The U-Shapes of Occupational Mobility*

Fane Groes
University of Copenhagen

Philipp Kircher
University of Pennsylvania

Iourii Manovskii
University of Pennsylvania

Abstract

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).

These facts are not implied by existing theories of occupational mobility that mostly treat occupations as horizontally differentiated sets of tasks. We suggest that it might be productive to think of occupations as forming vertical hierarchies. Workers who are unsure of their abilities learn about them by observing their output realizations. Employment opportunities in each occupation are scarce inducing competition among workers for them. Complementarities in the production function between worker's ability and productivity of an occupation induce sorting of workers into occupations according to their expected ability. We present an equilibrium model of occupational choice with these features and show analytically that it is consistent with patterns of mobility described above.

JEL Classification: E24, E25, J24, J31

Keywords: Occupational mobility, Learning, Labor markets

*Preliminary and incomplete. First version: June 2008. This version: February 2009. We would like to thank Martin Browning, Mette Ejrnæs and seminar participants at the 2008 Copenhagen Microeconometrics Summer Workshop and the Search & Matching workshop at the University of Pennsylvania for their comments. This research has been supported by the National Science Foundation Grants No. SES-0617876 and No. SES-0752076.

1 Introduction

When a worker switches detailed occupational categories (technician, engineer, manager) he or she moves to an observationally different technology. The fraction of workers switching occupations is remarkably large. Kambourov and Manovskii (2008) document that close to 20% of workers in the U.S. switch occupations in a given year. Moreover, these gross flows are much larger than the net flows that is needed to account for the changing sizes of occupations.What is it that induces workers to undertake all these occupational changes? We provide new evidence on the patterns of occupational mobility and suggest that the standard theories of occupational mobility are not consistent with these facts. We then proceed to develop a new theory of occupational mobility.

There are two commonly used classes of models of occupational mobility. The first one, described in, e.g., McCall (1990) and Neal (1999), is based on match-specific occupational sorting. Occupations are perceived as identical (e.g., not different with respect to skill requirements), but workers find out the quality of their specific match to an occupation over time. Match-specific sorting occurs when workers realize that their match-specific shock is bad and abandon the match in favor of (the search for) a better one. The predictions from this theory are based on selection: Since those workers that are content with their match stay, this theory predicts that the probability of switching occupation declines with tenure in that occupation, which is consistent with the data. Moreover, since good matches survive longer, wages and tenure are positively correlated in the cross-section of workers - an observation that is also consistent with the data.

A closer look at the data that we take in this paper, however, reveals that the fundamental selection mechanism in these match-specific sorting models is not consistent with the data. Virtually any model in which productivities are drawn independently for each worker-occupation-match rather than representing a permanent trait of either the occupation or the worker would predict that the probability of switching occupation is negatively related to wages which indicate match quality. Instead, we find a strong evidence that the probability of switching occupation is U-shaped in wages: not only is it people with wages lower than the occupational average, but also those with wages above the average that are more likely to switch.

The second class of existing models focuses on net mobility, which is explained by fluctuating demands for services of different occupations. They generally also imply that it is either only the people on the lower part of wage distribution within an occupation or only in the upper part of the distribution that tend to switch in response to a change in demand conditions, rather than workers on both ends of the spectrum. This is the property of the classic Roy

(1951) model (and its extensions in, e.g., Moscarini (2001)). The models in Kambourov and Manovskii (2005, 2009a) generically have a similar prediction. The represent a version of the island economy model of Lucas and Prescott (1974) where islands are interpreted as occupations and workers accumulate occupation-specific human capital. Human capital is destroyed upon switching occupations which implies that, if workers with different levels of human capital are perfectly substitutable in occupational production function, it is the low human capital, and hence, low wage, workers that switch first if an occupational demand declines. If occupational demand rises, no one leaves the occupation.

We will show below that in the data most occupations exhibit U-shapes in mobility. On top of this, however, when occupation experience an *increase* in demand, workers in the lower part of wage distribution of that occupation tend to *leave* it. None of the existing theories is consistent with this pattern. The data further imply that occupational switching is non-random. A worker who is in the upper tail of the wage distribution in some occupation and decides to switch to another occupation, on average moves to an occupation with higher mean earnings. A worker who is in the lower tail of the wage distribution in some occupation and decides to switch to another occupation, on average moves to an occupation with lower mean earnings.[1] Once again, existing models do not generate such pattern. The reason is that the literature has treated occupations as horizontally differentiated sets of tasks. We think, however, that it might be productive to think of occupations as also forming vertical hierarchies.

In our theory, workers have different innate abilities. Workers and employers learn about these abilities by observing the output realizations. In difference to, e.g., Johnson (1978), Miller (1984)and Papageorgiou (2007), the speed of learning is independent of the occupation the individual is working in, which allows us to consider more than two occupations without loosing tractability. This turns out to be important to understand the U-shapes in the switching pattern. Employment opportunities in each occupation are scarce - for example because other factor inputs to production are fixed or exhibit increasing costs when more employment is created.

With scarce employment opportunities workers compete for jobs. With complementarities in the production function between workers' ability and productivity of an occupation, the more able workers will in equilibrium occupy the jobs in more productive occupations. As agents learn that they are either too good or too bad for a given profession they switch to a more appropriate one, which induces the U-shapes. Those workers that are talented move to more

[1] Nevertheless, a worker who leaves an occupation from the top of its wage distribution on average experiences a decline of his wage growth upon a switch, while a switcher from the bottom of occupational distribution experience an increase in wage growth.

productive occupations, while those that are less talented switch to less productive occupations. Even those workers that switch to lower productivities benefit relative to staying. If they would attempt to stay they would block a better suited worker from the job. In a competitive labor market this opportunity cost translates directly into low wages for the inappropriate worker. In fact that wage is below the wage in a less productive profession for which the opportunity costs are not that high.[2] A similar logic applies with free entry when jobs in more productive occupations have higher capital costs: With complementarities in production only workers with high ability will be willing to pay the cost of creating a job in a highly productive occupation.

Extensions of this idea that allow for changing occupational productivities reveal that occupations with rising productivity indeed expand their high-ability workforce while shedding lower-ability workers in order to match the skill of their workforce to the productivity of the jobs. Similarly, occupations with declining productivity increase their low-ability workforce and loose the high-ability workers to better occupations. These insights obtain with fixed production factors as well as when entry is not fully elastic. In another extension we take into account that even in our vertically differentiated view of occupations a switch requires a new set of skills which induces costs to occupational switching (see e.g. Shaw (1984, 1987); Kambourov and Manovskii (2009b)). For example, engineers that move up to manage small groups need to adjust their human resource skills, while those that move down to become technicians need to adjust their applied skills. We extend our analysis to allow for occupation-specific human capital accumulation and retraining costs and show that U-shapes still arise. Since our findings challenge the importance of selection for wage growth because both bad *and good* workers leave occupations, human capital is the obvious remainder that can account for the positive relation between wages and occupational tenure.[3]

Of course, we do not think that the simple vertical sorting mechanism that we propose accounts for the full extent of occupational mobility. Both vertical and horizontal moves arise in the labor market, i.e. some occupations are considered better than others while some are just different. And among those that can be ranked 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. We do think, however, that the mechanism we emphasize should be an important part of any comprehensive theory of occupational mobility.

The remainder of the paper is organized as follows. In Section 2 we describe the set of new facts that characterize occupational mobility. In Section 3 we present the model that is

[2]The wage offer might in fact become negative, which we might interpret as firing.

[3]Our theory does generate returns to general experience as workers are able to sort better after learning.

consistent with the facts we document. Section 4 presents relevant extensions to our theory and Section 5 discusses it's contributions vis a vis the existing literature. Section 6 concludes.

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 Intergrated 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 0.995 and 0.005 percentile for each year of the selected sample described below.

We use the Danish rather that the U.S. data for two reasons. First, the sample size is much larger. One of our objectives is to documents the patterns of occupational mobility depending on the position of the individual in the wage distribution within her occupation. Large sample that ensures representative sample *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 US data (see Kambourov and Manovskii (2009b)).

2.1.1 Sample selection

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

We only select male workers in order to minimize the impact of the fertility decision on labor market transitions. Due to data limitations the sample is restricted to full time workers in the private sector. In the period 1995 to 1998 we do not observe the workplace of public employees and, to be able to use tenure information, we choose to include only the privately employed (rather than further restricting the time dimension of the data). The part-time workers are excluded because they do not have as dependable wage information. The sample is restricted to employees because we do not observe earnings for the self employed.

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 truncate the 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.

Finally, since we study occupational mobility between consecutive years, the sample only includes workers with valid occupation data in the year after we use them in the analysis (e.g., we use information from 2002 for this purpose).

Descriptive statistics of our sample are provided in the Table 1. Column 1 is the sample described above. Column 2 is for the sample where there is at least 10 workers in each occupation in each year. Column 3 is for the sample with at least 10 workers in each occupation, year, and years after graduation category and column 4 is for the sample with at least 100 workers in each occupation and year. Some of these samples are restricted to a minimum of workers in order to ensure a distribution e.g. within occupation and year. The samples will be used in the analyzes below.

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. Figure 1(a) is a non-parametric plot (from a kernel smoothed local linear regression with bandwidth 5) of the probability of switching out of an occupation as a function of a worker's position

Table 1: Summary statistics for the overall sample and subsamples

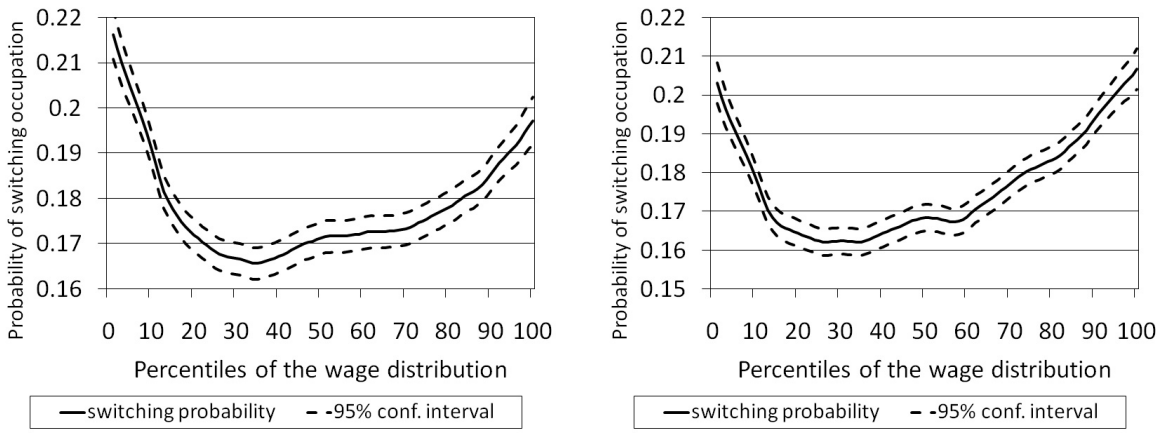
	Full sample	Over 10 per occupation and year	Over 10 per occupation, year, and experience	Over 100 per occupation and year
Number of observations	404800	402136	368520	375367
Number of occupations	353	229	143	105
Age	29.67	29.66	29.49	29.57
Occupational tenure	4.41	4.41	4.41	4.44
Occupational spell number	1.69	1.69	1.67	1.68
Occupational switchers	0.18	0.18	0.17	0.17
Employer tenure	2.36	2.36	2.33	2.35
Industry tenure	3.38	3.38	3.35	3.38
Years after graduation	6.49	6.49	6.40	6.48
Less than 12 years of school	0.04	0.04	0.04	0.04
Apprenticeship education	0.69	0.69	0.70	0.70
2 year university	0.10	0.10	0.10	0.10
Bachelor	0.10	0.10	0.09	0.08
Masters degree or above	0.07	0.07	0.06	0.06
Hourly wage in DKK in 1995	170.16	170.13	168.75	169.49
Married	0.30	0.30	0.30	0.30
Union	0.94	0.94	0.94	0.94
Number of children	0.71	0.71	0.70	0.71

in the wage distribution *in that occupation* in a given year. 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 wage deciles have the lowest probability of switching occupations.

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

$$lnw_{ijt} = X_{ijt}\beta + \epsilon_{ijt} \tag{1}$$

Where w_{ijt} is real hourly wage of an individual i in occupation j working in time t. The explanatory variables in X include dummies for calendar years, a three degree polynomial in general experience, a three degree polynomial in occupational tenure, a two degree polynomial of firm tenure, a three degree polynomial industry tenure, number of occupational spells, edu-



(a) wage distribution of raw wages within occupation and year

(b) wage distribution of wage residuals

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

cation, marital status, union membership, and regional dummies.[4] These wage regressions are estimated separately for each occupation.

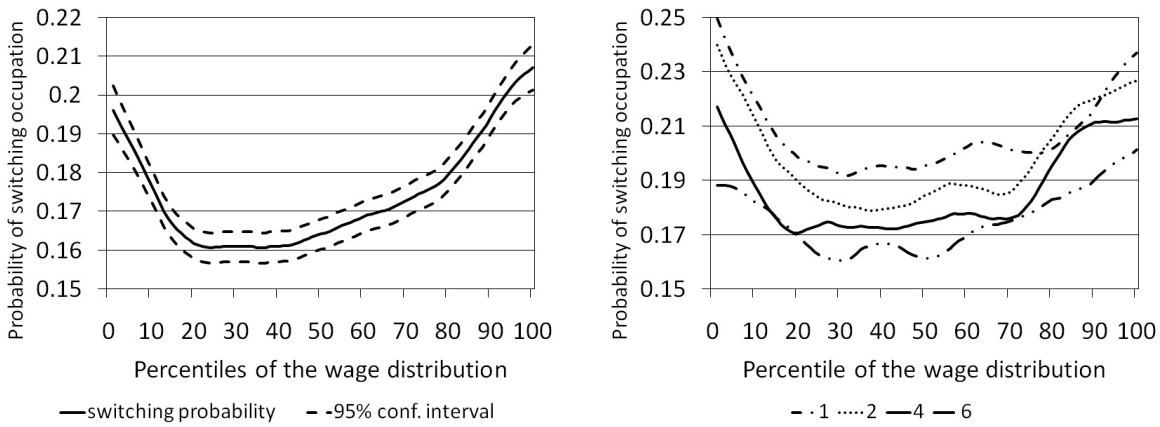
The U-shapes further hold if we look at wage percentiles within occupation, year, and years after graduation. Figure 2(a) plots the probability of switching occupation as a function of worker's position in the wage distribution of workers in the same occupation, calendar year, and years after graduation. Figure 2(b) separately graphs occupational mobility for 1, 2, 4, and 6 years after graduation. The figure shows U-shapes in occupational mobility for all years after graduation and shows that the level of mobility decreases with years after graduation for almost all percentiles of the within occupation, calendar year, and years since graduation wage distribution.[5]

An additional informative statistic is the percentage of occupation-year pairs that exhibit U-shapes. Computing this statistics requires enough workers in each occupation in each year to

[4]Figure A-1 in the appendix shows that excluding the regressors firm and industry tenure or excluding dummies for the occupational spell number in the wage regression does not change the qualitative result of the U-shape in occupational mobility.

[5]Figures 1 and 2 are both plots from a kernel smoothed local linear regression with bandwidth 5 for a random 60 percent of the data. In the Appendix Figures A-2 to A-5 we show that the U-shapes hold for bandwidths which are half and double of what we use in Figures 1 and 2.

An extension of looking at the patterns occupational mobility is looking at the workers who switch both occupation and firm at the same time. Graphs where workers switch both occupation and firm for similar wage distributions as in figure 1 and 2 are given in the appendix figure A-6. These graphs show that the probability of switching occupation and firm remains U-shaped however, the right tail of the wage distribution is not as increasing in the probability of switching as is the case when we do not condition on switching firm at the same time as switching occupation.



(a) wage distribution of raw wages within occupation, year, and years after graduation

(b) wage distribution of raw wages within occupation, year, and years after graduation for different years after graduation

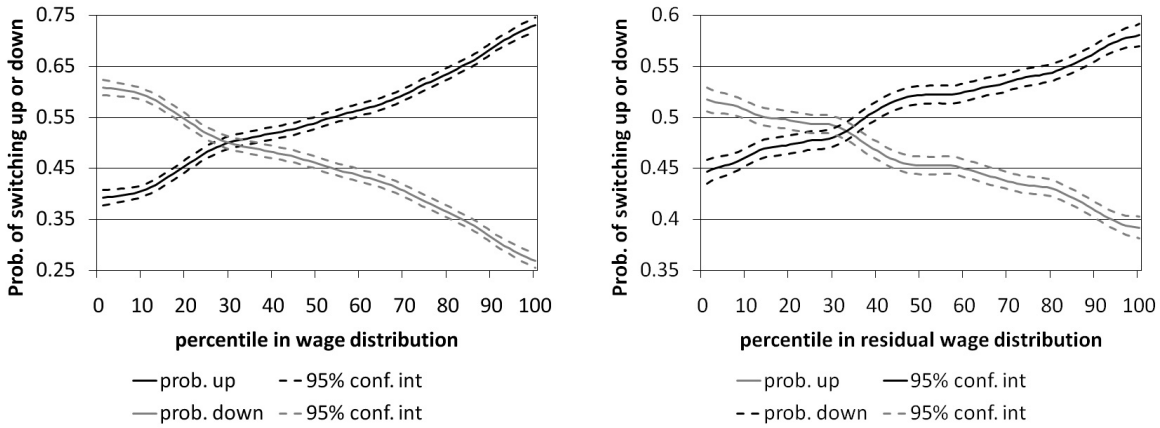
Figure 2: Non-parametric plot of probability of switching occupation by worker's percentile in the wage distribution within occupation, year, and years after graduation.

accurately predict a 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 and we divide the wage distribution of each occupation into quintiles. We define U-shapes in each occupation-year pair in two ways. First, we count an occupation in a given year as having a U-shape if the quintile with the highest probability of changing occupation is either quintile 1 or quintile 5. Second, we count an occupation in a given year as having a U-shape if, in addition, the quintile with the lowest probability of changing occupation is in the interior, i.e., quintile 2, 3, or 4. There are 598 occupation-year observations with at least 100 workers. 95 Percent of these have maximum probability of switching occupation in one of the extreme quintiles when the quintiles are based on raw wages. When the quintiles are defined on the wage residuals, 98% of occupations exhibit U-shapes according to this definition. In addition, 66% of the these occupations have a global minimum in the interior of the distribution of raw wages and 77% of the these occupations have a global minimum in the interior of the distribution of wage residuals.

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 wage *within* their occupation tend to switch to occupations with higher (lower) *average* wages. We first find the average wage of the 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.[6] To obtain the ranking based on residual wages, we use our selected sample to run a similar wage regression as equation 1 for each occupation where we include time dummies in the regression (without the intercept). We interpret the coefficients on these time dummies as the average occupational wage in a given year, adjusted for human capital accumulation of workers in the occupation as well as other worker characteristics such as education, regional dummies, and marital status. For this wage regression we include only occupations, which have more than 100 observation in total over the 8 year period 1995-2002.



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

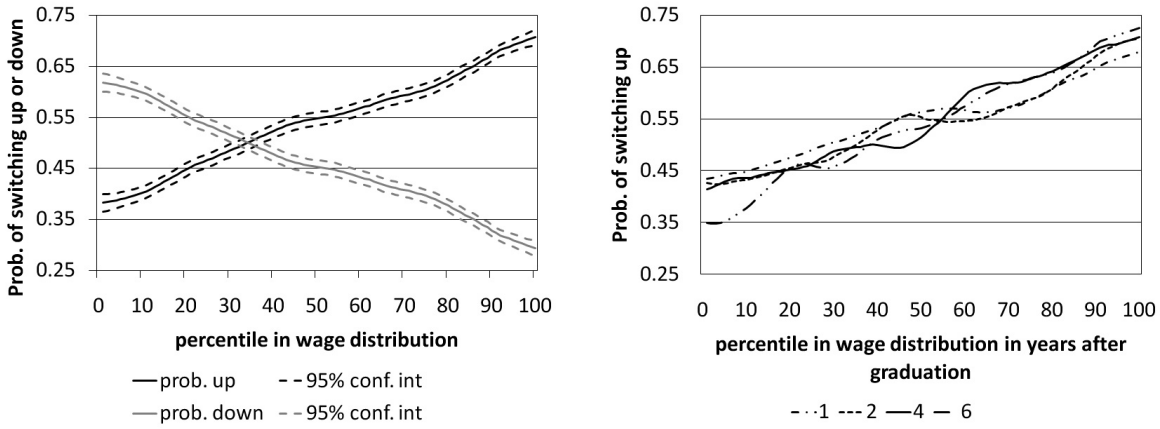
(b) wage distribution of wage residuals. Average wage in occupation from time constants in wage regression

Figure 3: Non-parametric plot of direction of occupational mobility, conditional on switching occupation.

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

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

residual wages within an occupation the worker is switching from. The evidence contained in these figures suggest that, conditional on switching occupation, the higher wage a person had in his occupation before the switch the higher is the probability that the worker will switch to an occupation with a higher average wage. Similarly, the lower wage a worker has in his occupation the higher is the probability that he will switch to an occupation with a lower average wage than in the occupation he switches from.



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

(b) wage distribution of raw wages within occupation, year, and year after graduation for individual years after graduation. Average wage in occupation from population.

Figure 4: Non-parametric plot of direction of occupational mobility, conditional on switching occupation.

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

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

As mentioned in the Introduction, these patterns are not implied by the existing theories of occupational mobility. Hence, in what follows we develop an alternative theory that is consistent with these features of the data. We confront additional implications of our theory with the data as we derive them.

3 The U-shapes of occupational mobility: Theory

The economy is set in a discrete-time infinite horizon setting, where workers choose employment in different occupations over time.

Workers: Each period a measure α 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 by factor $\beta \in (0, 1]$. Each worker has an ability level a_i that is drawn from a normal distribution with mean μ_a and variance σ_a. The amount of output that a worker can produce depends on his ability. In particular, he produces

$$X_i = a_i + \varepsilon_i \tag{2}$$

in a given period, where ε_i is a normally distributed noise term with mean zero and variance σ_ε. Workers don't know their precise ability, but observe the output they produce, even if they choose home production. We assume that the worker observes a first draw after finishing school, i.e. before the first time in the labor market, so that not all workers are identical when entering the labor force.[7]

While we think that not only ability but also occupation-specific human capital accumulation is an essential feature that leads to wage growth and that limits occupational switching, we first abstract from this to highlight the main insights of vertical occupational sorting in the simplest setting possible. We briefly return to this point in Section 4.

Occupations: There are a finite number of occupations, indexed by $k \in \{0, 1, ..., K\}$, in which workers can be employed. In each occupation the number of job opportunities is fixed to some measure γ_k that is constant over time. One can think of a limited measure γ_k of entrepreneurs who know how to implement the specific technology k, and each needs exactly one workers to operate the technology. The limited number of jobs in an occupation allows entrepreneurs to earn rents. We discuss entry of entrepreneurs in Section 4.

[7]The signal after school could have a different variance than the output realization - this would only complicate the notation slightly without altering the qualitative results.

Each unit of the good (or service) that is produced sells in the market at some exogenously given price P_k. We refer to the price of output also as the productivity of the occupation, and rank occupations in order of increasing productivity such that $P_K > ... > P_k > ... > P_0 = 0$. We interpret the lowest occupation as home production, which means that it is available to everybody.[8] An entrepreneur of type k who employs worker i thus obtains revenues

$$R_{ki} = P_k X_i.$$

This revenue function is supermodular, i.e. entrepreneurs in more productive occupations gain more from employing a more able worker than entrepreneurs in less productive occupations.[9]

Wages: We consider a competitive economy without matching frictions. The only frictions are information frictions in the sense that workers' abilities are not known. We assume that firms compete by posting output-contingent wages $w(X)$. An entrepreneur in occupation k who employs worker i has then an expected profit

$$\Pi_k = E(P_k X_i - w_k(X_i))$$

If an entrepreneur in occupation k can ensure himself some expected profit Π_k in any period by employing some specific worker, he can simply offer wage contract

$$w_k(X) = P_k X - \Pi_k \tag{3}$$

to any arbitrary worker. The specific worker is still willing to work at this firm because his expected wage is unchanged, and any other worker who accepts the job does not make the firm worse off. Therefore, such a "selling-the-shop" wage schedule has the effect that the firm does not need to know the type of the worker, but just needs to know how much profit it wants to secure to itself. It then adjusts the worker's wages according to (3) through performance-dependent boni/penalties in order to achieve this profit. We can therefore reinterpret the model as the workers offering a payoff Π_k to the entrepreneurs in occupation k for the right to work there and retain the surplus that is created.

[8] Availability to everybody means that $\gamma_0 > \alpha T$. The home production option guarantees that workers who have negative ability can still obtain a non-negative payoff. Our assumption that output is also observed in home production implies that there is no differential in the speed of learning.

[9] We note here that it is trivial to account not only for price differences but also for differences in the productivity of output generation across occupations because we can reinterpret P_k as the combination of selling price and the firms contribution to output. For example, an equivalent interpretation of our setup is that prices in all occupations are identical, but workers in occupation k product $P_k a_i$ units of output. In this interpretation jobs in more productive occupations can be viewed as higher up in a hierarchy that produces a homogeneous good (i.e. one manager may be equally important to production as several of his subordinates).

For workers only the expected wage that they can earn in a given occupation matters. If firms obtain expected profits Π_k with the wage schedule in (3) and a worker holds a belief about his own ability with a *mean* of A_i in this period, then his expected wage when working in occupation k is

$$\overline{w}_k(A_i) = P_k A_i - \Pi_k. \tag{4}$$

Without any costs to switching occupations it is clear that workers will choose the occupation that offers the highest expected wage. (We discuss switching costs together with specific human capital in the extensions.) Therefore, if workers switch occupations this happens because the expected wage in the alternative occupation is higher than the expected wage that they can obtain in their old occupation given the new information about their ability. While this sounds like voluntary quits by the worker, one may easily think of this as layoffs: If a worker realizes that he is worse than expected, the expected wage that he obtains after leaving profit Π_k to the entrepreneur might be very low in the current occupation. Such a wage offer might be interpreted as firing.

We have taken the stance that firms offer output contingent contracts so that workers self-select in the appropriate occupations even if the firm has no information about the prior work history (and the revealed signals). If the firm does observe the prior work history and has symmetric information relative to the worker, it can equally well offer a fixed wage based on the expected ability. This would correspond to the expected wage in (4). While we take a stance favoring the former, the prediction of U-shapes is robust to this assumption. We show U-shapes both when conditioning on realized wages as well as when conditioning on the expected wages that arise if the firm bears the risk of employment.

To formalize the optimal choice by the workers, let $I_k(A, \Pi)$ be the following indicator function: $I_k(A, \Pi) = 1$ if the expected wage according to (4) in occupation k is higher than in any other occupation, and $I_k(A, \Pi) = 0$ otherwise. Clearly this indicator depends on the vector $\Pi = (\Pi_1, \Pi_2, ..., \Pi_K)$ of profits that have to be left to the firms.

Updating: Neither workers nor entrepreneurs are sure about a worker's true ability. Each worker observes his output X_i and updates his beliefs according to Bayes' law. We are agnostic about whether firms learn as well, i.e. whether they observe the output history of a worker or not. The driving force in this model is the workers belief about his *mean* ability A_i in a given period. Of interest in solving the model is

1. how a worker updates his belief about his individual *mean* ability. This determines how individuals change occupations.

2. how the belief about the *mean* ability is distributed across the population. This determines the equilibrium profits Π_k and the associated wage offers according to (4).

For the first point, it is convenient to use the concept of precision, which is the inverse of the variance. Let $\phi_a = 1/\sigma_a$ and $\phi_\varepsilon = 1/\sigma_\varepsilon$, and define $\phi_t = \phi_a + t\phi_\varepsilon$ as the cumulative precision. A worker's initial belief about his mean ability before any output realization is $A_i^0 = \mu_a$. Consider a worker who has prior A_i^t in any period $t \in \{0, 1, .., T\}$ of his life and observes output realization X_i. Standard results on updating of normal distributions establish that his posterior mean A_i^{t+1} is the precision-weighted average of his prior mean and the observation

$$A_i^{t+1} = \frac{\phi_t}{\phi_{t+1}} A_i^t + \frac{\phi_\varepsilon}{\phi_{t+1}} X_i. \tag{5}$$

The weight on the prior increases the more observations you have already observed in the past, i.e. the higher t is. Correspondingly, the weight on the innovation decreases with years in the labor market. Workers become more convinced over time of their ability. Since workers draw once before entering the labor market, A_i^1 is the prior at the beginning of the first period of work. So the worker's posterior belief about his *exact ability* a_i is a normal with mean A_i^{t+1} and a variance of $1/\phi_{t+1}$.

For agents with prior A^t the realization of output and the resulting posterior mean A^{t+1} is still random. We denote the distribution of this posterior by $G_t(A^{t+1}|A^t)$ and its density by $g_t(.|.)$. One can show that this posterior is normally distributed with mean A^t and precision $\phi_t\phi_{t+1}/\phi_\epsilon$.[10] It is not important that the update is normally distributed. The following qualitative properties suffice for the results we want to show: $g_t(.|A)$ is single-peaked and symmetric around its peak at A, and shifting the mean A simply shifts the entire distribution about the posterior horizontally in the sense that $g_t(A'|A) = g_t(A' + \delta|A + \delta)$ for any δ. We call this last property lateral adjustment.

For the second point, note that at any point in time there is a measure α of workers that have been in the labor force for $t \in \{1, ..., T\}$ periods. Call the measure of workers in cohort t that have a belief about their mean ability weakly below A by $F^t(A)$, which is a non-normalized normal distribution.[11] We call the sum of this measure over the cohorts

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

[11] We have $F^t(\infty) = \alpha$ since the size of cohort t is α. Let $\tilde{F}_t$ be the probability that any given worker has a belief about his mean ability below A in period t. Then $F^t = \alpha\tilde{F}_t$. At the beginning of period t the workers

$F(A) = \sum_{t=1}^{T} F^t(A)$. Note that this distribution is independent of the choices of the agents because workers learn about their type in any eventuality. This simplifies the specification of an equilibrium substantially.[12] For simplicity we will assume that there are enough workers with positive levels of ability to fill all the jobs.[13]

Equilibrium: We are considering a standard stationary competitive equilibrium in this matching market between occupations and workers. Stationary means that the entrepreneurs' profits $(\Pi_1, \Pi_2, ..., \Pi_K)$ and the associated wage offers according to (3) are constant over time. Equilibrium implies that the workers' decisions equate demand and supply, where Π_k can be interpreted as the price workers have to pay to take over a job in occupation k.

Definition 1 *An equilibrium is a vector of profits* $\Pi = (\Pi_0, ...\Pi_K)$ *with* $\Pi_0 = 0$ *such that markets clear, i.e. for all* $k > 0$

$$\int I_k(A, \Pi) dF = \gamma_k.$$

As is standard in competitive equilibrium theory, one can interpret the market profits Π as optimal decisions by the entrepreneurs. Decreasing the demanded profit (i.e. increasing the wage) is not optimal because already all entrepreneurs employ a worker. Increasing the demanded profit (i.e. decreasing the wage) does not attract any worker, because workers expect to be able to work at the market wages. The indicator function $I_k(A, \Pi)$ ensures that workers indeed take optimal decisions when determining market clearing.

3.1 Analysis of Sorting

The tractability of the model arises from the fact that every period workers can reoptimize and therefore their life-time optimal decision is also the decision that maximizes the payoffs in each period. Since the distribution of mean abilities remains constant, we can solve most aspects with the standard tools for the analysis of static matching models. We provide these results first. Then we turn to problem that workers face over time as their individual uncertainty induces

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

[12] Other work such as Jovanovic and Nyarko (1997) and Papageorgiou (2007) focuses on differential speed of learning, which substantially complicates the analysis and limits the analysis in these papers to two occupations only. Moreover, these papers do not consider the implications for the U-shapes of switching behavior on which our analysis is centered.

[13] The precise condition for this is $\alpha T - F(0) > \sum_{k=1}^{K} \gamma_k$. Otherwise entrepreneurs in the less productive occupations do not fill their positions and thus these low occupations will not be observed.

agents to switch occupations as they transit to the stationary economy. These individual uncertainty yields high gross mobility of workers between occupations, even though the net mobility is by assumption zero in steady-state. Since gross mobility dwarfs net mobility in magnitude, this seems to be an important starting point. We will briefly touch on net mobility in the extension section.

3.1.1 Preliminary results about the equilibrium

The model can be easily be solved. In a given period, a workers decision only depends on his prior A about his mean ability. The revenue function $R = PA$ is super-modular, i.e. $\partial^2 R/(\partial P \partial A) > 0$. A result from the matching literature going back to Becker (1973) is that under supermodularity entrepreneurs in more productive occupations match with workers with higher mean ability in equilibrium. This is easy to see in our setup. Firms with higher productivity clearly make higher profits. A worker will choose occupation $k \geq 1$ over occupation $k-1$ only if the expected wage according to (4) is higher in the former, i.e.

$$P_k A - \Pi_k \geq P_{k-1} A - \Pi_{k-1}.$$

This is equivalent to

$$A \geq \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}} := B_k, \tag{6}$$

where B_k is the mean ability at which a worker is exactly indifferent between the two occupations. This shows that workers with a higher belief about their mean ability choose higher occupations. These worker can always mimic the choices of workers with lower beliefs, they have to earn higher wages than those. And since they choose better occupations, better occupations can be identified by the fact that they pay on average higher wages.

Since we assumed that there are enough workers with positive mean ability, all but the home production occupation will obtain strictly positive profits in equilibrium. To fulfill market clearing, it has to hold for all $k > 0$ that

$$F(B_{k+1}) - F(B_k) = \gamma_k, \tag{7}$$

where $B_{K+1} = \infty$. Moreover, $B_1 = \Pi_1/P_1$ and the measure of employed workers has to equal the overall demand for workers, which determines Π_1.[14] Then (7) can be used successively for higher k to determine the profits for all higher occupations. This constructively gives existence and exact levels for the profits in all occupations.

[14] The condition is $\alpha T - F(B_0) = \sum_{k=1}^{K} \gamma_k$.

3.1.2 Occupational mobility conditional on expected ability

An important part of the previous analysis is that it gives the levels B_k that determine at which belief a worker switches to a different occupation.

Consider a worker who has worked for $t > 1$ years and had a prior of $A \in [B_k, B_{k+1})$ in his t'th year of work. That is, he chose occupation k in the last period he worked. He will switch to a higher occupation between t and $t+1$ if his posterior $A^{t+1} > B_{k+1}$. We denote the probability of such an upward switch out of occupation k by $s_k^+(t, A^t)$. Conditioning on A^t is identical to *conditioning on the expected wage* $\bar{w}$ in (4) because of the one-to-one mapping between the two. The switching probability is given by

$$s_k^+(t, A^t) = 1 - G_t(B_{k+1}|A^t).$$

Similarly, if $A^{t+1} < B_k$ then the worker will switch to a lower occupation with a lower mean wage. We denote the probability of such a downward switch out of occupation k by $s_k^-(t, A^t)$ and have

$$s_k^-(t, A^t) = G_t(B_k|A^t).$$

The total switching probability is then $s_k(t, A) = s_k^-(t, A) + s_k^+(t, A)$. The domain of these functions is $[B_k, B_{k+1})$ because only with these priors would a worker choose occupation k.

In the following we will adopt the following convention, where our properties always refer to the second argument and not to the cohort indicator. Fix the cohort indicator t, then

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

Definition 3 (Strict U-shapes) *A function $f(t, .)$ is strictly U-shaped if it is U-shaped and its negative $-f(t, .)$ is single-peaked.*

U-shapes capture the qualitative feature that switching probabilities increase toward each of the ends of the domain, i.e. in the context of $s(t, .)$ switching becomes more likely for workers with low and high expected wages (abilities). Strict U-shapes additionally ensure that the switching probability increases monotonically from its interior minimum toward the extremes of the domain.

We will first consider the overall switching probability of a worker with prior $A \in [B_k, B_{k+1})$ in his t'th year of his work life

$$s_k(t, A) = G_t(B_k|A) + 1 - G_t(B_{k+1}|A).$$

Consider any occupation $k \in \{1, ..., K-1\}$ that is not at the extreme end of the spectrum. Since the distribution $g_t(A'|A)$ is symmetric and single-peaked, 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 5 illustrates this. The solid curve is the distribution of the posterior mean of an agent with prior $\overline{B}_k := \frac{B_k + B_{k+1}}{2}$. 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 distribution of the posterior mean for a worker starting with a prior above $\overline{B}_k$. It is more likely that his posterior lies above B_{k+1} compared to the solid curve, and this increase in the upper tail outweighs the decrease in the lower tail below B_k.

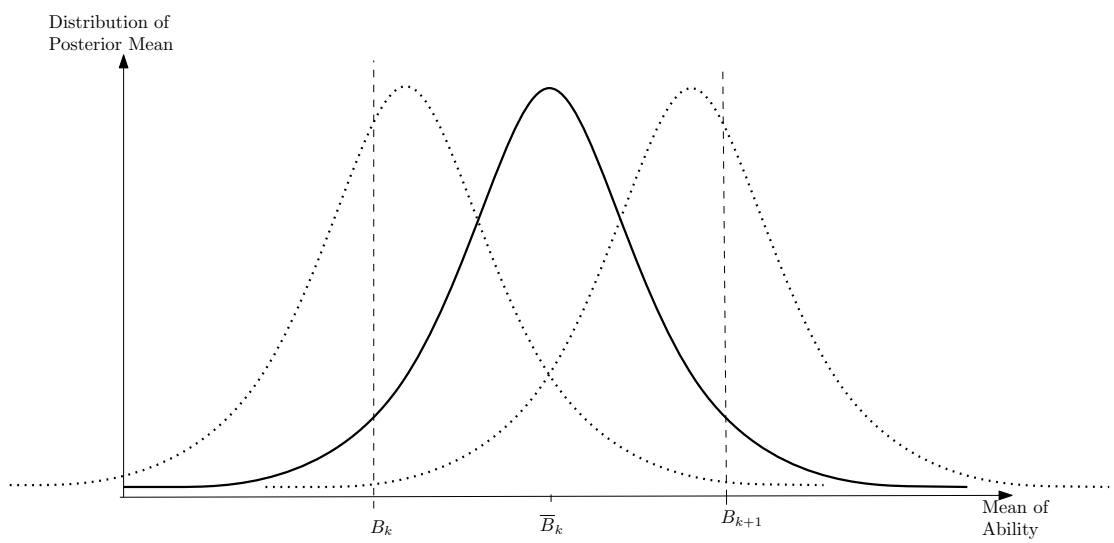


Figure 5: Illustration of the proof of Propositions 4 and 5.

Proposition 4 *In each occupation $k \in \{1, ..., K-1\}$ and for each cohort t, the switching probability $s_k(t, A)$ is strictly U-shaped in A.*

Proof. Let $\delta_k = (B_{k+1} - B_k)/2$ be half of the distance of interval $[B_k, B_{k+1})$, and recall that $\overline{B}_k = B_k + \delta_k$. Any other belief A can be written in terms of the distance δ from $\overline{B}_k$. Then

$$\begin{aligned} s_k(t, \overline{B}_k) - s_k(t, \overline{B}_k + \delta) &= G_t(B_k|\overline{B}_k) - G(B_k|\overline{B}_k + \delta) + G_t(B_{k+1}|\overline{B}_k + \delta) - G_t(B_{k+1}|\overline{B}_k) \\ &= G_t(-\delta_k|0) - G(-\delta_k - \delta|0) + G_t(\delta_k - \delta|0) - G_t(\delta_k|0) \\ &= \int_0^{\delta} \left[g_t(-\delta_k - \varepsilon|0) - g_t(\delta_k - \varepsilon|0)\right] d\varepsilon, \end{aligned} \tag{8}$$

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 (8) is strictly

negative for any $\varepsilon > 0$. Therefore, this interval is strictly negative for $\delta > 0$. When $\delta < 0$ the integrand of (8) is positive for all relevant ε but the range is negative, and so the integral becomes negative. Integral (8) decreases in the absolute value $|\delta|$. ■

For the extreme occupations of home production $k = 0$ and of $k = K$ the negative $-s(t, .)$ is also single-peaked, but the minimum is at the extreme of the domain, in the case of home production workers at the top are most likely to switch while in the case of the highest occupation workers at the bottom are most likely to switch.[15] The U-shapes are likely to persist when we condition on belief A but not on cohort t, yet theoretically there are cases where this does not hold. The reason is that at the same expected ability older workers have more precision and switch less. If young workers are mainly in the middle of the interval of mean abilities associated with a given occupation, while old workers are more at one side, this composition effect between cohorts can lead workers with interior abilities to switch more than those with abilities that are a bit more to the side. It is possible to construct examples where this happens in some occupation.

Next, we describe the direction of switching. Consider some occupation $k \in \{1, ..., K\}$. 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 tell them (and existing and prospective employers) that they are appropriate for better occupations. In Figure 5 this is visible 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 the ones most likely to find out that they are not as good as they thought and should move to a less productive occupation. As we mentioned before, such a switch might manifest itself through firing if the employer learns the same 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 captures this intuition about switching behavior. It characterizes the probability for upward and downward switches *conditional* on switching. If the switching probability $s_k(t, A) >$, then the conditional probability of switching up is $s_k^+(t, A)/s_k(t, A)$, and similar for downward switches.

Proposition 5 *In occupation $k \in \{1, ..., K-1\}$, among workers of experience t that switch the higher ability workers are more likely to switch up and the lower ability workers are more likely to switch down: $s_k^+(t, A)/s_k(t, A)$ is increasing and $s_k^-(t, A)/s_k(t, A)$ is decreasing in A.*

Proof. We can write $s_k^+(t, A) = 1 - G_t(B_{k+1}|A) = 1 - G_t(B_{k+1} - A|0)$, where the second equality follows from lateral adjustment. This is clearly increasing in A. A similar argument establishes

[15] Proposition 5 provides a more general formal proof for this.

that $s_k^-(t, A)$ is decreasing in A. This immediately implies that $s_k^-(t, A)/(s_k^-(t, A) + s_k^+(t, A))$ is increasing, while 1 minus this term is decreasing. ■

The analysis so far has conditioned on the prior A^t, which is equivalent to condition on the expected wage $\overline{w}_k(A^t)$ in (4). This is the easiest benchmark to establish in this environment.

3.1.3 Occupational mobility conditional on the realized wage

If the firm is not completely symmetrically informed about the workers ability, it is optimal to induce self-selection by the worker by offering the output-contingent wages $w_k(X_i)$ in (3) via boni or penalties for good and bad performance. Performance pay serves therefore a selection mechanism to attract people with the desired skills rather than an incentive device. An econometrician might not be able to elicit the belief A^t or the associated expected wage. Rather, he only observes the *realized wage* $w_k(X_i)$ that already includes performance boni or penalties. In analogy to our earlier definition about switching probabilities, we will denote the switching probabilities of a worker of cohort t who earned a wage w in the period t of his work life as $S_k(t, w)$. Similarly, $S_k^+(t, w)$ denotes the probability of upward switches and $S_k^-(t, w)$ the probability of downward switches. The domain of these functions is the entire real line since realized wages can take any value. We will establish the following two results.

Proposition 6 *In each interior occupation $k \in \{1, ..., K-1\}$ and for each cohort t, the switching probability $S_k(t, w)$ is U-shaped in w.*

Proof. See Appendix A1.1. ■

Figure 6 illustrates the logic behind the result. Given the wage w, we can back out from (3) the output realization $X(w)$ which is positively related with the wage. A worker with prior A will switch if his posterior mean exceeds the upper bound B_{k+1}. For given output $X(w)$ those workers with $A > A_w = (B_{k+1} - (1-\alpha)X(w))/\alpha$ switch upward where $\alpha = \phi_t/\phi_{t+1}$ is the weight in updating according to (5). Since the prior A is below B_{k+1} for workers who chose occupation k, no worker switches up if $X(w)$ below B_{k+1}. By a similar logic for $X(w)$ above B_k no worker switches down, so that the switching probability is minimal in the interior. The range of prior means A for which the workers switch upward becomes larger as the wage increases, and for high enough wages even the lowest type with prior B_k would switch upward and the switching probability becomes one. Similarly, when wages are low enough all workers will switch down and again the switching probability has a local (and global) maximum of one.[16]

[16] The reason why the switching probability might not be strictly U-shaped has to do with an inference effect. Consider a wage w at which all workers with prior mean above A_w switch upward. At a higher wage w' the

Weak U-shapes arise even if we do not condition on cohort identifier t, i.e. we only condition on the wage a person received in a given period in an occupation. Clearly for some intermediate wages the switching probability is less than one, while for very low and for very high wages any worker that chose occupation k is induced to switch.

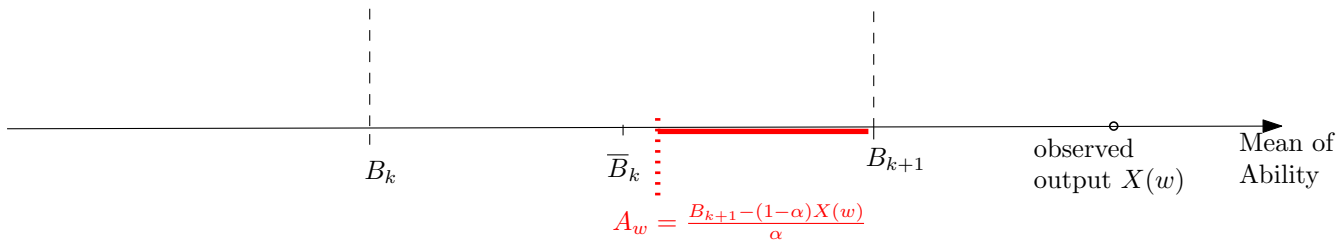


Figure 6: Illustration of the proof of Propositions 6 and 7.

As in Section 3.1.2 we again obtain the following directions for switching similar to those in the data.

Proposition 7 *In occupation $k \in \{1, ..., K-1\}$, among workers of experience t that switch the higher wage workers are more likely to switch up and lower wage workers are more likely to switch down. That is, $S_k^+(t,w)/S_k(t,w)$ is increasing in w and $S_k^-(t,w)/S_k(t,w)$ is decreasing in w.*

Proof. Conditional on switching means that output $X(w) = (w - \Pi_k)/P_k$ is either below B_k, in which the case the worker switches downward for sure. Or $X(w)$ is above B_{k+1}, in which case the move is upward because the belief about mean ability has improved. ■

3.1.4 Wage changes associated with occupational switching: Theory

The model has the immediate feature that cohorts with more years in the labor market receive on average higher wages. This is an immediate effect of learning, which allows workers to sort themselves into more appropriate occupations.

As a secondary result of our analysis we also obtain predictions about the behavior of wages before and after a change of occupation. It is clear that a worker who switches gets higher wages in $t+1$ in the new occupation than he would get if he had stayed in his previous occupation.

range of priors at which workers will switch extends, i.e. $A_{w'}$ is lower than A_w. While this extends the region $[A_w, B_{k+1})$ in which workers switch and decreases the region $[B_k, A_{w'})$ where workers stay, it also changes the likelihood that a given worker at this wage is from the first interval relative to the second. It is possible that higher wages mean that the lower interval is more likely, which can lead to non-monotonicities.

But this does not tell us about the relationship between his new wage in $t+1$ and his old wage in t.

We are interested in the question whether a worker who earned wage w in some period and then switches his occupation is likely to earn more after the switch then in the period before the switch. Let $E_k^+(w'|w)$ be the expected wage for a worker who earned wage w in occupation k in the last period and now switched upward. Let $E_k^-(w'|w)$ the expected wage for a worker who switched down.

We will show that a person who switches downward has on average an improvement in the wage, while for a person that switches upward it is ambiguous whether the wage will be higher or lower. The result is based on the following logic. A worker switches down if his observed wage falls below his expected wage. Yet his belief about his own ability is the average between his prior and his update, and therefore his next wage is likely to be not as bad as the wage that led him to revise the belief downward. Moreover, by reoptimizing his occupation he finds an even better fit. For workers that change upward this is not necessarily the case: The high wage that they observe and that led them to revise their beliefs upward is an outlier, and next periods wages are not likely to be that high. Still the reoptimizing of the fit with the occupation might raise the wage. Whether this second effect is sufficient depends on the exact parameters.

Proposition 8 *Consider workers that switch occupation between two years. For those who switch down the average wage before the switch is lower than the average wage after the switch. For workers who switch up this is ambiguous.*

Proof. Consider a worker with wage w in occupation k and associate output $X(w) = (w + \Pi_k)/P_k$. He switches downward only if $A \geq B_k$ but $X(w) < B_k$, and thus $X(w) < A$. If he stayed in occupation k, then his expected wage according to (4) after switching is $P_k A' - \Pi_k$ where $A' = \alpha A + (1-\alpha)X(w)$ and $\alpha = \phi_t/\phi_{t+1}$. We have

$$\begin{aligned} P_k A' - \Pi_k &> w \\ \Leftrightarrow P_k \alpha A + P_k(1-\alpha)X(w) - \Pi_k &> w \\ \Leftrightarrow \alpha A &> \alpha X(w), \end{aligned} \tag{9}$$

which we showed to be true. Moreover, a worker only switches if this improves his expected wage relative to staying in the previous occupation, and therefore $E^-(w'|w) > P_k A' - \Pi_k > w$. This proves that a downward move is on average associated with an improvement of the wage. The logic does not apply to upward shifts, because in this case $X(w) > A$. Therefore inequality (9) is no longer true and the wage would on average go down relative to the previous period if

the worker remained in k. Whether the reallocation improves the wage enough relative to this wage decrease depends on the exact difference $P_{k+1} - P_k$. ■

While the wage for workers who switch down improves relative to the previous period and might decline for workers that switch up, we nevertheless obtain the following ranking of wages in any given year:

Proposition 9 *Consider occupation $k \in \{1, ..., K-1\}$. After any given year some workers in this occupation switch up to $k' > k$, some switch down to $k' < k$, and the rest stay. In the next year the average wage for those who switch up is higher than for those who stay, which is higher than for those who switched down.*

Proof. People switch because they change their belief about their mean ability. In any given period people with higher mean ability do on average better, because at worst they can choose exactly the same occupations as those with lower ability and earn higher profits. ■

3.1.5 Wage changes associated with occupational switching: Evidence

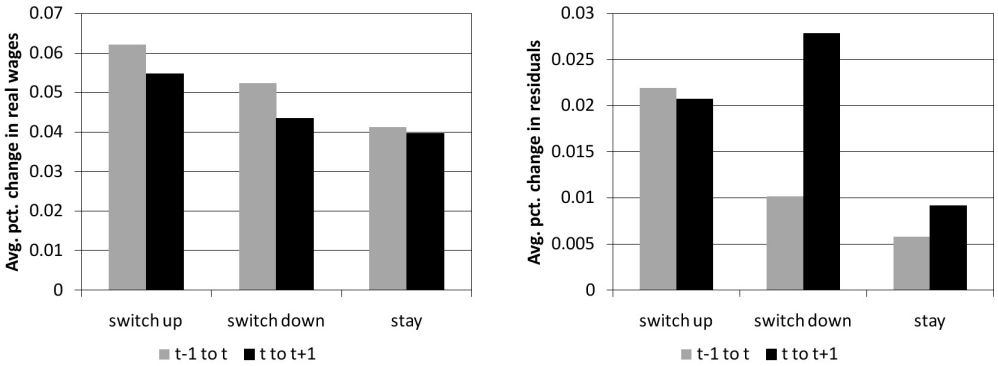
Closer investigation of the data supports these conclusions about wage dynamics. For workers that switch to a lower ranked occupation the wages indeed improve relative to their last wage in the previous occupation. For workers that switch to higher ranked occupations the effect is ambiguous, and in general they take slight wage cuts. Nevertheless in any given period the wages of switchers vs stayers are ordered as in Proposition 9.

Consider a further restriction of our sample to workers who do not change occupations from year $t-1$ to t. From period t to $t+1$ some of these workers switch to higher ranking occupations, some switch to lower ranking occupations, and some stay. Figure 7 illustrates that workers who switch to higher ranking occupations have large real wage percentage increases both before and after the switch relative to workers who switch down or stay. In figure 7(b) we plot the percentage changes in wage residuals, rather than wages, for the three groups. The wage residuals are from wage regression 1 including occupational dummies and occupational dummies interacted with tenure in the occupation and excluding occupational spell number. We take the exponential of the residuals from the wage regression and find percentage change for the three groups. For people switching occupation between period t and $t+1$ we subtract the mean of the old occupation from period t and add the the mean of the new occupation in period $t+1$ to the residuals change. We do this to compare workers residuals in a their new occupation (with higher mean if they switched up) to the residual of their old occupation.

We employ this procedure to control the wage changes of the three groups for composition effects of any of the included control variables in the wage regression. For example, if engineers

have higher wage growth than architects and architects have lower probability of switching occupation than engineers, then occupational switchers will have higher wage changes simply because they originate mainly from a different occupation than stayers. This is the compositional effect we would like to control for.

Figure 7(b) shows that when we control for observables in the wage changes by using residuals instead of raw wages, the change in residuals show that workers who switch to higher ranking occupations experience a smaller residual wage increase after the switch than they did in the year prior to switch. Furthermore, the workers who switch to lower ranking occupations experience a larger increase in residual wages after the switch than a year before. Workers who stay in their occupation in all three years have small residual percentage changes both between $t-1$ to t and between t to $t+1$. The appendix figure A-7 shows the change in wages and residuals in absolute levels of the same people from figure 7.



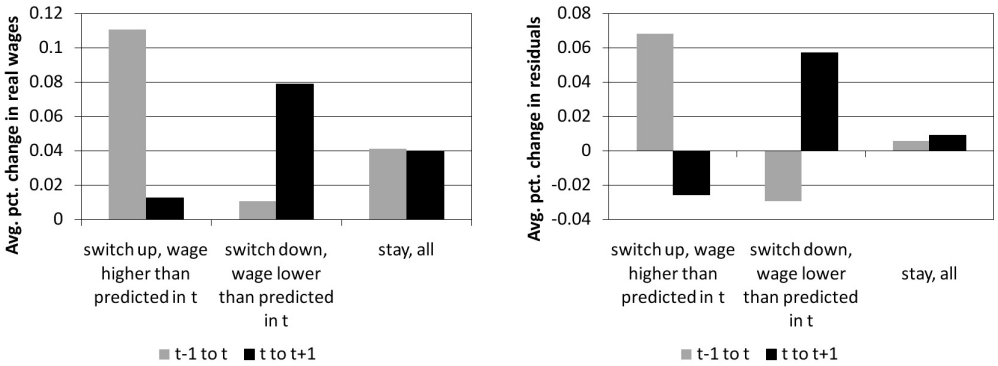
(a) Percentage wage change for workers who switch occupation up, down, or stay

(b) Percentage change in wage residuals for workers who switch occupation up, down, or stay

Figure 7: Change in wages and wage residuals for workers who switch occupation up or down between years t and $t+1$ and stay in their occupation between $t-1$ and t. Stayers are in the same occupation in years $t-1$, t, and $t+1$

Our theory implies that workers who receive a wage higher than their expected wage will update upward beliefs about their ability level. Thus, these are the workers who are possible candidates of switching to higher ranking occupations. Similarly, workers who receive a wage that is lower than their expected wage, will adjust downward the expectation of their ability and are the candidates of switching to lower ranking occupations. In figure 8 we use the workers' predicted wages in period t and compare the prediction to their actual wages. For workers who switch to higher occupation between years t and $t+1$ we include only those who had an actual wage in year t, which was higher than their predicted wage. For workers who switch to lower

occupations between years t and $t+1$ we use only the workers who had actual wages in period t, which were lower than what the wage regression predicts for them in year t. These are the occupational switchers our theory describes.



(a) Percentage wage change for workers who switch occupation up, down, or stay, using differences between actual and predicted wages in period t

(b) Percentage change in wage residuals for workers who switch occupation up, down, or stay, using differences between actual and predicted wages in period t

Figure 8: Change in wages and wage residuals for workers who switch occupation up or down between years t and $t+1$ and stay in their occupation between $t-1$ and t. Stayers are in the same occupation in years $t-1$, t, and $t+1$. Using differences between actual and predicted wages in period t

Figure 8(a) illustrates that workers who switch occupation up between years t and $t+1$ and who had higher actual wages than their predicted wages in year t experience a much lower wage increase after the switch than a year before. Furthermore, workers who switch occupations down between t and $t+1$, and who had an actual wage lower than their predicted wage in year t, experience a large wage increase when they switch occupation and only a slight wage increase in the year before the switch. These results are consistent with our theory.

Figure 8(b) plots the change in wage residuals for the same sample of workers as in Figure 8(a). Controlling for the composition effect of observables from the wage regression including occupational dummies further supports the prediction of wage changes around an occupational switch. Workers who switch to a higher ranking occupation between years t and $t+1$ and who had higher actual wages than their predicted wages in year t experience an increase in residuals one period before the switch and a decrease in residual wages upon the switch. For workers who switch to a lower ranking occupation and who had an actual wage lower than their predicted wage in year t experienced a decrease in residuals a year prior to the switch and a small increase in residual wages upon the switch. Workers who stayed in the same occupation in all three years had almost no change in their residuals between any years. The appendix

figure A-8 shows the change in wages and residuals in absolute levels of the same people from figure 8.

4 Extensions

In this section we discuss three extensions. First, we introduce changes to the productivity of occupations. Second, we allow for entry of firms. Third, we allow for human capital and switching costs (but leave productivities constant).

4.1 Changing Occupational Productivities

We denote calendar time by τ and index occupations by a name $r \in \{0, 1, ..., K\}$ rather than their rank in terms of productivity, with $r = 0$ still being home production. We continue to assume that prices $P_r^\tau > 0$ are a (realization of a possibly stochastic) function of calendar time for all occupations $r > 0$. We assume still that the measure of entrepreneurs in an occupation remains constant. Let $r_\tau(k)$ be the name of the occupation that has a productivity that is higher than that in k other occupations. Since workers optimal occupational choice still coincides with the choices that maximizes their utility in the current period and since the distribution F of beliefs remains stationary, we can solve the model period by period as outlined in the previous section. In each period we can assign prices $P_k = P^\tau_{r_\tau(k)}$ and solve for the period equilibrium profits and cutoffs via the same equations (6) and (7) from the previous section. This delivers the boundaries B_k for this period. The lower and upper boundaries for the beliefs of workers in occupation $r_\tau(k)$ in this period are then $\underline{B}^\tau_{r_\tau(k)} = B_k$ and $\overline{B}^\tau_{r_\tau(k)} = B_{k+1}$. We assume strict ranks of occupations in all periods and denote by Γ_r^τ the measure of all jobs that have weakly lower output prices (i.e. do not belong to more productive occupations) than the jobs in occupation r in period τ. We call Γ_r^τ the *position* of occupation r in the distribution of productivities.

When the positions for all occupations remain constant between two consecutive periods, the switching behavior of workers $s_r(t, A)$ is exactly as outlined in the previous section.[17] When the position of a specific occupation r stays constant for two periods, i.e. $\Gamma_r^\tau = \Gamma_r^{\tau+1}$, it is easy to show that still the cutoffs that determine who stays in the occupation remain constant, i.e. $\underline{B}_r^\tau = \underline{B}_r^{\tau+1}$ and $\overline{B}_r^\tau = \overline{B}_r^{\tau+1}$, and so this switching behavior of workers still remains unchanged. Moreover, in this case it follows directly from lateral adjustment in updating that workers with the highest and lowest belief have equal switching probabilities. This changes when the relative rankings change.

[17] The function $S_r(t, w)$ is constant only if also the prices remain constant for both periods.

Proposition 10 *When an occupation improves its position, $\Gamma_r^{\tau+1} > \Gamma_r^{\tau}$, the workers with the lowest prior mean in occupation r are more likely to switch then than their counterparts with the highest priors, and the ability of the workforce improves in the sense of first order stochastic dominance relative to the previous period. For a declining occupations with $\Gamma_r^{\tau+1} < \Gamma_r^{\tau}$ the opposite is the case.*

Proof. We will consider the case $\Gamma_r^{\tau+1} > \Gamma_r^{\tau}$; the other case follows by analogous arguments. In period τ the workers with the highest belief in occupation r have belief $\overline{B}_r^{\tau}$ and those with the lowest belief have $\underline{B}_r^{\tau}$. The shift in the position implies that $\underline{B}_r^{\tau+1} > \underline{B}_r^{\tau}$ and $\overline{B}_r^{\tau+1} > \overline{B}_r^{\tau}$. Workers stay in occupation r if their posterior belief is in $[\underline{B}_r^{\tau+1}, \overline{B}_r^{\tau+1})$. Since this interval is closer to $\overline{B}_r^{\tau}$ than to $\underline{B}_r^{\tau}$ the likelihood that the update falls in this interval is higher for the high worker types. The fact that the mean abilities of the workers that choose occupation r get higher in the sense that $\underline{B}_r^{\tau+1} > \underline{B}_r^{\tau}$ and $\overline{B}_r^{\tau+1} > \overline{B}_r^{\tau}$ implies first order stochastic dominance of the ability distribution. ■

Changes in the position of an occupation have direct consequences for the wages that are paid. Clearly, since the workforce becomes better the improvement is associated with rising wages. Also we obtain predictions for the wages of stayers, i.e. of those workers that do not change occupations. The conditions in the following propositions are fulfilled for example when two occupations of equal size switch productivities while the productivities of all other occupations stay the same, but also hold under other reasons for changes in position induced by shifts of multiple occupations.

Proposition 11 *If the position of an occupation increases sufficiently in the sense that $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$, then workers that stay in this rising occupation all earned wages above the occupation average in period τ. For a sufficient decline $\Gamma_r^{\tau+1} \leq \Gamma_r^{\tau} - \gamma_r$ workers that stay in this declining occupation earned wages below the occupation average.*

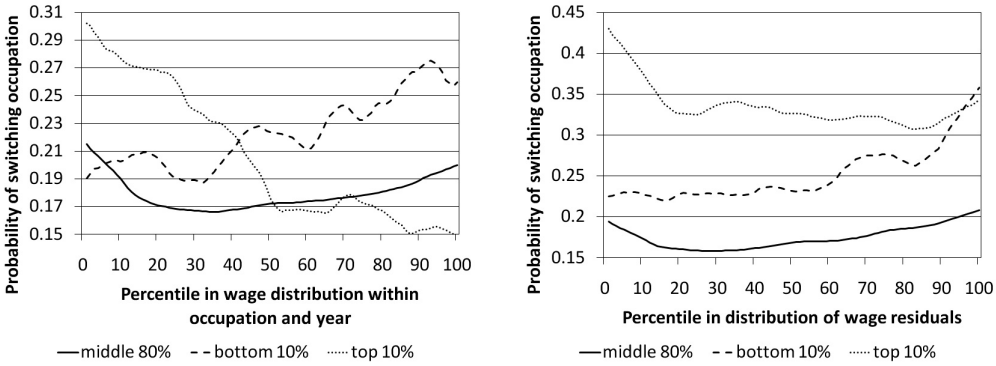
Proof. The rising occupation attracts workers with mean ability in $[\underline{B}_r^{\tau}, \overline{B}_r^{\tau})$ in τ. The condition $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$ implies that the lower bound in the next period is higher than the upper bound in τ, i.e. $\underline{B}_r^{\tau+1} \geq \overline{B}_r^{\tau}$. Since the average wage $\bar{w}_r^{\tau}(\overline{B}_r^{\tau})$ according to (3) of the worker with the highest belief $A = \overline{B}_r^{\tau}$ is above the occupation average in period τ, workers that earn below average wages earn a wage below $\bar{w}_r^{\tau}(\overline{B}_r^{\tau})$. They therefore have output observations X that are below $\overline{B}_r^{\tau}$. Therefore no worker with below average wages improves his posterior above $\overline{B}_r^{\tau}$, and therefore none of them improves his posterior into the range $[\underline{B}_r^{\tau+1}, \overline{B}_r^{\tau+1})$. In contrast, some of the workers with above average wages improve their posteriors into $[\underline{B}_r^{\tau+1}, \overline{B}_r^{\tau+1})$ and are suited for the rising occupation. A similar argument applies to the declining occupation. ■

The result is driven by the fact that only those workers stay whose posterior improves in line with the increase in occupational importance and who, thus, remain suitable for this occupation. Only workers with above average wages fulfill this criterion. Even if we relax $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r$ a bit such that high ability workers remain even after an output realization below their expected average, the statement still remains true because these retained workers still earn more than the occupational average. If we relax this condition further some workers with high prior will stay even when they earn wages below the occupational average because their posterior is still sufficiently above initial period's lower bound $\underline{B}_r^{\tau}$. In general, the more $\Gamma_r^{\tau+1}$ improves over Γ_r^{τ}, the higher the lower bound of wages of the workers who still remain in the occupation. Similarly for a declining occupation: The more the position $\Gamma_r^{\tau+1}$ drops below Γ_r^{τ}, the lower the higher bound on wages of the workers who still remain in the occupation.

4.1.1 Mobility in response to shocks: 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 is by studying occupations with different growth rates of the average wage. The average wage of an occupation is found in the same two ways as in section 2.3. First, we find the average wage of the full time private sector workers in a given occupation in a given year. Alternatively, we find the average wage of an occupation in a given year is by using our selected sample to run a wage regression for each occupation where we include time dummies in the regression. We use use the coefficients on the time dummies in the regression as the average residual occupational wage in a given year.

Next, 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 9(a) plots three groups of occupations, separated by the growth rates in raw 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 function of the workers' position in wage distribution in their occupation in year t. Figures 9(a) and 9(b) show that workers in the lowest growing occupations between t and $t+1$ have higher probability of leaving their occupation between t and $t+1$ if the are from the upper end of the occupational wage distribution in year t. Workers in the fast growing occupations have



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

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

Figure 9: Non-parametric plot of direction of occupational mobility, conditional on switching occupation.

higher probability of changing occupation if they are in the low end of the wage distribution in their occupation. Workers in occupation, which grows faster than the slowest 10 percent but slower than the fastest 10 percent, have a probability of changing occupation that is U-shaped in their in their wage percentile.[18]

4.2 Free Entry into Occupations

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

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

[18]The results is robust to calculating average wage change of the occupation only from workers who stay in the occupation between t and $t + 1$.

The drawback of having only fixed costs $\overline{c}_k$ is the response of the market when productivities change over time. Among the occupations that hire workers, those with lower productivity have to have lower fixed costs because otherwise they would not be competitive and would not hire any workers. As long as the rank of occupations does not change the analysis is straightforward. Yet if such an occupation changes rank with the next higher occupation, then it has higher productivity and lower costs and the other occupation completely disappears. There are various reasons why we don't expect this to occur: Prices might change in response to output changes or costs might change in response to the number of jobs in the occupation. The second might reflect the fact that resources into production become scarce when more entrepreneurs produce. Alternatively, it can be interpreted as cost heterogeneity among entrepreneurs and $c(\gamma_k)$ reflects the costs of the marginal entrant: The more entrepreneurs enter the less able the marginal one is.[19] We integrate this idea into the model by assuming that $c(.)$ is increasing and convex. If prices are always high enough to cover the fixed cost, then some Inada conditions on the second component ensure that even with changing productivities no occupation completely vanishes, but the level of its operation might substantially vary.[20][21]

An equilibrium is now a tuple $\Pi = (\Pi_0, ..., \Pi_K)$ of profits and a tuple $\gamma = (\gamma_0, ...\gamma_K)$ of entry levels such that all conditions in Equilibrium Definition 1 are satisfied and additionally it holds that $\Pi_k = C(\gamma_k)$ for all $k > 0$. All results regarding switching behavior from Section 3 apply, only that now the cutoffs B_k are determined in a way that incorporates optimal entry. It is easy to solve for these cutoffs by considering the following set of equations in analogy to (6) and (7)

$$\frac{C(\gamma_k) - C(\gamma_{k-1})}{P_k - P_{k-1}} = B_k, \tag{10}$$

$$F(B_k) - F(B_{k-1}) = \gamma_k, \tag{11}$$

for all $k > 0$.

Equation system (10) and (11) allows us to determine the size of each occupation in each

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

[20] In particular, it is easy to verify that the following conditions ensure employment in all occupations $k > 0$ in all periods. Assume that $c'(0) = 0$ and there is some constant $\psi > 0$ and employment level $e = [\alpha T - F(\psi)]/K$ such that $\lim_{\gamma \to \varepsilon} c'(\gamma) = \infty$. This ensures that no occupation employs more than e workers. Moreover, let $P > 0$ be the lowest price that can ever arise in any occupation (apart from home production). Then $\psi P > \max_k \overline{c}_k$ ensures that it is optimal to have at least some employment in each occupation at each point in time because worker with ability ψ never gets employment and therefore could be hired for free.

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

period even in the case when productivities are changing as in the previous Subsection 4.1. We can now define an improving occupation in the sense of Proposition 10 as one that improves its position at both the high and the low end, i.e. $\Gamma_r^{\tau+1} > \Gamma_r^{\tau}$ and $\Gamma_r^{\tau+1} - \gamma_r^{\tau+1} > \Gamma_r^{\tau} - \gamma_r^{\tau}$.[22] A sufficient increase in the sense of Proposition 11 still means $\Gamma_r^{\tau+1} \geq \Gamma_r^{\tau} + \gamma_r^{\tau}$. With these extended definitions the propositions remain valid. If on the other hand an occupation with increasing productivity expands so much in size that the measure of jobs with strictly lower productivities $\Gamma_r - \gamma_r$ actually decreases, it starts to employ not only more high ability but also more low ability workers. When we consider a smooth increase in the productivity of occupation m and hold the other productivities fixed, it is easy to see that the expansion of the workforce is continuous but the position switches upward when it overtakes another profession, at which point indeed both upper and lower position Γ_r and $\Gamma_r - \gamma_r$ increase jointly and the ability of the work force improves substantially in the sense of first order stochastic dominance.

4.3 Human Capital and Switching Costs

Here we briefly introduce human capital and switching costs in the basic environment of Section 4.1. Whenever a worker wants to switch into occupation k he has to pay cost κ_k. This captures application effort, retraining costs and time the worker is not on the job. Moreover, a worker who has already worked ι consecutive years in occupation k has human capital $h_k(\iota)$ in the next period. We assume that human capital is zero without experience, and the human capital function is weakly increasing. If a worker switches occupation, he looses his human capital and $\iota = 0$. The output of a worker with ι years of occupational experience in occupation k is in analogy to (2)

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

Wages are still determined by (3) given the profit Π_k that firms want to obtain. The main difference to the previous sections is that workers solve a dynamic programming problem when deciding on the optimal occupation decision. Since human capital is a deterministic function, a worker who observes his output can back out $\tilde{X}_i = a_i + \varepsilon_i$, and therefore learning is not affected by human capital accumulation and the distribution of mean abilities F remains unchanged. We again consider a stationary equilibrium where firms' equilibrium profits Π_k remain constant over time. We define the precise notion of an equilibrium for this setup in Appendix A1.2.

It is straightforward to show that our Propositions 6 and 7 carry over to this setting. For any intermediate occupation $k \in \{1, ..., K-1\}$ there is an upper and lower bound on the prior

[22] Again superscripts indicate the period.

mean A among the agents that choose these occupations. Given that the expected wage per period is

$$P_k(A + h_k(\iota)) - \Pi_k$$

workers with very low priors will not have enough periods of employment left to recover the losses if they don't choose home production. For agents with very high priors it clearly dominates to choose the highest occupation. Given that priors in the intermediate occupations are bounded, after very high wage observations any worker that chose an intermediate occupation will update his prior above the upper bound and choose occupation K, so that for very high occupations the probability to switch upward is one. Similarly, for very low wage observations the probability to switch downward is one. For some intermediate wages some agents do not change occupations, introducing an interior minimum.

Our observation that wages rise for agents who switch to lower occupations still holds for those agents in their last period T of their work life who switch because their prior went down. If they would have stayed their wage would be higher than the wage that made them switch, because the posterior is only partly influenced by the low wage observation that induced the switch. Given that they switch their wage in the new occupation must be higher than if they had stayed even after accounting for the loss due to switching costs and human capital destruction. In earlier years of their working life workers can recoup the loss due to switching costs and human capital destruction over several periods, and immediate wages can decrease because these costs materialize immediately while the benefits are spread out. Whether immediate wage gains arise for such earlier cohorts, therefore, depends on the exact parameters.

5 Connection to existing models

5.1 Basic Search Models

Work on occupation-specific mobility is based on the assumption that workers sort themselves according to the fit of the worker to the occupation. The standard assumption is that all occupations are essentially the same, only that a worker might fit better to some occupation than to another. This is usually modeled as a shock which the worker learns over time (McCall (1990), Neal (1999)). In such models low wages are an indication of a bad match, and low wage workers are the ones who leave in order to find a better match. In contrast to such "horizontal" heterogeneity of occupations we pursue the idea of "vertical" heterogeneity in which some occupations are more productive than others. Also workers are heterogeneous, and there is complementarity between workers and occupations. In such a world a bad fit can be

characterized by underqualification *or overqualification* of a worker for a particular job. This means that not only low wage workers leave an occupation, but also very qualified workers with high wages.

5.2 Roy model

The idea that occupations might be vertically ordered goes back at least to Roy (1951). In the basic version of the Roy model according to the formalization in Heckman and Honore (1990) there are two occupations 1 and 2. Each worker is endowed with a two-dimensional skill set (s_1, s_2) that describes his skill in each occupation. A worker observes his skills perfectly but they are unobserved by the econometrician. The skill-endowment in the population is governed by some two-dimensional type distribution. The output in occupation i can be sold for price P_i to which we refer as the occupation's productivity. The wage of a worker with skills (s_1, s_2) is $P_1 s_1$ in occupation 1 and $P_2 s_2$ in occupation 2. Each worker chooses the occupation where he earns the highest wage.

Figure 10 illustrates the implications of the Roy model. The dotted curve illustrates the skill distribution. In general this can be some arbitrary cloud. The specific version drawn is one of absolute advantage in which a person with a high skill in one occupation also has a high skill in the other occupation. The solid curve is the indifference curve between the two sectors: All workers whose type lies below that line prefer occupation 1 while all workers whose types lie above prefer occupation 2. Since wages are linear in skills, the line goes through the origin. The distinguishing feature is that there are enough jobs in each occupation, and each worker who wants to work in an occupation can do so and earn the prevailing wages. Occupational switching arises only if prices change and the solid curve shifts, i.e. the model focuses on gross mobility.

In our model workers are characterized by their ability a that is common across all sectors. Of main concern is the learning about this ability. Yet with only two sectors and known abilities our model can be compared to the Roy model. The main difference to the Roy model is that the skill distribution is (a, a) and thus goes through the origin, while the indifference curve does no longer go through the origin since the profits that entrepreneurs earn in each of the occupations introduces an intercept. These profits are due to the scarcity of the production opportunities, which introduces competition among workers for jobs and sets our approach apart from the Roy model. We illustrate the features of our model in Figure 11.

For given prices P_1 and P_2 the models are similar since that Figures (10) and (11) are rotations of one another. In this sense one can interpret our model as an extension of the Roy

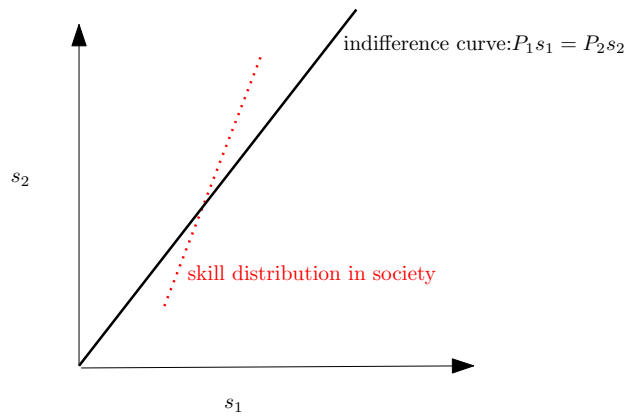


Figure 10: Illustration of the Roy model. s_i and P_i: skill level and price of output in occupation $i \in \{0, 1\}$.

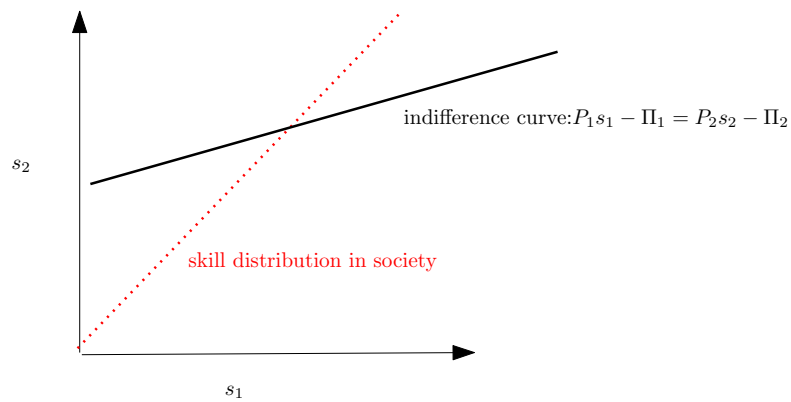


Figure 11: Illustration of our model. $s_i = a$: skill level in occupation $i \in \{0, 1\}$. P_i and Π_i: Price of output and profit in occupation $i \in \{0, 1\}$.

model to multiple occupations and learning about one's type which induces net mobility even when productivities are not changing. When prices P_1 and P_2 are changing, our model still resembles the Roy model when there are fixed costs of entry of entrepreneurs into occupations because each worker can simply "rent" at job at the entry cost.

When the number of entrepreneurs is fixed due to a limited stock of knowledgeable entrepreneurs (or if the production costs of the marginal entrepreneur is increasing in the mass of entrants) then our model differs with respect to the standard Roy model when prices change. For example if price P_1 goes up, in the Roy model the black indifference curve becomes flatter and therefore more workers choose jobs in occupation 1 - i.e. the low productivity workers from sector one change to sector two. In contrast, in our model in Figure 11 both the slope and the intercept of the indifference curve change: Jobs in occupation 1 become more attractive,

but since their supply is limited their price will rise and the intercept between the dotted and solid curve remains. Workers behavior will change substantially once P_1 becomes so large that the ranking between occupations change, i.e. when occupation 1 becomes more productive than occupation 1. Then good workers sort themselves into the now better occupation 2 while worse workers select themselves into occupation 1. Occupations that move up in the productive hierarchy increase their high-skilled workforce but reduce their low-skilled workforce, while in the Roy model all workers stay in an occupation that becomes more productive.

The difference in predictions is driven by differing assumptions about the scarcity of production factors. In the Roy model, there is abundance of production opportunities in each occupation. This turns the economy into an individual worker's decision problem that is independent of the other workers. If an occupation becomes more productive while the others stay unchanged, than each worker will view the more productive occupation as more attractive than before. No worker will quit this occupation, and some will enter because it now dominates their previous occupation. In contrast, in our model scarcity production factors implies that the opportunity cost of employing some type of worker is endogenous and depends on which other worker types are available in the economy. When productivity of an occupation increases then it is not only the productivity of its existing workforce that increases, but also the productivity of alternative workers that are not currently employed there. Exactly at the point when one occupation exceeds another in terms of productivity, the opportunity cost of forgoing alternative workers exceeds the increase in productivity of the existing workforce because of the complementarities between workers and firms. While human capital and match-specific factors will prevent an extreme exchange of the workforce between the occupations in reality, the efficiency effect of better sorting is still likely to lead to shedding of bad workers and expansion of good workers in particularly fast-growing occupations.

5.3 Island Economies

The scarcity of production factors in our approach is similar to the setup in Lucas and Prescott (1974). In their language each occupation is called an island. The prices on each island are determined competitively given the scarcity of the production factors. This leads to an efficient allocation of resources in our model as well as in theirs. In contrast to their model we have heterogeneous workers and a supermodular production function, which leads to sorting of specific workers to specific islands. The learning in our environment leads to the specific correlations of wages and switching behavior that seems consistent with the data that we document.

5.4 Career Progressions

Jovanovic and Nyarko (1997), Sichernam and Galor (1990) suggest that some occupatons form rungs of a career ladder. Workers spend time on the lower rungs accumulating skills that allow them to perform effectively at higher rungs. Our setup and these theories share the idea that occupations maybe vertically ranked. However, while thier models describe only the upward mobility or theory generates mobility in both directions.

6 Conclusion

Using administrative panel data on 100% of Danish population we document a new set of facts characterizing the patterns of occupational mobility. We find that a worker's probability of switching occupation is U-shaped in her position in the wage distribution in her 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).

These facts are not implied by existing theories of occupational mobility that mostly treat occupations as horizontally differentiated sets of tasks. We suggest that it might be productive to think of occupations as forming vertical hierarchies. Workers who are unsure of their abilities learn about them by observing their output realizations. Employment opportunities in each occupation are scarce inducing competition among workers for them. Complementarities in the production function between worker's ability and productivity of an occupation induce sorting of workers into occupations according to their expected ability. We present an equilibrium model of occupational choice with these features and show analytically that it is consistent with patterns of mobility described above.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

APPENDICES

A1 Omitted Proofs and Derivations

A1.1 Proof of Proposition 6

Consider an agent at the beginning of his t'th year in the labor market who has prior A about his mean ability and who chose occupation k this period. After observing wage w he can infer by (3) his output $X(w) = (w + \Pi_k)/P_k$ in the current period. Given that the worker chose occupation k, his prior is in $[B_k, B_{k+1})$. His posterior is according to (5) $A' = \alpha A + (1-\alpha)X(w)$, where the weight $\alpha = \phi_t/\phi_{t+1}$ depends on his labor market experience t. He will switch only if his posterior either exceeds B_{k+1} or is below B_k. Therefore, for any $X(w) \in [B_k, B_{k+1})$ or respectively for wages $w \in [P_k B_k - \Pi_k, P_k B_{k+1} - \Pi_k)$ the switching probability for workers is zero, and therefore the minimum of $S_k(t, w)$ is in the interior of the domain. For wages above $P_k B_{k+1} - \Pi_k$ workers will switch upward if $\alpha A + (1-\alpha)X(w) > B_{k+1}$. Even the worker with the lowest belief $A = B_k$ will switch if $\alpha B_k + (1-\alpha)X(w) > B_{k+1}$ or equivalently if $w > P_k(B_{k+1} - \alpha B_k)/(1-\alpha) - \Pi_k$. Therefore, toward the upper end of the domain the switching probability becomes one and therefore we have a local (and global) maximum. Similarly, for all low wages below $w < P_k(B_k - \alpha B_{k+1})/(1-\alpha) - \Pi_k$ the switching probability is also one, only that in this case workers switch to lower occupations.

A1.2 Equilibrium definition with human capital and switching costs

The output-contingent wages of workers are still given by (3), where output is now determined by (12). The expected wage for a worker in occupation k with prior mean A and experience ι in this occupation is therefore in analogy to (4)

$$\bar{w}_k(A, \iota) = P_k[A + h_k(\iota)] - \Pi_k.$$

For any given profit vector $\Pi = (\Pi_0, ..., \Pi_K)$ workers can forecast their expected wages in all occupations for given prior and given experience. A worker can then evaluate his optimal choice of occupation by simple backward induction. A worker's state vector at the beginning of each period is (t, k, ι, A) : his year in the labor market t, the occupation k he was last employed in, the consecutive years of experience in this occupation ι and his belief about his mean ability A. Newborns start with home production as their previous occupation. In the last year of his life the worker optimizes

$$V(T, k, \iota, A) = \max\left\{\bar{w}_k(A, \iota), \max_{m \neq k}\{\bar{w}_m(A, 0) - \kappa_m\}\right\},$$

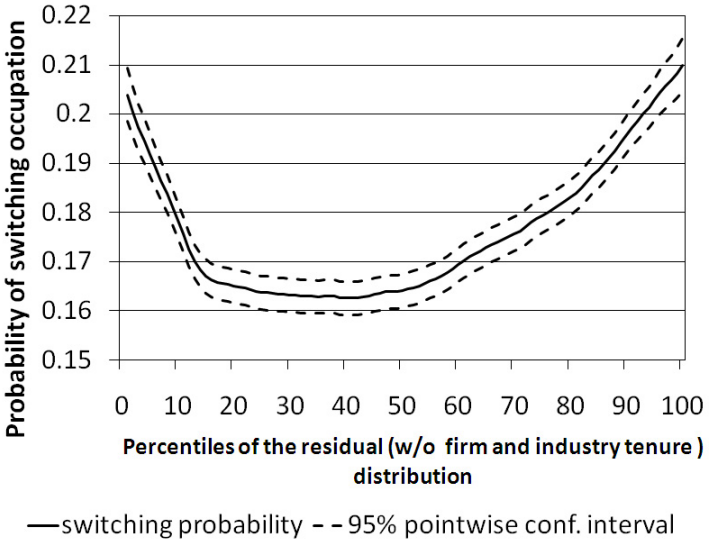
i.e. he chooses whether to stay in his previous occupation or to switch to a new occupation with zero experience and pay the switching costs. This gives a decision rule $d(T, k, \iota, A|\Pi) \in \{0, ..., K\}$ regarding the occupation that the worker chooses given the profits that firms make. Similarly,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

$$V(t, k, \iota, A) = \max \left\{ \begin{array}{c} \bar{w}_k(A, \iota) + \beta E_{A'} V(t+1, k, \iota+1, A'), \\ \max_{m \neq k} \{\bar{w}_m(A, 0) - \kappa_m + \beta E_{A'} V(t+1, m, 1, A')\} \end{array} \right\},$$

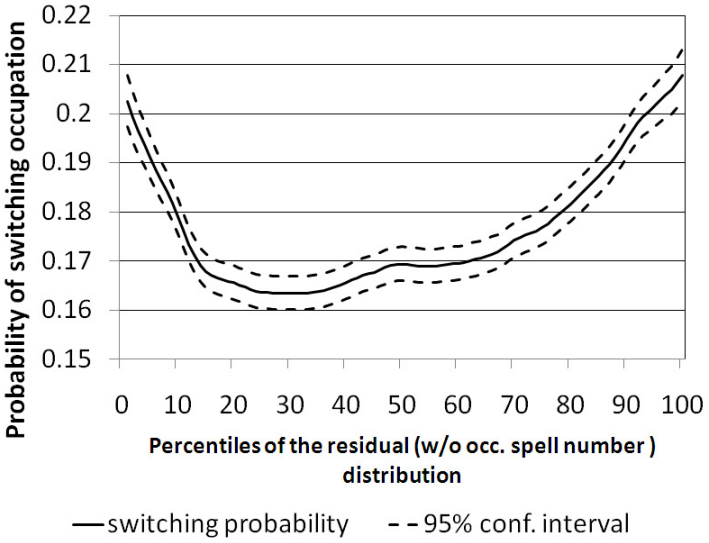
where $\beta \in (0, 1]$ is the discount factor and A' is the update about the worker's mean ability. The solution to this problem gives again a decision rule $d(t, k, \iota, A|\Pi) \in \{0, ..., K\}$. It is straightforward to show that for given profit vector Π these decision rules are unique for almost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

Definition 12 *An equilibrium is a vector of profits* $(\Pi_0, ...\Pi_K)$ *such that* Π_0 *and* $v_k(\Pi) = \gamma_k$ *for all* $k > 0$.

A2 Data Appendix



(a) residual distribution from wage regression not including firm and industry tenure



(b) residual distribution from wage regression not including occupational spell number

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

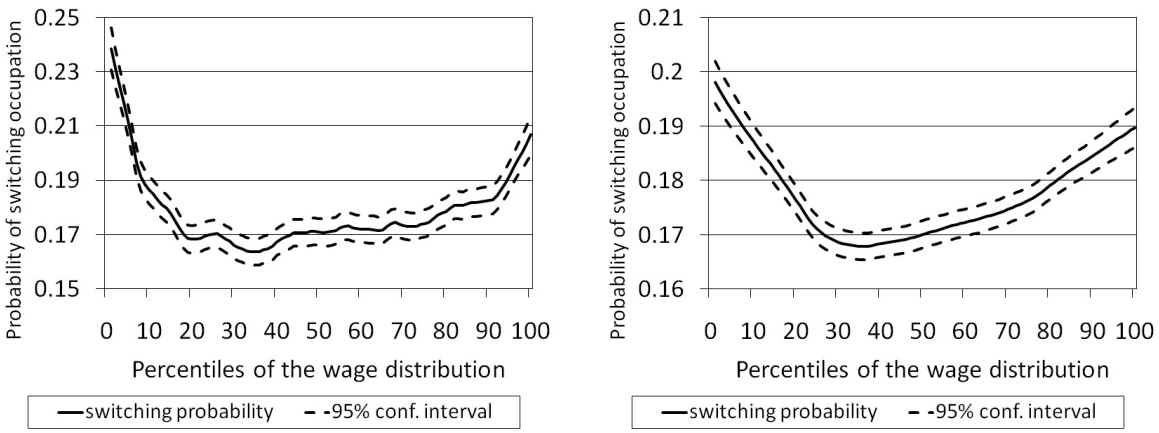


Figure A-2: Non-parametric plot of probability of switching occupation by worker's percentile in the wage distribution within occupation and year for half and double bandwidth.

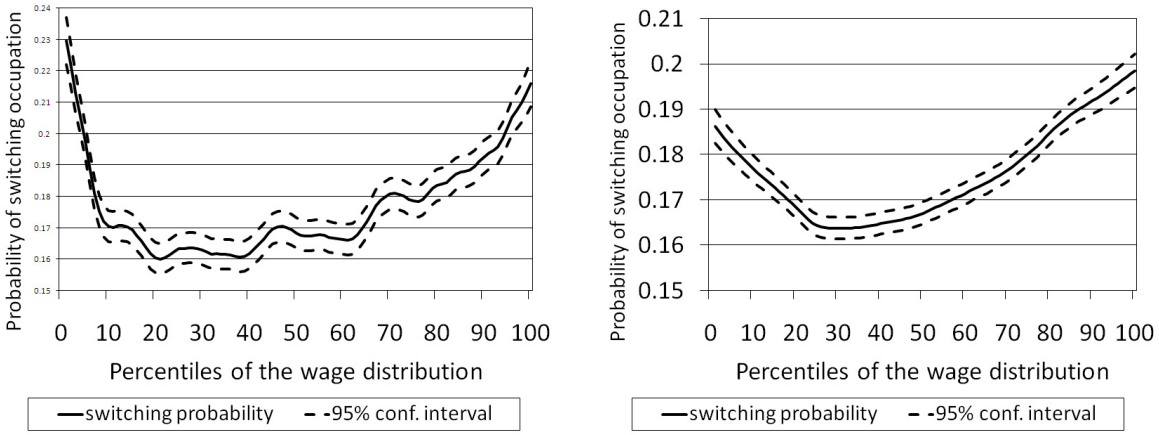


Figure A-3: Non-parametric plot of probability of switching occupation by worker's percentile in the wage residuals for half and double bandwidth.

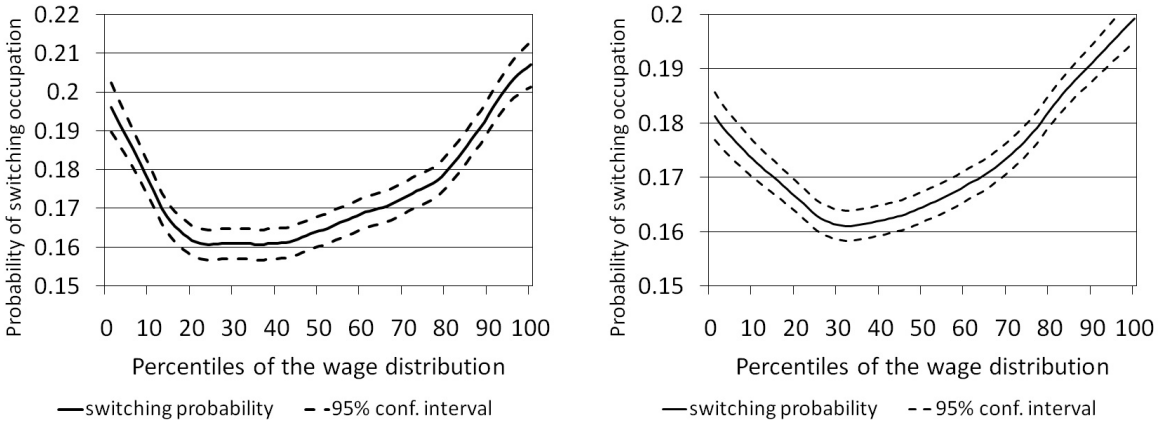


Figure A-4: Non-parametric plot of probability of switching occupation by worker's percentile in the wage within occupation, year, and years after graduation for half and double bandwidth.

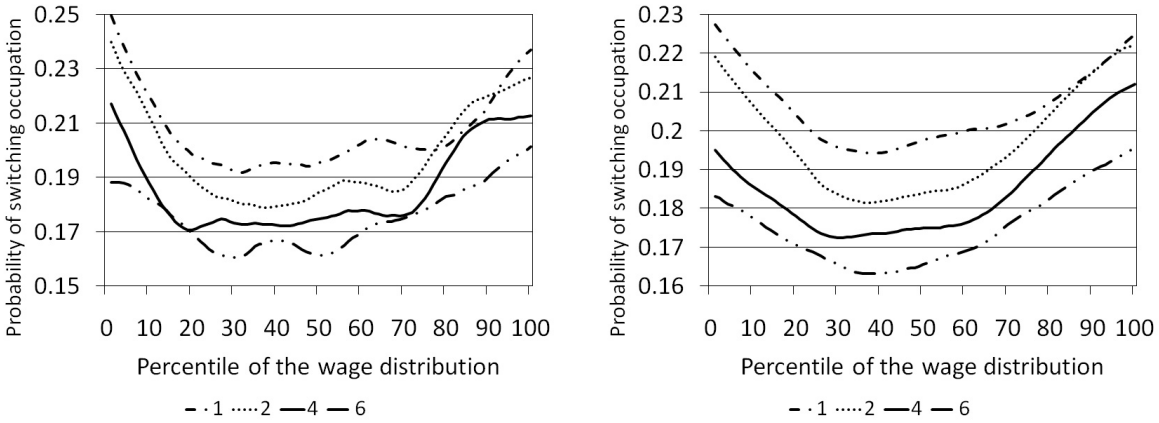
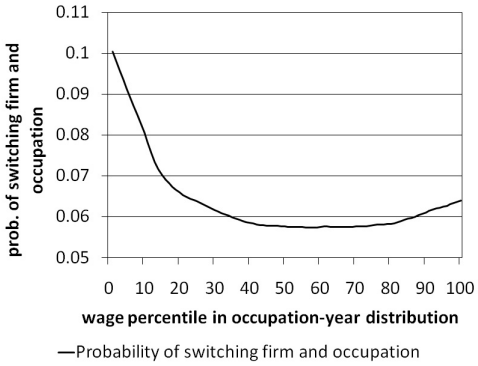
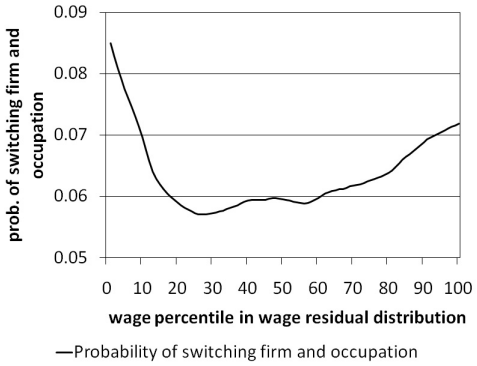


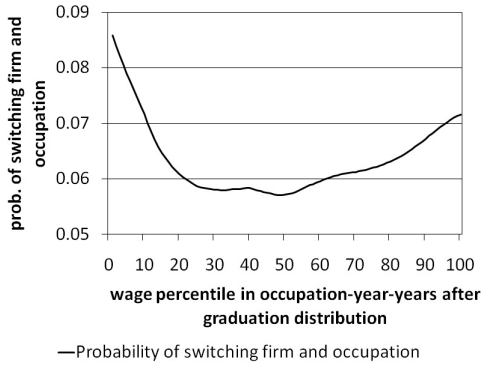
Figure A-5: Non-parametric plot of probability of switching occupation by worker's percentile in the wage within occupation, year, and 1, 2, 4, and 6 years after graduation for half and double bandwidth.



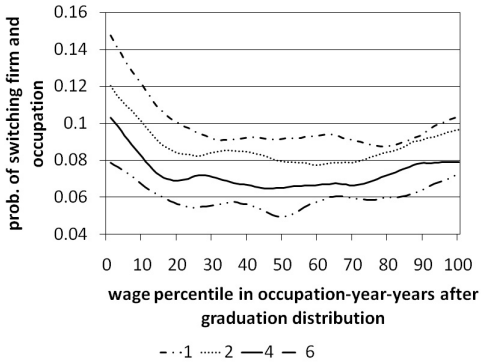
(a) wage distribution of raw wages within occupation and year



(b) wage distribution of wages residual

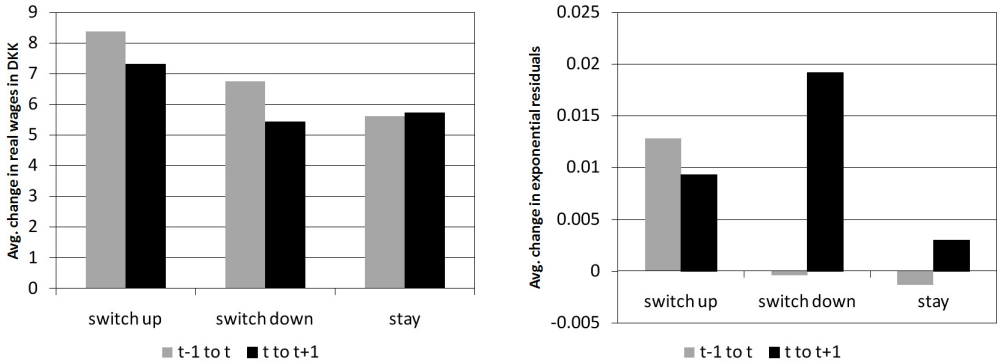


(c) wage distribution of raw wages within occupation, year, and year after graduation



(d) wage distribution of raw wages within occupation, year, and year after graduation for 1, 2, 4, and 6 years after graduation

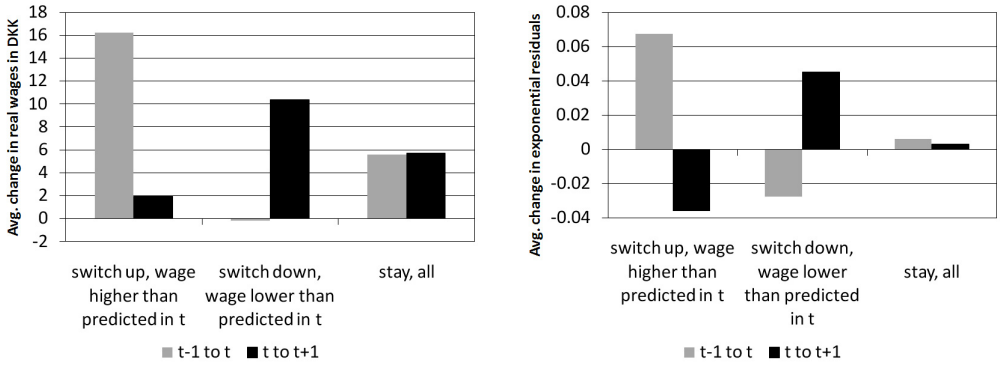
Figure A-6: Non-parametric plot of probability of switching occupation AND firm by worker's percentile in the wage distribution.



(a) Change in real wages in Danish Kroner for workers who switch occupation up, down, or stay

(b) Change in exponential of wage residuals for workers who switch occupation up, down, or stay

Figure A-7: Change in wages and wage residuals for workers who switch occupation up or down between years t and $t+1$ and stay in their occupation between $t-1$ and t. Stayers are in the same occupation in years $t-1$, t, and $t+1$



(a) Change in real wages in Danish Kroner for workers who switch occupation up, down, or stay, using differences between actual and predicted wages in period t

(b) Change in exponential of wage residuals for workers who switch occupation up, down, or stay, using differences between actual and predicted wages in period t

Figure A-8: Change in wages and wage residuals for workers who switch occupation up or down between years t and $t+1$ and stay in their occupation between $t-1$ and t. Stayers are in the same occupation in years $t-1$, t, and $t+1$. Using differences between actual and predicted wages in period t