition for workers. This is similar to Papageorgiou (2012), although we abstract from a number of elements (search frictions, differential speed of learning) to be able to provide a clear and easily comprehensible insight on sorting across many occupations.

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. Older workers switch occupation less frequently, and their wage-distribution is more dispersed than that of younger workers. Finally, the equilibrium nature of the model implies that occupations with sharply increasing productivity will retain their high earners but shed their low earners, and the opposite holds for occupations with a substantial decline in productivity.

Therefore, vertical re-sorting as a consequence of changes in perceived ability seems a promising avenue to account and control for endogeneous occupational mobility. This view of the labor market has a long tradition, even though in the context of occupational mobility horizontal sorting and match-specific shocks have arguably received more attention. Vertical sorting is a basic feature of the famous Roy (1951) specification with absolute advantage, and its combination with learning about workers’ permanent ability has been explored, e.g., in Johnson (1978), Miller (1984), Gibbons and Katz (1992), Jovanovic and Nyarko (1997), Gibbons and Waldman (1999), and Papageorgiou (2012). We discuss the similarities and the differences from the Roy (1951) model explicitly. The main distinguishing feature arises in the presence of occupational productivity shocks, where decision-theoretic models imply that a rising occupational productivity will make the occupation more attractive for all workers, while in our equilibrium model it becomes more attractive only for the more productive workers who compete with and drive out the less able workers. In contrast to other equilibrium models, the main advantage of our formulation is its simplicity, which allows us to easily handle many occupations which is necessary to derive many of the important results. In line with all the work cited in this paragraph, we abstract from an explicit notion of firms. Our main reason is that in the data the pattern of occupational switching is similar for workers who stay with the same firm as for those who switch firms.

In the next section we present our key empirical findings that occupational switching is U-shaped and directional, the main indicators that suggest the use of the absolute advantage

---

3. We compare our model to the simplest and most popular in applied work version of the Roy model - a comparison that we think has substantial pedagogical merit. We also abstract from the presence of search frictions, introduced into the Roy model by Moscarini (2001), which help explain the excess of gross over net mobility.

4. U-shapes arise in the model only for intermediate occupations, where less able workers can leave for lower ranked occupations and more able worker can switch to better ranked occupations. With only two occupations, high ranked workers in the top occupation have no-where better to go and low-ranked workers in the bottom occupation again have no-where lower to go, which limits mobility and allows only for one-sided selection.

5. Our analysis in Section 2.4.2 and the Online Appendix OA14.4 suggests that firm affiliation might not be of first order importance for understanding the basic patterns of occupational mobility.
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. In Section 4 we extend the model to allow for changing occupational productivity, and confront its implications with the data.

Due to space limitations, we discuss a number of extensions of the model and various alternative explanations for our findings in the Appendix available online. In particular, in the Online Appendix OA14 we introduce specific and general human capital accumulation into the model. General human capital accumulation helps the model match the observation that on average workers switch to more productive occupations with age. While we show that the model with human capital accumulation tends to feature similar patterns of mobility as our simpler benchmark model, these extensions will likely be important for future work that will incorporate the selection model that we propose in the empirical investigation of wage and human capital accumulation patterns. We also highlight existing econometric techniques that might be suitable given our analysis. While vertically differentiated view of occupations appears key to explaining why high-wage workers within an occupation switch more than medium-wage workers, in the Online Appendix we discuss that learning about ability is only one possible determinant of occupational mobility. For example, human capital accumulation alone provides an alternative explanation for upward occupational mobility, but since a non-trivial fraction of workers experience downward occupational mobility both within and across firms it would have to be combined with another element. Learning implies that workers sometimes find out that they are less able than anticipated, but general shocks to workers’ ability would provide another plausible explanation. Both have very similar implications, but the former gives a more natural reason why older workers change occupations less. In the Online Appendix we also discuss the role of compensating differentials and the connection of our theory with the literature on the internal labor markets within firms.

The main message of this paper concerns the nature of occupational selection. Selection at the bottom of the within-occupation wage spectrum has long been emphasized and is very intuitive: Low wages are an indication that a person should be doing something else, who therefore has a tendency to leave the occupation. If that were the only source of selection, in the cross-section wages would increase with occupational tenure simply because only high-wage earners stay. In contrast, we highlight that selection is equally strong at the top of the within-occupation wage spectrum. This suggests that high wages are not a sign that a person is particularly well matched in his current occupation: in a world with vertically differentiated occupations we show that this is rather a sign of overqualification that induces workers to seek more suitable types of work. If this is the case, the dominant direction of mobility of high earners should be different from that of low earners, which is consistent with the pattern we document in the data. 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 the empirical assessment of 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 (for example, in a complementary work Papageorgiou (2011) finds evidence of substantial horizontal worker sorting across three broad occupational categories based on the comparative advantage). 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. 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 using these data, 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 et al. (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). 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 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 7th 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 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. The codes are described at http://www.ilo.org/public/english/bureau/stat/isco/isco88/major.htm, and in the Online Appendix OA19.
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. Thus, 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.

For the analysis in the body of the paper, we only select male workers in order to minimize the impact of fertility decision on labor market transitions. However, for completeness, we also report a full set of results on the sample of females in the Online Appendix OA7. 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. 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. Part-time workers are excluded because they do not have as dependable wage information and the majority 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. In order to have 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.
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.\textsuperscript{9} It also includes workers who re-enter the sample after having a missing firm, industry, or occupational spell.\textsuperscript{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 450,000 observations.

The results based on the Large Sample that includes approximately 1.3 million observations are reported in the Online Appendix OA4. They have the same qualitative features as the results based on the Small Sample. 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 \textit{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 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.\textsuperscript{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

\textsuperscript{9} We treat part time work as non-employment.

\textsuperscript{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.

\textsuperscript{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.
(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

The standard reduced-form wage regression

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

(2.1)

where $w_{ijt}$ is real hourly wage of an individual $i$ working in occupation $j$ in period $t$ and $\epsilon_{ijt}$ is the residual. 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 lagged regional unemployment rates. These wage regressions are estimated separately for each occupation.\textsuperscript{12}

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.\textsuperscript{13}

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

---

\textsuperscript{12} Online Appendix Figure OA-1 shows that excluding firm and industry tenure or dummies for the sequence number of the occupational spell from the wage regression does not qualitatively affect the results. Online Appendix Figures OA-2 to OA-4 illustrate that the U-shaped pattern of mobility is robust to alternative bandwidths choices.

\textsuperscript{13} Because tenure predicts mobility (inversely), labor market transitions tend to be correlated over time for the same individual. Workers who just switched are at the highest risk of switching again. Therefore, annual data like IDA, which miss multiple intra-annual transitions, are likely to dampen U-shapes, as they count any number of transitions as if they were one.
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 2nd degree polynomial in worker’s percentile in the wage distribution within occupation and year, i.e. \( Pr(switch) = \Phi(\alpha + \beta \cdot perc + \gamma \cdot perc^2) \). The partial effect of the wage percentile on the probability of switching occupation is \( \frac{\partial Pr(switch)}{\partial perc} = \varphi(\alpha + \beta \cdot perc + \gamma \cdot perc^2)(\beta + 2\gamma \cdot perc) \). The U-shaped pattern implies that this derivative evaluated at \( perc = 0 \) must be negative, that is \( \varphi(\alpha)(\beta) < 0 \). Similarly, the U-shaped pattern also implies that the derivative evaluated at \( perc = 1 \) must be positive, i.e., \( \varphi(\alpha + \beta + \gamma)(\beta + 2\gamma) > 0 \).

In our Small Sample, 66% 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, 74% of occupations (employing 85% of workers) satisfy these criteria. In our Large Sample, 75% 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, 82% of occupations (employing 92% of workers) satisfy these criteria.\(^{14}\)

### 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.

\(^{14}\) For the reasons that would become apparent in Section 4, U-shaped pattern of occupational mobility is considerably more prevalent among occupations that do not experience large changes in relative productivity. For example, restricting attention to occupations in the interior 80% of wage growth, 85% of occupations (employing 91% of workers) satisfy both of these criteria when percentiles are defined in raw wages. If percentiles are defined in wage residuals, 92% of occupations (employing 95% of workers) satisfy these criteria.
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.\(^{15}\) To obtain the ranking based on residual wages, we use our selected sample to run a similar wage regression as in Equation 2.1 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 suggests that, conditional on switching occupations, the higher relative wage a person has in his occupation before the switch, the higher is the probability that he 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.

\(^{15}\) Note that this is a bigger sample than our selected sample. The results are, however, robust to only looking at the average wages in our selected sample.
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.

Figure 5
Non-parametric plot of direction of occupational mobility in terms of change of occupational percentiles, conditional on switching occupation, by worker’s percentile in the relevant wage distribution.

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

2.4 Discussion of Empirical Evidence

2.4.1 Magnitudes of Changes in Occupational Ranks. In Figure 5 we describe, for occupational switchers, the relationship between the worker’s rank in the wage distribution of his occupation and the change in the rank of the occupational average wage upon a switch.
The Figure indicates that workers above the 30th wage percentile of the wage distribution within their occupation tend to move to occupations with higher average wages than the occupation they came from, while workers below the 30th percentile on average move to lower ranking occupations. This point roughly corresponds to the crossing of the two lines in Figure 3.

2.4.2. U-shapes of Occupational Mobility within and between Firms. Our findings remain robust if we separately consider occupational switchers 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 his 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 6 and 7 summarize this evidence when the worker’s relative position in the wage distribution is determined based on raw wages. Online Appendix Figures OA-12 and OA-13 summarize this evidence when the worker’s relative position in the wage distribution is determined based on wages residuals.17 After developing our theory of occupational mobility, we combine it with a simple theory of firm mobility in the Online Appendix OA14.4 and show that the combined theory is quantitatively capable of accounting for the differences in the U-shapes of occupational mobility conditional on staying with or switching firms in Figures 6 and 7.

These results suggest that unemployment is not the main driver of the occupational switching. As Appendix Table A-2 indicates, 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

16. It does not appear possible to distinguish voluntary from involuntary occupational changes. Thus the terminology “occupation change” might be more accurate than occupation “switching,” especially for workers who change firm as we do not know whether the change was initiated by the worker.

17. We use the 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.
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 of raw wages.

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 wage changes persist five years after the occupational transition.\(^{18}\)

\(^{18}\) To construct five-year wage changes we restrict the sample to workers who remain in their occupations or switch to higher/lower ranked occupations in 1995, 1996, and 1997 and for who we can observe wages five years later in 2000, 2001, and 2002.
These patterns are similar for workers switching firms although the wage changes for this group of workers are somewhat larger.

2.4.3. The U-shapes of Occupational Mobility: Females. While our main analysis focuses on the sample of male workers to avoid possible impact of fertility decisions on labor market transitions, in the Online Appendix OA7 we provide a complete set of results on the U-shapes of mobility on the sample of females. The sample is constructed by imposing exactly the same restrictions as on the sample of males. To avoid cluttering the paper, we report the results for females only on the Large Sample, but note that all the results are fully robust to using the Small Sample as well.

We find that the results on the female sample parallel the results documented above on the sample of males. In particular, the probability of an occupational switch is U-shaped in the position of a female worker in the distribution of wages (either raw or residual) in her occupation. Interestingly, the U-shape is somewhat skewed to the right, so that women who are particularly successful in their occupations are even more likely to switch than the particularly unsuccessful ones, who, by the definition of the U-shape, are more likely to switch than the women in the middle of the occupational wage distribution. Occupational mobility on the sample of females is also directional. Relatively more successful women in an occupation tend to switch to higher ranked occupations, while relatively less successful women are more likely to switch to lower ranked occupations. The magnitudes of the jumps in occupational ranks conditional on occupational change are also similar in the samples of women and men. Finally, the U-shaped patterns of occupational mobility remain robust to conditioning on switching employers or remaining with the current one.

2.4.4. Alternative Occupational Classifications. In interests of space, we explore the sensitivity of our findings to alternative definitions of occupations in the Online Appendix OA8. In particular, we consider (1) 1-, 2-, and 3-digit occupational classifications as opposed to the 4-digit classification used in our main analysis, (2) mobility across occupational groups within which workers perform relatively similar tasks, (3) classifying all managers as one occupation or excluding them form the sample altogether, (4) excluding the “... not elsewhere classified” occupations from the analysis. In all these experiments we find that while the level of mobility is affected by the change in occupational classifications, the U-shaped pattern of mobility remains unchanged.

19. In addition to helping 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, the latter experiments are relevant to the contention of Moww and Kalleberg (2010) that in U.S. data the polarization of occupations is due to a relatively small number of occupations, of which “Managers and Administrators, Not Elsewhere Classified” is an important one.
2.4.5. 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. To investigate the potential effect of the measurement error on our findings 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 potential switching point. The shape and direction of mobility for these workers is reported in Figures OA-14 through OA-17 in the Online Appendix OA6. We find that our results remain robust. Similar results are obtained on the samples of occupational switchers within and across firms. We discuss additional evidence on the role of measurement error in the Online Appendix OA9.

2.4.6. 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, Figure 8(a) illustrates that it is poor matches in the bottom part of the wage distribution in an industry that are more likely to be destroyed. These patterns are similar to those documented in the US data by Bils and McLaughlin (2001) who found that at cyclical frequencies industry switchers come from the lower part of the wage distribution in the industry they leave.20

In Figure 8(b) we plot the direction of industry mobility. The results indicate that, conditional on switching industries, workers are much more likely to move into industries that on average pay more to their workers. There is however a weaker relationship (as compared to Figure 3(a) for occupations) between the relative position of the worker in the wage distribution in his industry and the rank of the industry he is switching to.

The difference between occupation and industry switching might be due to the fact (also pointed out by Bils and McLaughlin (2001)) that industries have less of a natural ladder. Workers who find out that they are very talented may not change industries, but are likely to switch to more demanding occupations, explaining the flat right-hand side of Figure 8(a). Similarly, if workers move up to better occupations within their industry, the move across industries might be driven by other considerations.21

20. Similar to those authors, we also find that industry switchers tend to enter in the bottom of the wage distribution in the destination industries.

21. One interpretation is that some industries are paying more than other across all levels of worker skills, and therefore independently of the current ability all workers queue for jobs in the higher-paying industries. Such a queuing theory would explain why workers across the wage-spectrum move to higher-paying industries over time. In contrast, better occupations might only pay more to very high-qualified applicants, and only those at the
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.

2.4.7. 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 tend to switch to occupations with higher average wages. Workers with low wages relative to their occupational average tend to 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 with his ability.

top within a given occupation might qualify, which is the view that we will formalize below in the theory section.
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 Online Appendix 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 $\beta$. Each worker has an innate ability level $a_i$ that is drawn at the beginning of his life from a normal distribution with mean $\mu_a$ and variance $\sigma_a^2$. For the baseline model without human capital accumulation we assume that this ability remains constant throughout a 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

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

in a given period $t$ of his life, where $\varepsilon_{i,t}$ is a normally distributed noise term with mean zero and variance $\sigma_\varepsilon^2$. 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.\(^{22}\)

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 $\phi_t := \phi_a + t\phi_\varepsilon$.\(^{23}\) 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 $\phi_t$, 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

\(^{22}\) In general we allow this error term to be distributed with a different variance $\sigma_\varepsilon^2$ that might not coincide with the variance of the labor market error term $\sigma_\varepsilon^2$. While our exposition is presented for $\sigma_\varepsilon^2 = \sigma_\varepsilon^2$, the more general expressions can be obtained with minor modifications mentioned below.

\(^{23}\) If the first ability signal before the worker enters the labor market has variance $\sigma_0^2 \neq \sigma_\varepsilon^2$, this can be accommodated by adjusting the cumulative precision to $\phi_t = \phi_a + \phi_0 + (t-1)\phi_\varepsilon$, where $\phi_0 = 1/\sigma_0^2$. 
A precision-weighted average of the prior mean and the output observation:

\[ A_{i,t+1} = \frac{\phi_t}{\phi_{t+1}} A_{t,t} + \frac{\phi_{t+1}}{\phi_{t+1}} X_{i,t}. \tag{3.3} \]

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, (3.3) 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 define the following two distributions. 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 \( A_t \). 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 \( \delta \). This is all we need for most proofs. 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 (3.3) and is independent of any choices that agents make. Therefore, the measure of agents with belief below \( A \) across 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 \( \gamma_k \) of available jobs. We treat the number of jobs as exogenous in this exposition, yet the Online Appendix OA16 discusses how endogenous entry can be accommodated (limited entry and associated competition among workers for scarce jobs will be most important in Section 4 to explain the mobility patterns when occupational productivities change).

24. Conditional on knowing the true ability \( a \) of a worker, the output \( X_t \) is distributed normally with mean \( a \) and precision \( \phi_t \), i.e. \( X_t \sim N(a,\phi_t) \). 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 \( X \sim N(A_t, \phi_t \phi_t/\phi_{t+1}) \). We are not interested in the output per se, but in the update \( A_{t+1} = \frac{\phi_{t+1}}{\phi_{t+1}} \phi_{t+1} A_t + \phi_{t+1} X_t \) as a function of output. This linear combination implies that the posterior distribution \( G_t(A_{t+1}|A_t) \) is a normal with mean \( A_t \) and precision \( \phi_t \phi_{t+1}/\phi_{t+1} \), i.e. \( A_t \sim N(A_t, \phi_t \phi_{t+1}/\phi_{t+1}) \).

25. 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.

26. 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 \( t \phi_t \). Since \( a \) is not known, an agent with prior \( \mu_a \) faces realizations of \( X \) that are normal with mean \( \mu_a \) and precision \( t \phi_t \). Since the update is \( A^t = (t\phi_t \bar{X} + \phi_a \mu_a)/\phi_t \), \( F^t \) is normal mean \( \mu_a \) and precision \( \phi_a \phi_t/(t \phi_t) \).
Each unit of the good (or service) that is produced sells in the market at some exogenously given price $P_k$. Therefore, worker $i$ employed in a job of type $k$ generates revenue

$$R_{ki} = P_k X_i.$$  \hspace{1cm} (3.4)

Equivalently, we can interpret $P_k$ as the productivity in terms of efficiency units of the labor (at a common sale 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: $\sum_{k=1}^{K} \gamma_k < T - F(0)$).\hspace{1cm} 27

**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 $\Pi_k$ it can simply offer the wage contract

$$w_k(X) = P_k X - \Pi_k$$  \hspace{1cm} (3.5)

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 that the firm has as much information as the workers, since workers would self-select. The relevant sorting criterion for risk-neutral workers is their expected wage given their belief $A$ about their mean ability:

$$W_k(A) = P_k A - \Pi_k.$$  \hspace{1cm} (3.6)

Alternatively, if the firm has the same information as the worker it can directly pay expected wages according to (3.6). 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 (3.6)), but observed wages differ according to the specification and could potentially lead to different assessments of observed wage 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 (3.5) and (3.6) to provide both incentives for self-selection as well as insurance to workers, and our arguments can easily be extended to show that our main propositions hold for any such convex combination.

27. Otherwise the lowest occupations would not attract any workers and would simply not be observed in the data.
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 (3.5) [or (3.6)]. It is notationally more convenient to focus on the profits. Stationary means that the entrepreneurs’ profits \( \Pi = (\Pi_1, \Pi_2, \ldots, \Pi_K) \) 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:

\[
B_k \equiv \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}}, \quad \text{for } k \in \{1, \ldots, K\}.
\]

(3.7)

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 \( B_{K+1} \equiv \infty \). 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 \( \gamma_k \) available in this occupation:

\[
\gamma_k = F(B_{k+1}) - F(B_k), \quad \text{for } k \in \{1, \ldots, K\}.
\]

(3.8)

The system (3.7) and (3.8) can easily be solved recursively: Summing (3.8) across all \( k \) and noting that \( F(B_{K+1}) \) equals the total population \( T \), we get \( \sum_{k=1}^{K} \gamma_k = T - F(B_1) \) which determines \( B_1 \). Then successive application of (3.8) yields the remaining cutoff levels \( (B_2, \ldots, B_K) \).

Since zero productivity in the lowest occupation implies zero profit, (3.7) then delivers the profits of the firms in the various occupations \( (\Pi_1, \ldots, \Pi_K) \). To sum up:

**Definition 1.** An equilibrium is a vector of profits \( \Pi = (\Pi_1, \ldots, \Pi_K) \) with \( \Pi_0 = 0 \) and a vector of optimal worker cutoff level \( (B_1, B_2, \ldots, B_K) \) such that equations (3.7) and (3.8) hold.

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 (3.6). 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

\[S_{k,t}(W)\]
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 (3.5), then denote the wage by lower-case letter $w$ and the corresponding switching probabilities by $s_{k,t}^{-}(w), \ s_{k,t}^{+}(w)$ and $s_{k,t}^{−}(w)$. To analyze these switching probabilities formally, we adopt the following definition.

**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 $S_{k,t}^{-}(·)$ 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 (3.6). 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

$$S_{k,t}^{-}(W) = G_t(B_k | A), \quad (3.9)$$
$$S_{k,t}^{+}(W) = 1 - G_t(B_{k+1} | A) \quad \text{and} \quad (3.10)$$
$$S_{k,t}^{+}(W) = G_t(B_k | A) + 1 - G_t(B_{k+1} | A). \quad (3.11)$$

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 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 9 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 $A = B_k := (B_k + B_{k+1})/2$. 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 $B_k$. 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 occupations. 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 A2.

For interior occupations, U-shapes are likely to persist even when we do not condition on cohort $t$. For the extreme occupations of $k = 0$ 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 case of the lowest occupation 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. 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 9 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.\(^{30}\)

**Proposition 2 (Direction of Sorting).** Consider workers of cohort \( t \) in interior occupation \( k \) that switch. Among these, higher wage workers are more likely to switch up and lower wage 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 (3.6); see Appendix A2 for (3.5). Recall that a worker earns wage \( W \) only if he has belief \( A = (W + \Pi_k)/P_k \). By (3.10) 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. These results also hold in terms of wage residuals when we control for more variables than just labor market experience. For example, we might also want to condition on occupational tenure. While conditioning on a particular occupational tenure changes the distribution of workers across the various ability levels, it does not change the insight on the shape of sorting since workers who are closer to a cutoff will still switch for smaller increments in information than workers in the middle, and it neither changes the robust insights on the direction because workers close to the upper cutoff are still more likely to get shocks that lead them to switch upwards than other workers.

\(^{30}\) It is easy to show that \( S_{k,t}(W) > 0 \) for all \( W \). Yet it can be that case that \( s_{k,t}(w) = 0 \), and 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, a pattern documented in the Online Appendix OA10, but also visible in, e.g., Figure 2(b) in the main text. 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 (3.3) by which \( A_t \) evolves.

Younger workers of cohort \( t \) switch more 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).

Under our learning process, if wages are paid according to expected ability (3.6), 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 (3.6) diverge for older workers. Since the ability of young workers is not very precise, they get similar wages. As information gets revealed in the production 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 (3.5) is analytically less tractable since the variance in the distribution of mean ability is confounded with the variance in the output process.

31. 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.

32. It is easy to verify that the wage schedule \( W(A) = \max_k W_k(A) \) are convex in expected ability. So mean-preserving spreads of the distribution of expected ability increase the variance of wages. The result follows since younger workers know less about their ability: \( F_{t'}(A) \) is a mean-preserving spread of \( F_{t}(A) \) when \( t' > t \). To see this, note that Footnote 26 showed that \( F_{t} \) is normal with mean \( \mu_A \) and precision \( \frac{\phi_{t'\epsilon} + \phi_{t\epsilon} \phi_a}{\phi_a} \), and the latter is monotonically decreasing in \( t \).

33. Under (3.5), 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 higher productivities \( P_k'' > P_k \), which tends to increase the variance.
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 (3.6) but depends on parameters when wages are set according to (3.5). More robust predictions of the model that are independent of whether the wage setting process is assumed to be given by (3.5) or (3.6) 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.* Workers’ beliefs about their mean ability are strictly ranked as follows: 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 $B_k$. The result follows immediately because expected wages are increasing in the belief.

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 $t$, and any occupations $k$ and $k'$ that are not too large or too far apart (i.e., $\max\{|B_{k+1} - B_k|, |B_{k+1} - B_k|\} \leq \sqrt{\phi_{t+1}/(\phi_t + \phi_t)}$). 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 $t + 1$. The average wage of the stayers in $k'$ is lower than the average wage of downward switchers (i.e., if $k > k'$), but is higher than the average wage of upward switchers (i.e., if $k < k'$).

*Proof.* See Appendix A2.

34. Wages set according to (3.6) strictly increase in expected ability. Since a worker only switches up if the assessment of his ability improved more than the assessment of 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 (3.5) 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.
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. The qualifier that the occupations are not too far apart ensures that updates are within one standard deviation of the distribution of updates, which guarantees that the distribution in this range is concave. It is a sufficient condition used in the proof, but is not necessary.

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 a year or five years after the switch. Consider all workers in occupation \(k\) in period \(t\). In period \(t+1\) (or \(t+5\)) 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 year \(t+1\) or \(t+5\) 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 10(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 10(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 \(t\). The comparison of wages five years after the switch indicates that these effects are highly persistent.\(^{35}\)

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 11(a) we construct the weighted average of period \(t+1\) wage ratios of switchers into occupation \(k'\) over stayers in occupation \(k'\). Figure 11(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.

\(^{35}\) The standard errors of the wage ratios in Figures 10 and 11 range from 0.00083 to 0.00197. Due to their small size, they are invisible in the Figures.
4. CHANGING OCCUPATIONAL PRODUCTIVITIES

Most occupations are rather stable in their occupational ranking over time. Nevertheless, some occupations such as computer programmers in the 1990s have seen substantial wage increases for
their workforce, while other occupations such as textile machine operators have seen substantial wage declines.\textsuperscript{36} The associated changes in occupational rank are an indication of changing productivity at the occupational level. In light of the model, these occupations would require a different workforce as their productivity changes.

In this section we analyze how changes in occupational productivity affect occupational mobility. First, we extend the model to allow 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. Second, we show that this pattern is indeed present in the data for the occupations with fastest wage growth, which is the sign of fast productivity growth in the model. 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. Finally, we describe how our model relates to a stylized version of the famous Roy (1951) model that is often used as the workhorse model to analyze worker mobility in the presence of sectoral productivity shocks. This part shows why at least in the most basic setting the equilibrium nature of our model is crucial in predicting these patterns.

4.1. Mobility in Response to Changing Occupational Productivity: Theory

Consider our basic model, but assume from one period to the next the productivity of occupation \( k \) changes from \( P_k \) to \( P'_k > P_k + 1 \).\textsuperscript{37} This changes the ranking of occupations. Workers in occupation \( k \) who realize that they are better than expected and would have changed occupation will now stay, while workers whose ability remains constant and that would have stayed might now leave because of the competitive pressure of other workers that enter this occupation. In the absence of switching costs it is not relevant whether this change is temporary or permanent. We obtain the following result for the switching patterns 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.\textsuperscript{38}

\textbf{Proposition 5.} Consider an occupation \( k \) with a rise in productivity such that \( P'_k > P_{k+1} \). If occupation \( k+1 \) is no smaller than occupation \( k \) and wages are set according to (3.6), the

36. Kambourov and Manovskii (2009a) measure the magnitude of changes in occupational productivities.
37. For a more general treatment of a model where multiple occupations may have changes in productivity, see the Online Appendix OA15. The proposition in this section carries over essentially unchanged to occupations that change their rank upward or downward relative to the other occupations. Also, note that more productive workers sort into more productive occupations, and therefore an improved productivity that does not change the ranks of occupations only changes the wages but not the ability cutoffs at which workers select in one occupation or another. This would change if there were strictly positive switching costs as then also the level of wages matters.
38. For wages set according to (3.5) we can prove the following. Consider an occupation \( k \) with a rise in productivity such that \( P'_k > P_{k+1} \). If occupation \( k+1 \) is no smaller than occupation \( k \) and wages are set according to (3.6), then only some of the workers with wages above the occupational mean stay, while all lower wage workers leave. For a decline in productivity, all workers above the mean leave and only some below stay. For the proof please see the more general version of this result in the appendix.
Consider an occupation $k$ with a decline in productivity such that $P_k' < P_{k-1}$. If occupation $k-1$ is no smaller than occupation $k$ and wages are set according to (3.6), the probability of switching out of occupation $k$ increases with higher wages for workers in the same cohort.

**Proof.** We prove the result for a rising occupation; analogous steps establish the result for a declining occupation. Workers are initially in occupation $k$ because their prior $A$ was in set $[B_k, B_{k+1})$. To stay in occupation $k$, their posterior now has to be above $B_k'$, which is larger than $B_{k+1}$ since occupation $k$ is no larger than occupation $k + 1$. Thus, workers with a higher prior are closer to the region where they stay in $k$, 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 $G_t$).

4.2. 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 average wage growth between years $t$ and $t + 1$. 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 12 plots three groups of occupations, separated by the growth rates of average wages 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 12(a) and 12(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 the Online Appendix Figure OA-10. They clearly exhibit the same qualitative patterns.

4.3. The Relation to the Roy Model in the Presence of Shocks

Our model is related to the popular Roy (1951) model. That model was designed to investigate the impact of shocks on the sorting of workers and their wages. It might be instructive to highlight briefly the commonalities and differences. To see the commonality, consider the following version of the Roy model formalized in Heckman and Honore (1990), simplified to two occupations 1 and 2 with output prices $P_1$ and $P_2$, respectively. Each worker is endowed with a two-dimensional skill set $(s_1, s_2)$ that describes his output in each occupation. Each worker chooses the occupation where he obtains the highest payoff, i.e., he chooses Occupation 1 if $P_1 s_1 > P_2 s_2$. Figure 13(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. The dotted line indicates the skill distribution in the special case where this is a straight line. Its slope indicates absolute advantage in the sense that a worker who is good in one occupation is also good in the other. 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 13(b), and all skills are on the diagonal line since $(s_1, s_2) = (A, A)$. Without shocks it simply constitutes a rotation of the basic Roy model. In that case the main additional feature of

39. The later sections in Heckman and Honore (1990) also discuss in detail identification with multiple occupations.
40. In contrast to absolute advantage, relative advantage indicates a line where workers that are better in one occupation are not as skilled in the other. Note that most applied work goes beyond a straight line to consider a distribution of skills, such as a joint-normal where the direction of the correlation of skills gives an indication of absolute or relative advantage.
skill distribution in society

(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 $\Pi_i$; price of output and profit in occupation $i \in \{1, 2\}$

**Figure 13**

Comparison between Roy model and our model for fixed productivities $P_1$ and $P_2$

(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 14**

Comparison between Roy model and our model when productivity $P_1$ increases

our model is learning about skills, which implies that an agent’s position on the solid line is not fixed over time, which generates mobility.

Another important difference between the models arises in the presence of shocks to prices (occupational productivities), which is the only source of mobility in the basic Roy model. If the low productivity occupation $P_1$ improves, then the solid line becomes steeper. In the case of absolute advantage, the worst workers in Occupation 2 will leave and become the highest wage workers in Occupation 1, as depicted in Figure 14(a). Note that there is no competition among workers for jobs: A worker can create a job for himself in either occupation independently of the choices of other workers, so some workers that used to choose Occupation 1 now create a job for
themselves in Occupation 2, but none of the works who used to choose Occupation 2 would now create their job in Occupation 1. This is a general feature of the basic Roy model even if skills are not just concentrated on one line.\(^{41}\) This is very different in our setting. The immediate effect of an increase in \(P_1\) in our model is similar: the line that divides who selects into which occupation becomes steeper, as indicated in step A in Figure 14(b). This would be the only change if profits would remain constant (or if workers could create jobs for themselves at cost \(\Pi_k\)). Yet since jobs in each occupation are limited, it cannot be that more workers can choose a given occupation, which is reflected in an increase of profits of the firms and therefore a decrease in the intercept as depicted by step B in Figure 14(b). The competition among workers for the fixed number of jobs means that exactly the same skills select into each occupations as before, unless the solid curve becomes steeper than the skill distribution (which happens when occupations reverse rank, i.e., \(P_1 > P_2\)). In this case the matching pattern reverse: while high ability workers used to select into occupation 2, they now select into occupation 1. Therefore, only those workers who learned that they are sufficiently able stay in occupation 1 while the rest is driven out by incoming high-ability workers, while only those workers that learned that they are not very able stay in occupation 2 while the rest leaves for the now more attractive occupation 1. These are the effects described in Section 4.1. Clearly, if there are costs to switching occupations the change of the workforce would not be as abrupt as discussed here. Also, as occupations become more productive they might indeed expand in size, but if it becomes increasingly expensive to create more positions the results of this discussion still apply (see the Online Appendix OA16 for details).

5. 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

\(^{41}\) Nevertheless, in a Roy model with three or more occupations, a simultaneous productivity shock to multiple occupations might give rise to very general patterns, the study of which is beyond the scope of this paper.
on their ability after observing their output and re-sort themselves according to the update. 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.\footnote{In the Online Appendix OA14.4 we integrate the notion of jobs into the model, and find that a theory of firm mobility driven by occupational moves (rather than occupational mobility driven by firm moves) seems to account well for the data patterns that we document when conditioning on staying with or switching firms.}

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 accounts well for the observed pattern. We also show that standard upward career progression due to human capital accumulation can easily be integrated into the framework, yet by itself 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 suggest that the model might also generate the right quantitative magnitudes, and might provide a fruitful way to think about selection issues in the presence of occupational mobility. Taking the model and its extension to human capital in Section OA14.2 and the Online Appendix OA17 as given, adaptations of existing econometric methods allow to control for selection and to estimate the human capital process.\footnote{As discussed in Section OA14.2, we provide direct evidence on occupational switching assumed in partial equilibrium with related payoff structure by Gibbons et al. (2005). They argue that lagged variables that are valid instruments for occupational choice within the structure.}

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\footnote{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 (2006) and for convergence when the short side of a market gets matched with near certainty see Eckhout and Kircher (2010)).} – 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, which might also require a broader notion of skills beyond the one-dimensional measure that we use in our basic model.\textsuperscript{45} For 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.

\textsuperscript{45} Note that the parameters of the stationary environments in Section 3 and Section OA16 such as occupational productivity $P_k$, profits $\Pi_k$ and human capital accumulation functions $H(t)$ and $h_k(t)$ can be consistently estimated using the methodology proposed by Gibbons et al. (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) or if switching occurs also across hierarchies, further investigation is necessary.
### TABLE A-1

*Summary statistics for the Large and Small samples and subsamples*

<table>
<thead>
<tr>
<th></th>
<th>Small Sample</th>
<th>Large Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample A</td>
<td>Sample B</td>
</tr>
<tr>
<td>Number of observations</td>
<td>485,859</td>
<td>449,517</td>
</tr>
<tr>
<td>Number of occupations</td>
<td>243</td>
<td>154</td>
</tr>
<tr>
<td>Age</td>
<td>29.58</td>
<td>29.42</td>
</tr>
<tr>
<td>Occupational tenure</td>
<td>4.21</td>
<td>4.21</td>
</tr>
<tr>
<td>Occupational spell number</td>
<td>1.73</td>
<td>1.71</td>
</tr>
<tr>
<td>Occupational switchers</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>Employer tenure</td>
<td>3.88</td>
<td>3.84</td>
</tr>
<tr>
<td>Employer switchers</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>Industry tenure</td>
<td>3.32</td>
<td>3.30</td>
</tr>
<tr>
<td>Years after graduation</td>
<td>6.37</td>
<td>6.28</td>
</tr>
<tr>
<td>12 years of school or less</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>13 years of school or more</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>Hourly wage in DKK in 1995</td>
<td>168.47</td>
<td>167.30</td>
</tr>
<tr>
<td>Married</td>
<td>0.30</td>
<td>0.29</td>
</tr>
<tr>
<td>Number of children</td>
<td>0.70</td>
<td>0.68</td>
</tr>
</tbody>
</table>

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.

A1. APPENDIX TABLES
TABLE A-2
Frequency of various occupational transitions and the associated wage changes

<table>
<thead>
<tr>
<th>Type of occupational transition between years $t$ and $t + 1$</th>
<th>Year $t + 1$</th>
<th>Year $t + 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Switch up</td>
<td>Switch down</td>
</tr>
<tr>
<td>All Workers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage change</td>
<td>1.93 (0.13)</td>
<td>-0.94 (0.14)</td>
</tr>
<tr>
<td>Fraction of all workers</td>
<td>0.104</td>
<td>0.084</td>
</tr>
<tr>
<td>Firm stayers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage change</td>
<td>0.70 (0.15)</td>
<td>-0.64 (0.16)</td>
</tr>
<tr>
<td>Fraction of firm stayers</td>
<td>0.080</td>
<td>0.064</td>
</tr>
<tr>
<td>Firm switchers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage change</td>
<td>2.08 (0.32)</td>
<td>-3.30 (0.34)</td>
</tr>
<tr>
<td>Fraction of firm switchers</td>
<td>0.195</td>
<td>0.160</td>
</tr>
</tbody>
</table>

Note – The table contains 1-year and 5-year wage changes for workers experiencing various types of occupational transitions net of the 1-year or 5-year wage change of the corresponding group of occupational stayers. The wage change is measured in year $t + 1$ or $t + 5$ relative to year $t$ conditional on a transition between years $t$ and $t + 1$. Standard errors in parenthesis.
A2. OMITTED PROOFS AND DERIVATIONS

Remainder of Proof of Proposition 1:

Proof. Consider first wage setting process (3.6) 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 $\delta$ from $\overline{B}_k$. Then

$$S_{k,t}(P_k \overline{B}_k - \Pi_k) - S_{k,t}(P_k(\overline{B}_k + \delta) - \Pi_k)$$

$$= G_t(B_k \overline{B}_k) - G(B_k \overline{B}_k + \delta) + G_t(B_{k+1} \overline{B}_k + \delta) - G_t(B_{k+1} \overline{B}_k) \quad \text{(A1)}$$

$$= G_t(-\delta(0) - G(-\delta - \delta(0)) + G_t(\delta(\delta(0) - G_t(\delta(0)$$

$$\int_0^\delta [g(-\delta - \varepsilon(0) - g(\delta - \varepsilon(0))] \, d\varepsilon,$$  

(A2)

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 $\varepsilon$ 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 (3.5). A sufficient condition for U-shapes is that switching probabilities are minimal in the interior and maximal at the boundaries. We will show that they are zero in the interior and one at the boundaries. To see this, note that we observe the worker of cohort $t$ in occupation $k$ at wage $w$, (3.5), this 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 (3.3) 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 $w \in [B_k P_k - \Pi_k, B_{k+1} P_k - \Pi_k)$. Therefore, for such intermediate wages the switching probability $s_{k,t}(w) = 0$, which constitutes a local minimum.

We complete the proof by by showing that for very low or very high wages within an occupation the subsequent 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.46

Remainder of Proof of Proposition 2:

46. It does not establish strict U-shapes for $s_{k,t}$, 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.
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 $X(w) \geq B_{k+1}$, 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 $k' > k$. In year $t+1$ we compare their wages, conditional on choosing $k'$. All workers that we compare have some belief in $[B_k, B_{k'+1}]$ 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 $B_{k'+1} - B_k$ is not too large since normal distributions are concave around their mean. In particular, a normal distribution is concave within one standard deviation of its mean. The update has standard deviation $\sqrt{\phi_{t+1}/(\phi_t + \phi_{t+1})}$, so we require $B_{k'+1} - B_k < \sqrt{\phi_{t+1}/(\phi_t + \phi_{t+1})}$. Since by market clearing $F(B_{k'+1}) - F(B_k) = \sum_{j=k}^{k'+1} \gamma_j$ this holds if the measure of firms in the occupations between $k$ and $k'+1$ is not too large.

The workers’ update $A_{t+1}$ is distributed symmetrically around $A_t$. Since $k' > k$, 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 $A_t \in [B_{k'}, B_{k'+1}]$). 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 condition for concavity of the update is that $B_{k'+1} - B_k$ is not too large (i.e., $B_{k'+1} - B_k < \sqrt{\phi_{t+1}/(\phi_t + \phi_{t+1})}$). When $k' < k$, the density of the update is still higher but the derivative is lower, which directly implies that the distribution for switchers first order stochastically dominates the distribution for stayers.
Acknowledgement. We would like to thank the editor and two anonymous referees for their comments and suggestions. We are also grateful to Martin Browning, Mette Ejrnæs, Pieter Gautier, Jonathan Heathcote, Hans Christian Konsted, Dale Mortensen, Giuseppe Moscarini, Theodore Papageorgiou, Kjetil Storesletten, and many seminar audiences for their comments. This research has been supported by the National Science Foundation Grant No. SES-0922406 and the European Research Council Grant No. 284119.

REFERENCES


