Multidimensional Skill Mismatch

Working Paper 2018-02

FATIH GUVENEN, BURHAN KURUSCU, SATOSHI TANAKA, DAVID WICZER

March 28, 2018

Stony Brook University

Multidimensional Skill Mismatch*

Fatih Guvenen[†] Burhan Kuruscu[‡] Satoshi Tanaka[§] David Wiczer[¶]

March 27, 2018

Abstract

What determines the earnings of a worker relative to his peers in the same occupation? What makes a worker fail in one occupation but succeed in another? More broadly, what are the factors that determine the productivity of a worker-occupation match? To help answer questions like these, we propose an empirical measure of multidimensional skill mismatch, which is based on the discrepancy between the portfolio of skills required by an occupation and the portfolio of abilities possessed by a worker for learning those skills. This measure arises naturally in a dynamic model of occupational choice and human capital accumulation with multidimensional skills and Bayesian learning about one's ability to learn skills. Not only does mismatch depress wage growth in the current occupation, it also leaves a scarring effect—by stunting skill acquisition—that reduces wages in future occupations. Mismatch also predicts different aspects of occupational switching behavior. We construct the empirical analog of our skill mismatch measure from readily available US panel data on individuals and occupations and find empirical support for these implications. The magnitudes of these effects are large: moving from the worst- to the best-matched decile can improve wages by 11% per year for the rest of one's career.

JEL Codes: E24, J24, J31.
Keywords: Multidimensional skills, skill mismatch; match quality; Mincer regression; ASVAB; O*NET; occupational switching

*For comments and discussions, we thank Joe Altonji, Alessandra Fogli, Tim Kautz, Philipp Kircher, Jeremy Lise, Fabien Postel-Vinay, Rob Shimer, Kjetil Storesletten, José-Víctor Ríos-Rull, Carl Sanders, and Uta Schönberg as well as the participants at the 2011 and 2015 SED conferences, 2014 NBER Summer Institute, 2014 Barcelona GSE Summer Forum, 2015 CIREQ Conference on Information Frictions and at the Universities of Copenhagen, Hitotsubashi, Tokyo, California at Santa Barbara and Western Ontario. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Banks of Minneapolis or St. Louis. Guvenen acknowledges financial support from the National Science Foundation. Kuruscu acknowledges financial support from Social Sciences Humanities Research Council of Canada.

[†]University of Minnesota, FRB of Minneapolis, and NBER; `guvenen@umn.edu`

[‡]University of Toronto; `burhan.kuruscu@utoronto.ca`

[§]University of Queensland; `s.tanaka@uq.edu.au`

[¶]Stony Brook University; `david.wiczer@stonybrook.edu`

1 Introduction

What determines the earnings of a worker relative to his peers in the same occupation? What makes a worker fail in one occupation but succeed in another? More broadly, what are the factors that determine the productivity of a worker-occupation match? Each of these questions highlights a different aspect of the career search process that all workers go through in the labor market.

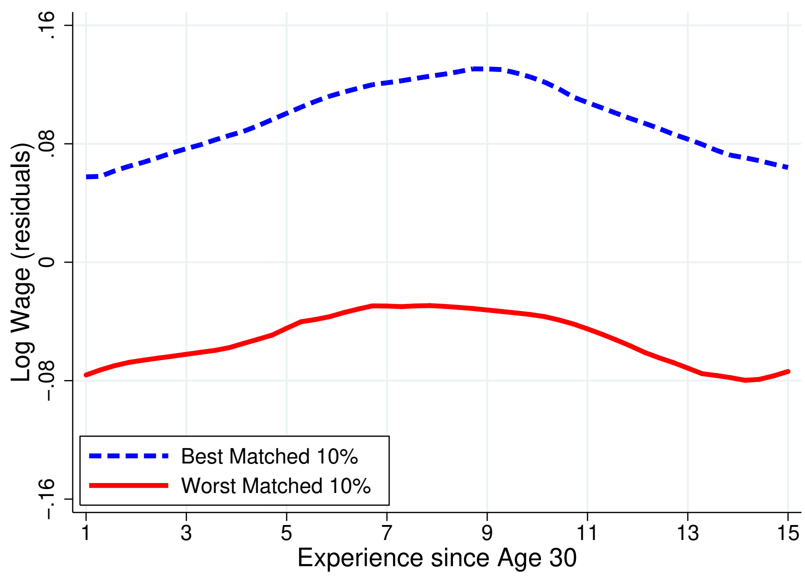
To explain the differences in outcomes of worker-occupation (job) matches, economists often appeal to the idea of "match quality," that is, some unobservable match-specific factor that determines the productivity of a match after controlling for the observable characteristics of the worker and the job. A long list of papers, going as far back as Jovanovic (1979) and Mortensen and Pissarides (1994), have shown that allowing for such an idiosyncratic match quality can help explain a wide range of labor market phenomena, such as how wages and job separations vary by job tenure, among others (see Rogerson et al. (2005) for a survey of this literature). While theoretically convenient, mapping this abstract notion of match quality onto empirical constructs that can be easily measured has proved elusive. Consequently, in empirical work, match quality is often treated as a residual, whose value is pinned down by making the model fit data on various labor market outcomes.[1]

In this paper, we propose an empirical measure of match quality that can be constructed by combining micro data on workers and on their occupations. It turns out to be convenient to measure the *lack* of match quality, or what we call *skill mismatch*. We model an occupation as a set of tasks to be completed, and our measure of skill mismatch is based on the discrepancy between the portfolio of skills required by an occupation (for performing the tasks that produce output) and the portfolio of abilities possessed by a worker for learning those skills. If the vector of required skills does not align well with the vector of a worker's abilities, the worker is mismatched, being either overqualified (positive mismatch) or underqualified (negative mismatch) along different dimensions of this vector.

Our notion of skill mismatch is multidimensional, which is motivated by a great deal of psychometric and educational research emphasizing multiple intelligences that act and develop independently of each other. In fact, developmental psychologist Howard

[1]Examples of this approach include Miller (1984), Flinn (1986), Jovanovic and Moffitt (1990), Moscarini (2001), and Nagypal (2007).

Figure 1 – Wage Gap Between the Best- and Worst-Matched Workers Persists For Many Years.



Note: Workers are grouped by their rank in the average of our mismatch measure over all the occupations they held before age 30. Residual wages are obtained by regressing log real wages on demographics, polynomials for occupation tenure, employer tenure, worker experience, a worker's ability measure, an occupational skill requirement measure, and their interactions with occupation tenure, and dummy variables for one-digit-level occupations and industries. Tenure variables are instrumented following Altonji and Shakotko (1987). See Section 4 for details of those variables. To obtain two lines, we run local polynomial regressions with residual wages on labor market experience for each group of workers, with a rule-of-thumb bandwidth.

Gardner, who first proposed the theory of multiple intelligences in his 1983 book found particular motivation for this idea in the proliferation of occupations (see, e.g., Gardner (2011, p. xxii)). Of course, economists are no strangers to the idea of multidimensional skills. After all, a long list of papers have built on the Roy model—which features multiple skills and comparative advantage—to study wages and occupational choice. Our paper follows this tradition by proposing a measure of mismatch in a world with multiple skills.

Before delving into the details of the paper, we highlight one of our main findings: workers who are poorly matched with their occupations earn lower wages even many years after they have left the occupation. In other words, not only does skill mismatch depress wages at the current occupation, it also leaves a lasting negative impact on the future career path of the worker. To give an idea about the potential magnitudes of this effect, Figure 1 plots the average (residual) wage paths for two groups of workers that differ in their skill mismatch averaged over all occupations they held before age 30 (controlling for a long list of worker and occupation characteristics). After age 30, the best-matched 10% (blue dashed line) earn consistently higher wages than the worst-matched 10% (red solid line), for a cumulative gap of $128,000 over 15 years (in 2002 dollars). The more rigorous calculations we conduct in the empirical analysis later below confirm these large losses from skill mismatch.

The empirical measure of skill mismatch we propose naturally emerges from a structural model of occupational choice, multidimensional skill accumulation, and Bayesian learning about abilities to acquire skills. In this model, output is produced at economic units called "occupations," which combine a vector of distinct skills supplied by their workers. The technology operated by an occupation is given by a vector of *skill requirements*, which determines both how much output a worker with a given ability (vector) produces and how much human capital he accumulates. This technology is specified such that for every worker there is a unique optimal amount of investment in each skill type depending on his abilities, and thus an optimal/ideal occupation choice. Skill mismatch is defined as a situation where the worker is not working his ideal occupation.

Our first contribution is to derive a Mincer-style wage equation that shows that the current wage is negatively related to mismatch at a worker's both current and past occupations. Because the derivation of this equation does not require any assumptions about what creates mismatch, the empirical wage regressions we estimate later in the paper provide support for the long-lasting effects of mismatch on wages—even after a worker leaves the mismatched occupation—without having to take a stance on the precise friction that gives rise to mismatch.

To investigate the relationship between skill mismatch and occupational switching, we do need to take a stand on what creates mismatch. We explore one plausible source, the idea that workers have imperfect information about their true abilities (to acquire different types of skills) at the time they join the labor market.[2] Skill mismatch will happen as young workers misjudge their learning abilities and choose occupations with skill requirements that do not align with their skill portfolio. Each period workers update their beliefs about their true abilities in a Bayesian fashion, which causes them to optimally switch occupations. As beliefs converge closer to the truth, leading to a decline in occupational switching rates and skill mismatch, consistent with the data.

[2] Different versions of this friction has been studied in a number of previous papers, see, for example, Farber and Gibbons (1996); Gibbons et al. (2005); Antonovics and Golan (2012); Papageorgiou (2014). One difference however is that these studies assumed learning was about comparative advantage whereas here it is about the ability to accumulate skills.

In the empirical analysis, we study four key predictions of this model. First, as noted earlier, we find that mismatch reduces both the level and the growth rate of wages with tenure. Second, current wages also depend negatively on *cumulative* mismatch in previous occupations, consistent with the model's prediction. Third, the probability of switching occupations increases with mismatch in the data, as predicted by the model: higher mismatch indicates current beliefs are farther away from truth, so a given wage realization causes a larger Bayesian updating of beliefs. Fourth, occupational switches are directional: workers who are overqualified in a given skill dimension tend to switch to occupations that are more skill intensive in that dimension, and vice versa for a skill they are underqualified in. Furthermore, conditional on switching, the reduction in mismatch is proportional to the level of mismatch in previous occupation.

To study these implications of skill mismatch, we use the 1979 National Longitudinal Survey of Youth (NLSY79) for information on workers' occupations and wage histories. NLSY79 respondents were also given an occupational placement test—the Armed Services Vocational Aptitude Battery (ASVAB)—that provides detailed measures of occupation-relevant skills and abilities.[3] In addition, respondents also reported various measures of noncognitive skills, which we use to describe a worker's ability for socially interactive work. For comparability with existing work, we focus on male workers. Turning to the skill requirements of each occupation, we use data from the U.S. Department of Labor's O*NET project.[4] Combining these two sources of information allows us to compute our mismatch measure for each worker in our NLSY sample. In the most detailed case, we measure mismatch along three skill dimensions: (cognitive) math skills, (cognitive) verbal skills, and (noncognitive) social skills.

We incorporate these (contemporaneous and cumulative) mismatch measures into a Mincerian wage regression along with flexible interactions with occupational tenure (and a large set of other controls). Consistent with our theory, we find that the coefficient on mismatch is robustly negative and that its interaction with occupational tenure is

[3]We interpret workers' test scores as corresponding to (noisy measures of) abilities in our model. Although these scores probably reflect a combination of abilities and accumulated skills, this distinction is probably not critical for our purposes, because the two are highly correlated at young ages. Huggett et al. (2011) estimate that this correlation exceeds 0.85 at age 20. Since these tests are taken at the beginning of workers' careers, we interpret them as abilities.

[4]The reader might wonder why workers do not choose their ideal occupation if they know their ASVAB scores for each ability type. This is because, first, the NLSY respondents were not told their exact test score—they were given a fairly wide range; and second, these test scores are themselves noisy measures of individuals' true underlying abilities as discussed further later.

robustly negative. The estimates imply that, after 10 years of occupational tenure, the average wage of workers in the top decile of the mismatch distribution is 6.8% lower than those in the bottom decile (i.e., best-matched workers). Perhaps even more important, cumulative past mismatch has a similarly large effect: wages are 9.4% lower for workers in the top decile of cumulative mismatch distribution compared with those in the bottom decile. These long-lasting effects of mismatch are consistent with the human capital accumulation channel captured by our model; it would be missed by theories that postulate that match quality only affects the current match. Overall, eliminating all mismatch would raise the average wages of workers by 11% every year.

Turning to occupational switching behavior, the data reveal patterns consistent with our model. First, estimating a hazard model for occupational switching shows that the probability of a switch is increasing in mismatch. The magnitudes are also fairly large: the switching probability is about 3.4 percentage points higher for a worker at the 90th percentile of the mismatch distribution relative to another worker at the 10th percentile. This gap is about one-fifth of the average switching probability in our sample. Second, we follow workers across occupational transitions to see if they tend to "correct" previous mismatches. Indeed, they do: if a worker is overqualified in his current occupation along a certain skill dimension, the next occupation, on average, has higher skill requirements *in that dimension* (as well as in other skill dimensions, but to a lesser extent). Furthermore, consistent with our theory, we find that the step size of these switches are proportional to the level of current mismatch.

A multidimensional measure of mismatch has important empirical implications. For example, if a worker who is very talented in one type of skill currently works in an occupation requiring another skill intensively, he would be considered mismatched even though both the worker and the occupation would be described as high skill on average. It also allows us to see the potentially different effects of being over- or underqualified along different dimensions, and we show that there are important qualitative differences. For example, mathematical mismatch contributes more to the level of wages, whereas verbal mismatch affects the growth of wages with occupational tenure.

Finally, we extend our wage regressions to distinguish mismatch for overqualified and underqualified workers. We find that both those who were overqualified and those who were underqualified in previous occupations have lower wages today. This implication is consistent with our model, but is inconsistent with a standard Ben-Porath model with multidimensional skills, as we discuss in Section 2.2.

The paper proceeds as follows. In Section 2 we present our model. In Section 3 we describe our data, and Section 4 describes our methodology and how we create our mismatch measures. Section 5 presents the results, and finally, we conclude in Section 6.

1.1 Related Literature

Our paper contributes to an active and growing literature that studies the skill content of occupations. One strand of this literature uses data on occupation characteristics (e.g., from the O*NET) and explores how these occupational skill requirements are related to the wages and career trajectories of workers employed in those occupations. Notable examples of this approach include Ingram and Neumann (2006), Poletaev and Robinson (2008), Gathmann and Schönberg (2010), Bacolod and Blum (2010), and Yamaguchi (2012). A second strand of this literature focuses exclusively on worker-side information to study occupations and skill mismatch. Perry et al. (2014) reviews various attempts to quantify skill mismatch using this approach. A recent notable example of this approach is Fredriksson et al. (2015), which uses data on Swedish workers and defines mismatch as the gap between the skills of a new hire and his experienced peers in the same occupation and establishment, where the skills of the latter are taken as a measure of the skill requirements of the occupation.

The main difference of our approach from this earlier work is that we construct a mismatch measure that combines information from both sides—worker skills and occupational skill requirements. We then show that this measure is a significant predictor of current and future wages, even after controlling for worker skills and occupational skill requirements separately.

Our paper also has useful points of contact with the literature that studies career outcomes in the presence of comparative advantage or employer/worker learning or both. Because different sectors can reward skills differently, as workers learn about their skills they switch toward sectors that maximize their comparative advantage (Gibbons et al. (2005), Antonovics and Golan (2012), Gervais et al. (2016), and Papageorgiou (2014)). We show that our empirical mismatch measure contains critical information about this switching propensity and direction, along with implications for workers' current and future wages.[5]

[5]In a slightly different context, Farber and Gibbons (1996) and Altonji and Pierret (2001) investigate the extent of employer learning about worker abilities. They show that an interaction term between ability and job tenure included in a Mincer wage regression is positive—so wages reflect abilities more

A precursor to our paper is Groes et al. (2015) who study how mismatch affects occupational switching by using a one-dimensional measure of mismatch based on the deviation of a worker's own wage from the average of his peers in the same occupation. Our results on occupation switching is consistent with theirs. However, they did not examine the effect of mismatch on wages, which is a major focus of our paper.

Lise and Postel-Vinay (2016) and Sanders (2016) are two contemporaneous papers that share similarities with ours: they both estimate models with multidimensional skills, sorting, and human capital accumulation, using data similar to ours (NLSY79 and O*NET).[6] Sanders (2016)'s main focus is on the effects of uncertainty about one's skills, so he classifies job tasks based on how much workers may know about their skills to perform that task. Instead, we focus on different classes of skills, such as math, verbal, and social, and construct skill mismatch along these dimensions. Sanders does not directly examine the effects of mismatch on labor market outcomes. Lise and Postel-Vinay (2016) study a model where mismatch results from search frictions rather than from learning. Our empirical findings in Section 5 lend support to learning but does not provide evidence against the existence of search frictions as an additional source of mismatch. The second difference is about the focus: their approach infers mismatch through the lens of a structural model by matching data moments, whereas our approach is to develop a model-based measure of skill mismatch that can be constructed directly from micro data, and we then turn to examine if this measure is informative about wages and occupational switching as predicted by the model. These are two complementary approaches, each with its distinct advantages.

2 Model

In this section, we present a life-cycle model of occupational choice and human capital accumulation, which provides a natural framework to analyze how skill mismatch arises and, more importantly, how it affects current and future wages of a worker in different occupations.[7] The model delivers some testable implications, which we will empirically study in Section 5.

closely over time—which they interpret as evidence of employers learning about workers' abilities during their relationship.

[6] Speer (2017) is also methodologically relevant to our work by combining worker- and occupation-side skill measures to study how gender discrimination and layoffs shape occupational choice.

[7] Throughout the paper we use the terms "human capital" and "skill" interchangeably.

Briefly, the structure of the labor market builds upon Rosen (1972), wherein workers supply labor services to firms (where we think of each firm as a distinct occupation), and firms in turn offer different training/learning opportunities. Consequently, a worker's compensation from a job/occupation is not only a wage but also a rise in his skill level. Our model introduces two key features into this framework. First, human capital is multidimensional, and the rate at which a worker acquires each new skill depends both on his learning ability of that skill and the occupation's requirements for investment in that skill. Second, workers enter the labor market without full knowledge about their learning abilities; instead, they have a prior belief about their abilities, which they update over time in a Bayesian fashion. This information friction is the source of skill mismatch between workers and their occupations. We now turn to the details of the model.

2.1 Environment

Each worker lives for T periods and supplies one unit of labor inelastically in the labor market. The objective of a worker is to maximize the expected present value of earnings/wages:

$$\mathbb{E}_0\left[\sum_{t=1}^{T}\beta^{t-1}w_t\right],$$

where β is the subjective time discount factor.

Occupations and Technology. There is a continuum of occupations, each using n types of skills, indexed with $j \in \{1, 2, ..., n\}$. Occupations differ in their *skill intensity* (alternatively called *skill requirement*) of each skill type, denoted with the vector $\mathbf{r} = (r_1, r_2, ..., r_n) \geq 0$, which is fixed over time. With a slight abuse of notation, we will use $\mathbf{r}_t = (r_{1,t}, r_{2,t}, ..., r_{n,t})$ to denote the particular occupation chosen by the worker in period t. Just as skills are multidimensional, so are the abilities to accumulate them, which we stack in a vector, $\mathbf{A} \equiv (A_1, A_2, ..., A_n)$.

A worker starts period t with the skill portfolio $\mathbf{h}_t = (h_{1,t}, h_{2,t}, ..., h_{n,t})$, chooses an occupation $\mathbf{r}_t$, accumulates new skills at that occupation, and then produces output with these upgraded skills. The technology for skill accumulation is a key ingredient in our model and is given by

$$k_{j,t} \equiv h_{j,t} + (A_j + \varepsilon_{j,t})\, r_{j,t} - r_{j,t}^2/2, \tag{1}$$

where $k_{j,t}$ is the upgraded skill of type j, which is used in production in period t and determines next period's starting human capital level (as explained in a moment), and $\varepsilon_{j,t}$ is a random disturbance term (whose role will become clear later). One way to think about this structure is that, in a given period, workers first go through training and then produce output with their new/upgraded skills.

This specification has two important features. First, the occupation's skill requirement, $r_{j,t}$, affects human capital accumulation nonmonotonically—the linear positive term captures the benefit of an occupation whereas the negative quadratic captures the costs of skill acquisition, such as additional training required at high-skill jobs. As will become clear later, this nonmonotonicity ensures that optimal occupational choice is an interior one (i.e., workers do not all flock to the occupations with the highest $r_{j,t}$). Second, with the formulation in (1), the linear benefit term is proportional to the worker's ability $(A_j + \varepsilon_{j,t})$, whereas the cost term is independent of ability, which gives rise to comparative advantage and sorting by ability level—workers choose occupations with higher skill requirements only in dimensions where their ability is relatively high. These two features will come to play important roles in what follows.

The output of a worker in a given occupation is the sum of his end-of-period skills, and perfect competition among occupations for workers ensures that a worker's wage equals the output he produces:

$$w_t = \sum_j k_{j,t}(h_{j,t}, A_j; r_{j,t}, \varepsilon_{j,t}). \tag{2}$$

At first blush, this wage equation (2) may appear a bit peculiar as there is no explicit price multiplying each skill (as would be the case in a standard human capital model, e.g., Ben-Porath (1967) or Rosen (1972)), which makes it seem as if there is no variation in the value of skills across occupations or even a relative price of each skill at the aggregate level. This is not the case however: the $k_{j,t}$'s already embed how productive each (newly produced) skill is in each occupation (through equation (1)). So, two workers with the same human capital portfolio at the beginning of the period, $\mathbf{h}_t$, who work at different occupations during the period, will end up with different amounts of $k_{j,t}$'s and therefore receive different levels of compensation. Thus, a worker's wage depends on his human capital vector $\mathbf{h}_t$, learning ability, $\mathbf{A}$, occupation $\mathbf{r}_t$, and stochastic disturbance, $\varepsilon_{j,t}$. The occupation-specific skill intensity, $\mathbf{r}_t$, plays a dual role, determining the relative price of each skill in each occupation in addition to also affecting the rate of skill accumulation in each occupation.[8]

Finally, the beginning-of-period human capital in period $t+1$ is equal to $k_{j,t}$ adjusted for depreciation:

$$h_{j,t+1} = (1-\delta)\,k_{j,t} = (1-\delta)\left(h_{j,t} + (A_j + \varepsilon_{j,t})\,r_{j,t} - r_{j,t}^2/2\right), \tag{3}$$

where the depreciation rate δ is the same for all skill types and occupations.

2.2 Empirical Wage Equation and Skill Mismatch

We can now derive our first key equation, which shows how a worker's wage depends on his abilities and the history of occupations he has been employed at. To this end, first, combining equations (1)–(2) and rearranging yields:

$$w_t \;=\; \sum_{j=1}^{n}\left(h_{j,t} + \frac{A_j^2}{2} - \frac{(A_j - r_{j,t})^2}{2}\right) + \sum_{j=1}^{n} r_{j,t}\varepsilon_{j,t}. \tag{4}$$

This expression makes clear that a worker's wage depends positively on his (beginning-of-period) human capital and his ability, and negatively on $(A_j - r_{j,t})^2$, which simply the deviation between his ability level and his occupation's skill requirement—or what we will call, skill *mismatch*. Next, using equation (3) and repeatedly substituting for human capital (and setting $\delta \equiv 0$ for clarity) yields:[9]

$$h_{j,t} \;=\; h_{j,1} + \frac{A_j^2}{2}\,(t-1) - \sum_{s=1}^{t-1}\frac{(A_j - r_{j,s})^2}{2} + \sum_{s=1}^{t-1} r_{j,s}\epsilon_{j,s}, \tag{5}$$

which shows that human capital grows with experience at a rate that is proportional to ability (second term) and, more importantly for our purposes, is depressed by the degree of mismatches at all his past occupations. Finally, substituting this expression into (4) delivers the key empirical wage equation that we will use in the empirical analysis:

$$w_t \;=\; \sum_j h_{j,1} + \underbrace{\frac{1}{2}\sum_{j=1}^{n} A_j^2 \times t}_{\text{ability}\times\text{experience}} \;-\underbrace{\frac{1}{2}\sum_{j=1}^{n}\sum_{s=1}^{t}(A_j - r_{j,s})^2}_{\text{mismatch}} + \sum_{j=1}^{n}\sum_{s=1}^{t} r_{j,s}\varepsilon_{j,s}. \tag{6}$$

[8] In this sense, our model fully captures the spirit of comparative advantage as in the Roy model but does so in a way that both retains tractability, which will allow us to derive a number of key predictions and deliver implications of mismatch as will become clear in a moment. The generality afforded by this formulation is also in the same spirit as the "skill weights" approach in Lazear (2009).

[9] The analogous expression with $\delta > 0$ is given in Appendix A.

Two remarks are in order. First, the wage inherits the two properties of human capital noted above; in particular, its growth is proportional to a weighted average of the worker's ability portfolio, and it is depressed by the history of past mismatches.[10] This long-lasting negative effect of past mismatches on future wages is the one of the key implications of the model that we will test in the data. Second, notice that we did not yet specify how workers choose occupations or what causes mismatch. So, the empirical implications of this wage equation do not depend on those specifics. Instead, they come from the specification of technology that determines how wages and human capital accumulation depend on abilities and occupations.[11]

Turning to the wage equation, we rearrange it slightly to distinguish between the effect of current mismatch and past cumulative mismatch on wages. To this end, let t^c denote the period in which the worker switched to his current occupation (i.e., $r_{j,s} = r_{j,t^c}$ for $s \geq t^c$), so his tenure in the current occupation is $t - t^c + 1$. We can rewrite equation (6) as

$$\begin{aligned} w_t \;=\; & \sum_j h_{j,1} + \underbrace{\frac{1}{2}\sum_j A_j^2 \times t}_{\text{ability}\times\text{experience}} - \underbrace{\frac{1}{2}\sum_j \left(A_j - r_{j,t^c}\right)^2}_{\text{current mismatch}} \times \underbrace{(t - t^c + 1)}_{\text{current tenure}} \\ & - \underbrace{\frac{1}{2}\sum_{s=1}^{t^c-1}\sum_j \left(A_j - r_{j,s}\right)^2}_{\text{cumulative past mismatch}} + \sum_{s=1}^{t}\sum_j r_{j,s}\varepsilon_{j,s}, \end{aligned} \tag{7}$$

which shows the separate effects of current mismatch and past cumulative mismatches, which we will map into the data in Section 5.

2.3 Information Structure and Bayesian Learning

Each worker draws ability A_j from a normal distribution at the beginning of his life: $A_j \sim \mathcal{N}(\mu_{A_j}, \sigma^2_{A_j})$. The worker does not know his true A_j but observes a signal, $\hat{A}_{j,1} = A_j + \eta_j$, where $\eta_j \sim \mathcal{N}(0, \sigma^2_{\eta_j})$. So, his prior beliefs are normally distributed with mean $\hat{A}_{j,1}$ (unbiased) and precision $\lambda_{j,1} \equiv 1/\sigma^2_{\eta_j}$.

[10] With positive depreciation, mismatches that are farther in the past will be discounted in calculating cumulative mismatch. See equation (11) in Appendix A.

[11] It might seem that if the model is correctly specified and we were to estimate the wage equation (6) from data, we should get coefficients of 1/2 on the ability and mismatch terms. But this would only be true if A_j and $r_{j,s}$ were cardinal variables that had exact analogues in the data, which is not the case. Therefore, the model's prediction is more about the significance of these variables for determining wages rather than about particular values for the estimated coefficients.

Each period, the worker observes $A_j + \varepsilon_{j,t}$, where $\varepsilon_{j,t} \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon_j}\right)$. This is equivalent to saying the worker observes each of his skills each period and knows his initial skill vector, $\mathbf{h}_1$. Then, given his current beliefs, the worker updates his belief about A_j. The worker's belief at the beginning of each period is normally distributed. Let $\hat{A}_{j,t}$ be the mean and $\lambda_{j,t}$ be the precision of this distribution at the beginning of period t and λ_{ε_j} be the precision of $\varepsilon_{j,t}$. After observing $A_j + \varepsilon_{j,t}$, the worker updates his belief according to the following recursive Bayesian formula:

$$\hat{A}_{j,t+1} = \frac{\lambda_{j,t}}{\lambda_{j,t+1}} \hat{A}_{j,t} + \frac{\lambda_{\varepsilon_j}}{\lambda_{j,t+1}} \left(A_j + \varepsilon_{j,t}\right), \tag{8}$$

where $\lambda_{j,t+1} = \lambda_{j,t} + \lambda_{\varepsilon_j}$.[12]

2.4 Workers' Problem

Given the current beliefs about his abilities, the problem of the worker in period t is given as follows:

$$V_t(\mathbf{h}_t, \hat{\mathbf{A}}_t) = \max_{\{r_{j,t}\}} \mathbb{E}_t \Big[\sum_j k_{j,t} + \beta V_{t+1}(\mathbf{h}_{t+1}, \hat{\mathbf{A}}_{t+1}) \Big],$$

subject to (1), (3), and (8). Since occupations are represented by a vector of skill intensities, this problem yields a choice of occupation in the current period, which then determines not only current wages but also future human capital levels. The expectation in the worker's problem is taken with respect to the distribution of his beliefs about A_j (for $j = 1, ..., n$), given by $\mathcal{N}(\hat{A}_{j,t}, 1/\lambda_{j,t})$, and the distribution of $\varepsilon_{j,t}$, given by $\mathcal{N}(0, \sigma^2_{\varepsilon_j})$.

Proposition 1. *The optimal solution to the worker's problem is characterized by the following two functions:*

1. *Occupational choice:* $r_{j,t} = \hat{A}_{j,t}$*;*

[12] Notice that to the extent that abilities are correlated, each new observation about one ability also provides information about other abilities, which we implicitly assumed away here. Taking this additional information into account is possible but will turn the learning problem into a multidimensional Kalman filtering problem, which makes the solution of the model analytically less tractable, so we do not consider it here. It is a potential extension in future work. That said, the extent to which this cross-updating of beliefs empirically matters depends on the magnitude of the correlation between abilities (c.f., Guvenen (2007)), which is low for 2 out of the 3 ability pairs we study later and high between the other pair (see Table I. Finally, we also assume that a worker's occupation and its skill requirements does not influence the quality of signal a worker gets about his ability. These assumptions allow us to derive the analytical results.

2. *Value function:*

$$V_t(\mathbf{h}_t, \hat{\mathbf{A}}_t) = \Big(\sum_{s=t}^{T} \beta^{s-t}\Big)\Big(\sum_{j=1}^{n} \big(h_{j,t} + \hat{A}_{j,t}^2/2\big)\Big) + B_t(\hat{\mathbf{A}}_t),$$

where B_t is a known time-varying function that does not affect the worker's choices.[13]

Three remarks about this solution are in order. First, given that $r_{j,t} = \hat{A}_{j,t}$, skill *mismatch* in dimension j can be written either as $(A_j - \hat{A}_{j,t})^2$ or $(A_j - r_{j,t})^2$. In the empirical section, we use the worker's test scores that proxy A_j's and his occupation's skill intensities that correspond to $r_{j,t}$'s in order to construct our mismatch measure. Second, since A_j's enter into the worker's objective function linearly, the solution only depends on the worker's expectation of A_j, which is $\hat{A}_{j,t}$. Third, the worker's human capital and wage depend both on his belief $\hat{A}_{j,t}$ and also his true ability A_j and the shock $\varepsilon_{j,t}$. Thus, his realized wage and human capital will be different from his own expectations of these two variables.

Relationship to the Standard Human Capital Framework

Before moving further, it is instructive to point out three key ways our model differs from the standard Ben-Porath formulation. The first two should be obvious by now. First, we introduce multidimensional skills and abilities; and second, here, skill accumulation depends not only on a worker's learning abilities (A_j) but also with his occupation. This latter feature is in the spirit of Rosen (1972).

Third, and less obvious, the incentives that drive skill acquisition are quite different here compared to the Ben-Porath framework. To see this, notice that in the Ben-Porath model (assuming perfect information and one type of skill, and interpreting r_t as human capital investment), the current wage would be $w_t = h_t - r_t^2/2$, and the next period's human capital would be $h_{t+1} = h_t + Ar_t$. Thus, the choice of r_t that maximizes the current wage is zero, whereas the one that maximizes future human capital is infinite, so there is a strong intertemporal trade-off that determines the choice of r_t. In contrast, in our model, the analogues of the two equations are $w_t = h_t + (Ar_t - r_t^2/2)$ and $h_{t+1} = h_t + (Ar_t - r_t^2/2)$, where both equations have the common term: $Ar_t - r_t^2/2$. Because of this symmetry, the same interior choice of r_t maximizes both current wage and future human capital—the intertemporal trade-off disappears.[14] This makes our setup more similar to a learning-

[13] The proof of this proposition for the general case with positive depreciation is in Appendix A.

[14] We show this result more formally in Appendix A.2.

by-doing model where more work today raises both the current wage and future human capital.

Fourth, in our framework, both overqualified and underqualified workers (i.e., those with $r_j \neq A$) experience lower rates of skill accumulation, which is clear from the law of motion for h_{t+1}. In contrast, in the Ben-Porath version, if a worker is underqualified today (i.e., $A < r_j$), he would earn lower wages today but higher wages in the future because he over-accumulates skills (compared to his optimal occupation) in the higher r occupation (recall $h_{t+1} = h_t + Ar_t$). Consequently, in a wage regression, negative *past* mismatch (reflecting the extent to which a worker was underqualified in the past) will have a positive effect on the current wage. In contrast, in our setup both positive and negative past mismatches have a negative effect on current wages. Our empirical wage regressions will speak to these different predictions of our formulation compared to the standard human capital framework.

Two Issues With Implementation

Before concluding this section, we discuss two practical issues that arise when the wage equation (7) is estimated. The first one is that the error term in (7) is correlated with mismatch measures because the mean of a worker's posterior beliefs about his abilities is correlated with past shocks.[15] As a result, our estimates of the coefficient on mismatch will be biased. Fortunately, as we state formally in Lemma 1, we can determine the sign of the bias, which will become useful in interpreting our results.

Specifically, first, current mismatch and past shocks are positively correlated. If we observe a high (low) wage in a period due to positive (negative) shocks, we will also observe a high (low) mismatch. Because we expect wages and mismatch to be negatively correlated and wages and errors to be positively correlated, the direction of the bias will be towards zero and the true effect of mismatch on wages should be stronger than the effect we estimate in our empirical analysis.[16] The following lemma summarizes this result.

Lemma 1. *Let* $M_{j,t} \equiv \sum_{s=1}^{t} (A_j - r_{j,s})^2$ *and* $\Omega_{j,t} \equiv \sum_{s=1}^{t} r_{j,s}\varepsilon_{j,s}$. *Then,* $Cov(M_{j,t}, \Omega_{j,t}) > 0$. *Therefore, the estimated coefficient of mismatch provides a lower bound (in magnitude) for the true effect.*

[15] This is a direct implication of Bayesian learning and can be seen by repeatedly substituting equation (8) backward.

[16] "True" mismatch effect is the one when workers are randomly assigned to occupations and therefore mismatch happens exogenously.

Another issue that concerns the empirical estimation of the wage equation is that we do not directly observe A_j's. Instead, we will use workers' ASVAB test scores, which are noisy signals about their true abilities. To illustrate how this might affect our estimates, let $\widetilde{A}_j \equiv A_j + \nu_j$, where $\nu_j \sim \mathcal{N}(0, \sigma^2_{\nu_j})$, denote the test scores. To see how using $\widetilde{A}_j$ instead of A_j in the estimation affects our results, insert $A_j = \widetilde{A}_j - \nu_j$ into (6), which gives

$$w_t = \sum_j h_{j,1} + \frac{1}{2}\sum_j \widetilde{A}_j^2 \times t - \frac{1}{2}\sum_{s=1}^{t}\sum_j \left(\widetilde{A}_j - r_{j,s}\right)^2 + \sum_{s=1}^{t}\sum_j r_{j,s}\left(\varepsilon_{j,s} - \nu_j\right).$$

In the following lemma, we show that estimating this equation delivers estimates of the coefficients on both the ability term and mismatch that are biased toward zero.

Lemma 2. [Measurement Error and Attenuation Bias] *Let* $\widetilde{\Delta}_{j,t} = \widetilde{A}_j^2 \times t$ *and* $\widetilde{M}_{j,t} \equiv \sum_{s=1}^{t}\left(\widetilde{A}_j - r_{j,s}\right)^2$, $\widetilde{\Omega}_{j,t} \equiv \sum_{s=1}^{t} r_{j,s}\left(\varepsilon_{j,s} - \nu_j\right)$. *Then,*

1. $Cov\left(\widetilde{\Delta}_{j,t}, \widetilde{\Omega}_{j,t}\right) < 0$*: Therefore, the estimated coefficient of ability-experience interaction provides a lower bound for the true effect.*

2. $Cov\left(\widetilde{M}_{j,t}, \widetilde{\Omega}_{j,t}\right) > 0$*: Therefore, the estimated coefficient of mismatch provides a lower bound for the true effect.*

Lemmas 1 and 2 establish that the coefficients we obtain in the empirical analysis provide *lower bounds* on the effects of mismatch on wages.

2.5 Occupational Switching

We now turn to workers' occupational switching decisions and how they relate to past and current mismatch. Note that workers' beliefs are unbiased at any point in time, so mean beliefs over the population are equal to mean abilities. However, each worker will typically over- or under-estimate his abilities in a given period. Over time, beliefs will become more precise and converge to his true abilities. Thus, workers choose occupations with which they are better matched and mismatch declines. The following lemma formalizes this simple result.

Lemma 3. [Mismatch by Labor Market Experience] *Average mismatch is given by* $E[(A_j - r_{j,t})^2] = 1/\lambda_{j,t}$*. Since the precision* $\lambda_{j,t}$ *increases with labor market experience, average mismatch declines with experience.*

The occupational switching decision is closely linked to mismatch. To illustrate this point, assume that an occupational switch occurs if a worker chooses an occupation whose skill intensities fall outside a certain neighborhood of the skill intensities of his previous occupation in at least one skill dimension. More formally, letting $\kappa_j > 0$ be a positive number, an occupational switch occurs in period t if $r_{j,t} > r_{j,t-1} + \kappa_j$ or $r_{j,t} < r_{j,t-1} - \kappa_j$ for some j. The following two propositions characterize the patterns of occupational switches.

Proposition 2. [Probability of Occupational Switching] *The probability of occupation switching increases with current mismatch and declines with age.*

Mismatch is higher when the mean of a worker's belief is further away from his true ability. In that case, conditional on labor market experience, each observation causes a bigger update of the mean of a worker's belief. Because occupational switching is related to the change in the mean belief, the probability of switching increases with mismatch.[17] Moreover, conditional on mismatch, if the precision of beliefs is higher, the probability of switching occupations will be lower since each observation will update the belief by a smaller amount. Since the precision of beliefs increases and the worker's occupational choice converges to his ideal occupation with experience (i.e., mismatch declines), the probability of switching occupation declines. In Table IX, we show empirical evidence in support of this prediction.

We now turn to some key predictions of our model for the "direction" of occupational switching, which we test in the data in Section 5.4. In particular, we can show that occupational switches tend to be in the direction of reducing existing mismatch. That is, workers who are overqualified (positive mismatch) in a certain skill j will, on average, switch to an occupation with a higher requirement of skill j, thereby reducing the amount by which they are overqualified. And the opposite applies for skill dimensions along which they are negatively mismatched (underqualified). The following proposition formalizes this result.

Introducing some notation, let $\pi^{\text{up}}_{j,t} \equiv \Pr(r_{j,t+1} - r_{j,t} > \kappa_j)$ denote the probability that a worker's occupation next period will have skill requirement j that is higher than his current occupation. We refer to this as "moving up." Similarly, define the probability of moving down: $\pi^{\text{down}}_{j,t} \equiv \Pr(r_{j,t+1} - r_{j,t} < -\kappa_j)$.

[17] Since $r_{j,t} = \hat{A}_{j,t}$, notice that occupational switch would occur if $\hat{A}_{j,t} - \hat{A}_{j,t-1} > \kappa_j$ or $\hat{A}_{j,t} - \hat{A}_{j,t-1} < -\kappa_j$ for some j.

Lemma 4. **[Direction of Occupational Switches]** *If the worker is overqualified in skill j, that is, $A_j - r_{j,t} > 0$, then:*

1. the probability of moving up in skill j is larger than the probability of moving down: $\pi_{j,t}^{up} > \pi_{j,t}^{down}$, and

2. the probability of moving up in skill j increases with the extent of overqualification: $\frac{\partial \pi_{j,t}^{up}}{\partial (A_j - r_{j,t-1})} > 0$.

A worker has positive mismatch (is overqualified) for his occupation in skill dimension j if he chose an occupation with a lower skill-j intensity than his ability. This would happen if he underestimates his ability in dimension j. For such a worker, a new observation, on average, increases his expectations of his ability, and as a result, he becomes more likely to switch to an occupation with a higher skill-j intensity. While the proposition is stated in terms of upward mobility of overqualified workers, the opposite is also true: underqualified workers are more likely to move to occupations with lower skill intensities.

In addition to predicting the likelihood of switching, the model also delivers a prediction for the *size* of the change in the skill space *conditional* on switching, and how it varies with the level of mismatch in the current occupation. The following proposition states the result.

Proposition 3. *Conditional on switching occupations, the reduction in mismatch in skill j is proportional to the level of mismatch in previous occupation.* $r_{j,t} - r_{j,t-1} \sim N\left(\frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(A_j - r_{j,t-1}\right), \left(\frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\right)^2 \sigma_{\epsilon_j}^2\right)$.

In addition to providing a clear testable prediction (which we will test in 5.4), this proposition also provides a way to distinguish between information frictions from search frictions as the source of mismatch. More concretely, with undirected search (e.g., as in Lise and Postel-Vinay (2016)), an overqualified worker will switch to occupations that require higher skills as he samples more occupations, *but* conditional on switching the reduction in mismatch will be independent of his initial level of mismatch. Thus, in our model the average change in skills upon switching is an increasing function of initial mismatch, whereas with undirected search it is a flat line. We will revisit this discussion in the empirical analysis.

3 Data

The main source of data for our analysis is the NLSY79, which tracks a nationally representative sample of individuals who were 14 to 22 years of age on January 1, 1979. It contains detailed information on earnings, employment, and occupational titles for each job, of each worker. In addition, all respondents took the ASVAB test at the start of the survey. The respondents were also given a behavioral test to elicit their social attitudes (e.g., self-esteem, willingness to engage with others, etc.).

Using these test scores, we construct three ability measures for each worker—*verbal*, *math*, and *social*. We then link these ability measures to the skill requirements of his main occupation (expressed in a way that is comparable to these abilities). The latter are constructed from O*NET data on occupations (as described later). Combining these two pieces of information allows us to create a measure of skill mismatch for each worker-occupation pair. We now provide a brief description of the data sources and methods, and relegate further details of sample selection, variable construction, and sample statistics to Appendix C.

3.1 NLSY79

We use the Work History Data File of the NLSY79 to construct yearly panels from 1978 to 2010, providing up to 33 years of labor market information for each individual. We restrict our analysis to males and focus on the nationally representative sample of 3,003 individuals. We exclude individuals who were already working when the sample began so as to avoid the left truncation in their employment history. Such truncation would pose problems for our empirical measures, which require the complete work history to be recorded for each individual. We further drop individuals that are weakly attached to the labor force. The complete description of our sample selection is in Appendix C. Our final sample runs from 1978 through 2010 and includes 1,992 individuals and 44,591 individual-year observations.

Measurement error in occupational switching has received particular attention, and we address it by dropping transitions that immediately revert (which often indicates incorrect coding in the middle year) and conditioning occupation switches on simultaneous employer switches. The latter condition corresponds to one of the filters used by Kambourov and Manovskii (2009b), which Visschers and Carrillo-Tudela (2014) show is quite important because the majority of miscoded occupational switches are within employers.

Annual occupational mobility in our sample is 15.94% compared with 18.48% reported in Kambourov and Manovskii (2008) who use the Panel Study of Income Dynamics (PSID) for the period 1968–1997. Appendix C contains further details.

Data on Workers' Abilities

The version of the ASVAB taken by NLSY79 respondents had 10 component tests.[18] We focus on the following 4 components on verbal and math abilities, which can be linked to skill counterparts: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge. We followed Altonji et al. (2012) to process the ASVAB scores. In particular, because age differences can have a systematic effect on the score and respondents ranged from age 14 to 21, we normalize the mean and variance of each test score by their age-specific values following these authors.

The NLSY79 included three attitudinal scales, which measure a respondent's noncognitive abilities. We focus on two of these measures: the Rotter Locus of Control Scale and Rosenberg Self-Esteem Scale. Both were administered early in the sample, 1979 and 1980, respectively. The Rosenberg scale measures a respondent's feelings about oneself, his self-worth and satisfaction. The Rotter scale elicits a respondent's feelings about his autonomy in the world, the primacy of his self-determination rather than chance. These scores were previously used by Heckman et al. (2006), and Bowles et al. (2001) review evidence on the effects of noncognitive abilities on earnings. Just as with the ASVAB scores, we equalized the mean and variance across ages. We call this dimension of a noncognitive ability *social* ability hereafter.

Occupational Skill Requirements

The U.S. Department of Labor's O*NET project aims to characterize the mix of knowledge, skills, and abilities that are used to perform the tasks that make up an occupation. It includes information on 974 occupations, which can be mapped into the 292 occupation categories included in the NLSY79. For each of these occupations, analysts at O*NET give a score for the importance of each of 277 descriptors.[19] We use 26 of these descriptors that are most related to the ASVAB component tests—a choice

[18] These 10 components are arithmetic reasoning, mathematics knowledge, paragraph comprehension, word knowledge, general science, numerical operations, coding speed, automotive and shop information, mechanical comprehension, and electronics information.

[19] We use the analysts database, version 4.0, which does not include data from survey respondents and should yield a more consistent picture across occupations.

dictated by our measures that relate ASVAB to O*NET and described below—and another 6 descriptors related to the social skills. The complete list is in Appendix C.2.

3.2 Creating Verbal, Math, and Social Components

Information about workers' abilities and occupational skill requirements in verbal and math fields are aggregated in two steps. First, we convert the O*NET skills into 4 ASVAB test categories using the cross-walk created by the Defense Manpower Data Center (DMDC).[20] The DMDC selected 26 O*NET descriptors that were particularly relevant and assigned each a *relatedness score* to each ASVAB category test. For each ASVAB category test, we create an O*NET analog by summing the 26 descriptors and weighting them by this relatedness score. The result is that each occupation gets a set of scores that are comparable to the ASVAB categories, each a weighted average of the 26 original O*NET descriptors.

Second, after normalizing each dimension's standard deviation to be one, we reduce these 4 ASVAB categories into 2 composite dimensions, verbal and math, by applying Principal Component Analysis (PCA). The verbal score is the first principle component of Word Knowledge and Paragraph Comprehension, and the math score is that of Math Knowledge and Arithmetic Reasoning. Because the scale of these principal components is somewhat arbitrary, we convert all four scores (verbal worker ability, math worker ability, verbal occupation requirement, math requirement) into percentile ranks among individuals or among occupations.[21]

Likewise, we create a single index of social ability both on workers' and occupations' sides. From the O*NET, we reduce the six O*NET descriptors to a single dimension by taking the first principal component after scaling each dimension's standard deviation to be one. For the worker's side, we first take the negative of the Rotter scale, because a lower score implies more feeling of self-determination. After scaling both NLSY79 measures to have a standard deviation of one, we take the first principal component. Both occupation- and worker-side data are then converted into percentile rank scores.

[20]To increase the ASVAB's general appeal, the ASVAB Career Exploration Program was established by the U.S. Department of Defense to provide career guidance to high school students. As part of the program, they created a mapping between ASVAB test scores and O*NET occupation requirements (OCCU-Find). The mapping is available at: http://www.asvabprogram.com/downloads/Technical_Chapter_2010.pdf.

[21]The rank scores of skills among occupations are calculated by weighting each occupation by the number of observations of individuals in that occupation in NLSY79.

Table I – Correlations among Ability and Skill Requirement Scores

	(a) Worker Ability			(b) Occupational Skill Requirement		
Workers' Ability	Verbal	Math	Social	Verbal	Math	Social
Verbal	1.00			0.37	0.34	0.35
Math	0.78	1.00		0.44	0.40	0.35
Social	0.30	0.27	1.00	0.13	0.11	0.16

Note: The correlations reported in panel (a) are computed using data on 1,992 individuals in our sample and those in panel (b) are computed using 44,591 worker-occupation pairs observed during the sample period.

Table I reports (a) the correlation between workers' verbal, math, and social ability scores for 1,992 individuals in our sample, and (b) the correlation between workers' abilities and the corresponding skill requirements in their current occupation for 44,591 observations in our sample. In the left panel (a), the highest correlation between ability scores is 0.78 (math and verbal), and the others are 0.27 and 0.30, indicating that each component carries useful independent information. Panel (b) provides a rough measure of sorting of workers into occupations by skills. While the correlations are fairly low (reflecting both mismatch but also measurement error) they are all positive. Workers with strong math skills tend to sort into occupations with generally high skill requirements. A worker's social skills have a relatively low correlation with occupation requirements along every dimension.

4 Empirical Methodology

In this section, we introduce our empirical measure of (contemporaneous) skill mismatch as well as *cumulative mismatch*—to analyze the persistent effects of past mismatch on current wages. We also present two additional statistics—called *positive* and *negative mismatch*—to analyze the effects of over- and underqualification at a given occupation. These measures are then incorporated into the wage equation developed in Section 2 and estimated in the next section.

4.1 An Empirical Measure of Skill Mismatch

(Contemporaneous) Mismatch

Specifically, as in Section 2.2, $\tilde{A}_{i,j}$ is the *measured* ability of individual i in skill dimension j, and $\tilde{r}_{c,j}$ is the *measured* skill requirement of occupation (or *c*areer) c in the same dimension. Let $q(\tilde{A}_{i,j})$ and $q(\tilde{r}_{c,j})$ denote the corresponding percentile ranks of the worker ability and the occupation skill requirements. To define our measure, we take the difference in each skill dimension j between worker abilities and occupational requirements. We sum the absolute value of each of these differences using weights $\{\omega_j\}$ to obtain:[22]

$$m_{i,c} \equiv \sum_{j=1}^{n} \left\{\omega_j \times \left|q(\tilde{A}_{i,j}) - q(\tilde{r}_{c,j})\right|\right\}.$$

The weights are chosen to be the factor loadings from the first principal component, normalized to sum to 1.[23] To help understand magnitudes in our analysis, we rescale our mismatch measure so that its standard deviation is equal to 1. Table B.1 in Appendix B shows descriptive statistics for the mismatch measure, which reveal that the prevalence of mismatch is not specific to a particular educational group, race, or industry.

Cumulative Past Mismatch

A key idea that we will explore in this paper is whether a poor match between a worker and his current occupation can have persistent effects that last beyond the current job. To this end, we construct a measure of *cumulative* mismatch as follows. Consider a worker who has worked at p different occupations as of period t, whose indices are given by the vector $\{c(1), c(2), \ldots, c(p)\}$. The tenure in each of these matches is given by the vector $\left\{\hat{T}_{c(1)}, \hat{T}_{c(2)} \ldots, \hat{T}_{c(p-1)}, T_{c(p),t}\right\}$ where $\hat{T}_{c(s)}$ denotes total tenure in the past occupation $c(s)$, and $T_{c(p),t}$ is the tenure in the current occupation at period t. These must

[22] We use absolute deviations instead of another metric like quadratic deviations, as would be suggested by the quadratic mismatch terms that appear in Equation 7. This is because our measures $q(\tilde{A}_{i,j})$ and $q(\tilde{r}_{c,j})$ are ordinal rather than cardinal. In Section 2, we derived the quadratic form knowing the cardinal values. Given that we can only measure ranks, absolute deviations are the more robust measure of distance. Having said this, we also tried with quadratic distance and the results were substantively unchanged.

[23] That is, we apply principal component analysis (PCA) to the set of absolute values of differences, $\left\{\left|q(\tilde{A}_{i,j}) - q(\tilde{r}_{c,j})\right|\right\}_{j=1}^{n}$, and obtain the first principle component. The weights for the first principle component through PCA turned out to be $(\text{verbal}, \text{math}, \text{social}) = (0.43, 0.43, 0.12)$. We do not know a priori the relative importance of each skills dimension to wages, which could have been a preferable basis for weighting. However, our results were little changed when we used other reasonable weights, like the one which sets an equal weight for all dimensions.

add up to total experience of the worker at period t: $\hat{T}_{c(1)} + \hat{T}_{c(2)} + \cdots + \hat{T}_{c(p-1)} + T_{c(p),t} = E_t$. Cumulative mismatch is defined as the *average* mismatch in the $p-1$ previous occupations:

$$\overline{m}_{i,t} \equiv \frac{m_{i,c(1)}\hat{T}_{c(1)} + m_{i,c(2)}\hat{T}_{c(2)} + \cdots + m_{i,c(p-1)}\hat{T}_{c(p-1)}}{\hat{T}_{c(1)} + \hat{T}_{c(2)} + \cdots + \hat{T}_{c(p-1)}} = \frac{\sum_{s=1}^{p-1} m_{i,c(s)}\hat{T}_{c(s)}}{\sum_{s=1}^{p-1} \hat{T}_{c(s)}}. \tag{9}$$

Each past mismatch value is weighted by its corresponding $\hat{T}_{c(s)}$, so the duration the worker was exposed to an occupation determines its influence on average. This variable is the empirical analogue of the cumulative mismatch term in equation (7) and represents the lingering effect of previous mismatches on the current wage. If occupational match quality only had an effect within a given match (as in, e.g., Jovanovic (1979) or Mortensen and Pissarides (1994)), this variable would have no effect on later wages. On the other hand, if dynamic decisions—such as human capital accumulation—are important and mismatch depresses it, then poor matches in past occupations can reduce current wages.

Positive vs. Negative Mismatch

Equation (7) in Section 2 tells us mismatch may reduce a worker's wages for two reasons: a worker's ability may exceed the occupational requirement, and/or his ability does not meet the occupational requirement. To analyze these positive and negative effects of mismatch separately, we introduce two additional measures. We call them *positive mismatch* and *negative mismatch*, which are defined as

$$m^{+}_{i,c} \equiv \sum_{j=1}^{n} \omega_j \max\left[q(\tilde{A}_{i,j}) - q(\tilde{r}_{c,j}), 0\right], \text{ and } \; m^{-}_{i,c} \equiv \sum_{j=1}^{n} \omega_j \min\left[q(\tilde{A}_{i,j}) - q(\tilde{r}_{c,j}), 0\right],$$

respectively. These definitions mean that $m_{i,c} = m^{+}_{i,c} + \left(-m^{-}_{i,c}\right)$. That is, we decompose our mismatch measure into a part where some of the worker's abilities are overqualified (positive mismatch) and a part where some of them are underqualified (negative mismatch). We can also define *positive cumulative mismatch* and *negative cumulative mismatch* based on these two measures by applying the definition of cumulative mismatch in Section 4.1.

4.2 Empirical Specification of the Wage Equation

Based on our theory in Section 2, we augment the standard Mincer wage regression with measures of mismatch to investigate whether current or cumulative mismatch (or both) matters for current wages. To the extent that current mismatch matters for the level of wages, it can be viewed as a useful, measurable proxy for occupational match quality, which has been treated as an unobservable component of the regression residual by much of the extant literature.[24] Furthermore, if cumulative mismatch or the interaction between match quality and tenure turns out to matter for current wages, then this would provide evidence that match quality affects human capital accumulation and life-cycle wage dynamics.

We will estimate the empirical analogue of the wage equation (7) with a few modifications to consider the wage equation for individual i who is working with employer l in occupation c at time t:

$$\begin{aligned}\ln w_{i,l,c,t} = & \underbrace{\gamma_1 m_{i,c}}_{\text{current mismatch}} + \underbrace{\gamma_2 \left(m_{i,c} \times T_{i,c,t}\right)}_{\text{current mismatch}\times\text{tenure}} + \underbrace{\gamma_3 \overline{m}_{i,t}}_{\text{cumulative past mismatch}} \\ & + \gamma_4 \overline{A}_i + \gamma_5 \left(\overline{A}_i \times T_{i,c,t}\right) + \gamma_6 \overline{r}_c + \gamma_7 \left(\overline{r}_c \times T_{i,c,t}\right) \\ & + \Phi_1 \left(J_{i,l,t}\right) + \Phi_2 \left(T_{i,c,t}\right) + \Phi_3 \left(E_{i,t}\right) + \alpha_4 OJ_{i,t} + X'_{i,t}\beta + \theta_{i,l,c,t}. \end{aligned} \tag{10}$$

In the above equation, we have our mismatch measure, $m_{i,c}$, and its interaction term with occupational tenure, $m_{i,c} \times T_{i,c,t}$. We also include our cumulative mismatch measure, $\overline{m}_{i,t}$, to capture the potential scarring effects of previous mismatches on current wages. This form comes from Equation (7) in our model, which suggests there should be dynamic effects from mismatch.

In the second line of Equation (10), $\overline{A}_i$ is the ability of worker i averaged across skill dimensions, and $\overline{r}_c$ is the skill requirement of occupation c averaged over skill dimensions.[25] We also include their interactions with occupational tenure. These variables are important to include because we might worry that our match quality measures are just proxies for an individual effect from worker or occupation. Equation (7) would suggest

[24] See, for example, Altonji and Shakotko (1987); Topel (1991); Altonji and Williams (2005); Kambourov and Manovskii (2009b).

[25] More precisely, $\overline{A}_i$ is the average of the percentile rank scores of the measured worker's abilities, $\left\{q(\widetilde{A}_{i,j})\right\}_{j=1}^{n}$, and $\overline{r}_c$ is that of the measured occupational requirements, $\left\{q\left(\widetilde{r}_{c,j}\right)\right\}_{j=1}^{n}$. Both $\overline{A}_i$ and $\overline{r}_c$ are again converted into percentile rank scores among individuals or among occupations.

that we include $\bar{A}_i \times E_{i,t}$ instead of $\bar{A}_i \times T_{i,c,t}$. We have also estimated the equation including $\bar{A}_i \times E_{i,t}$ term instead, and obtained very similar results (See Appendix J.5). We have chosen to present the results with $\bar{A}_i \times T_{i,c,t}$ in the regression as our baseline because we are principally concerned with the coefficient on $m_{i,c} \times T_{i,c,t}$ and, by including $\bar{A}_i \times T_{i,c,t}$, we want to convince readers that the mismatch terms are not simply capturing ability, as $m_{i,c}$ is correlated with $\bar{A}_i$ by construction.

In the last line of Equation (10), we have employer tenure, $J_{i,l,t}$, occupational tenure, $T_{i,c,t}$, labor market experience, $E_{i,t}$, and a dummy variable that indicates a continuing job, $OJ_{i,t}$, where Φ_1, Φ_2, and Φ_3 denote polynomials.[26] Finally, when estimating Equation (10), we include one-digit level occupation and industry dummies and a vector of education education and demographic characteristics, $X_{i,t}$.

The last term, $\theta_{i,l,c,t}$, corresponds to the accumulated informational noise in the model and as shown above in Lemma 1, the correlation of this term with mismatch biases the estimated coefficient on mismatch terms towards zero. To the extent that this correlation is large, the true effect of mismatch on wages will be larger than our estimates in the next section indicate.

In addition, $\theta_{i,l,c,t}$ may also include unobserved individual- and match-specific factors and a serial correlation structure. To deal with serial correlation, we allow for an AR(1) structure for the residual, $\theta_{i,l,c,t} = \rho_\theta \theta_{i,l,c,t-1} + \varepsilon_{i,l,c,t}$.[27] Although this structure is not as general as an unrestricted correlation structure, it provides large gains in efficiency (see Cameron and Miller (2015) for a detailed discussion of this point). We will conduct sensitivity analyses allowing for more general error structures later below.

Instrumenting Tenure Variables

As Proposition 2 established, the probability of occupational switching is a function of match quality. Consequently, occupational tenure is endogenous a wage regression because both tenure and wages are functions of match quality.[28] This particular endogeneity problem is well recognized in the literature, going back to Altonji and Shakotko (1987), who proposed and implemented an instrumental variables estimator to avoid the resulting bias. To be clear, if our empirical mismatch measure captured all relevant

[26] We use a second-order polynomial for $\Phi_1(\cdot)$ and third-order polynomials for $\Phi_2(\cdot)$ and $\Phi_3(\cdot)$.

[27] This assumption is common in the income dynamics literature and was also used in Kambourov and Manovskii (2009b).

[28] This endogeneity bias is separate from the bias discussed in Lemma 1, which showed how the mismatch term was correlated with errors.

aspects of mismatch as well as unobservable individual heterogeneity that mattered for occupational tenure, including this mismatch measure—which we do—would take care of the endogeneity problem. But clearly, our measure is a proxy and very likely misses some aspects of mismatch that matters for both wages and tenure, so to preempt an endogeneity bias, we follow Altonji and Shakotko (1987)'s IV approach to instrument for experience and tenure variables. This method has been used extensively in the previous literature (see, among others, Topel (1991) and Altonji and Williams (2005)). Because our regressors also include variables that interact with tenure, we create a corresponding instrument replacing tenure with its instrument. Employer tenure, labor market experience, and the dummy variable for a continuing job are also instrumented in the same manner.[29]

4.3 Workers' Information Set

Before concluding this section, it is important to discuss why workers in our NLSY sample might be uncertain about their abilities, as assumed in our model, even after they have taken the ASVAB, Rotter, and Rosenberg tests. There are at least two reasons for this uncertainty. First, and most important, the NLSY respondents were not told their rank in the test, but were rather given a relatively broad range where their score landed. For example, a respondent knew he scored 10 out of 25 on mathematics knowledge, but was only told that his score corresponded to a rank between 20th and 40th percentiles. Just as in our theoretical model, this is a noisy signal centered around the true mean. As the econometrician, we see the entire NLSY79 sample, so we can compute the worker's precise rank.

Second, as the econometrician, we can process these test scores to extract more information than what the respondents could do. For example, we removed age affects from the test scores, which affects the scores the respondents see but is probably not economically relevant. Similarly, by taking the first principal component from several related tests, we are, statistically speaking, uncovering the underlying ability from several tests that are individually noisy measures. Not knowing the population-level correlations, the respondents could not possibly do the same analysis.

[29]Similar to an occupational match component, an employer match component is potentially correlated with employer tenure and the dummy variable of a continuation of a job. An individual-specific component is potentially correlated with labor market experience.

5 Empirical Results

In this section, we discuss the empirical evidence using our mismatch measures. We will first relate mismatch to wages by incorporating it into a Mincer regression and then study its relationship to switching probability and the direction of switching. We find that the mismatch-tenure interaction and cumulative past mismatch are quite important in the determination of wages. Mismatch also increases the probability of a switch, and once one does switch, it predicts whether a worker will move up or down in the skills required by his occupation. Furthermore, the reduction in mismatch upon switching is proportional to the level of mismatch in the previous occupation.

5.1 Mismatch and Wages

Equation (7) of our model suggested a direct link between mismatch, the history of mismatch and wages. With this motivation in mind, we operationalized it in the regression in Equation (10). Table II presents the main coefficients from these wage regressions and the rest are in Appendix B. The first column includes our measure of mismatch into a standard wage regression. The next adds its interaction with occupational tenure. The full, baseline specification is in the third column, where we introduce our measure of cumulative mismatch. As we discussed in the previous section, we instrument all the tenure variables in the columns labeled "IV-GLS" and show the results without instrumenting in the "GLS" columns. In columns (3) and (6), there are fewer observations because estimating cumulative mismatch requires that the worker held at least one previous occupation.

In column (1) of Table II, contemporaneous mismatch has an estimated coefficient of –0.0205 (and is significant at 1% level), indicating a strong effect on wages. To give an economic interpretation to this coefficient, recall that we have normalized the standard deviation of mismatch to 1, so wages are predicted to be about 4.1% (2.05% $\times$ 2) lower for workers whose mismatch is one standard deviation above the mean relative to those one standard deviation below it.

In the next column, we introduce a mismatch interaction with occupation tenure. Now part of the level effect is replaced by a negative tenure effect (also significant at 1% level). Mismatch depresses initial wages and also leads to slower wage growth over the duration of the match. After 5 years, the overall depression in wages due to slower growth exceeds the losses due to the initial impact.

Table II – Wage Regressions with Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0205***	-0.0091*	-0.0054	-0.0196***	-0.0136***	-0.0128**
Mismatch × Occ Tenure		-0.0020***	-0.0022**		-0.0010**	-0.0007
Cumul Mismatch			-0.0374***			-0.0383***
Worker Ability (Mean)	0.3015***	0.3044***	0.4095***	0.3094***	0.3101***	0.3987***
Worker Ability × Occ Tenure	0.0115***	0.0109***	0.0072**	0.0090***	0.0088***	0.0081***
Occ Reqs (Mean)	0.0977***	0.0962***	0.0878***	0.1383***	0.1374***	0.1426***
Occ Reqs × Occ Tenure	0.0144***	0.0144***	0.0182***	0.0084***	0.0084***	0.0092***
Observations	41596	41596	30599	41596	41596	30599

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation. More detailed regression results are in Appendix B.

In column (3), we introduce cumulative mismatch while keeping all the regressors from column (2). Cumulative mismatch has a significant and negative effect on wages. The tenure effect of mismatch in the current match is unaffected, though the initial level effect becomes smaller and insignificant. To help interpret the size of these coefficients, Table III computes the implied wage losses using specification (3). Looking at the effect of current mismatch, we see that the 90th percentile worst-matched workers face 8.1% lower wages after 10 years of occupational tenure compared with a perfectly matched worker. The difference between the 90th percentile and the 10th percentile of *mismatch* is about 4.1% after 5 years of occupational tenure and widens to 6.8% after 10 years. Comparing the 90th percentile to the 10th percentile of *cumulative mismatch*, we see a wage difference of 9.4%.

For comparison purposes, the last three columns of Table II reports the GLS estimates of the same specifications in the first three columns. Notice that the coefficient on the mismatch and tenure interaction is quite different between IV-GLS and GLS. As we discussed in Section 4.2, the return to tenure is biased because it is correlated with unobservable match quality. The instruments reduce the return to occupational tenure itself by a factor of about 3 (see Table B.2 in Appendix B), precisely because the OLS estimate on tenure takes some variation from the mismatch-tenure interaction term. When we instrument tenure, we purge its correlation with match quality so it is instead ascribed to the interaction between mismatch and occupational tenure, making its coefficient larger.

Table III – Wage Losses from Mismatch & Cumulative Mismatch

Mismatch Degree (High to Low)	*Mismatch* Effect 5 years	10 years	15 years	*Cumul. Mismatch* Effect
90%	-0.049	-0.081	-0.114	-0.123
	(0.011)	(0.018)	(0.030)	(0.014)
50%	-0.021	-0.036	-0.050	-0.065
	(0.005)	(0.008)	(0.013)	(0.007)
10%	-0.008	-0.013	-0.018	-0.029
	(0.002)	(0.003)	(0.005)	(0.003)

Note: Wage losses (relative to the mean wage) are computed for each percentile of each measure in the above table. Standard errors are in parentheses.

Cost of Mismatch

Finally, we estimate the overall cost of mismatch. For this purpose, we compute the predicted wages in our full regression and the counter-factual predicted wages if all mismatch terms were set to zero and we compute the average wages for each case. We find that the overall cost of mismatch is substantial: the average wages would have been 11% higher if mismatch could be eliminated.[30] We would like to emphasize here that this estimate does not reveal why mismatch exists but it provides evidence for the lingering effects of mismatch through human capital accumulation.

Table IV illustrates the cost of mismatch at different percentiles of the predicted and counterfactual wage distributions. The first two columns are in 2002 dollars and the 3rd is the percent difference. Overall, the cost of mismatch is typically higher at higher percentiles of the wage distribution.

We could also try to summarize the overall contribution of mismatch to wage growth in a calculation akin to Topel and Ward (1992), who ascribe 40% of cumulative wage growth during the first ten years to job-to-job transitions. We could further break these transitions into those that improve occupational mismatch and those that do not. We

[30]We can provide a back-of-the-envelope calculation to illustrate this point. From Column 3 of Table II, we have the following coefficients: Mismatch: -0.0054; Mismatch×Occ. Tenure: -0.0022; and Cum. Mismatch: -0.0374. The average value of Mismatch is 1.56 in Table B.1. The average Cumulative Mismatch is 1.9 and the average of Mismatch×Occ. Tenure is 10. Based on these, we can compute the wage loss due to each component: Mismatch = -0.0054×1.56 = -0.0084; Mismatch×Occ. Tenure = -0.0022×10 = -0.0220; and Cum. Mismatch = -0.0374×1.9 = -0.0711. And the total effects is -0.1015. The most important effect comes from cumulative mismatch, which reduces wages by 7.11%, followed by Mismatch×Tenure interaction.

Table IV – Cost of Mismatch Over the Wage Distribution

	No Mismatch	Actual	Difference (%)
10%	10.43	9.63	8.3
50%	16.20	14.51	11.7
90%	26.97	24.61	9.7

Note: The table reports wages at different percentiles.

first replicated their exercise in the NLSY 79, using their definition of excess wage growth:

$$\Delta w_E = w_{t+1} - w_{t-2} - E[w_{t+1} - w_t|\cdot] - E[w_{t-1} - w_{t-2}|\cdot]$$

where a job transition occurred between t and $t-1$. The expectations terms, $E[w_{t+1} - w_t]$, $E[w_{t-1} - w_{t-2}]$, for wage growth are taken with the predicted values from our baseline regression excluding mismatch, skill and occupational requirements terms. Essentially, excess wage growth at time t is the residual wage growth above that predicted by the trend.

Overall, our data shows essentially the same total excess wage growth as Topel and Ward (1992). Job changes bring an average of 8.5% excess wage growth and account for 35% of growth during the first 10 years. When we look only at job transitions that also include an occupation switch and divide these switches that increase mismatch from those that decrease mismatch, we find that occupational transitions that improve mismatch have average excess wage growth of about 12%. These switches that reduce mismatch account for 14% of the cumulative wage growth in the first 10 years of experience. To put this another way, of the 35% of wage growth attributable to all job changes, improvements in mismatch account for 40% of that growth.

Three Dimensions of Skill Mismatch

When presenting the theoretical results of our model, we aggregated over skills, which helped keep it tractable. However, we could create an analogous wage equation by splitting dimensions apart and allowing them to enter Equation 7 separately. Here, we consider this notion empirically, and look for wage implications dimension-wise. In Table V, we report the results when we include each component mismatch measure in our regressions. The component mismatch measure in skill j is defined as the difference in the rank scores of ability and occupational requirement, $m_{i,c,j} \equiv \left| q(\tilde{A}_{i,j}) - q(\tilde{r}_{c,j}) \right|$. As before, we scale each dimension to have a standard deviation of one so that they are

Table V – Wage Regressions with Mismatch by Components

	(1) IV-GLS	(2) IV-GSL	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch Verbal	-0.0103***	0.0082	0.0133	-0.0104***	0.0032	0.0051
Mismatch Math	-0.0108***	-0.0172**	-0.0193**	-0.0096**	-0.0178***	-0.0193***
Mismatch Social	-0.0027	0.0005	0.0055	-0.0021	-0.0012	0.0034
Mismatch Verbal × Occ Tenure		-0.0031***	-0.0043***		-0.0022***	-0.0028***
Mismatch Math × Occ Tenure		0.0010	0.0020*		0.0013**	0.0021**
Mismatch Social × Occ Tenure		-0.0005	-0.0007		-0.0001	-0.0006
Cumul Mismatch Verbal			-0.0148***			-0.0137**
Cumul Mismatch Math			-0.0241***			-0.0260***
Cumul Mismatch Social			-0.0088**			-0.0078*
Verbal Ability	-0.0622*	-0.0652*	0.0029	-0.0132	-0.0181	-0.0006
Math Ability	0.3660***	0.3711***	0.4260***	0.3209***	0.3254***	0.4027***
Social Ability	0.0997***	0.1003***	0.1078***	0.1046***	0.1046***	0.1230***
Verbal Ability × Occ Tenure	0.0128***	0.0129***	0.0083	0.0066**	0.0070**	0.0083**
Math Ability × Occ Tenure	-0.0051	-0.0060	-0.0066	-0.0016	-0.0024	-0.0045
Social Ability × Occ Tenure	0.0061**	0.0060**	0.0071**	0.0059***	0.0060***	0.0056**
Occ Reqs Verbal	0.0190	0.0209	-0.0420	0.0761	0.0832	0.0211
Occ Reqs Math	0.1313*	0.1244*	0.1772**	0.1176**	0.1080*	0.1731**
Occ Reqs Social	-0.1115***	-0.1084***	-0.1014***	-0.1064***	-0.1047***	-0.0966***
Occ Reqs Verbal × Occ Tenure	-0.0055	-0.0061	0.0034	-0.0184***	-0.0202***	-0.0123
Occ Reqs Math × Occ Tenure	0.0138	0.0153	0.0077	0.0189***	0.0213***	0.0120
Occ Reqs Social × Occ Tenure	0.0095***	0.0088***	0.0115***	0.0112***	0.0108***	0.0134***
Observations	41596	41596	30599	41596	41596	30599

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation. More detailed regression results are in Appendix B.

comparable.

Looking at math and verbal skills in Table V, we see a pattern emerge: mismatch in either dimension has a negative effect on wages but math mismatch reduces the level of wages without a significant growth rate effect, whereas the opposite is true for verbal, which has a small level effect but a strong growth rate effect. The interaction term for verbal mismatch implies a 13% wage gap after 10 years of tenure between the top and bottom deciles of mismatch. Math mismatch implies a 5% initial wage difference between 90th and 10th percentiles that attenuates over time. Social mismatch has a weaker effect overall, though still negative.

Interestingly, the same difference between math and verbal skills is seen in the effects of *ability* on wages (lower panel of Table V): in the first three columns, math ability has a large level effect (ranging from 37% to 43% across specifications) but little tenure effect, whereas verbal ability has a higher tenure effect (ranging from 0.8 to 1.3% per year) on wages, though the effect of verbal ability on wages is insignificant in our most general specification (column 3). Social ability has both significant level and tenure effects, with the tenure effect exceeding the level effect in two years. One interpretation of these results might be that math ability is easier to observe by employers and the market and so are priced immediately, whereas verbal and social abilities capture some more subtle traits that are revealed more slowly over time, leading to a growth rate effect.[31]

The effect of cumulative mismatch is negative in all three dimensions and statistically significant at 5% level or higher. The cumulative effect for verbal skills is equivalent to about 7 years of mismatch in the current occupation, combining the immediate and tenure effects; cumulative social mismatch is equivalent to about 20 years of mismatch in the current occupation, mainly because the effects of current mismatch are small.

Positive and Negative Mismatch

Next, we investigate the effects of positive and negative mismatch, as defined in Section 4.1, on wages. Recall that, as we discussed in Section 2.4, in our model both workers who are overqualifed and those who are underqualified experience lower skill accumulation and lower wages. We also noted that this is not the case in the standard Ben-Porath version of our model where a worker who is underqualified at his current occupation (i.e., negative mismatch) worker earns a lower wage today but a higher wage in the future since his human capital accumulation is still faster than what what would be at this optimal occupation.

To examine these different implications, we rerun our benchmark wage regressions but now splitting the mismatch term into the positive and negative components. Column (1) of Table VI presents the simplest version—the analogue of Column 1 in Table II: positive and negative mismatch both reduce wages.[32] However, the effect is not symmetric as predicted by our model; the coefficient on negative mismatch is about 4 times larger than that of positive mismatch. In Column (2), we add the interaction terms with tenure, which are statistically significant for both types of mismatch.

[31] This view is consistent with Altonji and Pierret (2001)'s interpretation of public learning about unobserved abilities.

[32] Recall that negative mismatch adds all skill dimensions for which the worker is underqualified (so by definition it is a negative number), and the positive estimated coefficients imply that negative mismatch reduces wages. Furthermore, we do not include terms for the level of worker abilities or occupational requirements because in breaking apart the absolute value of the mismatch measure, we would encounter problems of collinearity between the positive and negative mismatch measure and those terms.

Table VI – Wage Regressions with Positive and Negative Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Positive Mismatch	-0.0072*	0.0087	0.0141*	-0.0074**	-0.0015	0.0010
Negative Mismatch	0.0305***	0.0172***	0.0231***	0.0276***	0.0137***	0.0195***
Pos. Mismatch × Occ Tenure		-0.0028***	-0.0032***		-0.0011*	-0.0005
Negs. Mismatch × Occ Tenure		0.0023***	0.0014		0.0024***	0.0022***
Cumul Positive Mismatch			-0.0145***			-0.0220***
Cumul Negative Mismatch			0.0159***			0.0085*
Observations	41596	41596	30599	41596	41596	30599

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation. More detailed regression results are in Appendix B.

Finally, in Column (3), we add cumulative positive and negative mismatch, which speaks more directly to the distinction between our model and the Ben-Porath specification. Two points are worth noting. First, both types of cumulative mismatch reduce wages as predicted by our model (and in contrast to the Ben-Porath version). Further, the magnitudes of the two coefficients are quite close to each other (-0.0145 for positive and 0.0159 for negative), which is consistent with the symmetric loss assumed in our model. Second, the effects of contemporaneous positive and negative mismatch are consistent with the model: they both reduce wages when the level and tenure effects are combined. The point estimate of the level of positive mismatch is positive but it is barely significant (at 10% only). Even with that positive coefficient (0.0141), the negative estimated tenure effect (-0.0032) implies that after 5 years of tenure the overall effect is turns negative. Overall, the signs and the statistical significance (4 out of 6 in column 3) of the estimated coefficients are consistent with the predictions of our model and provide empirical support for the technology for skill accumulation and wages at different occupations as we specified in our model (see equation (1)).

To summarize, the results in Tables II to VI collectively speak to the importance of skill mismatch for current wages as well as for generating long-lasting scarring effects on future wages. We now turn address various issues about the sensitivity of these results to various assumptions we made in the benchmark analysis.

Table VII – Wage Regressions with Different Assumptions on Errors

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) IV-CLU	(5) IV-CLU	(6) IV-CLU
Mismatch	-0.0205***	-0.0091*	-0.0054	-0.0271***	-0.0145*	-0.0054
Mismatch × Occ Tenure		-0.0020***	-0.0022**		-0.0020**	-0.0024***
Cumul Mismatch			-0.0374***			-0.0355***
Worker Ability (Mean)	0.3015***	0.3044***	0.4095***	0.2466***	0.2475***	0.3408***
Worker Ability × Occ Tenure	0.0115***	0.0109***	0.0072**	0.0166***	0.0161***	0.0140***
Occ Reqs (Mean)	0.0977***	0.0962***	0.0878***	0.1529***	0.1528***	0.1576***
Occ Reqs × Occ Tenure	0.0144***	0.0144***	0.0182***	0.0155***	0.0154***	0.0161***
Observations	41596	41596	30599	44591	44591	33072

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. Baseline regression with all of the regressors. IV-GLS corresponds to our benchmark result in column (3) of Table II. IV-CLU uses standard errors clustered at the individual level. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Full results are in Appendix Table D.1.

5.2 Robustness of the Wage Regression Results

We begin with two potentially important technical assumptions—one about the error structure and another one about including fixed effects—and investigate the effects of relaxing each one. We then present several robustness checks and extensions to our baseline wage regressions.

Serial Correlation. In panel data, wage regression errors are often serially correlated. This may bias downward the standard errors on our estimates unless we correct for it. Therefore our benchmark regressions use GLS to allow for the autoregressive structure for the error term. While this method imposes some restrictions, it is more efficient than the leading alternative, panel robust standard errors.[33]

To allow the correlation structure on errors to be mostly unrestricted, we estimate panel-robust standard errors, clustering at the individual level.[34] All of our mismatch coefficients remain still highly significant despite larger standard errors than if we assume errors have an AR(1) structure as in GLS. The point estimates lend further credence to our estimator: were we to misspecify the error structure, it would bias our point estimates; instead, they they change little across the different methods. Appendix D provides estimates for our other specifications, corresponding to Tables II and V. There,

[33] Cameron and Miller (2015) make a powerful case for FGLS and conclude: "It is remarkable that current econometric practice with clustered errors ignores the potential efficiency gains of [F]GLS."

[34] This estimator was introduced by Arellano (1987). See Cameron and Trivedi (2005) for the details.

Table VIII – Wage Regressions with Fixed Effects

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) IV-GLS-FE	(5) IV-GLS-FE	(6) IV-GLS-FE
Mismatch	-0.0205***	-0.0091*	-0.0054	-0.0097***	-0.0030	-0.0038
Mismatch × Occ Tenure		-0.0020***	-0.0022**		-0.0019***	-0.0020**
Cumul Mismatch			-0.0374***			-0.0363***
Worker Ability × Occ Tenure	0.0115***	0.0109***	0.0072**	0.0126***	0.0121***	0.0105***
Occ Reqs (Mean)	0.0977***	0.0962***	0.0878***	0.0066	0.0056	-0.0131
Occ Reqs × Occ Tenure	0.0144***	0.0144***	0.0182***	0.0145***	0.0144***	0.0162***
Worker Ability (Mean)	0.3015***	0.3044***	0.4095***			
Observations	41596	41596	30599	41596	41596	30599

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. Baseline regression with all of the regressors. IV-GLS corresponds to our benchmark result in column (3) of Table II. IV-GLS-FE includes fixed effects at the individual level. Baseline regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Fixed effects regressions do not include time-invariant terms, demographics or mean worker ability.

the specification with individual components is affected the most, since we have larger number of regressors and thus, the lost efficiency from this method is more consequential.

Fixed Effects. One might worry that our estimates merely reflect fixed differences across people: perhaps low wage workers simply choose occupations haphazardly. To address this, we include individual-level fixed effects and repeat all of our estimates. The results for the baseline are in Table VIII. Appendix E contains detailed output. The fixed effect specification, while it controls for individual level unobservables, significantly restricts the amount of information available to our estimates: essentially we are only using information from the occupations switches to identify the wage effect of the level of mismatch. Nevertheless, the effects of our mismatch and cumulative mismatch measures are quantitatively similar to the benchmark case without fixed effects.

Physical Skill. In addition to verbal, math, and social dimensions, we experimented with a physical mismatch measure whose construction and results are presented in Appendix F. However, these results do not show a meaningful relationship between mismatch in physical skills and wages. The physical dimension has a very small loading in our mismatch composite measure. When it is included in the regression separately, the coefficients are generally insignificant.

College Graduates. In Appendix G, we split the sample by education level (college and noncollege) and reestimate our benchmark wage regression for each group. We find that the negative effects of skill mismatch are larger for college graduates. This is true especially of the cumulative mismatch measure, which nearly doubles in magnitude.

Among the dimensions, social and verbal have a particularly pronounced effect among this subsample.

Younger versus Older Workers. Age is another dimension along which the effects from mismatch may differ, though the direction is indeterminate. On one hand, our model suggests that workers switch to better matches over time. Thus, a larger portion of mismatch observed among the old may be noise and therefore attenuate the coefficient on *current* mismatch. On the other hand, our theory suggests that the longer a worker experiences mismatch, the more it affects his human capital and earnings, which would show up as a more negative coefficient on average *cumulative* mismatch among the relatively old.

In order to test these two conjectures, we interact our mismatch terms with a dummy for workers less than 35 (see Table H.1 in Appendix H). We do not find any evidence for smaller effect of *current* mismatch among older workers. On the other hand, we find evidence for older workers being more affected by average *cumulative* mismatch. We also computed the overall cost of mismatch for young and old workers. For young workers, the average predicted wage is 8.5 log points lower than without mismatch. For old workers, the average predicted wage is 12.5 log points lower than without mismatch. The higher cost for old people comes mostly from the larger effect of cumulative mismatch for that group.

Earnings. It is well understood that, in micro survey data, earnings are typically measured more precisely than wages, which often contain significant measurement error.[35] With this in mind, we complement our benchmark analysis (that uses wages) with an analogous analysis that uses earnings as the left-hand-side variable in a Mincer-style regression. The results reported in Appendix I confirm the robustness of the benchmark regression.

Log Specification. Because our mismatch measure does not have an inherent cardinality, it's possible that a transformation of its values would be suitable. Therefore, we also tried regressions with the logarithm of the mismatch measure. We replaced mismatch and cumulative terms in our benchmark framework (Equation (10)) with a log transformation of the mismatch measure. The effects of mismatch on wages obtained

[35]See Bound et al. (2001) for a thorough survey of the evidence. This finding is often attributed to the fact that most workers are salaried rather than paid on an hourly basis, and actual hours are found to be difficult to recall.

through these regressions are very similar to those in Table III. For these results, see Table J.1 in Appendix J.

Higher-Order Terms for Ability and Skill Requirements. One potential concern with our benchmark result is that our mismatch measure is capturing nonlinear effects of worker's mean ability or occupation's mean skill requirements rather than mismatch itself. In order to make sure these are not the case, we include quadratic terms of worker's mean ability and occupation's mean skill requirements in the benchmark framework and run regressions. Contrary to our doubt, the results of these regression are rarely different from those in our benchmark regressions. The results from this exercise are in Table J.3 in Appendix J.

Ability-Experience Interaction Term. Our benchmark regression did not include an interaction term between ability and experience, which is an commonly used explanatory variable in Mincerian regressions as it is predicted to matter in many versions of the human capital model. To examine if this omission could be affecting our results, we reestimated the benchmark wage regression by including an ability-experience interaction, in addition to the ability-occupation tenure interaction that was already present. The results reported in Table J.5 of Appendix J) show that the ability-experience coefficient is positive, large (about 1% per year of experience) and statistically significant at 1% level at the expense of the coefficient on ability-occupational tenure interaction, which is now small and statistically insignificant (whereas it was significant at 1% level in the benchmark case; see Table II). However, this addition had has virtually no effect on the main coefficients of interest, which change very little, both in terms of magnitude and significance, from the benchmark case.

5.3 Mismatch and Occupational Switching

So far we have focused on the impact of mismatch on wages. We now turn to the second key question we raised in the introduction and implied by the model. What is the effect of mismatch on the probability of occupational switching? In the model, highly mismatched workers have larger errors in their beliefs. This also leads to a a higher probability of an occupational switch because Bayesian learning implies a larger expected correction to their beliefs. Proposition 2 formalized this logic, and we now look for this potential effect in the data.

We estimate a linear probability model for occupational switching on the same set of

Table IX – Regressions for the Probability of Occupational Switch

	(1) LPM-IV	(2) LPM-IV	(3) LPM-IV	(4) LPM	(5) LPM	(6) LPM
Mismatch	0.0135***			0.0066***		
Mismatch Verbal		0.0076***			0.0053**	
Mismatch Math		0.0074***			0.0025	
Mismatch Social		0.0007			-0.0012	
Positive Mismatch			0.0134***			0.0087***
Negative Mismatch			-0.0130***			-0.0028
Worker Ability (Mean)	-0.0370***	-0.0370***	-	-0.0208*	-0.0211*	-
Worker Ability × Occ Tenure	-0.0003	-0.0003	-	0.0019**	0.0020**	-
Occ Reqs (Mean)	-0.0333**	-0.0334**	-	-0.1225***	-0.1223***	-
Occ Reqs × Occ Tenure	-0.0052***	-0.0052***	-	0.0106***	0.0106***	-
Observations	41,596	41,596	41,596	41,596	41,596	41,596

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated as robust Huber-White sandwich estimates. More detailed regression results are in Appendix B.

regressors as in our wage regressions, again instrumenting for occupational tenure. We are most interested in the coefficient on mismatch in the current occupation. Table IX displays our baseline estimates.

Notice that the effect of current mismatch on the probability of switching occupations is always positive and significant at the 1% level, with the exception of social mismatch in column (2). To give a better idea about the magnitudes implied by these coefficients, in Table X we compute the occupational switching probabilities across the mismatch distribution using the specifications in Columns (1) and (2). A worker who is in the 90th percentile of the mismatch distribution is 3.4 percentage points more likely to switch occupations than an otherwise comparable worker in the 10th percentile, a difference corresponding to about 21% of the average switching rate.

Splitting mismatch into components (last three columns of Table X), we see that the gap in switching probability between 90th and 10th percentile is approximately 2 percentage points for verbal and math skills, but is close to zero for social skills. Thus, consistent with what we found for wages, contemporaneous social mismatch seems to only have a modest effect on outcomes once we account for math and verbal skills.

In Column 3 of Table IX, we see that the effects are roughly symmetric, with increased switching probability similarly associated with positive and negative mismatch. Workers

TABLE X – Effect of Mismatch on Occupational Switching Probability

Mismatch Degree (High to Low)	*Mismatch* Effect	Effect by *Component*		
		Verbal	Math	Social
90%	0.0407	0.0209	0.0203	0.0020
	(0.0069)	(0.0078)	(0.0076)	(0.0062)
50%	0.0178	0.0082	0.0076	0.0008
	(0.0030)	(0.0030)	(0.0029)	(0.0024)
10%	0.0065	0.0014	0.0013	0.0001
	(0.0011)	(0.0005)	(0.0005)	(0.0004)

Note: Each cell reports the change in the probability of switching occupations. Standard errors are in parentheses.

whose skills are worse than their occupations or better are both more likely to switch occupations. This is consistent with the "U-shape" of Groes et al. (2015) , that workers either under- or overqualified for their occupation are more likely to switch.

5.4 Switch Direction

Not only do mismatched workers switch occupations more frequently, but their switches are also directional. They switch to improve their match quality and the magnitude of their mismatch predicts the magnitude of the correction. Workers who are overqualified tend to switch to occupations with higher skill requirements and the converse for underqualified workers.

As Lemma 4 lays out, our model of learning suggests that switches ought to be directed because as they learn, their beliefs tend towards their true ability. Furthermore, Proposition 3 shows that the magnitude of the change is related to the magnitude of mismatch. This is a prediction of our model that contrasts with search models of mismatch. If search frictions induced mismatch, workers will switch towards better matches, but the size of the switch will be independent of the size of mismatch.

Our parametric and nonparametric evidence shows that switches tend to correct past mismatch and the magnitude of the correction depends on the magnitude of mismatch. Nonparametrically, we see this in panels (a) through (c) of Figure 2. We plot on the vertical axis the change in skill requirement for every worker who switches occupation and on the horizontal we plot the last positive or negative mismatch in that skill. Specifically, the vertical axis is the difference between the j^{th} skill requirement in the last occupation and that in the current one, i.e., $q(\tilde{r}_{c(p),j}) - q(\tilde{r}_{c(p-1),j})$. The horizontal axis, positive and negative mismatch *in skill j,* is defined as in Section 4.1, but using only one dimension at a time.[36] To give the scatter plots some shape in Figure 2, we fit local polynomial regressions for positive and negative mismatch.

As shown in these panels, the upward-sloping curves on both sides of zero mean that individuals who are overqualified in skill j (the right half of the axes) tend to choose their next occupation with a higher skill requirement, whereas the opposite is true for individuals who are underqualified. The positive relationship means that the more mismatched the worker was in the last occupation, the larger the change in requirements of that skill is in the next switch. Panel (d) plots the same relationship by aggregating across all three skill types, which again shows the same patterns.[37]

Figure 2 only documents a univariate relationship between changes in requirements as a function of current mismatch in the same skill dimension. To investigate richer dependencies, we regress the change (upon switching occupations) in skill requirement j on positive and negative mismatch in all three skill dimensions for the worker's last occupation.[38] We also include all of the worker characteristics from our wage regression on positive and negative mismatch, e.g. demographics, employer and occupational tenure.

Columns (1) through (3) of Table XI report the coefficient estimates from this regression. Column (4) reports the case where the average change in skills is regressed on positive and negative mismatch.

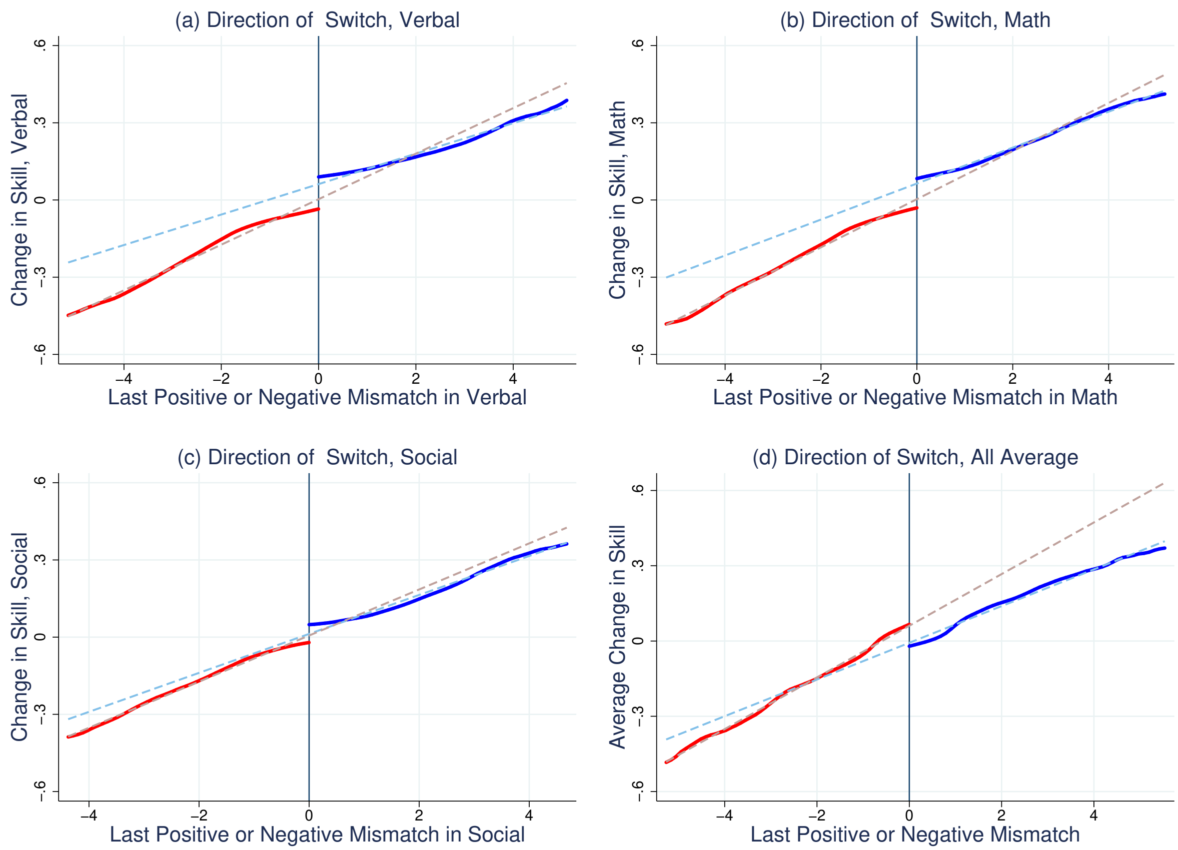
There are several takeaways from this table. Most importantly, the positive coefficients on all regressors confirm the main message of Figure 2: the skill change upon switching is an increasing function of current mismatch, so switching works to reduce skill mismatch. This is true even when we control for mismatch along other dimensions. For example, Column (1) tells us that a worker will choose his next occupation to have a higher verbal skill requirement if he is currently overqualified in verbal dimension (first row), but even more so if he is currently overqualified in math skills (coefficients of 0.0316 vs 0.0599). Mismatch in the social dimension has little impact (coefficient of 0.0061) on verbal requirements, but a strong effect on changes in social skill requirements. These

[36] $m^{+}_{i,c(p-1),j} \equiv \max\left[q(\tilde{A}_{i,j}) - q(\tilde{r}_{c(p-1),j}), 0\right]$ and $m^{-}_{i,c(p-1),j} \equiv \min\left[q(\tilde{A}_{i,j}) - q(\tilde{r}_{c(p-1),j}), 0\right]$.

[37] Again, we restrict our observations to those who have strictly positive mismatch (to the right of the axis) and those who have strictly negative mismatch (to the left of the axis). Unlike positive or negative mismatch in skill j, observations don't split into either the positive or negative side in this case. That is, a number of observations show up on both sides.

[38] As before, skill requirement is measured in terms of percentile rank.

Figure 2 – Non-Parametric Plots of Direction of Switch



Note: We run local polynomial regressions with a simple rule-of-thumb bandwidth (solid lines). On the X-axis, we have the value of the last positive or negative mismatch measure. On the Y-axis, a change in a skill is computed as the difference in the rank score of the skill in the current occupation and the one in the last occupation. An average change is computed as the mean of the changes in the rank scores in all skills.

results echoes the same theme as before that math and verbal skills are distinct, yet closely connected, whereas social skills have their own dynamics. The coefficients also show an asymmetry that was apparent in Figure 2: with the exception of math, workers who are underqualified in a dimension reduce that skill requirement by more than an overqualified worker increases his skill requirement.

To provide some interpretation of the estimated coefficients, we compute the effect of positive and negative mismatch in skill j on the change in that skill for 90th, 50th, and 10th percentile rank of each measure in Table XII using (diagonal entries from the) regression results. For example, a worker with high positive verbal mismatch, in the 90th percentile, will choose an occupation 9.95 percentiles higher in verbal skill requirements. A similarly underqualified worker in the 10th percentile of negative mismatch reduces

Table XI – Regressions for Direction of Switch

Dependent variable →	(1) Verbal	(2) Math	(3) Social	(4) All Average
Last Pos. Mismatch, Verbal	*0.0316******	0.0097**	0.0143***	
Last Neg. Mismatch, Verbal	*0.0838******	0.0536***	0.0216***	
Last Pos. Mismatch, Math	0.0599***	*0.0898******	0.0021	
Last Neg. Mismatch, Math	0.0558***	*0.0893******	0.0076	
Last Pos. Mismatch, Social	0.0061*	0.0046	*0.0774******	
Last Neg. Mismatch, Social	0.0264***	0.0166***	*0.1043******	
Last Positive Mismatch				0.0751***
Last Negative Mismatch				0.1143***
Observations	6,594	6,594	6,594	6,594

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure before switch, employer tenure before switch, work experience before switch, and dummies for one-digit-level occupation and industry for the last job held. Standard errors are estimated as robust Huber-White sandwich estimates. More detailed regression results are in Appendix B.

Table XII – Effect of Last Mismatch on Change in Skills

Last mismatch percentile in skill j		Predicted Percentile Change in Skill j Verbal	Math	Social
Positive	90%	9.95	27.91	22.87
		(1.36)	(1.47)	(1.04)
	50%	4.03	10.70	9.71
		(0.55)	(0.56)	(0.44)
	10%	0.67	1.88	1.65
		(0.09)	(0.10)	(0.07)
Negative	90%	–1.59	–1.77	–1.82
		(0.10)	(0.11)	(0.07)
	50%	–9.07	–10.02	–10.31
		(0.57)	(0.60)	(0.39)
	10%	–24.25	–26.35	–29.59
		(1.53)	(1.58)	(1.12)

Note: These values are changes in percentile rank scores in each skill dimension. Standard errors are in parentheses.

his verbal skill requirements by 24.25 percentiles when switching.

6 Conclusion

In this paper, we propose an empirical measure of multidimensional skill mismatch that is implied by a dynamic model of skill acquisition and occupational choice. Mismatch arises in our model due to workers' imperfect information about their learning abilities, which causes them to choose occupations that are either above or below their optimal level. As workers discover their true abilities workers better allocate themselves toward their optimal careers.

Our empirical findings provide support to the notion of mismatch proposed in this paper. In particular, we find that mismatch predicts wages even with a long list of controls, including worker fixed effects and abilities and occupation requirements constructed from ASVAB and O*NET. Furthermore, mismatch has a long-lasting impact on workers' wages, depressing them even in subsequent occupations. This latter finding is consistent with the human capital channel that is embedded in our theoretical model.

A second set of findings shows workers choose to switch occupations so as to reduce their skill mismatch. This is true even when we split mismatch into its components. The magnitudes involved are also quite large, revealing large adjustments for workers in the skill space upon switching. These findings are consistent with our model of learning. Comparing skill dimensions, we find social skills behave somewhat differently from math and verbal skills. Although social ability appears to matter for wages, mismatch between a worker and an occupation along this dimension has a weaker relationship to wages than math or verbal mismatch.

These findings should only serve to motivate further work on the mechanisms involved in learning, human capital accumulation, and occupational choice. The empirical evidence we presented suggests a strong link between human capital accumulation and lifetime earnings, but fully quantifying its effects will require a structural quantitative model. Such a model will also allow us to conduct policy experiments and quantify their impact on lifetime welfare.

References

Altonji, Joseph G. and Charles R. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

_ **and Nicolas Williams**, "Do Wages Rise with Job Seniority? A Reassessment," *Industrial & Labor Relations Review*, 2005, *58* (3), 370–397.

_ **and Robert A. Shakotko**, "Do Wages Rise with Job Seniority?," *Review of Economic Studies*, 1987, *54* (3), 437–459.

_ **, Prashant Bharadwaj, and Fabian Lange**, "Changes in the Characteristics of American Youth: Implications for Adult Outcomes," *Journal of Labor Economics*, 2012, *30* (4), 783–828.

Antonovics, Kate and Limor Golan, "Experimentation and Job Choice," *Journal of Labor Economics*, 2012, *30* (2), 333–366.

Arellano, Manuel, "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics*, 1987, *49* (4), 431–34.

Bacolod, Marigee P. and Bernardo S. Blum, "Two Sides of the Same Coin: U.S. "Residual" Inequality and the Gender Gap," *Journal of Human Resources*, 2010, *45* (1), 197–242.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75* (4), 352–365.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement error in survey data," in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Elsevier, 2001, chapter 59, pp. 3705–3843.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne, "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 2001, *39* (4), 1137–1176.

Cameron, A Colin and Douglas L Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

_ **and Pravin K Trivedi**, *Microeconometrics: Methods and Applications*, Cambridge university press, 2005.

Farber, Henry S. and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Flinn, Christopher J., "Wages and Job Mobility of Young Workers," *Journal of Political Economy*, 1986, *94* (3), S88–S110.

Fredriksson, Peter, Lena Hensvik, and Oskar Nordström Skans, "Mismatch of Talents? Evidence on Match Quality, Job Mobility and Entry Wages," *mimeo*, 2015.

Gardner, Howard, *Frames of Mind: The Theory of Multiple Intelligences*, 3rd ed., Basic Books, 2011.

Gathmann, Christina and Uta Schönberg, "How General is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 2010, *28* (1), 1–49.

Gervais, Martin, Nir Jaimovich, Henry E. Siu, and Yaniv Yedid-Levi, "What should I be when I grow up? Occupations and unemployment over the life cycle," *Journal of Monetary Economics*, 2016, *83* (C), 54–70.

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

Groes, Fane, Philipp Kircher, and Iourii Manovskii, "The U-Shapes of Occupational Mobility," *Review of Economic Studies*, 2015, *82* (2), 659–692.

Guvenen, Fatih, "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, June 2007, *97* (3), 687–712.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411–482.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality," *American Economic Review*, 2011, *101* (7), 2923–2954.

Ingram, Beth F. and George R. Neumann, "The Returns to Skill," *Labour Economics*, 2006, *13* (1), 35–59.

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

_ **and Robert Moffitt**, "An Estimate of a Sectoral Model of Labor Mobility," *Journal of Political Economy*, 1990, *98* (4), 827–852.

Kambourov, Gueorgui and Iourii Manovskii, "Rising Occupational and Industry Mobility in the United States: 1968–97," *International Economic Review*, 2008, *49* (1), 41–79.

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

Lazear, Edward P., "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 2009, *117* (5), 914–940.

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional Skills, Sorting, and Human Capital Accumulation," *Working Paper*, 2016. University College London.

Michaud, Amanda M. and David Wiczer, "Occupational Hazards and Social Disability Insurance," *Working Paper*, 2014. Federal Reserve Bank of St. Louis.

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

Mortensen, Dale T. and Christopher A. Pissarides, "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 1994, *61* (3), 397–415.

Moscarini, Giuseppe, "Excess Worker Reallocation," *Review of Economic Studies*, 2001, *68* (3), 593–612.

Nagypal, Eva, "Learning by Doing vs. Learning about Match Quality: Can We Tell Them Apart?," *Review of Economic Studies*, 2007, *74* (2), 537–566.

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

Papageorgiou, Theodore, "Learning Your Comparative Advantages," *Review of Economic Studies*, 2014, *81* (3), 1263–1295.

Parent, Daniel, "Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics," *Journal of Labor Economics*, 2000, *18* (2), 306–323.

Pavan, Ronni, "Career Choice and Wage Growth," *Journal of Labor Economics*, 2011, *29* (3), 549–587.

Perry, Anja, Simon Wiederhold, and Daniela Ackermann-Piek, "How Can Skill Mismatch be Measured? New Approaches with PIAAC," *Methods, Data, Analyses*, 2014, *8* (2), 137–174.

Poletaev, Maxim and Chris Robinson, "Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000," *Journal of Labor Economics*, 2008, *26* (3), 387–420.

Rogerson, Richard, Robert Shimer, and Randall Wright, "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 2005, *43* (4), 959–988.

Rosen, Sherwin, "Learning and Experience in the Labor Market," *Journal of Human Resources*, 1972, *7* (3), 326–342.

Sanders, Carl, "Skill Accumulation, Skill Uncertainty, and Occupational Choice," *mimeo*, August 2016.

Speer, Jamin D., "Pre-Market Skills, Occupational Choice, and Career Progression," *Journal of Human Resources*, 2017, *52* (1), 187–246.

Topel, Robert, "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, 1991, *99* (1), 145–176.

Topel, Robert H and Michael P Ward, "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 1992, *107* (2), 439–79.

Visschers, Ludo and Carlos Carrillo-Tudela, "Unemployment and endogenous reallocation over the business cycle," *Working Paper*, 2014. Universidad Carlos III de Madrid.

Ware, John E., Mark Kosinski, and Susan D. Keller, *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*, Health Institute, New England Medical Center, 1995.

Yamaguchi, Shintaro, "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 2012, *30* (1), 1–53.

Supplemental Online Appendix

NOT FOR PUBLICATION

Appendix

A Proofs and Derivations

A.1 Baseline Model with Depreciation

Proof. **Derivation of Human Capital Decision and Wage Equation**

Using

$$h_{j,t} = (1-\delta)\left(h_{j,t-1} + (A_j + \varepsilon_{j,t-1})\, r_{j,t-1} - r_{j,t-1}^2/2\right),$$

we obtain

$$h_{j,t} = (1-\delta)\left(h_{j,t-1} + \frac{A_j^2}{2} - \frac{(A_j - r_{j,t-1})^2}{2} + r_{j,t-1}\epsilon_{j,t-1}\right)$$

and repeatedly substituting for human capital, we obtain

$$h_{j,t} = (1-\delta)^{t-1} h_{j,1} + \sum_{s=1}^{t-1} (1-\delta)^{t-s} \left(\frac{A_j^2}{2} + \frac{(A_j - r_{j,s})^2}{2} + r_{j,s}\epsilon_{j,s}\right).$$

Inserting this expression into the following wage equation

$$w_t = \sum_j \left(h_{j,t} + \frac{A_j^2}{2} - \frac{(A_j - r_{j,t})^2}{2}\right) + \sum_j r_{j,t}\varepsilon_{j,t},$$

gives

$$\begin{aligned}
w_t &= (1-\delta)^{t-1} \sum_j h_{j,1} \\
&+ \sum_{s=1}^{t-1} (1-\delta)^{t-s} \sum_j \left(\frac{A_j^2}{2} - \frac{(A_j - r_{j,s})^2}{2} + r_{j,s}\epsilon_{j,s}\right) \\
&+ \sum_j \left(\frac{A_j^2}{2} - \frac{(A_j - r_{j,t})^2}{2}\right) + \sum_j r_{j,t}\varepsilon_{j,t} \\
&= (1-\delta)^{t-1} \sum_j h_{j,1} + \frac{1}{2}\left(\sum_j A_j^2\right)\left(\sum_{s=1}^{t} (1-\delta)^{t-s}\right) \\
&+ \frac{1}{2}\sum_{s=1}^{t} (1-\delta)^{t-s} \sum_j (A_j - r_{j,s})^2 \\
&+ \sum_{s=1}^{t} (1-\delta)^{t-s} \sum_j r_{j,s}\epsilon_{j,s} \qquad (11)
\end{aligned}$$

□

Setting the depreciation rate to zero, we would obtain:

$$h_{j,t} = h_{j,1} + \frac{A_j^2}{2}(t-1) - \sum_{s=1}^{t-1} \frac{(A_j - r_{j,s})^2}{2} + \sum_{s=1}^{t-1} r_{j,s}\epsilon_{j,s}$$

and

$$w_t = \sum_j h_{j,1} + \underbrace{\frac{1}{2}\left(\sum_j A_j^2\right) \times t}_{\text{ability}\times\text{experience}} - \underbrace{\frac{1}{2}\sum_{s=1}^{t}\sum_j (A_j - r_{j,s})^2}_{\text{mismatch}} + \sum_{s=1}^{t}\sum_j r_{j,s}\varepsilon_{j,s}.$$

Proof. (Proposition 1) We solve the worker's problem backwards:

$$V_t\left(\mathbf{h}_t, \hat{\mathbf{A}}_t\right) = \max_{\{r_{j,t}\}} E_t \quad \left[\sum_j \left(h_{j,t} + (A_j + \epsilon_{j,t})\, r_{j,t} - \frac{r_{j,t}^2}{2}\right)\right.$$
$$+ \qquad \left.\beta V_{t+1}\left(\mathbf{h}_{t+1}, \hat{\mathbf{A}}_{t+1}\right)\right]$$

subject to

$$\hat{A}_{j,t+1} = \frac{\lambda_{j,t}}{\lambda_{j,t+1}}\hat{A}_{j,t} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t+1}}(A_j + \epsilon_{j,t})$$

and

$$h_{j,t+1} = (1-\delta)\left(h_{j,t} + (A_j + \epsilon_{j,t})\, r_{j,t} - \frac{r_{j,t}^2}{2}\right) \text{ for all } j.$$

The worker's problem in the last period of his life is

$$V_T\left(\mathbf{h}_T, \hat{\mathbf{A}}_T\right) = \max_{\{r_{j,T}\}} E_T\left[\sum_j \left(h_{j,T} + (A_j + \epsilon_{j,T})\, r_{j,T} - \frac{r_{j,T}^2}{2}\right)\right],$$

which, due to linearity of the objective in A_j's, can be written as

$$V_T\left(\mathbf{h}_T, \hat{\mathbf{A}}_T\right) = \max_{\{r_j\}} \sum_j \left(h_{j,T} + \hat{A}_{j,T} r_{j,T} - \frac{r_{j,T}^2}{2}\right).$$

The optimal solution is the same as in the previous section

$$r_{j,T} = \hat{A}_{j,T}.$$

Substituting the solution, we obtain

$$V_T\left(\mathbf{h}_T, \hat{\mathbf{A}}_T\right) = \sum_j \left(h_{j,T} + \frac{\hat{A}_{j,T}{}^2}{2}\right).$$

Now look at the problem in period $T-1$:

$$\begin{aligned} V_{T-1}\left(\mathbf{h}_{T-1}, \hat{\mathbf{A}}_{T-1}\right) &= \max_{\{r_{j,T-1}\}} E_{T-1}\left[\sum_j \left(h_{j,T-1} + \left(A_j + \epsilon_{j,T-1}\right) r_{j,T-1} - \frac{r_{j,T-1}^2}{2}\right)\right. \\ &+ \left.\beta V_T\left(\mathbf{h_T}, \hat{\mathbf{A}}_\mathbf{T}\right)\right] \end{aligned}$$

subject to

$$\hat{A}_{j,T} = \frac{\lambda_{j,T-1}}{\lambda_{j,T}}\hat{A}_{j,T-1} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,T}}\left(A_j + \epsilon_{j,T-1}\right)$$

and

$$h_{j,T} = (1-\delta)\left(h_{j,T-1} + \left(A_j + \epsilon_{j,T-1}\right) r_{j,T-1} - \frac{r_{j,T-1}^2}{2}\right) \text{ for all } j.$$

Substituting the law of motion for human capital, we can write as

$$\begin{aligned} V_{T-1}\left(\mathbf{h}_{T-1}, \hat{\mathbf{A}}_{T-1}\right) &= \max_{\{r_{j,T-1}\}} E_{T-1}\left[\sum_j \left(h_{j,T-1} + \left(A_j + \epsilon_{j,T-1}\right) r_{j,T-1} - \frac{r_{j,T-1}^2}{2}\right)\right. \\ &+ \left.\beta \sum_j (1-\delta)\left(h_{j,T-1} + \left(A_j + \epsilon_{j,T-1}\right) r_{j,T-1} - \frac{r_{j,T-1}^2}{2}\right)\right] \\ &+ E_{T-1}\left[\beta \sum_j \hat{A}_{j,T}{}^2/2\right] \end{aligned}$$

$$\begin{aligned} V_{T-1}\left(\mathbf{h}_{T-1}, \hat{\mathbf{A}}_{T-1}\right) &= \max_{\{r_{j,T-1}\}} \left(1 + \beta\left(1-\delta\right)\right) \sum_j \left(h_{j,T-1} + \hat{A}_{j,T-1} r_{j,T-1} - \frac{r_{j,T-1}^2}{2}\right) \\ &+ E_{T-1}\left[\beta \sum_j \hat{A}_{j,T}^2/2\right]. \end{aligned}$$

The solution gives $r_{j,T-1} = \hat{A}_{j,T-1}$. And

$$\begin{aligned} V_{T-1}\left(\mathbf{h}_{T-1}, \hat{\mathbf{A}}_{T-1}\right) &= \left(1 + \beta\left(1-\delta\right)\right) \sum_j \left(h_{j,T-1} + \frac{\hat{A}_{j,T-1}^2}{2}\right) \\ &+ E_{T-1}\left[\beta \sum_j \hat{A}_{j,T}^2/2\right] \end{aligned}$$

Continuing backwards, we obtain

$$V_t(\mathbf{h}_t, \hat{\mathbf{A}}_t) = \Big(\sum_{s=t}^{T} \left(\beta\left(1-\delta\right)\right)^{s-t}\Big)\Big(\sum_{j=1}^{J}\left(h_{j,t} + \hat{A}_{j,t}^2/2\right)\Big) + B_t(\hat{\mathbf{A}}_t),$$

Since B_t only involves beliefs and does not depend on r_j this term does not affect the worker's decision rules. $\square$

Proof. (Lemma 1) Given the history $(A_j+\epsilon_{j,1}, A_j+\epsilon_{j,2}, ..., A_j+\epsilon_{j,t-1})$, the worker's belief at the beginning of period t is given as

$$\begin{aligned}\hat{A}_{j,t} &= \frac{\lambda_{j,1}}{\lambda_{j,t}}\hat{A}_{j,1} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(A_j+\epsilon_{j,1}+A_j+\epsilon_{j,2}+...+A_j+\epsilon_{j,t-1}\right)\\ &= \frac{\lambda_{j,1}}{\lambda_{j,t}}\hat{A}_{j,1} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}(t-1)A_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,t-1}\right),\end{aligned}$$

where

$$\lambda_{j,t} = \lambda_{j,1} + (t-1)\lambda_{\epsilon_j}.$$

Since $\hat{A}_{j,1}$ is normally distributed with $\mathcal{N}\left(A_j, \sigma^2_{\eta_j}+\sigma_{A_j^2}\right)$. Then, $\hat{A}_{j,t}$ will be normally distributed since

$$\begin{aligned}\hat{A}_{j,t} &= \frac{\lambda_{j,1}}{\lambda_{j,t}}\left(A_j+\eta_j\right) + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}(t-1)A_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,t-1}\right)\\ &= A_j + \frac{\lambda_{j,1}}{\lambda_{j,t}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,t-1}\right)\end{aligned}$$

From this expression, we obtain

$$\hat{A}_{j,t} - A_j = \frac{\lambda_{j,1}}{\lambda_{j,t}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,t-1}\right)$$

which implies that $E\left[\hat{A}_{j,t}-A_j\right]=0$. Inserting $\hat{A}_{j,t}-A_j$ from expression above into $M_{j,t} = \sum_{s=1}^{t}\frac{(A_j-r_{j,s})^2}{2}$, we obtain

$$M_{j,t} = \sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,s-1}\right)\right)^2$$

and

$$\Omega_{j,t} = \sum_{s=1}^{t}\left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,s-1}\right)\right)\epsilon_{j,s}.$$

Note that $\mathrm{Cov}\left(M_{j,t},\Omega_{j,t}\right) = E\left(M_{j,t}\Omega_{j,t}\right) - E\left(M_{j,t}\right)E\left(\Omega_{j,t}\right)$. Furthermore, $E\left(\Omega_{j,t}\right)=0$ since $\epsilon_{j,s}$ is uncorrelated with all the terms in $\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1}+\epsilon_{j,2}+...+\epsilon_{j,s-1}\right)$ and $E\left(\epsilon_{j,s}\right)=0$. Thus, $\mathrm{Cov}\left(M_{j,t},\Omega_{j,t}\right) = E\left(M_{j,t}\Omega_{j,t}\right)$. An important point to notice is that $M_{j,t}\Omega_{j,t}$ includes multiplication of $\epsilon_{j,s}$, ϵ_j^2, η_j, and η_j^2, and ${\epsilon_j}^3$. Note that both ϵ_j and η_j are normal with mean zero. And, for a normal random variable x with mean zero, $E\left(x^n\right)$ is zero if n is an odd number and positive if n is an even number. Thus, we have $E\left({\epsilon_{j,s}}^m\eta_j^n\right)=0$ if either m or n is an odd number and $E\left({\epsilon_{j,s}}^m\eta_j^n\right)>0$ if both m and n are even numbers. As a result, only terms that remain are the positive ones. Thus, $E\left(M_{j,t}\Omega_{j,t}\right)>0$. $\square$

Proof. (Lemma 2) Note that $\tilde{\Delta}_{j,t} = \left(A_j^2 + \nu_j^2 + 2A_j\nu_j\right)t$. Using

$$r_{j,s} \quad = \hat{A}_{j,s} = \quad A_j + \frac{\lambda_{j,1}}{\lambda_{j,t}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right),$$

we also obtain

$$\widetilde{M}_{j,t} = \sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right) - \nu_j\right)^2$$

and

$$\widetilde{\Omega}_{j,t} = \sum_{s=1}^{t}\left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right)\right)\left(\varepsilon_{j,s} - \nu_j\right).$$

First, note that $\mathrm{Cov}\left(\widetilde{\Delta}_{j,t}, \widetilde{\Omega}_{j,t}\right) = E\left[\widetilde{\Delta}_{j,t}\widetilde{\Omega}_{j,t}\right]$ since $E\left[\widetilde{\Omega}_{j,t}\right] = 0$. Using the fact that odd moments of the normal distribution are zero, we obtain

$$\mathrm{Cov}\left(\widetilde{\Delta}_{j,t}, \widetilde{\Omega}_{j,t}\right) \quad = \quad -A_j^3\nu_j t^2 - 2\left(A_j\nu_j t\right)^2 < 0.$$

Second, similarly $\mathrm{Cov}\left(\widetilde{M}_{j,t}, \widetilde{\Omega}_{j,t}\right) = E\left[\widetilde{M}_{j,t}\widetilde{\Omega}_{j,t}\right]$. Rewrite

$$\begin{aligned}
\widetilde{M}_{j,t} &= \sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right) - \nu_j\right)^2 \\
&= \sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right)\right)^2 \\
&- 2\nu_j\sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right)\right) \\
&+ \nu_j^2 t
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\Omega}_{j,t} &= \sum_{s=1}^{t}\left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right)\right)\varepsilon_{j,s} \\
&- \nu_j\sum_{s=1}^{t}\left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1}\right)\right).
\end{aligned}$$

Noting that

$$M_{j,t} = \sum_{s=1}^{t}\left(\frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1}\right) - \nu_j\right)^2$$

and

$$\Omega_{j,t} = \sum_{s=1}^{t}\left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}}\eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}}\left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1}\right)\right)\varepsilon_{j,s},$$

we can write

$$
\begin{aligned}
\widetilde{M}_{j,t}\widetilde{\Omega}_{j,t} &= M_{j,t}\Omega_{j,t} \\
&- \nu_j \left(\sum_{s=1}^{t} \left(\frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right)^2 \right) \left(\sum_{s=1}^{t} \left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right) \right) \\
&- 2\nu_j \left(\sum_{s=1}^{t} \left(\frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right) \right) \left(\sum_{s=1}^{t} \left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right) \varepsilon_{j,s} \right) \\
&+ 2\nu_j^2 \left(\sum_{s=1}^{t} \left(\frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right) \right) \left(\sum_{s=1}^{t} \left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \ldots + \epsilon_{j,s-1} \right) \right) \right) \\
&= \nu_j^2 t \sum_{s=1}^{t} \left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1} \right) \right) \varepsilon_{j,s} \\
&- \nu_j^3 t \sum_{s=1}^{t} \left(A_j + \frac{\lambda_{j,1}}{\lambda_{j,s}} \eta_j + \frac{\lambda_{\epsilon_j}}{\lambda_{j,s}} \left(\epsilon_{j,1} + \epsilon_{j,2} + \ldots + \epsilon_{j,s-1} \right) \right).
\end{aligned}
$$

Notice that in the expressions above, ν_j^n is multiplied by other random variables which are not correlated with ν_j^n. Thus, the expectations of all these expressions are zero. Then we are left with

$$
E\left[\widetilde{M}_{j,t}\widetilde{\Omega}_{j,t}\right] = E\left[M_{j,t}\Omega_{j,t}\right].
$$

We already know from the proof of Lemma 1 that $E\left[M_{j,t}\Omega_{j,t}\right] > 0$. □

Proof. (Lemma 3)

$$
\begin{aligned}
Var\left[\hat{A}_{j,t} - A_j\right] &= \frac{\lambda_{j,1}^2}{\lambda_{j,t}{}^2}\sigma_{\eta_j}^2 + \frac{\lambda_{\epsilon_j}^2}{\lambda_{j,t}{}^2}(t-1)\,\sigma_{\epsilon_j}^2 \\
&= \frac{\lambda_{j,1}}{\lambda_{j,t}{}^2} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}{}^2}(t-1) \\
&= \frac{1}{\lambda_{j,t}},
\end{aligned}
$$

Since $E\left[\hat{A}_{j,t} - A_j\right] = 0$, $E\left[\left(\hat{A}_{j,t} - A_j\right)^2\right] = Var\left[\hat{A}_{j,t} - A_j\right] + \left(E\left[\hat{A}_{j,t} - A_j\right]\right)^2 = Var\left[\hat{A}_{j,t} - A_j\right] = \frac{1}{\lambda_j(t)}$. Note that $\lambda_{j,t} = \lambda_{j,1} + (t-1)\,\lambda_{\epsilon_j}$ increases with experience. Thus, $E\left[\left(\hat{A}_{j,t} - A_j\right)^2\right]$, which is equal to $E\left[\left(r_{j,t} - A_j\right)^2\right]$, declines with age. □

Proof. (Proposition 2) Note that the probability of switching occupation in period t is given by

$$
\Pr\left(\mathbf{r_t} \neq \mathbf{r_{t-1}}\right) = 1 - \prod_j \text{Prob}\left(r_{j,t-1} - \kappa_j \leq r_{j,t} < r_{j,t-1} + \kappa_j\right)
$$

$$= 1 - \prod_j \text{Prob}\left(\hat{A}_{j,t-1} - \kappa_j \leq \hat{A}_{j,t} < \hat{A}_{j,t-1} + \kappa_j\right).$$

Now look at the term $\text{Prob}\left(\hat{A}_{j,t-1} - \kappa_j \leq \hat{A}_{j,t} < \hat{A}_{j,t-1} + \kappa_j\right)$. Inserting

$$\hat{A}_{j,t} = \frac{\lambda_{j,t-1}}{\lambda_{j,t}} \hat{A}_{j,t-1} + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}} \left(A_j + \epsilon_{j,t-1}\right)$$

and

$$\lambda_{j,t} = \lambda_{j,t-1} + \lambda_{\epsilon_j},$$

we obtain

$$\begin{aligned}
& \text{Prob}\left(\hat{A}_{j,t-1} - \kappa_j \leq \hat{A}_{j,t} < \hat{A}_{j,t-1} + \kappa_j\right) \\
= \; & \text{Prob}\left(-\kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} \leq \left(A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}\right) < \kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}}\right) \\
= \; & \text{Prob}\left(-\kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - \left(A_j - \hat{A}_{j,t-1}\right) \leq \epsilon_{j,t-1} < \kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - \left(A_j - \hat{A}_{j,t-1}\right)\right).
\end{aligned}$$

Letting F be the cumulative distribution function of a normal variable with mean zero, and noting that normal distribution with mean zero is symmetric around zero, $F(M + x) - F(-M + x)$ declines with $|x|$. Since x^2 and $|x|$ move in the same direction, probability of staying in an occupation declines with mismatch $\left(A_j - \hat{A}_{j,t-1}\right)^2$.

Now evaluate the unconditional probability of staying in an occupation:

$$\begin{aligned}
& \text{Prob}\left(\hat{A}_{j,t-1} - \kappa_j \leq \hat{A}_{j,t} < \hat{A}_{j,t-1} + \kappa_j\right) \\
= \; & \text{Prob}\left(-\kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} \leq \left(A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}\right) < \kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}}\right) \\
= \; & \text{Prob}\left(-\kappa_j \sqrt{\frac{\lambda_{j,t-1}\lambda_{j,t}}{\lambda_{\epsilon_j}}} \leq \frac{A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}}{\sqrt{\frac{\lambda_{j,t}}{\lambda_{j,t-1}\lambda_{\epsilon}}}} < \kappa_j \sqrt{\frac{\lambda_{j,t-1}\lambda_{j,t}}{\lambda_{\epsilon_j}}}\right)
\end{aligned}$$

Remember that $A_j - \hat{A}_{j,t-1}$ is normally distributed with mean zero and variance $1/\lambda_{j,t-1}$ (see proof of Lemma 3). Thus, $\left(A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}\right)$ is normally distributed with mean zero and variance $\frac{1}{\lambda_{j,t-1}} + \frac{1}{\lambda_\epsilon} = \frac{\lambda_{j,t}}{\lambda_{j,t-1}\lambda_\epsilon}$. Thus, $\left(A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}\right) \times \sqrt{\frac{\lambda_{j,t-1}\lambda_\epsilon}{\lambda_{j,t}}}$ is normally distributed with mean zero and variance one. Since both $\lambda_{j,t}$ and $\lambda_{j,t-1}$ increases with age, probability of staying in the last period's occupation increases and probability of switching decreases with age. $\square$

Proof. (Lemma 4) Probability of switching to an occupation with higher skill-j intensity is

$$\begin{aligned}
\pi_{j,t}^{up} &= \text{Prob}\left(\hat{A}_{j,t} > \hat{A}_{j,t-1} + \kappa_j\right) \\
&= \text{Prob}\left(\epsilon_{j,t} > \kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - \left(A_j - \hat{A}_{j,t-1}\right)\right) \\
&= \text{Prob}\left(\epsilon_{j,t} > \kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - (A_j - r_{j,t-1})\right).
\end{aligned}$$

Note that probability of switching to an occupation with higher skill-j intensity increases with $(A_j - r_{j,t-1})$. Thus, to the extent that the worker is overqualified, he will switch to a higher skill occupation. The probability of switching to a lower skill occupation is given by

$$\begin{aligned}
\pi_{j,t}^{down} &= \text{Prob}\left(A_{j,t} < A_{j,t-1} - \kappa_j\right) \\
&= \text{Prob}\left(\epsilon_{j,t} < -\kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - \left(A_j - \hat{A}_{j,t-1}\right)\right) \\
&= \text{Prob}\left(\epsilon_{j,t} < -\kappa_j \frac{\lambda_{j,t}}{\lambda_{\epsilon_j}} - (A_j - r_{j,t-1})\right).
\end{aligned}$$

Using these two equations above, it is easy to observe that $\pi_{j,t}^{up} > \pi_{j,t}^{down}$ when $A_j - r_{j,t-1} > 0$. Similar proof can be made for the case $r_{j,t-1} - A_j > 0$. But we skip it for the sake of brevity. □

Proof. (Proposition 3) Using equation 8, we obtain

$$\hat{A}_{j,t} - \hat{A}_{j,t-1} = \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\left(A_j - \hat{A}_{j,t-1} + \epsilon_{j,t-1}\right).$$

Using $r_{j,t} = \hat{A}_{j,t}$ implies

$$r_{j,t} - r_{j,t-1} = \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}(A_j - r_{j,t-1}) + \frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\epsilon_{j,t-1}.$$

Thus, conditional on $A_j - r_{j,t-1}$, $r_{j,t} - r_{j,t-1}$ is normally distributed with mean $\frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}(A_j - r_{j,t-1})$ and variance $\left(\frac{\lambda_{\epsilon_j}}{\lambda_{j,t}}\right)^2 \sigma_{\epsilon_j}^2$. □

A.2 Baseline Model vs. Ben-Porath Version

In the baseline model of Section 2, the occupation/human capital choice problem is a static one. To see this most clearly, we consider a simplified version of the model with a single skill and where ability is observed. In this model, the lifetime problem not only reduces to a static one, but the decision rule does not change over time. Next we show that the Ben-Porath version of the same model features an occupation/human capital choice that changes every period, underscoring the dynamic nature of the decision. The following derivations establish these results. These results extend straightforwardly to the case with multiple skills and Bayesian learning at the expense of much more complicated algebra. (These results are available upon request).

Baseline Model with No Learning

Let us write the problem of the individual sequentially, starting from the last period of life.

$$V_T\left(h_T\right) = \max_{r_T}\left(h_T + Ar_T - \frac{r_T^2}{2}\right).$$

Now insert this into the following problem

$$V_{T-1}\left(h_{T-1}\right) \quad = \quad \max_{r_{T-1}}\left(h_{T-1} + Ar_{T-1} - \frac{r_{T-1}^2}{2}\right) + \beta\left(h_T + Ar_T - \frac{r_T^2}{2}\right)$$

s.t.

$$h_{j,T} = (1-\delta)\left(h_{j,T-1} + Ar_{T-1} - \frac{r_{T-1}^2}{2}\right).$$

Inserting the law of motion into the objective function, we obtain

$$V_{T-1}\left(h_{T-1}\right) \quad = \quad \max_{r_{T-1}}\left(1+\beta(1-\delta)\right)\left(h_{T-1} + Ar_{T-1} - \frac{r_{T-1}^2}{2}\right) + \beta\left(Ar_T - \frac{r_T^2}{2}\right)$$

The problem in period $T-1$ does not depend on any period-T variables or affect any decision in period T. This is because the occupation choice/human capital accumulation decision does not depend on the stock of human capital—it only depends on the workers' ability level, which does not change over time. Consequently, the decision rule in period $T-1$ would be the same if we just ignored period T and just solved

$$\max_{r_{T-1}}\left(h_{T-1} + Ar_{T-1} - \frac{r_{T-1}^2}{2}\right).$$

It can be shown that the solution is the same in all periods and given by

$$r_t = A \qquad \forall t. \tag{12}$$

Ben-Porath with No Learning

In Ben-Porath the wage equation is the same as in our baseline model (with a single skill): $w_t = h_t - \frac{r_t^2}{2}$. The lifetime maximization problem is

$$\max_{\{r_t\}_{t=1}^T} \mathbb{E}_0 \left[\sum_{t=1}^{T} \beta^{t-1} (h_t - \frac{r_t^2}{2}) \right]$$
$$\text{s.t.}$$
$$h_{t+1} = (1-\delta)(h_t + A r_t)$$

This model corresponds to the standard Ben-Porath formulation with a quadratic cost function. Following the same steps as above, the solution can be shown to be:

$$r_t = \sum_{s=1}^{T-t} (\beta(1-\delta))^s A,$$

which unlike the solution in (12) is not constant and changes every period. This is due to the intertemporal trade-off noted above inherent in the Ben-Porath model.

Mismatch in the Ben-Porath Model

Note that both positive and negative past mismatch reduces current wage in our model. Ben-Porath model on the other hand has different implications for positive and negative past mismatch. In particular, it implies that positive past mismatch (worker being over-qualified in past occupations) reduces the current wage and negative past mismatch (worker being under-qualified in past occupations) increases the current wage. Abstracting from multi-dimensions for simplicity, note that the optimal occupational choice of a worker in period t under perfect information, denoted by r_t^*, is given by

$$r_t^* = \sum_{s=1}^{T-t} (\beta(1-\delta))^s A.$$

Under Bayesian Learning, assuming the same information structure as in our model, the worker's optimal choice is given by

$$r_t = \sum_{s=1}^{T-t} (\beta(1-\delta))^s \hat{A}_t,$$

where $\hat{A}_t$ is the mean belief of worker's ability in period t.

Substituting $h_t = (1-\delta)(h_{t-1} + (A + \epsilon_{t-1}) r_{t-1})$ repeatedly into $w_t = h_t - \frac{r_t^2}{2}$ and assuming zero depreciation gives

$$w_t = h_1 + A \sum_{s=1}^{t-1} r_s + \sum_{s=1}^{t-1} \epsilon_s r_s - \frac{r_t^2}{2}.$$

By adding and subtracting $A\sum_{s=1}^{t-1} r_s^*$, we obtain the following wage equation

$$w_t = h_1 + A\sum_{s=1}^{t-1} r_s^* - A\sum_{s=1}^{t-1}(r_s^* - r_s) + \sum_{s=1}^{t-1}\epsilon_s r_s - \frac{r_t^2}{2}.$$

Note here that a worker would be overqualified (underqualified) in his occupation in period s if $r_s^* > r_s$ ($r_s^* < r_s$), i.e. his occupation's skill requirement is below (above) his ideal occupation's skill requirement. Thus, if a worker is overqualified in the past, he earns lower wages today as in our model. Key difference comes when a worker is underqualified in the past. In our model, he still earns less wages today. But in the Ben-Porath model, he earns more wages since he would be accumulating more human capital than he would in his ideal occupation.

B Additional Tables

B.1 Descriptive Statistics for Mismatch Measure

Table B.1 – Descriptive Statistics for Mismatch Measure

	Mismatch	
Group Name	Mean	Std. Dev.
All Observations	1.56	1
By Educational Group		
Less than High School	1.50	1.13
High School	1.63	1.00
Some College	1.52	0.96
By Race		
Hispanic	1.61	1.05
Black	1.52	1.02
Non-Black, Non-Hispanic	1.56	0.99
By Industry		
Agriculture, Forestry, Fisheries	1.59	1.15
Mining	1.63	1.02
Construction	1.70	1.13
Manufacturing	1.57	1.00
Transportation, Communications, Util.	1.47	0.90
Wholesale and Retail Trade	1.57	0.93
Finance, Insurance and Real Estate	1.37	0.86
Business and Repair Services	1.63	0.99
Personal Services	1.68	1.08
Entertainment and Recreation Services	1.92	1.20
Professional and Related Services	1.46	0.94
Public Administration	1.49	0.97

B.2 Regression Tables

Table B.2 – Wage Regressions with Mismatch (Full Results)

	(1)	(2)	(3)	(4)	(5)	(6)
	IV-GLS	IV-GLS	IV-GLS	GLS	GLS	GLS
Mismatch	-0.0205***	-0.0091*	-0.0054	-0.0196***	-0.0136***	-0.0128**
	(0.0031)	(0.0052)	(0.0062)	(0.0030)	(0.0042)	(0.0050)
Mismatch × Occ Tenure		-0.0020***	-0.0022**		-0.0010**	-0.0007
		(0.0007)	(0.0009)		(0.0005)	(0.0007)
Cumul Mismatch			-0.0374***			-0.0383***
			(0.0043)			(0.0042)
Worker Ability (Mean)	0.3015***	0.3044***	0.4095***	0.3094***	0.3101***	0.3987***
	(0.0252)	(0.0253)	(0.0303)	(0.0208)	(0.0208)	(0.0252)
Worker Ability × Occ Tenure	0.0115***	0.0109***	0.0072**	0.0090***	0.0088***	0.0081***
	(0.0028)	(0.0028)	(0.0036)	(0.0019)	(0.0019)	(0.0026)
Occ Reqs (Mean)	0.0977***	0.0962***	0.0878***	0.1383***	0.1374***	0.1426***
	(0.0217)	(0.0217)	(0.0254)	(0.0188)	(0.0188)	(0.0221)
Occ Reqs × Occ Tenure	0.0144***	0.0144***	0.0182***	0.0084***	0.0084***	0.0092***
	(0.0027)	(0.0027)	(0.0034)	(0.0019)	(0.0019)	(0.0024)
Emp Tenure	-0.0107***	-0.0106**	-0.0081	0.0034	0.0034	0.0062
	(0.0041)	(0.0041)	(0.0054)	(0.0033)	(0.0033)	(0.0044)
Emp Tenure^2 × 100	0.0572***	0.0571***	0.0538*	-0.0138	-0.0136	-0.0255
	(0.0219)	(0.0219)	(0.0322)	(0.0184)	(0.0184)	(0.0274)
Occ Tenure	0.0014	0.0050	0.0052	0.0288***	0.0307***	0.0299***
	(0.0049)	(0.0051)	(0.0065)	(0.0039)	(0.0040)	(0.0051)
Occ Tenure^2 × 100	-0.0812**	-0.0841**	-0.0963*	-0.1388***	-0.1416***	-0.1302***
	(0.0414)	(0.0414)	(0.0566)	(0.0353)	(0.0353)	(0.0486)
Occ Tenure^3 × 100	0.0012	0.0013	0.0021	0.0018**	0.0019**	0.0017
	(0.0010)	(0.0010)	(0.0015)	(0.0009)	(0.0009)	(0.0013)
Experience	0.0550***	0.0551***	0.0431***	0.0402***	0.0402***	0.0308***
	(0.0055)	(0.0055)	(0.0094)	(0.0052)	(0.0052)	(0.0085)
Experience^2 × 100	-0.1427***	-0.1434***	-0.0707	-0.0917**	-0.0921**	-0.0380
	(0.0375)	(0.0375)	(0.0574)	(0.0358)	(0.0358)	(0.0529)
Experience^3 × 100	0.0013*	0.0013*	0.0001	0.0006	0.0006	-0.0003
	(0.0008)	(0.0008)	(0.0011)	(0.0007)	(0.0007)	(0.0010)
Old Job	-0.0328***	-0.0328***	-0.0319***	-0.0326***	-0.0326***	-0.0308***
	(0.0064)	(0.0064)	(0.0076)	(0.0060)	(0.0060)	(0.0071)
< High School	-0.0744***	-0.0752***	-0.0678***	-0.0643***	-0.0648***	-0.0571***
	(0.0123)	(0.0123)	(0.0148)	(0.0121)	(0.0121)	(0.0145)
4-Year College	0.2642***	0.2639***	0.2321***	0.2519***	0.2518***	0.2220***
	(0.0097)	(0.0097)	(0.0117)	(0.0096)	(0.0096)	(0.0115)
Hispanic	0.0060	0.0059	-0.0033	0.0060	0.0059	-0.0009
	(0.0141)	(0.0141)	(0.0168)	(0.0139)	(0.0139)	(0.0166)
Black	-0.0700***	-0.0697***	-0.0815***	-0.0674***	-0.0673***	-0.0788***
	(0.0120)	(0.0120)	(0.0141)	(0.0118)	(0.0118)	(0.0139)

Constant	4.0074***	3.9945***	4.0332***	4.0050***	3.9988***	4.0429***
	(0.0255)	(0.0259)	(0.0375)	(0.0246)	(0.0248)	(0.0348)
Observations	41596	41596	30599	41596	41596	30599

All regressions include occupation and industry dummies.

$***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$.

Standard errors in parentheses are estimated via GLS assuming AR(1) autocorrelation.

Table B.3 – Wage Regressions with Mismatch by Components (Full Results)

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch Verbal	-0.0103***	0.0082	0.0133	-0.0104***	0.0032	0.0051
	(0.0040)	(0.0069)	(0.0081)	(0.0039)	(0.0056)	(0.0067)
Mismatch Math	-0.0108***	-0.0172**	-0.0193**	-0.0096**	-0.0178***	-0.0193***
	(0.0040)	(0.0068)	(0.0081)	(0.0039)	(0.0055)	(0.0066)
Mismatch Social	-0.0027	0.0005	0.0055	-0.0021	-0.0012	0.0034
	(0.0031)	(0.0054)	(0.0064)	(0.0031)	(0.0043)	(0.0052)
Mismatch Verbal × Occ Tenure		-0.0031***	-0.0043***		-0.0022***	-0.0028***
		(0.0009)	(0.0012)		(0.0006)	(0.0008)
Mismatch Math × Occ Tenure		0.0010	0.0020*		0.0013**	0.0021**
		(0.0009)	(0.0012)		(0.0007)	(0.0008)
Mismatch Social × Occ Tenure		-0.0005	-0.0007		-0.0001	-0.0006
		(0.0007)	(0.0009)		(0.0005)	(0.0006)
Cumul Mismatch Verbal			-0.0148***			-0.0137**
			(0.0057)			(0.0056)
Cumul Mismatch Math			-0.0241***			-0.0260***
			(0.0055)			(0.0054)
Cumul Mismatch Social			-0.0088**			-0.0078*
			(0.0043)			(0.0042)
Verbal Ability	-0.0622*	-0.0652*	0.0029	-0.0132	-0.0181	-0.0006
	(0.0367)	(0.0367)	(0.0439)	(0.0298)	(0.0298)	(0.0361)
Math Ability	0.3660***	0.3711***	0.4260***	0.3209***	0.3254***	0.4027***
	(0.0371)	(0.0371)	(0.0443)	(0.0303)	(0.0303)	(0.0363)
Social Ability	0.0997***	0.1003***	0.1078***	0.1046***	0.1046***	0.1230***
	(0.0224)	(0.0224)	(0.0268)	(0.0181)	(0.0181)	(0.0218)
Verbal Ability × Occ Tenure	0.0128***	0.0129***	0.0083	0.0066**	0.0070**	0.0083**
	(0.0043)	(0.0044)	(0.0055)	(0.0031)	(0.0031)	(0.0040)
Math Ability × Occ Tenure	-0.0051	-0.0060	-0.0066	-0.0016	-0.0024	-0.0045
	(0.0044)	(0.0044)	(0.0054)	(0.0032)	(0.0032)	(0.0039)
Social Ability × Occ Tenure	0.0061**	0.0060**	0.0071**	0.0059***	0.0060***	0.0056**
	(0.0027)	(0.0027)	(0.0034)	(0.0019)	(0.0019)	(0.0025)
Occ Reqs Verbal	0.0190	0.0209	-0.0420	0.0761	0.0832	0.0211
	(0.0742)	(0.0747)	(0.0910)	(0.0608)	(0.0610)	(0.0747)
Occ Reqs Math	0.1313*	0.1244*	0.1772**	0.1176**	0.1080*	0.1731**
	(0.0681)	(0.0686)	(0.0837)	(0.0557)	(0.0560)	(0.0685)
Occ Reqs Social	-0.1115***	-0.1084***	-0.1014***	-0.1064***	-0.1047***	-0.0966***
	(0.0248)	(0.0248)	(0.0295)	(0.0208)	(0.0208)	(0.0248)
Occ Reqs Verbal × Occ Tenure	-0.0055	-0.0061	0.0034	-0.0184***	-0.0202***	-0.0123
	(0.0101)	(0.0103)	(0.0138)	(0.0070)	(0.0071)	(0.0096)
Occ Reqs Math × Occ Tenure	0.0138	0.0153	0.0077	0.0189***	0.0213***	0.0120
	(0.0092)	(0.0094)	(0.0125)	(0.0064)	(0.0065)	(0.0087)
Occ Reqs Social × Occ Tenure	0.0095***	0.0088***	0.0115***	0.0112***	0.0108***	0.0134***
	(0.0032)	(0.0032)	(0.0042)	(0.0022)	(0.0022)	(0.0030)

Emp Tenure	-0.0108***	-0.0108***	-0.0075	0.0036	0.0035	0.0065
	(0.0041)	(0.0041)	(0.0054)	(0.0033)	(0.0033)	(0.0044)
Emp Tenure$^2\times$ 100	0.0576***	0.0580***	0.0497	-0.0166	-0.0157	-0.0264
	(0.0219)	(0.0219)	(0.0322)	(0.0184)	(0.0184)	(0.0274)
Occ Tenure	-0.0024	0.0015	0.0014	0.0244***	0.0262***	0.0265***
	(0.0051)	(0.0053)	(0.0067)	(0.0040)	(0.0041)	(0.0053)
Occ Tenure$^2\times$ 100	-0.0761*	-0.0798*	-0.0861	-0.1244***	-0.1277***	-0.1209**
	(0.0413)	(0.0414)	(0.0567)	(0.0353)	(0.0353)	(0.0487)
Occ Tenure$^3\times$ 100	0.0012	0.0013	0.0019	0.0015*	0.0016*	0.0016
	(0.0010)	(0.0010)	(0.0015)	(0.0009)	(0.0009)	(0.0013)
Experience	0.0568***	0.0568***	0.0439***	0.0420***	0.0420***	0.0323***
	(0.0055)	(0.0055)	(0.0094)	(0.0052)	(0.0052)	(0.0085)
Experience$^2\times$ 100	-0.1516***	-0.1518***	-0.0749	-0.1019***	-0.1021***	-0.0481
	(0.0374)	(0.0374)	(0.0572)	(0.0357)	(0.0357)	(0.0528)
Experience$^3\times$ 100	0.0015**	0.0015**	0.0001	0.0008	0.0008	-0.0001
	(0.0007)	(0.0007)	(0.0011)	(0.0007)	(0.0007)	(0.0010)
Old Job	-0.0326***	-0.0326***	-0.0325***	-0.0324***	-0.0324***	-0.0313***
	(0.0064)	(0.0064)	(0.0076)	(0.0059)	(0.0059)	(0.0070)
< High School	-0.0615***	-0.0620***	-0.0506***	-0.0524***	-0.0525***	-0.0418***
	(0.0125)	(0.0125)	(0.0150)	(0.0123)	(0.0123)	(0.0147)
4-Year College	0.2578***	0.2580***	0.2239***	0.2466***	0.2469***	0.2146***
	(0.0098)	(0.0098)	(0.0119)	(0.0097)	(0.0097)	(0.0117)
Hispanic	0.0078	0.0088	0.0039	0.0072	0.0082	0.0050
	(0.0142)	(0.0142)	(0.0169)	(0.0139)	(0.0139)	(0.0166)
Black	-0.0626***	-0.0619***	-0.0690***	-0.0616***	-0.0611***	-0.0686***
	(0.0123)	(0.0123)	(0.0145)	(0.0121)	(0.0121)	(0.0143)
Constant	4.0006***	3.9868***	4.0175***	3.9958***	3.9907***	4.0237***
	(0.0266)	(0.0272)	(0.0388)	(0.0256)	(0.0258)	(0.0358)
Observations	41596	41596	30599	41596	41596	30599

All regressions include occupation and industry dummies.

$***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$.

Standard errors in parentheses are estimated via GLS assuming AR(1) autocorrelation.

Table B.4 – Wage Regressions with Positive and Negative Mismatch (Full Results)

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Positive Mismatch	-0.0072*	0.0087	0.0141*	-0.0074**	-0.0015	0.0010
	(0.0038)	(0.0063)	(0.0077)	(0.0037)	(0.0051)	(0.0063)
Negative Negative	0.0305***	0.0172***	0.0231***	0.0276***	0.0137***	0.0195***
	(0.0037)	(0.0061)	(0.0071)	(0.0037)	(0.0050)	(0.0059)
Positive Mismatch × Occ Tenure		-0.0028***	-0.0032***		-0.0011*	-0.0005
		(0.0009)	(0.0011)		(0.0006)	(0.0008)
Negative Mismatch × Occ Tenure		0.0023***	0.0014		0.0024***	0.0022***
		(0.0008)	(0.0011)		(0.0006)	(0.0008)

Cumul Positive Mismatch			-0.0145***			-0.0220***
			(0.0054)			(0.0053)
Cumul Negative Mismatch			0.0159***			0.0085*
			(0.0051)			(0.0050)
Emp Tenure	-0.0106**	-0.0105**	-0.0081	0.0034	0.0032	0.0054
	(0.0042)	(0.0042)	(0.0055)	(0.0034)	(0.0034)	(0.0045)
Emp Tenure$^2 \times$ 100	0.0556**	0.0554**	0.0521	-0.0131	-0.0110	-0.0146
	(0.0221)	(0.0221)	(0.0325)	(0.0186)	(0.0186)	(0.0277)
Occ Tenure	0.0123**	0.0165***	0.0156**	0.0350***	0.0384***	0.0391***
	(0.0048)	(0.0050)	(0.0063)	(0.0039)	(0.0040)	(0.0051)
Occ Tenure$^2 \times$ 100	-0.0513	-0.0552	-0.0695	-0.1115***	-0.1188***	-0.1288***
	(0.0417)	(0.0417)	(0.0573)	(0.0356)	(0.0356)	(0.0492)
Occ Tenure$^3 \times$ 100	0.0006	0.0007	0.0014	0.0013	0.0014	0.0019
	(0.0010)	(0.0010)	(0.0015)	(0.0009)	(0.0009)	(0.0013)
Experience	0.0527***	0.0529***	0.0422***	0.0394***	0.0393***	0.0312***
	(0.0056)	(0.0056)	(0.0096)	(0.0053)	(0.0053)	(0.0087)
Experience$^2 \times$ 100	-0.1321***	-0.1333***	-0.0669	-0.0910**	-0.0911**	-0.0463
	(0.0381)	(0.0381)	(0.0585)	(0.0364)	(0.0364)	(0.0539)
Experience$^3 \times$ 100	0.0011	0.0012	0.0000	0.0006	0.0007	-0.0001
	(0.0008)	(0.0008)	(0.0011)	(0.0007)	(0.0007)	(0.0010)
Old Job	-0.0336***	-0.0337***	-0.0327***	-0.0331***	-0.0332***	-0.0314***
	(0.0065)	(0.0065)	(0.0077)	(0.0060)	(0.0060)	(0.0071)
< High School	-0.1390***	-0.1397***	-0.1424***	-0.1303***	-0.1311***	-0.1380***
	(0.0122)	(0.0122)	(0.0148)	(0.0121)	(0.0121)	(0.0146)
4-Year College	0.3421***	0.3413***	0.3183***	0.3285***	0.3273***	0.3067***
	(0.0095)	(0.0095)	(0.0115)	(0.0094)	(0.0094)	(0.0113)
Hispanic	-0.0321**	-0.0319**	-0.0414**	-0.0313**	-0.0314**	-0.0417**
	(0.0143)	(0.0143)	(0.0171)	(0.0141)	(0.0141)	(0.0169)
Black	-0.1389***	-0.1382***	-0.1591***	-0.1365***	-0.1359***	-0.1609***
	(0.0119)	(0.0119)	(0.0143)	(0.0118)	(0.0118)	(0.0141)
Constant	4.1698***	4.1533***	4.1903***	4.1751***	4.1653***	4.2080***
	(0.0233)	(0.0237)	(0.0362)	(0.0227)	(0.0229)	(0.0336)
Observations	41596	41596	30599	41596	41596	30599

All regressions include occupation and industry dummies.

$***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$.

Standard errors in parentheses are estimated via GLS assuming AR(1) autocorrelation.

Table B.5 – Regressions for Probability of Occupational Switch (Full Results)

	(1) LPM-IV	(2) LPM-IV	(3) LPM-IV	(4) LPM	(5) LPM	(6) LPM
Mismatch	0.0135***			0.0066***		
	(0.0023)			(0.0018)		
Mismatch Verbal		0.0076***			0.0053**	
		(0.0028)			(0.0022)	
Mismatch Math		0.0074***			0.0025	
		(0.0028)			(0.0022)	
Mismatch Social		0.0007			-0.0012	
		(0.0022)			(0.0017)	
Positive Mismatch			0.0134***			0.0087***
			(0.0027)			(0.0022)
Negative Mismatch			-0.0130***			-0.0028
			(0.0027)			(0.0021)
Worker Ability (Mean)	-0.0370***	-0.0370***		-0.0208*	-0.0211*	
	(0.0134)	(0.0134)		(0.0116)	(0.0116)	
Worker Ability × Occ Tenure	-0.0003	-0.0003		0.0019**	0.0020**	
	(0.0018)	(0.0018)		(0.0009)	(0.0009)	
Occ Reqs (Mean)	-0.0333**	-0.0334**		-0.1225***	-0.1223***	
	(0.0146)	(0.0146)		(0.0125)	(0.0125)	
Occ Reqs × Occ Tenure	-0.0052***	-0.0052***		0.0106***	0.0106***	
	(0.0018)	(0.0018)		(0.0009)	(0.0009)	
Emp Tenure	0.0039	0.0039	0.0040	-0.0064***	-0.0063***	-0.0071***
	(0.0029)	(0.0029)	(0.0029)	(0.0017)	(0.0017)	(0.0017)
Emp Tenure^2× 100	-0.0189	-0.0189	-0.0192	0.0301***	0.0300***	0.0327***
	(0.0161)	(0.0161)	(0.0161)	(0.0090)	(0.0090)	(0.0090)
Occ Tenure	0.0996***	0.0996***	0.0966***	-0.0547***	-0.0547***	-0.0495***
	(0.0038)	(0.0038)	(0.0037)	(0.0023)	(0.0023)	(0.0022)
Occ Tenure^2× 100	-0.5452***	-0.5453***	-0.5474***	0.3248***	0.3248***	0.3406***
	(0.0365)	(0.0365)	(0.0363)	(0.0185)	(0.0185)	(0.0184)
Occ Tenure^3× 100	0.0122***	0.0122***	0.0122***	-0.0066***	-0.0066***	-0.0070***
	(0.0010)	(0.0010)	(0.0010)	(0.0004)	(0.0004)	(0.0004)
Experience	-0.0626***	-0.0625***	-0.0619***	0.0132***	0.0133***	0.0134***
	(0.0028)	(0.0028)	(0.0028)	(0.0024)	(0.0024)	(0.0024)
Experience^2× 100	0.2355***	0.2355***	0.2312***	-0.1297***	-0.1299***	-0.1312***
	(0.0212)	(0.0212)	(0.0211)	(0.0171)	(0.0171)	(0.0172)
Experience^3× 100	-0.0036***	-0.0036***	-0.0035***	0.0027***	0.0027***	0.0027***
	(0.0005)	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0004)
Old Job	0.1506***	0.1506***	0.1508***	0.0069	0.0069	0.0063
	(0.0056)	(0.0056)	(0.0056)	(0.0051)	(0.0051)	(0.0051)
< High School	0.0413***	0.0414***	0.0505***	0.0090	0.0091	0.0180***
	(0.0087)	(0.0087)	(0.0085)	(0.0070)	(0.0070)	(0.0069)
4-Year College	-0.0433***	-0.0435***	-0.0551***	-0.0040	-0.0042	-0.0093**
	(0.0060)	(0.0060)	(0.0059)	(0.0047)	(0.0047)	(0.0046)

Hispanic	0.0044	0.0044	0.0104	0.0068	0.0067	0.0092
	(0.0095)	(0.0095)	(0.0094)	(0.0074)	(0.0075)	(0.0074)
Black	0.0108	0.0107	0.0207***	0.0029	0.0027	0.0104*
	(0.0080)	(0.0080)	(0.0079)	(0.0064)	(0.0064)	(0.0063)
Constant	0.1474***	0.1486***	0.0871***	0.3702***	0.3727***	0.2918***
	(0.0250)	(0.0252)	(0.0221)	(0.0206)	(0.0208)	(0.0182)
Observations	41,596	41,596	41,596	41,596	41,596	41,596

All regressions include occupation and industry dummies.

Robust standard errors in parentheses. $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$.

Table B.6 – Regressions for Direction of Occupational Switch (Full Results)

	(1) Average	(2) Verbal	(3) Math	(4) Social
Last Mismatch Positive	0.0751***			
	(0.0028)			
Last Mismatch Negative	0.1143***			
	(0.0030)			
Last Pos. Mismatch, Verbal		0.0316***	0.0097**	0.0143***
		(0.0043)	(0.0046)	(0.0046)
Last Neg. Mismatch, Verbal		0.0838***	0.0536***	0.0216***
		(0.0053)	(0.0056)	(0.0057)
Last Pos. Mismatch, Math		0.0599***	0.0898***	0.0021
		(0.0044)	(0.0047)	(0.0048)
Last Neg. Mismatch, Math		0.0558***	0.0893***	0.0076
		(0.0050)	(0.0053)	(0.0054)
Last Pos. Mismatch, Social		0.0061*	0.0046	0.0774***
		(0.0033)	(0.0035)	(0.0035)
Last Neg. Mismatch, Social		0.0264***	0.0166***	0.1043***
		(0.0037)	(0.0039)	(0.0039)
Employer Tenure	-0.0044	-0.0034	-0.0099*	-0.0055
	(0.0053)	(0.0056)	(0.0057)	(0.0046)
Employer Tenure$^2\times$ 100	0.0127	0.0052	0.0603*	0.0278
	(0.0319)	(0.0339)	(0.0343)	(0.0278)
Occupational Tenure	0.0011	0.0007	-0.0043	0.0005
	(0.0065)	(0.0069)	(0.0070)	(0.0057)
Occupational Tenure$^2\times$ 100	-0.0226	-0.0310	0.0924	-0.0067
	(0.0727)	(0.0772)	(0.0780)	(0.0633)
Occupational Tenure$^3\times$ 100	0.0008	0.0012	-0.0030	0.0003
	(0.0022)	(0.0023)	(0.0023)	(0.0019)
Experience	0.0025	0.0038	0.0041	0.0045
	(0.0037)	(0.0039)	(0.0039)	(0.0032)
Experience$^2\times$ 100	-0.0010	-0.0001	-0.0212	-0.0175
	(0.0321)	(0.0341)	(0.0344)	(0.0279)
Experience$^3\times$ 100	-0.0001	-0.0003	0.0003	0.0002
	(0.0008)	(0.0008)	(0.0008)	(0.0007)
Old Job	-0.0065	-0.0071	0.0063	-0.0057
	(0.0095)	(0.0101)	(0.0102)	(0.0083)
< High School	0.1256***	0.1290***	0.0764***	0.1163***
	(0.0091)	(0.0096)	(0.0097)	(0.0079)
4-Year College	-0.1607***	-0.1508***	-0.1142***	-0.1498***
	(0.0087)	(0.0092)	(0.0093)	(0.0075)
Hispanic	0.0651***	0.0694***	0.0096	0.0538***
	(0.0115)	(0.0122)	(0.0124)	(0.0100)
Black	0.1283***	0.1340***	0.0397***	0.1136***
	(0.0094)	(0.0100)	(0.0101)	(0.0082)

Constant	0.2849***	0.2395***	0.3001***	0.2754***
	(0.0262)	(0.0278)	(0.0281)	(0.0227)
Observations	6,594	6,594	6,594	6,594

All regressions include occupation and industry dummies.

Robust standard errors in parentheses. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

C Data

C.1 Panel Construction and Sample Selection

To construct annual panel data for our main analysis, we use NLSY79's Work History Data File, which records individuals' employment histories up to five jobs on a weekly basis from 1978 to 2010. Following the approach by Neal (1999) and Pavan (2011), we calculate total hours worked for each job within a year from the information of usual hours worked per week and the number of weeks worked for each job. Then, we define primary jobs for each individual for each year as the one for which an individual spent the most hours worked within the year. We construct panel data with annual frequency (from 1978 to 2010) from a series of observations of primary jobs and out-of-labor-force status for each individual. The annually reported demographic information and detailed information of employment (occupation, industry, and hourly wage) are merged with the panels.[39]

Occupational Codes Before we merge the occupation information with the panel data, we clean occupational titles. In NLSY79, every year individuals report their occupations for up to five jobs that they had since their last interviews. Also, NLSY79 provides a mapping between five jobs reported in the current interview and those reported in the last interview if any of them are the same. Using the mapping of jobs across interviews, we first create an employment spell for each job. We then assign, to each employment spell, an occupation code that is most often observed during the spell. This approach is similar to the one by Kambourov and Manovskii (2009b), in which an occupational switch is considered as genuine only when it is accompanied with a job switch. Since the classifications of occupations are not consistent across years, we converted all the occupational codes into the Census 1990 Three-Digit Occupation Code before the cleaning.[40]

Industry Codes To clean the industry codes, we take the similar approach as the one for occupation codes. Since industry codes are also reported in the different classifications across years, we use our own crosswalk to convert them into the Census 1970 One-Digit Industry Code.[41] After the conversion into the Census 1970 Code, we clean the industry titles by using

[39]More precisely, in NLSY79, the information except weekly employment status is reported annually from 1979 to 1994, and biannually, from 1996 to 2010.

[40]NLSY79 reports workers' occupational titles in the Census 1970 Three-Digit Occupation Code until 2000. After 2000, they are reported in the Census 2000 Three-Digit Occupation Code. All of those codes are converted to the Census 1990 Three-Digit Occupation Code using the crosswalks provided by the Minnesota Population Center (http://usa.ipums.org/usa/volii/occ_ind.shtml).

[41]Industry codes are reported in the Census 1970 Industry Classification Code before 1994, in the Census 1980 Industry Classification Code for the year 1994, and in the Census 2000 Industry Classification Code after 2000. The crosswalk is presented in Appendix C.3.

job spells. We use those one-digit level industry codes to create industry dummies used in our regression analysis.

Employer and Occupational Switches We identify an employer switch when the primary job for an individual is different from the one in the last year. We identify occupational switches when the occupation in his primary job is different from the one in last year's primary job.

Labor Market Experience, Employer and Occupational Tenure As we will discuss below, we drop the individuals who have already been in labor markets when the survey started. We then set individuals' experience equal to zero when a worker is entering labor markets, and increase it by one every year when the worker reports a job. Employer and occupational tenure increase by one every year when the individual reports a job and are reset to zero when switches happen.

Wages Worker's wages are measured by the usual rate of pay for the primary job at the time of interview. These wages include tips, overtime and bonuses before deductions and are converted to an hourly rate using usual hours worked when not reported as such. All the wages are deflated by the price index for personal consumption expenditures into real term in the 2000 dollars. We drop the observations if their wages are missing. We also drop the top 0.1% and the bottom 0.1% of observations in the wage distribution in each round of the interview. This trimming strategy doesn't affect the regression result.[42]

Sample Selection We follow the approach by Farber and Gibbons (1996) for the sample selection. We first limit our sample to the individuals who make their initial long-term transition from school to labor markets during the survey period: that is, we drop those who work more than 1,200 hours in the initial year of the survey. We also focus on the individuals who work more than 1,200 hours at least for two consecutive years during the survey period. The individuals who are in the military service more than two years during the period are also eliminated from our sample. For the individuals who go back to school from the labor force during the survey period, we assume they start their career from the point they reenter labor markets, and drop the observations before that time. Also, the observations after the last time an individual report a job are eliminated. Furthermore, we drop individuals who are weakly attached to the labor force: those who are out of the labor force more than once before they work at least 10 years after they started their career. If an individual is out of the labor force only for one year after he started his career, or if he has worked more than 10 years before he first dropped from the labor force, we only drop those observations. Finally, we restrict our sample to those who have a valid occupation and industry code, who have valid demographics information, who are equal to or above age 16 and not currently enrolled in a school, and who have valid ASVAB scores and valid wage information. The number of the remaining individuals and observations after applying each sample selection criterion are summarized in Table C.1.

Descriptive statistics for the sample are reported in Table C.2. Because of the nature of the survey, which starts with workers when they are young in the workforce, the sample skews younger. As a result, the mean length of employer tenure in our sample is relatively short, although this is a well-understood point about the NLSY79 in the literature.[43]

[42]Similarly, Pavan (2011) drops the top ten and the bottom ten observations in the entire sample.

[43]Both Parent (2000) and Pavan (2011) report mean employer tenure in the NLSY79 that ranges

Table C.1 – Sample Selection, NLSY79, 1978 - 2010

Criterion for Sample Selection	Remaining Individuals	Remaining Observations
Male Cross-Sectional Sample	3,003	99,099
Started career after the survey started	2,408	79,242
Work more than 1,200 hours for two consecutive periods	2,311	65,041
Not in the military service more than or equal to two years	2,261	63,529
Drop the observations before they go back to school	2,261	63,477
Drop the observations after the last time they worked	2,261	55,406
Drop those who are weakly attached to labor force	2,095	49,154
Valid occupation and industry code	2,094	48,352
Valid demographics information	2,093	48,328
Drop those below age 16 and not enrolled in school	2,093	48,314
Valid ASVAB scores	1,996	46,253
Valid wage information	1,992	44,721
Drop top 0.1% and bottom 0.1% in the wage distribution	1,992	44,591

Table C.2 – Descriptive Statistics of the Sample, NLSY79, 1978–2010

Statistics	All Sample	$\leq$ High School	$>$ High School
Total number of observations	44,591	21,618	22,973
Total number of individuals	1,992	954	1,038
Average age at the time of interview	33.79	32.85	34.67
Highest education $<$ high school	7.01%	14.45%	-
Highest education $=$ high school	41.48%	85.55%	-
Highest education $>$ high school	51.51%	-	100.00%
Highest education $\geq$ 4-year college	31.74%	-	61.62%
Percentage African-American	10.46%	13.81%	7.31%
Percentage Hispanic	6.53%	6.92%	6.16%
Occupational mobility	15.94%	18.10%	13.90%
Occupational tenure (mean)	6.50	6.17	6.81
Occupational tenure (median)	4.00	4.00	5.00
Employer (job) mobility	30.39%	31.97%	28.90%
Employer (job) tenure (mean)	3.61	3.56	3.65
Employer (job) tenure (median)	2.00	2.00	2.00
Average hours worked within a year	1983.8	1958.8	2007.2

Note: Occupational mobility is defined as the fraction of individuals who switch occupations in a year. The same definition for employer mobility.

Table C.3 – List of Skills in O*NET

Verbal and Math Skills			
1.	Oral Comprehension	2.	Written Comprehension
3.	Deductive Reasoning	4.	Inductive Reasoning
5.	Information Ordering	6.	Mathematical Reasoning
7.	Number Facility	8.	Reading Comprehension
9.	Mathematics Skill	10.	Science
11.	Technology Design	12.	Equipment Selection
13.	Installation	14.	Operation and Control
15.	Equipment Maintenance	16.	Troubleshooting
17.	Repairing	18.	Computers and Electronics
19.	Engineering and Technology	20.	Building and Construction
21.	Mechanical	22.	Mathematics Knowledge
23.	Physics	24.	Chemistry
25.	Biology	26.	English Language
Social Skills			
1.	Social Perceptiveness	2.	Coordination
3.	Persuasion	4.	Negotiation
5.	Instructing	6.	Service Orientation

C.2 O*NET skills

The complete list of skills is in Table C.3.[44] O*NET's occupational classification is more detailed than the codes in the NLSY79, which are based on the Three-Digit Census Occupation Codes. We average scores over O*NET occupation codes that map to the same code in the Census Three-Digit Level Occupation Classification.

C.3 The Crosswalk of Census Industry Codes

We used the crosswalk in Table C.4 to convert the Census 1970, 1980, and 1990 Three-Digit Industry Code to the Census 1970 One-Digit Industry Code. We use one-digit level industry titles to create industry dummies used in our regression analysis. From the Census 1980 Three-Digit Code to the Census 1970 One-Digit Code, we first aggregate the Census 1980 Three-Digit level into the Census 1980 One-Digit level. Then, we combine Wholesale (500-571) and Retail Trade (580-691) in the Census 1980 One-Digit Code to create the category 6, "Whole Sale and Retail Trade", in the Census 1970 One-Digit Code. For other one-digit-level industry titles, the Census 1970 and 1980 have the same classification.

Unlike the one between the Census 1970 and the Census 1980, the mapping is not straightforward from the Census 2000 to the Census 1970. Sometimes, the same industry titles in three-

from 3 to 3.3 years. The corresponding figure in our sample is 3.6 years, which is close.

[44] For each descriptor, there is both a "level" and an "intensity" score. The ASVAB Career Exploration Program, which we describe below, uses only intensity and so do we.

digit-level are put in different one-digit-level categories. For example, "Newspaper Publishers" (code number 647 in the Census 2000) is in "Information and Communication" category in the Census 2000, but is put in "Manufacturing" in the Census 1970. Therefore, we check all the three-digit industry titles both in the Census 1970 and 2000 Industry Code, and made necessary changes to create our own mapping. The obtained crosswalk is reported in Table C.4.

TABLE C.4 – The Crosswalk across the Census 1970, 1980, and 1990 One-Digit Industry Classification Code

1970 One-Digit Classification		1970	1980	2000
1.	Agriculture, Forestry, Fishing, and Hunting	017-029	010-031	017-029
2.	Mining	047-058	040-050	037-049
3.	Construction	067-078	60	077
4.	Manufacturing	107-398	100-392	107-399, 647-659, 678-679
5.	Transportation, Communications, and Other Public Utilities	407-499	400-472	57-69, 607-639, 667-669
6.	Wholesale and Retail Trade	507-699	500-691	407-579, 868-869
7.	Finance, Insurance and Real Estate	707-719	700-712	687-719
8.	Business and Repair Services	727-767	721-760	877-879, 887
9.	Personal Services	769-799	761-791	866-867, 888-829
10.	Entertainment and Recreation Services	807-817	800-802	856-859
11.	Professional and Related Services	828-899	812-892	677, 727-779, 786-847
12.	Public Administration	907-947	900-932	937-959

D Clustering

For the benchmark results in Table II, we use the feasible generalized least square method to compute standard errors allowing for possibilities of heteroskedasticity. However, Cameron and Miller (2015) point out, the standard errors can be underestimated by the Huber-White method if the error term and/or the explanatory variables are serially correlated. Therefore, in this section, we compute the cluster-robust standard errors considering the errors uncorrelated across individuals but correlated within individuals. The results with clustered standard errors are shown in Table D.1. The significance of mismatch and cumulative mismatch are preserved even when we use clustered standard errors.

TABLE D.1 – Wage Regressions with Mismatch, Clustered Standard Errors

	(1) IV-CLU	(2) IV-CLU	(3) IV-CLU	(4) OLS-CLU	(5) OLS-CLU	(6) OLS-CLU
Mismatch	-0.0271***	-0.0145*	-0.0054	-0.0254***	-0.0214***	-0.0147*
Mismatch × Occ Tenure		-0.0020**	-0.0024***		-0.0006	-0.0006
Cumul Mismatch			-0.0355***			-0.0364***
Worker Ability (Mean)	0.2466***	0.2475***	0.3408***	0.2588***	0.2585***	0.3426***
Worker Ability × Occ Tenure	0.0166***	0.0161***	0.0140***	0.0130***	0.0129***	0.0127**
Occ Reqs (Mean)	0.1529***	0.1528***	0.1576***	0.2096***	0.2095***	0.2224***
Occ Reqs × Occ Tenure	0.0155***	0.0154***	0.0161***	0.0070*	0.0069***	0.0061
Observations	44,591	44,591	33,072	44,591	44,591	33,072
R^2	0.355	0.355	0.313	0.371	0.371	0.332

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed as panel-robust estimates clustered within individuals but not across individuals. We follow Arellano (1987) for the calculation cluster-robust of standard errors.

Table D.2 – Wage Regressions with Mismatch by Components, Clustered Standard Errors

	(1) IV-CLU	(2) IV-CLU	(3) IV-CLU	(4) OLS-CLU	(5) OLS-CLU	(6) OLS-CLU
Mismatch Verbal	-0.0147*	0.0030	0.0139	-0.0150*	-0.0053	0.0027
Mismatch Math	-0.0130	-0.0171*	-0.0203*	-0.0109	-0.0172*	-0.0182*
Mismatch Social	-0.0049	-0.0034	0.0067	-0.0041	-0.0046	0.0017
Mismatch Verbal × Occ Tenure		-0.0028**	-0.0045***		-0.0015	-0.0026*
Mismatch Math × Occ Tenure		0.0006	0.0021*		0.0010	0.0020
Mismatch Social × Occ Tenure		-0.0002	-0.0010		0.0001	-0.0005
Cumul Mismatch Verbal			-0.0123			-0.0107
Cumul Mismatch Math			-0.0252**			-0.0274**
Cumul Mismatch Social			-0.0083			-0.0073
Verbal Ability	-0.0440	-0.0486	0.0081	0.0112	0.0066	0.0158
Math Ability	0.2949***	0.3001***	0.3405***	0.2510***	0.2547***	0.3238***
Social Ability	0.0837***	0.0836***	0.1017***	0.0855***	0.0853***	0.1137***
Verbal Ability × Occ Tenure	0.0125**	0.0126**	0.0088	0.0047	0.0051	0.0066
Math Ability × Occ Tenure	0.0006	-0.0003	0.0011	0.0037	0.0031	0.0023
Social Ability × Occ Tenure	0.0072**	0.0073**	0.0075**	0.0076**	0.0076**	0.0063
Occ Reqs Verbal	0.0771	0.0757	0.0913	0.1414	0.1482	0.1214
Occ Reqs Math	0.1112	0.1075	0.1065	0.1004	0.0917	0.1361
Occ Reqs Social	-0.0932**	-0.0894**	-0.0978**	-0.0817**	-0.0803**	-0.0827**
Occ Reqs Verbal × Occ Tenure	-0.0071	-0.0070	-0.0098	-0.0229	-0.0245	-0.0205
Occ Reqs Math × Occ Tenure	0.0164	0.0172	0.0175	0.0228	0.0249	0.0175
Occ Reqs Social × Occ Tenure	0.0100**	0.0092**	0.0131***	0.0109**	0.0106**	0.0135**
Observations	44,591	44,591	33,072	44,591	44,591	33,072
R^2	0.358	0.358	0.317	0.374	0.375	0.335

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed as panel-robust estimates clustered within individuals but not across individuals. We follow Cameron and Trivedi (2005) for the calculation of standard errors.

E Fixed Effect Regression

Given that NLSY79 is a panel data set, we are able to use the fixed-effect model. As discussed in the robustness Section 5.2, omitting worker fixed-effect may bias the estimates for the effects of mismatch in unknown direction. Therefore, we run wage regressions with fixed effects to check whether our main results are robust even when we control for unobserved heterogeneity by fixed effects. The full regression results are shown in Table E.1, which compares the baseline regression to that with fixed effects. The key results were presented in Table in Section 5.2. Note that, in these regressions, we do not include worker's mean ability term or demographic characteristics, which are time-invariant.

In Table E.1, columns 1-3 present our baseline specification with fixed effects, these are the estimates using instrumental variables and fixed effects and correspond to columns 4-6 in Table VIII of the main text. For comparison, we include the estimates without instrumenting tenure variables as we have in our other tables. Table E.2 uses the baseline specification with fixed effects, column 6 of Table VIII, and computes the implied wage effects.

TABLE E.1 – Wage Regressions with Mismatch and Fixed Effects

	(1) IV-GLS-FE	(2) IV-GLS-FE	(3) IV-GLS-FE	(4) GLS-FE	(5) GLS-FE	(6) GLS-FE
Mismatch	-0.0117***	-0.0051	-0.0044	-0.0105***	-0.0057	-0.0052
Mismatch × Occ Tenure		-0.0019***	-0.0020**		-0.0014**	-0.0018**
Cumul Mismatch			-0.0373***			-0.0362***
Worker Ability × Occ Tenure	0.0129***	0.0124***	0.0107***	0.0170***	0.0168***	0.0146***
Occ Reqs (Mean)	0.0074	0.0064	-0.0133	0.0313	0.0307	0.0170
Occ Reqs × Occ Tenure	0.0147***	0.0146***	0.0163***	0.0091***	0.0090***	0.0094***
Observations	41596	41596	30599	41596	41596	30599
R^2				0.166	0.166	0.114

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

TABLE E.2 – Wage Losses Predicted by the Regression with Fixed Effects

Mismatch Degree (High to Low)	*Mismatch* Effect 5 years	*Mismatch* Effect 10 years	*Mismatch* Effect 15 years	*Cumul. Log Mismatch* Effect
90%	-0.043	-0.073	-0.103	-0.122
	(0.014)	(0.021)	(0.031)	(0.026)
50%	-0.019	-0.032	-0.045	-0.065
	(0.006)	(0.009)	(0.014)	(0.014)
10%	-0.007	-0.012	-0.016	-0.029
	(0.002)	(0.003)	(0.005)	(0.006)

Note: Wage losses (relative to the mean wage) are computed for each percentile of each measure, using the result of the specification (3) in Table E.1. Standard errors are in parentheses.

F Physical Skills Dimension

In addition to verbal, math, and social dimensions, one might expect that match quality is affected by the physical requirements of an occupation and a worker's physical abilities. In this appendix we try to incorporate a physical dimension into our measure. Conceptually, it is difficult to measure a worker's physical abilities, as these are going to change quite a bit over his working life and often as a result of the occupations he chooses. This endogeneity makes it quite difficult to identify an underlying ability for physically demanding work, as we had identified in the other ability dimensions. Beyond this direct reverse causality, there is a strong correlation between healthy behaviors and income, whereas most high income jobs have only mild physical requirements.

When we introduce our proxy for worker's physical ability and the occupational physical requirements, as described below, it seems to have little to do with wages. When we use principal components to aggregate dimensions of mismatch, physical gets a very low weight, suggesting its variation is not well related to the rest of the variation in the dataset. While this independence from the other dimensions may have actually been useful, we found that physical mismatch also has little relationship to wages. Generally physical mismatch is insignificant when we include it into a Mincer regression, as we did with the others. All this is not to say that physical match quality is unimportant, but given the measurement hurdles, we were unable to find a solid relationship. Details of the process and findings are given below.

Health/Physical Scores in NLSY79

Participants in the NLSY79 were asked to take a survey when they turned age 40 to evaluate their health status. In particular, the survey includes questions about how much health limited the respondents' (i) moderate activities; (ii) ability to climb a flight of stairs; and (iii) types of work they can perform; as well as (iv) how participants rated their own health status (often referred to as EVGFP) and (v) whether pain interfered with their daily activities.[45] Each participant was then assigned a composite health score, called PCS-12, by combining their scores on each question. One difficulty in using this health composite score in our analysis is that it is measured after a significant period of working life, so differences across individuals may simply reflect the effects of occupations on workers (see Michaud and Wiczer (2014)).

[45]The survey is conducted by health care survey firm Quality Metric; see Ware et al. (1995) for details.

Physical Skills in O*NET

Table F.1 – List of Physical Skills in O*NET

Physical Skills			
1.	Arm-Hand Steadiness	2.	Manual Dexterity
3.	Finger Dexterity	4.	Control Precision
5.	Multi-limb Coordination	6.	Response Orientation
7.	Rate Control	8.	Reaction Time
9.	Wrist-Finger Speed	10.	Speed of Limb Movement
11.	Static Strength	12.	Explosive Strength
13.	Dynamic Strength	14.	Trunk Strength
15.	Stamina	16.	Extent Flexibility
17.	Dynamic Flexibility	18.	Gross Body Coordination
19.	Gross Body Equilibrium		

To create a physical measure of an occupation, we again turn to the O*NET, which contains 19 descriptors related to the physical demands of an occupation (e.g., whether it requires strength, coordination, and stamina). To reduce the 19 descriptors to a single index measure of physical skills, we take the first principal component over the 19 descriptors. For the worker's physical ability measure, we use the NLSY's PCS-12 score. Both physical ability and skill scores again are converted into rank scores among individuals or among occupations. Notice that the coefficients to mismatch change relatively little from our previous specification. This is because the loading on physical mismatch is relatively small. Principal components assigns loadings (0.42, 0.42, 0.12, 0.4) when constructing mismatch measure.

Wage Regression Results with Physical Component

Table F.2 – Four Skills: Wage Regressions with Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0215***	-0.0094*	-0.0055	-0.0214***	-0.0147***	-0.0146***
Mismatch × Occ Tenure		-0.0020***	-0.0020**		-0.0011**	-0.0004
Cumul Mismatch			-0.0398***			-0.0398***
Worker Ability (Mean)	0.2936***	0.2982***	0.4108***	0.3201***	0.3218***	0.4009***
Worker Ability × Occ Tenure	0.0133***	0.0124***	0.0117***	0.0078***	0.0074***	0.0108***
Occ Reqs (Mean)	0.1242***	0.1212***	0.1121***	0.1553***	0.1534***	0.1606***
Occ Reqs × Occ Tenure	0.0117***	0.0119***	0.0139***	0.0070***	0.0072***	0.0065***
Observations	35405	35405	26168	35405	35405	26168
R^2	0.241	0.241	0.217	0.256	0.256	0.232

Note: $*** \ p < 0.01$, $** \ p < 0.05$, $* \ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

Table F.3 – Four Skills: Wage Regressions with Mismatch by Components

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch Verbal	-0.0130***	0.0070	0.0132	-0.0125***	0.0039	0.0070
Mismatch Math	-0.0075*	-0.0134*	-0.0166*	-0.0081*	-0.0166***	-0.0207***
Mismatch Social	-0.0060*	-0.0055	-0.0015	-0.0056*	-0.0068	-0.0020
Mismatch Phys	-0.0017	-0.0049	-0.0024	0.0007	-0.0027	-0.0022
Mismatch Verbal × Occ Tenure		-0.0032***	-0.0042***		-0.0026***	-0.0031***
Mismatch Math × Occ Tenure		0.0008	0.0019		0.0013*	0.0025***
Mismatch Social × Occ Tenure		-0.0000	-0.0002		0.0002	-0.0003
Mismatch Phys × Occ Tenure		0.0005	0.0011		0.0005	0.0013**
Cumul Mismatch Verbal			-0.0188***			-0.0172***
Cumul Mismatch Math			-0.0232***			-0.0253***
Cumul Mismatch Social			-0.0117**			-0.0105**
Cumul Mismatch Phys						
Verbal Ability	-0.0947**	-0.0981**	-0.0146	-0.0511	-0.0565*	-0.0296
Math Ability	0.3757***	0.3843***	0.4271***	0.3381***	0.3451***	0.4162***
Social Ability	0.0894***	0.0891***	0.0945***	0.1067***	0.1060***	0.1240***
Phys Ability	0.1074***	0.1088***	0.1503***	0.1316***	0.1313***	0.1340***
Verbal Ability × Occ Tenure	0.0128***	0.0131***	0.0087	0.0078**	0.0084**	0.0101**
Math Ability × Occ Tenure	-0.0043	-0.0059	-0.0058	-0.0027	-0.0040	-0.0057
Social Ability × Occ Tenure	0.0073***	0.0075***	0.0079**	0.0053***	0.0055***	0.0038
Phys Ability × Occ Tenure	0.0001	-0.0001	0.0029	-0.0043**	-0.0043**	0.0032
Occ Reqs Verbal	-0.0258	-0.0223	-0.0974	0.0101	0.0184	-0.0405
Occ Reqs Math	0.1859**	0.1756**	0.2553***	0.1911***	0.1787***	0.2551***
Occ Reqs Social	-0.1106***	-0.1085***	-0.1089***	-0.1092***	-0.1089***	-0.1129***
Occ Reqs Phys	0.0140	0.0121	-0.0231	-0.0012	-0.0033	-0.0378
Occ Reqs Verbal × Occ Tenure	-0.0034	-0.0038	0.0065	-0.0136*	-0.0153**	-0.0099
Occ Reqs Math × Occ Tenure	0.0109	0.0126	0.0037	0.0140**	0.0166**	0.0093
Occ Reqs Social × Occ Tenure	0.0122***	0.0116***	0.0122**	0.0138***	0.0136***	0.0153***
Occ Reqs Phys × Occ Tenure	0.0040	0.0044	0.0012	0.0027	0.0032	0.0015
Observations	35405	35405	26168	35405	35405	26168
R^2	0.245	0.245	0.222	0.259	0.260	0.237

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

G College Graduates

Given that our mismatch measure is based on higher-order cognitive, and social abilities, it is natural that this measure is more relevant to individuals with a higher level of education doing occupations that place greater emphasis on higher-order skills. In this appendix, we restrict our sample to college graduates, and run wage regressions as we did in the main analysis. The results are presented in Table G.1. Most of the coefficients of mismatch, mismatch times tenure, and cumulative mismatch increased in their effect on wages compared to those in Table II. In particular, the coefficient on the cumulative mismatch in Column (3) almost doubled for the college sample compared to the baseline result.

It is also interesting to see where the increase of the effect is coming from. By breaking down the measures into components in Column (3) of Table G.2, we learn that it is verbal and social components which are particularly strong effects and contribute to the differences in wages among college graduates. In particular, the coefficient on cumulative social mismatch is four times larger than the one in our benchmark result. The results here show that mismatch is a more important wage determinant among college graduates, and that verbal and social components have especially large effects compared to the benchmark case.

Table G.1 – College Graduate: Wage Regressions with Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0273***	-0.0170*	-0.0083	-0.0267***	-0.0130*	-0.0112
Mismatch × Occ Tenure		-0.0018	-0.0012		-0.0023***	-0.0002
Cumul Mismatch			-0.0693***			-0.0766***
Worker Ability (Mean)	0.2758***	0.2808***	0.3913***	0.2964***	0.2995***	0.3684***
Worker Ability × Occ Tenure	0.0121**	0.0108**	0.0086	0.0086**	0.0073**	0.0157***
Occ Reqs (Mean)	0.1602***	0.1715***	0.1866***	0.2021***	0.2186***	0.2350***
Occ Reqs × Occ Tenure	0.0188***	0.0170***	0.0184***	0.0109***	0.0083**	0.0089**
Observations	21461	21461	15093	21461	21461	15093
R^2	0.198	0.198	0.177	0.210	0.210	0.188

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

Table G.2 – College Graduate: Wage Regressions with Mismatch by Components

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch Verbal	-0.0269***	-0.0089	-0.0035	-0.0249***	-0.0089	-0.0106
Mismatch Math	0.0015	-0.0057	-0.0061	-0.0002	-0.0009	-0.0010
Mismatch Social	-0.0074	-0.0127	0.0021	-0.0061	-0.0112	0.0009
Mismatch Verbal × Occ Tenure		-0.0030**	-0.0031		-0.0027**	-0.0014
Mismatch Math × Occ Tenure		0.0011	0.0020		0.0000	0.0013
Mismatch Social × Occ Tenure		0.0009	-0.0008		0.0009	-0.0008
Cumul Mismatch Verbal			-0.0335***			-0.0351***
Cumul Mismatch Math			-0.0347***			-0.0422***
Cumul Mismatch Social			-0.0339***			-0.0301***
Verbal Ability	-0.1041*	-0.0934	0.0390	-0.0213	-0.0136	0.0261
Math Ability	0.3850***	0.3835***	0.4309***	0.3244***	0.3224***	0.3975***
Social Ability	0.0738**	0.0678*	0.0421	0.0810***	0.0750**	0.0548
Verbal Ability × Occ Tenure	0.0180***	0.0156**	0.0089	0.0079	0.0058	0.0124*
Math Ability × Occ Tenure	-0.0130*	-0.0131*	-0.0121	-0.0083	-0.0084*	-0.0059
Social Ability × Occ Tenure	0.0108***	0.0118***	0.0131**	0.0110***	0.0121***	0.0134***
Occ Reqs Verbal	-0.0100	0.0047	-0.1501	-0.0204	-0.0163	-0.1165
Occ Reqs Math	0.2023**	0.1867*	0.3327***	0.2516***	0.2537***	0.3674***
Occ Reqs Social	-0.1011**	-0.0960**	-0.0797	-0.0879***	-0.0809**	-0.0813*
Occ Reqs Verbal × Occ Tenure	-0.0106	-0.0121	-0.0000	-0.0133	-0.0127	-0.0134
Occ Reqs Math × Occ Tenure	0.0213*	0.0238*	0.0112	0.0157*	0.0148	0.0104
Occ Reqs Social × Occ Tenure	0.0147***	0.0135***	0.0180**	0.0141***	0.0125***	0.0195***
Observations	21461	21461	15093	21461	21461	15093
R^2	0.201	0.201	0.181	0.213	0.214	0.193

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

H Young Workers

There are several reasons one might be concerned with the effect of mismatch separately on young and old workers. While age is not directly in the model, it is an interesting variable to consider in the data and closely related to work experience, which does appear in our theory. On one hand, learning suggests that mismatch observed among the old is more likely to be due to the noise in our measure and attenuate these coefficients. On the other hand, given that our measure of mismatch is not necessarily cardinal, its interactions with experience and tenure may be non-linear and thus may effect differently the high experience groups, whom we know have lost more human capital due to mismatch.

In this appendix, we create a dummy for workers less than 35 years old, though the results are broadly the same if we were to use a cutoff at 30. We interact this dummy with each of our mismatch terms. The results are presented in Table H.1. In the specification with only mismatch, Column 1, the coefficients imply that the overall effect of mismatch is almost 30% more severe for younger workers.

The effect of cumulative past mismatch is stronger for the older group among whom mismatch has had longer to have an effect, as shown in the coefficients on cumulative mismatch and the interaction with < 35 in Column 3. As the theory suggests, mismatch is cumulative over time, and this accumulation is partly going to be picked up in the coefficient on average mismatch, a quantity that should not be affected by experience. We validate this conjecture by including total mismatch instead of average cumulative mismatch, only the numerator from our measure for CMM. The new measure strictly increasing in labor market experience. In this regression, the coefficient on total cumulative mismatch is -0.129 and -0.009 on the coefficient for the < 35 interaction (both 99% significant). This means that the additional cost of cumulative mismatch among the old stems from their additional years accumulating these effects. For the same amount of experience, the effect among young is slightly stronger.

To summarize the difference between young and old in the effect of mismatch we can create counter-factual wage distributions setting the mismatch terms to zero for both young and old. For young workers, the average predicted wage is 8.5 log points lower than without mismatch. For old workers, the average predicted wage is 12.5 log points lower than without mismatch. As in our baseline result in Section 5.1, cumulative mismatch is most important to this result. Average cumulative mismatch is about the same in the two groups, 1.90 among old and 1.91 among young, but the implied coefficients are quite different, -0.0437 for old and -0.0297 for young. The larger effect of mismatch on old than young is in line with the findings in Fredriksson et al. (2015), albeit measured in Sweden with a different definition of mismatch.

Table H.1 – Young Workers: Wage Regressions with Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0178***	-0.0581***	-0.0395***	-0.0184***	-0.0109*	-0.0119*
Mismatch, < 35	-0.0049	0.1527***	0.1359***	-0.0020	-0.0022	-0.0018
Mismatch $\times$ Occ Tenure		0.0022***	0.0016		-0.0011**	-0.0009
Mismatch $\times$ Occ Ten, < 35		-0.0323***	-0.0354***		-0.0006	0.0003
Cumul Mismatch			-0.0437***			-0.0348***
Cumul Mismatch, < 35			0.0146**			-0.0074
Worker Ability (Mean)	0.3018***	0.2948***	0.4001***	0.3098***	0.3111***	0.4032***
Worker Ability $\times$ Occ Tenure	0.0116***	0.0119***	0.0084**	0.0090***	0.0087***	0.0076***
Occ Reqs (Mean)	0.0974***	0.1021***	0.0853***	0.1381***	0.1368***	0.1419***
Occ Reqs $\times$ Occ Tenure	0.0144***	0.0121***	0.0162***	0.0084***	0.0084***	0.0094***
Observations	41596	41596	30599	41596	41596	30599
R^2	0.243	0.215	0.188	0.258	0.258	0.229

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

I Effects of Mismatch on Earnings

In the model we presented in Section 2, wages and earnings are identical because we assume worker's labor supply is constant (fixed to 1) over lifecycle. However, in reality, wages and earnings could be significantly different as there is large heterogeneity in individuals' working hours. Therefore, it is worth to see whether our model's implications still hold when we use individuals' earnings data rather than the wage data in regressions. Looking at earnings rather than wages is advantageous in the light of measurement error due to misreporting. As is common in many survey-based datasets, because most workers do not earn an hourly wage, and actual hours are often difficult to recall, earnings are much more precisely reported than hourly wages. Therefore, in this appendix, we check the robustness of our results by using the earnings in place of wages.

In order to create an annual earnings measure, we use two income variables from NLSY79: total income from wage and salary and total income from farm or business. One shortcoming of using these variables is that, after 1994, the information is only available every 2 years. Therefore, we have to reduce the number of observations significantly when we run a regression using the earnings measure. Another important issue to take into account is that these variables pool income from different jobs. Thus, when an individual works for more than one occupation in a year, income from different occupations are pooled in one earnings measure. Therefore, when a worker reports more than one job, we relate a worker's annual earnings to the job in which the worker earned the largest amount of money in the year, which is calculated by the hourly rate of pay of that job times the number of hours the worker spent in that job in the year. Obtained annual earnings measure is deflated by the price index for personal consumption expenditures into real term in the 2000 dollars. Finally, obtained, real annual earnings are put as the left-hand-side variable in a Mincer regression after taking a natural logarithm.

Table I.1 reports results of earnings regressions with mismatch. It is worth to emphasize that those results are very similar to those in previous wage regressions (compare the results with Table II). In our most preferred specification, (3), the coefficient for mismatch times tenure is slightly larger than the one in the wage regression, and the one for cumulative mismatch is slightly smaller. However, in general, the results are almost same as those in wage regression.

Turning to regressions by components reported in Table I.2, again, the results didn't change from Table V in general. In the specification (3), the coefficient for cumulative social mismatch obtain a slightly larger value when we use earnings, while that for cumulative verbal mismatch loses its significance. However, over all, the results are in line with those in wage regressions. This fact confirms the robustness of our results even when we use annual earnings as the left-hand-side variable of a Mincer regression.

Table I.1 – Earnings Regressions with Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0315***	-0.0332	0.0351	-0.0286***	-0.0370***	-0.0407***
Mismatch × Occ Tenure		0.0005	-0.0320**		0.0024	0.0033
Cumul Mismatch			-0.0383**			-0.0170**
Worker Ability (Mean)	0.2747**	0.2745**	0.5062***	0.2417***	0.2414***	0.3344***
Worker Ability × Occ Tenure	-0.0167	-0.0166	-0.0471	0.0000	0.0004	-0.0063
Occ Reqs (Mean)	0.2028**	0.2027**	0.2872*	0.3248***	0.3253***	0.3434***
Occ Reqs × Occ Tenure	0.0105	0.0106	-0.0287	-0.0017	-0.0018	-0.0034
Observations	18473	18473	10766	18473	18473	10766
R^2				0.241	0.241	0.215

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are computed via GLS assuming AR(1) autocorrelation.

Table I.2 – Earnings Regressions with Mismatch by Components

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch Verbal	-0.0235**	0.0446	0.1583***	-0.0071	-0.0070	-0.0027
Mismatch Math	-0.0122	-0.0867***	-0.1233**	-0.0269***	-0.0364***	-0.0449***
Mismatch Social	0.0022	0.0195	0.0330	0.0058	0.0063	0.0077
Mismatch Verbal × Occ Tenure		-0.0207**	-0.0713***		-0.0000	-0.0027
Mismatch Math × Occ Tenure		0.0233**	0.0412**		0.0028	0.0065*
Mismatch Social × Occ Tenure		-0.0053	-0.0144		-0.0002	-0.0004
Cumul Mismatch Verbal			0.0314			-0.0048
Cumul Mismatch Math			-0.0543**			-0.0091
Cumul Mismatch Social			0.0032			-0.0063
Verbal Ability	0.0963	0.1193	-0.0490	-0.0019	-0.0014	-0.0569
Math Ability	0.2861*	0.2653	0.9055***	0.1834***	0.1827***	0.2874***
Social Ability	-0.0460	-0.0601	-0.4286*	0.1493***	0.1484***	0.2143***
Verbal Ability × Occ Tenure	-0.0662*	-0.0764*	-0.0243	-0.0084	-0.0086	0.0197
Math Ability × Occ Tenure	0.0109	0.0172	-0.1885**	0.0110	0.0114	-0.0233
Social Ability × Occ Tenure	0.0406*	0.0445**	0.1925***	-0.0042	-0.0039	-0.0078
Occ Reqs Verbal	0.0787	0.0370	0.9529	0.2790**	0.2898***	0.5410***
Occ Reqs Math	0.0070	0.0312	-1.1516*	-0.0207	-0.0294	-0.2729**
Occ Reqs Social	0.1081	0.1090	0.4641**	0.0995***	0.0966**	0.1153**
Occ Reqs Verbal × Occ Tenure	0.0039	0.0197	-0.3995*	-0.0192	-0.0230	-0.0801**
Occ Reqs Math × Occ Tenure	0.0219	0.0172	0.5446**	0.0177	0.0212	0.0742**
Occ Reqs Social × Occ Tenure	-0.0129	-0.0139	-0.1672**	-0.0005	0.0002	0.0076
Observations	18473	18473	10766	18473	18473	10766
R^2				0.242	0.243	0.217

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation.

J Additional Regression Tables for Robustness

Table J.1 – Wage Regressions with Log Mismatch

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0252***	-0.0051	-0.0015	-0.0236***	-0.0171***	-0.0185**
Mismatch × Occ Tenure		-0.0033***	-0.0032**		-0.0011	-0.0002
Cumul Mismatch			-0.0354***			-0.0351***
Worker Ability (Mean)	0.3017***	0.3045***	0.4026***	0.3084***	0.3088***	0.3924***
Worker Ability × Occ Tenure	0.0114***	0.0107***	0.0072**	0.0090***	0.0089***	0.0078***
Occ Reqs (Mean)	0.0973***	0.0953***	0.0858***	0.1382***	0.1375***	0.1420***
Occ Reqs × Occ Tenure	0.0144***	0.0144***	0.0184***	0.0083***	0.0083***	0.0093***
Observations	41596	41596	30599	41596	41596	30599
R^2	0.243	0.242	0.213	0.257	0.257	0.229

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation.

Table J.2 – Wage Losses from Log Mismatch & Cumulative Log Mismatch

Mismatch Degree	*Log Mismatch* Effect			*Cumul. Log Mismatch* Effect
(High to Low)	5 years	10 years	15 years	
90%	-0.020	-0.037	-0.055	-0.061
	(0.006)	(0.009)	(0.016)	(0.007)
50%	-0.005	-0.010	-0.015	-0.020
	(0.002)	(0.003)	(0.004)	(0.002)
10%	0.011	0.022	0.032	0.028
	(0.000)	(0.000)	(0.004)	(0.003)

Note: Wage losses (relative to the mean wage) are computed for each percentile of each measure, using the result of the specification (3) in Table J.1. Standard errors are in parentheses.

Table J.3 – Wage Regressions with Higher-Order Terms

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0219***	-0.0108**	-0.0067	-0.0210***	-0.0159***	-0.0148***
Mismatch × Occ Tenure		-0.0019***	-0.0021**		-0.0009*	-0.0005
Cumul Mismatch			-0.0386***			-0.0396***
Worker Ability (Mean)	0.1876***	0.1979***	0.2471***	0.2148***	0.2190***	0.2247***
Worker Ability^2	0.1105**	0.1030**	0.1610***	0.0869*	0.0834*	0.1691***
Worker Ability × Occ Tenure	0.0117***	0.0111***	0.0073**	0.0099***	0.0097***	0.0087***
Occ Reqs (Mean)	-0.3529***	-0.3517***	-0.3075***	-0.3070***	-0.3065***	-0.2544***
Occ Reqs^2	0.4498***	0.4471***	0.3978***	0.4549***	0.4536***	0.4075***
Occ Reqs × Occ Tenure	0.0144***	0.0143***	0.0182***	0.0064***	0.0064***	0.0077***
Observations	41596	41596	30599	41596	41596	30599
R^2	0.245	0.245	0.216	0.260	0.260	0.231

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation.

Table J.4 – Wage Losses Predicted by the Regression with Higher Order Terms

Mismatch Degree	*Mismatch* Effect			*Cumul. Mismatch* Effect
(High to Low)	5 years	10 years	15 years	
90%	-0.052	-0.083	-0.115	-0.127
	(0.011)	(0.018)	(0.030)	(0.014)
50%	-0.023	-0.036	-0.050	-0.067
	(0.005)	(0.008)	(0.013)	(0.007)
10%	-0.008	-0.013	-0.018	-0.030
	(0.002)	(0.003)	(0.005)	(0.003)

Note: Wage losses (relative to the mean wage) are computed for each percentile of each measure, using the result of the specification (3) in Table J.3. Standard errors are in parentheses.

Table J.5 – Wage Regressions with Worker's Ability Times Experience Term

	(1) IV-GLS	(2) IV-GLS	(3) IV-GLS	(4) GLS	(5) GLS	(6) GLS
Mismatch	-0.0197***	-0.0082	-0.0049	-0.0185***	-0.0120***	-0.0122**
Mismatch × Occ Tenure		-0.0020***	-0.0022**		-0.0011**	-0.0008
Cumul Mismatch			-0.0347***			-0.0365***
Worker Ability (Mean)	0.2064***	0.2061***	0.2994***	0.1947***	0.1943***	0.3079***
Worker Ability × Experience	0.0113***	0.0117***	0.0100***	0.0116***	0.0117***	0.0074***
Worker Ability × Occ Tenure	0.0021	0.0011	-0.0012	0.0017	0.0014	0.0034
Occ Reqs (Mean)	0.0978***	0.0962***	0.0878***	0.1374***	0.1363***	0.1421***
Occ Reqs × Occ Tenure	0.0145***	0.0144***	0.0182***	0.0085***	0.0085***	0.0093***
Observations	41596	41596	30599	41596	41596	30599
R^2	0.244	0.244	0.214	0.258	0.258	0.230

Note: $***\ p < 0.01$, $**\ p < 0.05$, $*\ p < 0.1$. All regressions include a constant, terms for demographics, occupational tenure, employer tenure, work experience, and dummies for one-digit-level occupation and industry. Standard errors are estimated via GLS assuming AR(1) autocorrelation.

Table J.6 – Wage Losses Predicted by the Regression with Worker's Ability Times Experience Term

Mismatch Degree	*Mismatch* Effect			*Cumul. Mismatch* Effect
(High to Low)	5 years	10 years	15 years	
90%	-0.048	-0.080	-0.113	-0.114
	(0.011)	(0.018)	(0.030)	(0.014)
50%	-0.021	-0.035	-0.050	-0.060
	(0.005)	(0.008)	(0.013)	(0.008)
10%	-0.008	-0.013	-0.018	-0.027
	(0.002)	(0.003)	(0.005)	(0.003)

Note: Wage losses (relative to the mean wage) are computed for each percentile of each measure, using the result of the specification (3) in Table J.5. Standard errors are in parentheses.