

Skill Uncertainty, Skill Accumulation, and Occupational Choice

Carl Sanders*

Washington University in St. Louis

September 29, 2011

Abstract

Workers entering the labor market are uncertain about their skill set. Standard human capital theory assumes workers have perfect information about their skills. In this paper, I argue that skill uncertainty can explain one type of worker moves that standard human capital theory cannot: moves between jobs where they perform different kinds of tasks. I consider workers who have a multi-dimensional bundle of labor market skills and begin their careers uncertain about their skill levels. I construct a model that links learning about skills to the tasks performed in jobs: the more intensely a job uses a particular skill, the more the workers learn about their true level of that skill. The model also contains a skill accumulation motive: as workers use a skill they gain additional amounts of it. A simplified version of the model suggests that if skill uncertainty were the dominant force workers would switch between jobs that use skills in different ratios but similar total levels. On the other hand, if skill accumulation were the dominant force they would switch between jobs that use similar ratios of skills but higher total levels. Linking data on workers from the National Longitudinal Study of Youth 1979 with occupational characteristics from the US Department of Labor O*NET database, I show that worker mobility across different task mixes is common and I estimate the model parameters. The current results indicate that skill uncertainty explains approximately 30% of worker mobility across different task ratios.

*Email: carlesanders@wustl.edu. Website: <http://sites.google.com/site/carlesanders>. Thanks to Christopher Taber, John Kennan, and Salvador Navarro for their help and feedback on this project. All remaining errors are mine.

1 Introduction

Workers entering the labor market are uncertain about their skill set. Most theories of labor markets abstract away from this uncertainty. In particular, in standard human capital theory the worker knows his current skill levels when he chooses the optimal path of investment. In this paper, I argue that allowing for skill uncertainty in addition to skill accumulation can rationalize patterns in job mobility that are difficult to explain with only human capital accumulation. My focus is the fact that workers, particularly young workers, often switch between jobs where they perform tasks which require skills that they have not used before. This is difficult to explain by skill accumulation, since previously accumulated skills are not useful for the new tasks. For example, consider a worker who switches from being a door-to-door salesman to a real estate salesman. This move is easily understandable in a human capital framework: the worker developed sales skills and then moved into a job where those skills were more highly rewarded. On the other hand, consider a worker who begins his career as an economist and then moves into being a professional athlete. The skills developed as an economist will not be much help in athletics. If the worker is very uncertain about his skills and needs to experiment with them to find out his true levels, however, it makes sense to attempt two different jobs where the skill requirements are very different because learning more quickly can allow him to make better decisions in the future.

In this paper, I look at the ability of skill uncertainty and experimentation to explain worker mobility across different types of jobs both theoretically and empirically. I create a model with workers who may not know their true underlying skills. Using the model, I show that workers who are uncertain about their skills will move across different mixes of tasks, and once they have learned their skills they will move between occupations that have similar mixes of tasks. Using data on worker occupational choice over time linked with measures of the tasks they perform, I show that this intuition is consistent with the data. Finally, I estimate the model parameters and run counterfactual simulations, which allow me to quantify the relative importance of skill uncertainty and skill accumulation for workers' early career task mobility.

The model is an extension of models of skill-specific human capital accumulation. Workers have a set of skills and develop them over their career. My innovation is to allow workers in the model to be uncertain about their skills and choose to accumulate information about them. I construct an equilibrium model where workers choose between occupations that differ in terms of the tasks performed. Workers do not know their true skill levels, but make decisions based on their current beliefs. In making their decision, workers consider the wages each occupation offers, the amount of skills they would accumulate, and the amount of information about their true skill levels the occupation would reveal. Occupations differ across all three of

these characteristics, and I analyze workers' incentives and optimal choices. If skill accumulation were the dominant force, mobility across occupations would be between occupations that use a similar ratio of tasks as the previous one but a higher total amount of tasks performed. On the other hand, if skill uncertainty were the dominant force, job mobility would consist of moves across different ratios of tasks but similar total levels.

To test the predictions of the model, I use panel data from the National Longitudinal Survey of Youth 1979 on workers' wages, experience, and occupational choices over time. To determine the tasks they performed on the job, I link the occupational records with measures of tasks performed within occupations taken from the US Department of Labor O*NET survey. Using the linked data, I can see worker transition between tasks over time. In the data, there is significant worker mobility between occupations that have different task mixes. Also, workers switch between occupations that use different task mixes early in their career before switching between occupations which use similar ones. This is consistent with the predictions of the model if workers have uncertainty about their initial skill levels.

Even if worker skill uncertainty is generally consistent with the data, this does not necessarily mean it is quantitatively important for worker mobility. To evaluate the relative importance of skill uncertainty versus skill accumulation, I estimate the underlying parameters of the model. Using these estimates, I simulate counterfactuals by changing the parameters governing skill uncertainty so simulated workers have perfect knowledge of their skills. Comparing the counterfactual simulation to the data, I estimate that skill uncertainty plays a significant role in explaining worker mobility across different task mixes. Current results indicate that eliminating skill uncertainty would reduce this mobility by approximately 30%.

The remainder of this paper is organized as follows: Section 2 summarizes the previous work on models of learning and skill accumulation, Section 3 analyzes the model, Section 4 describes the data, Section 5 and Section 6 show the empirical methodology and estimation, Section 7 summarizes the results, and Section 8 concludes.

2 Previous Literature

I argue that skill uncertainty can provide an explanation for why agents move across jobs that require different mixes of skills within a human capital framework. In this section, I discuss how my paper relates to the existing literature on worker mobility across tasks.

The literature of skill-specific human capital looks at more general forms of human capital than the "firm-specific" / "general" distinction that is typically made. In these models human capital is through a vector that has different possibilities for substitution across different areas of the labor market. For example, the Roy (1951) model can be thought of as allowing for a 2-

dimensional vector of skills where each skill is only used in one sector. Later extensions of the Roy model such as Heckman and Sedlacek (1985) allowed for a general vector of human capital that could be utilized at different rates across sectors.

With new data sets that allowed the observation of multiple types of individual ability, such as test scores for cognitive and non-cognitive ability, and data about the types of tasks workers perform in jobs, there is an emerging literature that studies the importance of different types of skills for labor market outcomes. Among these are Heckman et al. (2006), Poletaev and Robinson (2008), Abraham and Spletzer (2009), Borghans et al. (2010), and Robinson (2010), which look at the returns to different types of skills and the movement of workers across different tasks when they switch jobs. Models where human capital is occupation-specific can be seen as a case of the skill vector literature as well. The dynamic programming model of Keane and Wolpin (1997) allowed for investment into two different types of occupational human capital, white collar and blue collar, which could not be substituted with each other. Kambourov and Manovskii (2009) show that empirical evidence suggests that what has commonly assumed to be returns to firm-specific tenure are actually primarily explained by occupational-specific tenure, which is evidence that occupation-specific capital plays a large role in wage growth.

The majority of previous papers on multi-dimensional skills use static models. This paper contributes to a recent literature that extends the idea of multi-dimensional skills to a dynamic decision framework. Cunha et al. (2010) estimate a dynamic model of parental investment in children, where adult outcomes for the children depend on the levels of both cognitive and non-cognitive skills. Yamaguchi (2010) estimates a dynamic model of endogenous learning by doing with a multi-dimensional set of skills where workers choose which skills to invest in. My model is very similar to Yamaguchi's, but I allow for workers to ask based on beliefs about their skills and to rationally experiment across different tasks in order to learn them.

I also extend the literature of worker learning to a more general setting. The analysis of models where workers learn about their productivity began with Jovanovic (1979), who modeled the match between a worker and a firm as an experience good, so individual workers would need to spend time with a firm before determining whether it was a good match or not. Recent extensions of the model to full equilibrium settings are Pries and Rogerson (2005), Moscarini (2005), and Woodcock (2010). Much of the recent work on extending the Jovanovic model has emphasized other types of matches that the worker may be uncertain about. For example, one literature has modeled match effects between workers, firms, and also occupations. Miller (1984) and McCall (1990) study the optimal path of occupational mobility if workers can choose from risky or safe occupations but need to spend time in the risky occupations in order to learn their true productivity, and Papageorgiou (2009) estimates a model of occupational uncertainty directly. More recently, Groes et al. (2009) use a model of uncertainty about the occupational

match to explain the empirical U-shape of occupation mobility. Neal (1999) and ? use models with career matches, where a career consists of a series of occupations that are good substitutes for one another in terms of tasks performed and skills required.

The task-based approach taken here is effectively looking inside the "black box" of these occupational effects. The advantage is that it is possible to directly compare two occupations on the task dimensions. On the other hand, it completely restricts the worker-occupation match to not have any other forms of heterogeneity.

The other paper to analyze skill uncertainty and experimentation is Antonovics and Golan (2011). They create a model where workers can endogenously take lower wages in order to learn about their skills, and they study the predictions of the model and show it can fit some stylized facts from panel data on occupational choice, occupational characteristics, and wages. Their model does not allow for skill accumulation, and their focus is primarily theoretical. My contribution to their paper is to provide a model with human capital accumulation as well as skill uncertainty and experimentation and to attempt to empirically distinguish the two factors. In addition, I allow the worker's problem to be derived from a full equilibrium model and estimate the parameters of the model, allowing me to look at simulations of counterfactuals.

This paper is to my knowledge the only one to analyze the question of differentiating empirically between skill uncertainty and skill accumulation. However, this question is closely related to empirically distinguishing worker-firm match effects from increasing productivity due to firm-specific human capital. In both cases, looking at the relationship between wages and tenure (or in the case here, amount of time spent using skill) is not informative. This is because workers who have learned they are a bad match will have abandoned the firm, so the selection effect is that wages will be increasing over tenure. Here, looking at the cross-section of individuals who have spent a long time in a given occupation, their wages will be higher than those who have just entered the labor market, but it is unclear whether this is selection or skill accumulation. Farber (1994), Farber (1999), and Nagypal (2007) have studied methods of decomposing wage growth over time within a firm to these influences. My solution is different from theirs because I use data not just on which occupations workers chose, but also how similar those occupations are in tasks.

3 Model

The model augments the literature on skill-specific human capital accumulation with a learning motive. In contrast to models with only skill uncertainty or only skill accumulation, including both allows the model to show the differences the two motives imply for observed behavior. In particular, models with only skill uncertainty cannot match the fact that over time workers

tend to move from lower task levels to higher ones. On the other hand, models with only skill accumulation cannot explain moves across different task mixes to jobs where accumulated skills would be less productive.

The other contribution of this model to the skill uncertainty literature is modeling the worker's problem as an equilibrium outcome. Previous work has studied a reduced-form decision problem of an individual worker without analyzing whether that problem could exist in equilibrium. In my model setup, I am unable to reproduce the reduced forms that other papers use for the wage function. The main reason for this is that the worker's flow utility will tend to depend on higher moments of their belief distribution, not just the mean as other papers assume. While I model the full equilibrium of the economy, for now I do not estimate the parameters of the full equilibrium; instead, counterfactuals are only valid at the individual level, and they cannot affect aggregate quantities. Estimation of a full equilibrium that would allow for more robust counterfactuals is a topic for future research.

In this section, first assuming a fixed set of wage contracts facing workers I formalize their occupational choice problem. Second, I model the determination of the wage contracts and show how the occupation-specific prices change the incentives for the workers in equilibrium. Finally, I discuss a reparametrization of the problem in terms of the specialization and magnitude of tasks and show how it can be used to capture the distinction between occupational moves due to skill accumulation and moves due to skill uncertainty.

3.1 Worker Decision Problem

The economy consists of a size 1 mass of workers who choose each period $t = 1 \dots T$ to work in one of a fixed set of occupations, $j = 1 \dots J$. All markets are competitive. Workers have an endowment of a vector of skills, here assumed to be two-dimensional, and throughout I will refer to them as “cognitive” and “manual” skills and denote their values as s_c and s_m . Workers do not know these skills exactly, but instead have beliefs about them. Let b_c and b_m denote the probability distributions that capture a worker's beliefs about his skills. Workers have flow utility that is logarithmic in consumption,

$$u(c) = \log c$$

and there is no borrowing or saving, so workers consume all their income and denoting wages recieved as W ,

$$u(c) = u(W) = \log W \equiv w$$

where lower case will be used to denote logarithms.

Occupations differ according to their technology. While the determination the firm/worker contract will be explained in the “Equilibrium Wage Determination” section below, intuitively since occupations have different technologies the marginal product of a given worker will differ across them, which leads to different wages offered to the same worker in different sectors. I define an occupation as a vector (τ_c, τ_m) , which denotes the task intensity of the occupation for tasks that use cognitive skills, τ_c , and the intensity of tasks that use manual skills, τ_m . Any two firms with the same τ vector are considered the same occupation, and within occupations all firms are identical. The support of technology is normalized to the unit square $(\tau_c, \tau_m) \in [0, 1] \times [0, 1]$. Firms have the same information set as workers, and all contracting is ex ante one period wage posting. Denote the wage offered to a worker with beliefs b at an occupation with technological parameters (τ_c, τ_m) as

$$W(b_c, b_m; \tau_c, \tau_m)$$

From the functional forms assumed later, it will be approximately true that the “task” score reflects the elasticity of individual-level output with respect to individual-level skills. Firms then differ by their responsiveness to these individual-level skills. For example, there are firms that have cognitive task scores of 0; i.e. the cognitive skills of the worker are irrelevant to their productivity on the job. On the other hand, a firm with a manual task score of 1 means that a 1% increase in the worker’s manual skills would lead to approximately a 1% increase in that workers output. Since individual skills are not directly measureable here, one could also implicitly define a unit of skills as the amount of ability a worker would need to increase for a 1% increase in output at some fixed job, then define all other jobs relative to that one. An alternative specification of the worker’s problem could involve choosing the number of tasks to perform in some fixed timeframe and their skills allowing different speeds of task completion, but the model adopted here is closer to the existing literature. The advantage is that it leads to parsimonious functional forms for the wage equation.

The model analyzed below will assume that beliefs about skills are normally distributed,

$$s_c \sim \mathcal{N}(\mu_c, \sigma_c^2), s_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$$

for some constants $\mu_c, \mu_m, \sigma_c^2, \sigma_m^2$. This means the wage function can be written

$$W(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m)$$

which is the wage that a workers with beliefs (μ_c, σ_c^2) and (μ_m, σ_m^2) would receive at an occupation with technology (τ_c, τ_m) . At this point I do not make assumptions on the functional form of

the wage function, since it is determined in equilibrium. However, in equilibrium the function W will be smooth in all its arguments and parameters.

In this next section I will analyze a two-period version of the model in order to show its essential features. In the second period, the worker solves the static problem. The problem for the worker is to choose the occupation to work in to maximize wages. The value function in the second period is

$$V_2(\mu_c, \sigma_c, \mu_m, \sigma_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} w(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m)$$

where w denotes the logarithm of the wage function. For the worker, the current states are the beliefs about his skill levels, and the controls are the choice of occupation. There is no guarantee that the wage function is concave so that first order conditions would be sufficient to characterize the optimum, but by the extreme value theorem there is at least one value of (τ_c, τ_m) that achieves the maximum of the wage function on the unit square for some fixed values of the μ and σ . Assume this value is unique and denote it $(\tau_c^*(\mu, \sigma), \tau_m^*(\mu, \sigma))$, so that

$$V_2(\mu_c, \sigma_c, \mu_m, \sigma_m) = w(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c^*(\mu, \sigma), \tau_m^*(\mu, \sigma))$$

What this equation says is that the value of having a set of beliefs in the second period is equal to the wage the worker would receive when choosing the optimal occupation based on those beliefs. The optimal policy for the worker is to choose the wage-maximizing occupation in the final period.

The first period value function is

$$V_1(\mu_c, \sigma_c, \mu_m, \sigma_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} w(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m) + \beta E[V_2(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m) | \tau]$$

where the primes indicate the values in the second period. The interpretation of this is that the worker's value depends both on the amount of wages he will receive today, but also on the evolution of his beliefs between today and tomorrow. The contribution of this model is through the two channels that beliefs are changed: first, workers receive a signal about their true skill levels and update their beliefs to reflect this new information, and second, those true skill levels change through a skill accumulation process. Both of these are endogenous: choosing higher task intensities leads to both the worker receiving a higher amount of information as well as the true skills increasing more. This will shift the incentives of the worker in the first period away from static wage maximization.

There are no switching costs in this model. This is simply for computational reasons, their addition would require two additional dimensions for the state space, since the worker would

need to keep track of their occupation in the previous period, whereas now their current beliefs are a sufficient statistic for their entire work history.

The information accumulation process works as follows: between periods, the worker observes the realization of two random variables, X_c and X_m , which are defined by

$$X_c \equiv s_c^0 + \frac{\varepsilon_c}{\tau_c}, \quad X_m \equiv s_m^0 + \frac{\varepsilon_m}{\tau_m}$$

where

$$\varepsilon_c \sim \mathcal{N}(0, \sigma_{\varepsilon_c}^2), \varepsilon_m \sim \mathcal{N}(0, \sigma_{\varepsilon_m}^2), \varepsilon_c \perp \varepsilon_m$$

The X contain information about the true levels of the initial endowments of skills s^0 . The amount of information is dependent on how intensely the worker's chosen occupation uses that skill. For example, if the worker chooses a task that does not use cognitive skills at all, $\tau_c = 0$, then the variance of the ε_c random shock is infinite and it is impossible to tell what the s is. On the other hand, if workers could choose tasks with $\tau = \infty$, the variance of the ε would be 0 and the value of X the worker sees would be his true skill level. This learning process is also used by Antonovics and Golan (2011). In contrast with models that require the signal to be based on the wage, this signal could contain a significantly larger amount of information than just the wage observation. This is consistent with the idea that workers receive many more signals than just a wage realization, such as day-to-day feedback from a supervisor that is unobservable to an econometrician.

The updated priors come from rational Bayesian updating. The key part of the update is the weighting function,

$$\omega_c(\tau_c, \sigma_c) \equiv \frac{\sigma_c^2}{\sigma_c^2 + \frac{\sigma_{\varepsilon_c}^2}{\tau_c^2}} = \frac{\tau_c^2 \sigma_c^2}{\tau_c^2 \sigma_c^2 + \sigma_{\varepsilon_c}^2}$$

and symmetrically for manual skills. The term $(\tau_c \sigma_c)^2$ can be interpreted as the “signal,” the amount of information the worker gets conditional on choice of task intensity, and σ_{ε}^2 is the amount of “noise” inherent in the process that slows down learning. The weighting function ω_c is increasing in τ_c and σ_c (and τ_m and σ_m for ω_m , manual skills) and has a positive cross-partial derivative, $\frac{\partial^2 \omega}{\partial \tau \partial \sigma} > 0$.

The Bayesian update for the workers beliefs about the mean of their skills is

$$\mu'_c = \omega_c(\tau_c, \sigma_c) \cdot X_c + (1 - \omega_c(\tau_c, \sigma_c)) \cdot \mu_c$$

and symmetrically for manual skills. This is a weighted average of the worker's signal and his previous beliefs. As the worker takes higher task jobs, the weighting function increases, and the updated beliefs put more emphasis on the signal and less on the previous beliefs. Additionally,

workers with higher levels of uncertainty about their skills, σ_c , put more weight on the signal relative to a worker who has more information.

The Bayesian update precision of the worker's beliefs is

$$\sigma_c'^2 = \sigma_c^2 \cdot (1 - \omega_c(\tau_c, \sigma_c))$$

and symmetrically for manual skills. The decrease of the variance of the worker's beliefs in the next period is proportional to the weighting function. The variance will decrease faster as workers take more intense jobs, and similarly will decrease faster if they have higher levels of uncertainty.

The second way in which a worker's beliefs will change between periods is because of a change in their true skills levels. The workers have a skill accumulation process

$$s'_c = s_c + \tau_c^\alpha R_c (\gamma_0 + \gamma_1 t + \gamma_2 t^2)^{-1} - \delta_c, \quad 0 \leq \alpha \leq 1$$

where α is a shape parameter for the relationship between task intensity and human capital accumulation, and the γ represent the decrease in the efficacy of human capital accumulation over time. This equation says that a worker's true skills are increased each period by a factor that includes R_c , a positive constant known by the worker but not the econometrician, the intensity of task usage τ_c , and decreased by a depreciation term δ_c . Any period, workers who take higher intensity jobs will gain more skills relative to those who take low intensity jobs. I assume that the additional amount a worker can learn decreases with the time index t , so even workers in the highest job in the economy accumulate 0 additional skills as time goes to infinity. This assumption is required to make long-term wage growth eventually go to 0. This type of assumption is standard in learning-by-doing models such as Jovanovic and Nyarko (1996) and Nagypal (2007).

The full two period problem can be written

$$V_1(\mu_c, \sigma_c, \mu_m, \sigma_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} w(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m) + \beta E[V_2(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m) | \tau]$$

$$V_2(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} w(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m; \tau_c, \tau_m)$$

$$\mu'_c = \omega_c(\tau_c, \sigma_c) \cdot X_c + (1 - \omega_c(\tau_c, \sigma_c)) \cdot \mu_c + \tau_c^\alpha R_c (\gamma_0 + \gamma_1 t + \gamma_2 t^2)^{-1} - \delta_c$$

$$\sigma'_c = \sigma_c \cdot \sqrt{1 - \omega_c(\tau_c, \sigma_c)}$$

$$\omega_c(\tau_c, \sigma_c) \equiv \frac{\tau_c^2 \sigma_c^2}{\tau_c^2 \sigma_c^2 + \sigma_{\varepsilon_c}^2}$$

Extending the problem to any finite number of periods has the same form as the value function in the first period, and can be solved by backwards recursion.

Workers enter the labor market with an initial endowment of skills (s_c^0, s_m^0) that are drawn from distributions f_c and f_m . I assume that these distributions are independent. This assumption is not required to write the problem, but was assumed above in the Bayesian updating formulas. In the empirical work I need to observe the true skill intensities τ . If the skills are not independent, it is easy to orthogonalize the distributions and other parameters, but the τ must be rescaled as well, and the observed intensities would not correspond to the true intensities. I also do not observe anything that can identify the rotation parameters. Independence of initial skills on labor market entry is a strong assumption, and would not arise from a model of investment prior to labor market entry, such as Cunha et al. (2010).

At labor market entry workers receive an initial signal about their skills. Workers observe $X_c^0 \equiv s_c^0 + v_c$ and $X_m^0 \equiv s_m^0 + v_m$. If the variance of v are 0, then workers have perfect knowledge of their initial skills, and if the variance is infinite, workers have no more information than the econometrician. I do not assume that workers are completely uninformed about their skills, but I estimate the amount of information workers begin their careers with by estimation the variance of the initial signal.

3.2 Equilibrium Wage Determination

An innovation of this model is not assuming the functional form of the worker's decision problem, but instead treating it as determined by the equilibrium wage structure. Antonovics and Golan (2011) assume that the worker's flow utility is linear in the first moments of his beliefs, but do not derive it from an equilibrium model. My model is a true partial equilibrium model since I analyze the worker's problem holding equilibrium aggregates constant, but the fact that the problem arises in equilibrium puts restrictions on the space of possible worker decision problems.

The firm side of the economy consists of occupations $j = 1 \dots J$ that produce occupation-specific intermediate goods using workers as inputs. Define $S_c \equiv \exp(s_c)$ and symmetrically for manual skills, so that S_c and S_m are both always positive. Since there is uncertainty about the worker's skill levels, firms only choose workers to hire to maximize expected profits. Let each occupation produce output according to a Cobb-Douglas technology in labor and capital, where labor is a CES aggregator function of the two individual skills, and the technology parameters can differ across occupations. For occupation j

$$y_j = L_j^{\alpha_j} K_j^{1-\alpha_j}$$

$$L_j = \int (\tau_{cj} S_{ci}^\rho + \tau_{mj} S_{mi}^\rho)^{\frac{1}{\rho}} dG(S_c, S_m)$$

where G is the distribution function (and g denotes the density) that corresponds to the measure of workers with skills S that (endogenously) choose that occupation. The marginal product of worker i is then (with some abuse of notation)

$$\frac{\partial y_j}{\partial i} = \alpha_j L_j^{\alpha_j-1} K_j^{1-\alpha_j} \frac{\partial L_j}{\partial i}$$

From the individual's perspective, α_j, L_j and K_j are exogenous. Denote the combination of occupation-specific exogenous variables as v_j . Assuming independence of skills for simplicity of notation, the marginal product is then

$$\frac{\partial y_j}{\partial i} = v_j \frac{\partial L_j}{\partial i} = v_j (\tau_{cj} S_{ci}^\rho + \tau_{mj} S_{mi}^\rho)^{\frac{1}{\rho}} g_c(S_{ci}) g_m(S_{mi})$$

There is also a competitive constant returns to scale final goods market that purchases the intermediate goods and aggregates them into a final consumption good Y using CES technology

$$Y(y_1, \dots, y_J) = \left(\sum_{j=1}^J \kappa_j y_j^\gamma \right)^{\frac{1}{\gamma}}$$

where κ_j are exogenous technologically-determined share parameters and γ governs the elasticity of substitution between occupation-specific intermediate goods. The cost function for the final goods market is the cost of purchasing the intermediate goods on J competitive markets at prices p_j

$$c_Y(y_1, \dots, y_J) = \sum_{j=1}^J P_j y_j$$

so the profit function for the final goods market (normalizing the price of the final output good to 1) is

$$\pi_Y = \left(\sum_{j=1}^J \kappa_j y_j^\gamma \right)^{\frac{1}{\gamma}} - \sum_{j=1}^J P_j y_j$$

The optimal demands for each of the intermediate goods are given by the standard first order conditions, and the total level of output is set by the zero profit condition.

In the intermediate goods market, the expected profit function for occupation j is

$$\pi_j = P_j y_j - \int_{i_j} W dG(S_c, S_m)$$

where i_j is an index set of all the individuals who choose to work in occupation j for a given set

of prices and wages. Since firms can contract with individual workers, occupations must earn zero expected profit per contract, so the wage contract must be

$$W(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m) = P_j E \left[\frac{\partial y_j}{\partial i} \right] =$$

which is

$$W(\mu_c, \sigma_c, \mu_m, \sigma_m; \tau_c, \tau_m) = P_j v_j E \left[(\tau_{cj} S_{ci}^\rho + \tau_{mj} S_{mi}^\rho)^{\frac{1}{\rho}} \right]$$

and log wages are

$$w = \psi_j + \log E \left[(\tau_{cj} S_{ci}^\rho + \tau_{mj} S_{mi}^\rho)^{\frac{1}{\rho}} \right]$$

The wage a worker receives has three components: his productivity within the occupation, the aggregate price the final goods market pays to that occupation, and the occupation-specific fixed effect that comes from the aggregate levels of labor and capital that occupation employs. The aggregate price depends on the final goods technology, particularly the share parameter κ_j , which determines how important the intermediate good is in the production of the final consumption good. It also depends on the amount and skill composition of workers who want to choose that occupation at given prices: if more workers desire to work there than the final goods market demands, the prices adjust downward until the wage makes the supply of workers and the associated intermediate goods equal to the demand from the final goods market.

The empirical importance of this equilibrium exercise has to do with the importance of allowing for occupation-specific fixed effects. It would require extremely restrictive assumptions on this model for the wage function to not require data on occupation-specific measures of aggregate labor levels, capital levels, and technology. However, using occupation-specific fixed effects can allow for unobserved differences in labor, capital, and prices, and makes it feasible to only use data on the task requirements.

The occupation-specific prices play a similar role to the task-specific prices in Heckman and Sedlacek (1985). In Antonovics and Golan (2011), implicitly the output of every occupation has the same price. In their model, if one occupation had a higher return to both skills than another, no workers would work in the worse occupation. They constrain their occupation space so that no occupation uniformly dominates another. Here, prices regulate the supply of workers to high productivity jobs. If the output price were constant across all sectors, every worker would choose to be a surgeon, airline pilot, or nuclear physicist. But the oversupply of workers to those jobs lowers the market price and also lowers wages, decreasing the amount of workers who want to be in that occupation until equilibrium is reached.

It is possible to get closed form solutions for wages in the case of $\rho = 0$, a Cobb-Douglas

aggregator. In this case, one can show the wage function is of the form

$$W = E[S_c^{\tau_c}] E[S_m^{\tau_m}]$$

Recall I defined $S_c \equiv \exp(s_c)$ which implies that $S_c^{\tau_c} \equiv \exp(s_c \tau_c)$. s_c is normally distributed with parameters (μ_c, σ_c^2) , so $s_c \tau_c$ is normal with parameters $(\tau_c \mu_c, \tau_c^2 \sigma_c^2)$, and using the expected value of the log normal distribution,

$$E[\Delta y_j] = E[S_c^{\tau_c}] E[S_m^{\tau_m}] = \exp\left(\tau_c \mu_c + \frac{1}{2} \tau_c^2 \sigma_c^2\right) \exp\left(\tau_m \mu_m + \frac{1}{2} \tau_m^2 \sigma_m^2\right)$$

and the worker's log wage received becomes

$$w = \log(P_j E[\Delta y_j]) = \psi_j + \tau_c \mu_c + \tau_m \mu_m + \frac{1}{2} \tau_c^2 \sigma_c^2 + \frac{1}{2} \tau_m^2 \sigma_m^2$$

where the lower case ψ_j includes both the aggregate capital and labor at the occupation as well as the log occupation-specific price. The interpretation of this is that the worker receives wages from:

1. The occupation-specific goods price P_j : the higher the price of that good, the higher the wage, since it raises ψ_j .
2. Having higher skill levels: as μ increases, the wage increases conditional on choice of tasks τ .
3. Choosing the right tasks: if the μ is positive, higher tasks pay higher wages; if negative, lower tasks pay higher wages.
4. Higher levels of uncertainty: wages are an increasing function of the σ .

The first three make intuitive sense, however, the last feature of the model makes a strong assumption on the relationship between idiosyncratic uncertainty and wages. The intuition for why higher levels of uncertainty can lead to higher wages is simple: if bad realizations of worker skill do not hurt output very much but high realizations are very profitable, conditional on the mean of beliefs about skills a worker with a higher variance will be more productive on average, and thus will be paid higher wages.

The fact that wages are increasing in uncertainty is only a result of the functional form assumptions above, however. In the CES model with elasticities of substitution other than one, numerical experiments show that the direction of the effect of variances on wages can be either positive or negative. However, the CES wage function requires numerical integration, which

makes it computationally infeasible given the current complexity of the problem. To use the spirit of 1-3 above without restricting the response of wages to uncertainty in 4 and for computational feasibility, I use a linear approximation. The wage function I will use for the empirical work is

$$w = \psi_j + \tau_{cj} (\mu_{ci} + \eta_c \sigma_{ci}) + \tau_{mj} (\mu_{mi} + \eta_m \sigma_{mi})$$

where ψ is an occupation-specific effect and the η are estimated parameters that estimate the effect of uncertainty on wages. In future work estimating the full equilibrium wage function could allow for more direct estimates of the production function.

3.3 The Specialization/Magnitude Task Decomposition

In this section, I discuss a tool I use to differentiate skill accumulation from skill uncertainty. In previous sections, I have claimed skill accumulation motives can explain why workers take higher level jobs, but not jobs that differ in the skill mix. In this section I formalize this intuition in the model. To do this, I create a new measure of an occupation's task bundle that can be used to show when human capital moves could lead to moves across different task mixes and when skill uncertainty is the driving force.

The standard way that skill-based models view the task bundle in an occupation is Cartesian coordinates; with two skills, the task requirements for occupation j are $\tau_j = (\tau_c, \tau_m)$. Here I will use the polar representation of the occupation's task requirements in the empirical work. The reason for this is that the polar coordinates $\tau_j = (\theta, \|\tau\|)$ where $\theta = \arccos\left(\frac{\tau_c}{\|\tau\|}\right)$ and $\|\tau\| = \sqrt{\tau_c^2 + \tau_m^2}$ have a natural interpretation in terms of specialization and total levels of task requirements. The second coordinate, $\|\tau\|$ or the distance from the origin of the τ_j vector, is the total amount of task complexity the occupation requires. The θ , the angle from the x-axis, is the specialization of the occupation in the cognitive skills task. A $\theta = 0$ represents a worker who uses only cognitive skills, and $\theta = \frac{\pi}{2}$ is a worker who uses only manual skills, regardless of their total magnitude of tasks. See Figures 1 and 2 for a graphical representation of this transformation.

Using the intuition from the model that the task scores reflect the elasticity of output with respect to skills, consider a firm where output responds 3% to a marginal increase in cognitive skills and 4% to a marginal increase in manual skills. One potential way to quantify the concept “the overall level of responsiveness to skills” could be by adding together the elasticities, but this mathematically leads to a non-natural metric that is used to compare different jobs. The approach adopted uses the more natural Euclidean norm on two dimensions to look at the overall level of responsiveness, and then the angle of the polar coordinates reflects the amount which is due to cognitive skills.

The reason this decomposition is useful is because I interpret similar levels of θ as similar

“career” type of occupations. Consider the case of the career path of college biology major to medical student to resident to surgeon. It seems plausible that the mix of manual dexterity to cognitive ability required is approximately unchanged across this path. But in terms of magnitudes, the college student certainly has the occupation with the lowest total task intensity, with the surgeon having the most. I will interpret moves across different levels of θ as moves across different “career ladders”, and moves with similar θ and higher $\|\tau\|$ as moves up one career ladder. I provide a list of occupations with their θ and $\|\tau\|$ scores in Table 4.

While there are no results for the full model because of its complexity, it is possible to use the $\theta/\|\tau\|$ decomposition in a simplified static version of the problem. Write the static version of the problem with a linear approximation of the wage function without any terms reflecting uncertainty:

$$V(\mu_c, \mu_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} \psi(\tau_c, \tau_m) + \tau_c \mu_c + \tau_m \mu_m$$

and assume that $\psi(\tau_c, \tau_m) = p_0 - a_1 \tau_c^2 - a_2 \tau_m^2$, which is a quadratic approximation to the pricing function where I have set the linear terms to 0. The problem then becomes

$$V(\mu_c, \mu_m) = \max_{(\tau_c, \tau_m) \in [0,1]^2} p_0 - a_1 \tau_c^2 - a_2 \tau_m^2 + \tau_c \mu_c + \tau_m \mu_m$$

Factoring the right hand side,

$$\max_{(c_1, c_2) \in [0,1]^2} p_0 - a_1 \left[\tau_c^2 + \frac{a_2}{a_1} \tau_m^2 \right] + \tau_c \mu_c + \tau_m \mu_m$$

Without loss of generality I can rescale τ_m to $\hat{\tau}_m = \tau_m \times \sqrt{\frac{a_1}{a_2}}$, and the term within the maximum is

$$p_0 - a_1 [\tau_c^2 + \hat{\tau}_m^2] + \tau_c \mu_c + \sqrt{\frac{a_2}{a_1}} \hat{\tau}_m \mu_m$$

and now I define $\|\tau\| = \sqrt{\tau_c^2 + \hat{\tau}_m^2}$ and $\theta = \arccos\left(\frac{\tau_c}{\|\tau\|}\right)$, and then by definition $\tau_c = \|\tau\| \cos \theta$ and $\tau_m = \|\tau\| \sin \theta$, so this equation is

$$p_0 - a_1 \|\tau\|^2 + \|\tau\| \cos \theta \cdot \mu_c + \sqrt{\frac{a_2}{a_1}} \|\tau\| \sin \theta \cdot \mu_m$$

and the value function is

$$V(\mu_c, \mu_m) = \max_{(\theta, \|\tau\|)} p_0 - a_1 \|\tau\|^2 + \|\tau\| \left(\mu_c \cos \theta + \sqrt{\frac{a_2}{a_1}} \mu_m \sin \theta \right)$$

This maximization problem in two variables can be solved by a two-step process: first, choose the value of θ that maximizes the term inside the parentheses, then given that value choose the

optimal magnitude. Intuitively, the worker can choose the specialization based on the ratio of their beliefs regardless if their skill levels are high or low, and then choose the magnitude based on their overall skill levels. The value of θ that maximizes this is $\theta^* = \arctan\left(\frac{\mu_m}{\mu_c} \sqrt{\frac{a_1}{a_2}}\right)$, which is independent of $\|\tau\|$. If this were a repeated static model with human capital accumulation, that is, the model with the discount factor $\beta = 0$, then if μ_1 and μ_2 rise proportionally from skill accumulation the optimal choice of θ^* remains the same over time, even as skills grow, but the optimal value of $\|\tau\|$ increases. On the other hand, with uncertainty but no skill accumulation, there would be shocks to the relative values of $\frac{\mu_m}{\mu_c}$. This would lead to different values of θ , but on average the fact that μ_c and μ_m are not systemically increasing would mean the value within the parentheses above would remain about the same, which would mean the optimal choice of $\|\tau\|$ remained about the same. This is another way of saying that skill accumulation type moves are ones with similar θ and changing $\|\tau\|$, while learning moves are moves with changing θ and stable $\|\tau\|$. In Sections 5 and 6 I use this intuition to look at the evidence regarding switching across θ .

4 Data

The data I use comes from merging the National Longitudinal Survey of Youth 1979 (NLSY79) with the U.S. Department of Labor Occupational Characteristics Database, O*NET. The NLSY79 is a representative sample of US households that was administered yearly from 1979-1994 by the Bureau of Labor Statistics, and once every two years since. The Weekly Labor Status variables can be used to create a weekly history of workers and the firms they work for, along with some characteristics of the job such as the hourly wage, hours worked, occupation, and industry. Additionally, the NLSY contains demographic variables such as age, sex, race, education, marital status, and scores on the Armed Forces Qualification Test (AFQT).

I restrict the NLSY79 sample to white male high school graduates who never attended college from the core sample who have left school and have a firm attachment to the labor force. Specifically, I define a full time worker as one who works 30 or more total hours in a week at any employers. While there is no obvious definition for attachment to the labor force, I use criteria similar to Neal (1999) and Yamaguchi (2010) where workers who work full time at least 20 of the past 24 weeks are considered attached to the labor force, otherwise I do not use those observations. The time of labor market entry is then the first period where the worker is considered firmly attached to the labor force after leaving school. Additionally, I drop workers who are ever in military service, since it makes sense they will have significantly different career paths than non-military individuals. From these weekly observations, I create monthly variables that contain the average wage paid during the previous month, the actual labor force experience of the

worker, and the time since labor market entry as defined above. In addition, I merge in variables from the O*NET database of occupation characteristics.

The O*NET database contains occupational characteristics for a wide range of occupations across the US economy, at varying levels of detail. O*NET is an update of the older Dictionary of Occupational Titles (DOT), which was criticized for being a non-representative sample of occupations, for not following standard survey design, and for poor data quality.¹ The current design of O*NET has two stages, a worker survey stage and a job evaluation stage. In the survey stage, a nationwide random sample of workers complete a survey consisting of 8 different questionnaires, such as “Skills,” “Education and Training,” and “Generalized Work Activities.” An example question from the “Skills” questionnaire is shown in Figure 3.

In the evaluation stage, job evaluation experts are given the responses to read and fill out an “Abilities” questionnaire that is supposed to reflect the skill sets of the average individual in the occupation. All means and variances from both stages are released to the public, but not the individual-level responses.

I use the scores recorded in the O*NET database to construct objects that I interpret as the τ 's in the model. In particular, if I assume I am looking for two τ for each job, the cognitive and manual task intensity, I need to determine how to reduce the information in many responses on the survey down to two variables. To do this, I follow Yamaguchi (2010) and first group the different questions into similar types of questions, where one group are questions that are related to cognitive tasks and the other group are related to manual tasks. I then use Principal Component Analysis to derive a single factor for both groups. This procedure creates a score for cognitive and task complexity as a weighted average of the scores on the individual questions, where if two questions tend to have the same responses in all occupations they receive less weight. The main function of this procedure is to attempt to rank jobs in terms of their cognitive skill and manual skill usage. I take the scores and rescale them between 0 to 1, corresponding to the occupation that uses the least amount of that task in the data and the job that uses that task the most, respectively. Summary statistics for the O*NET database is shown in Table 1, listings of some selected occupations and their task scores are shown in Table 2, and histograms of the distributions of characteristics across populations are shown in Figures 4 and 5.

Looking at Figure 3, the question is made up of two parts: the importance of the skill, and the level of the skill. While this could possibly have interesting identifying information, in practice they scores are almost perfectly correlated and there isn't much to gain by using both for each question. Instead, I just use the “level” score for each variable. Quantitative results for the regressions below are unchanged using the other score.

I combine the NLSY79 and O*NET scores by merging the data sets using a crosswalk be-

¹See Miller (1980) for a critical review of the DOT by the National Research Council.

tween the occupation codes in the NLSY79 and the 2000 Standard Occupational Classification codes used in O*NET. To create this crosswalk, I took the titles from the occupation codes in the NLSY79 and searched the O*NET website for the corresponding title. In the majority of cases there was an exact match between titles, or a closely corresponding title. In the majority of non-matched cases, the occupation was one occupation in the NLSY codes but is two in O*NET or vice versa. The unmatched observations are counted as missing data. In future versions I will attempt other ways to get around the measurement problem of these missing occupations and the combined occupations, but over 70% of occupations were matched. The O*NET occupational classifications are more detailed than the NLSY occupation codes, containing up to 10 sub-classifications of some occupations. Compared to some studies of task-specific human capital such as Autor and Handel (2009), I have data at a higher degree of aggregation, which could induce problems if there are wide amounts of skill dispersion within occupations that I do not observe.

O*NET contains a good deal of information other than on “cognitive” and “manual” skills, which themselves are broad definitions. For example, there is information about “interpersonal” tasks like in Borghans et al. (2010). However, here I only consider the division of cognitive and manual, since those are the two most common divisions of tasks for looking at the wages of high school graduates. Other divisions (or more than two dimensions) are an area for future research.

The last step is to create monthly τ_c and τ_m scores for each individual. Since workers often work multiple jobs during a month, one possibility is to just take the task scores from the job worked at the most. Instead I create a week-weighted average of the scores, so if an individual works 2 weeks at a job where $(\tau_c, \tau_m)_1 = (.8, .2)$ and 2 weeks at a job where $(\tau_c, \tau_m)_2 = (.4, .6)$, the total score given the month will be $(\tau_c, \tau_m)_{total} = (.6, .4)$. This assumes that working half a month at a manually-intensive job and then half a month at a cognitively-intensive job is the same as working for a month year at a job that is an average of both. This does not follow from the model, but if the amount of information and skill accumulation that happens within a month is small it can serve as an approximation.

After data cleaning and restricting the sample, I work with 840 individuals from the NLSY from 1978-2008 with an average length in the sample of 25.2 years. For the O*NET sample, I am left with data from 440 occupations to merge into the NLSY, and after using the crosswalk I am able to create yearly average task scores for 35,208 of the 39,462 (89%) yearly observations where a worker was firmly attached to the labor market. For some occupations I was unable to match task scores, so these occupations are treated as missing when I create the yearly averages. If these occupations are missing data for some systematic reason this may be a problem, but most of the non-matches are where the recorded occupation are general categories like “Teacher,

Other” or “Manager.” Summary statistics for the NLSY data are shown in Table 3.

5 Evidence for Worker Learning

In the first empirical section, I will discuss some stylized evidence that workers experiment to learn about their skills. This can be interpreted as additional evidence beyond that of Antonovics and Golan (2011), who have a closely related model without skill accumulation. Skill accumulation in my model complicates matters, since both learning and skill accumulation generate endogenous occupational moves, and both can explain the decreasing transitions between occupations over time. In the model section I argued that moves across different specializations are consistent with a model where workers have to learn their skills. In this section, I analyze the empirical distributions and changes of the specialization and magnitude variables both across and within workers to see if these patterns are consistent with the model.

In order to differentiate between skill accumulation and skill uncertainty, I use the task data to create the specialization variable θ as a proxy for the career path that an occupation is in. As workers accumulate skills, I assume that it only moves them up the task magnitude measure $\|\tau\|$ and use the measures of θ to look at experimentation across different career paths. The rest of this section is organized as follows: first, I discuss the distribution of the specialization and magnitude statistics in the data and look at the rates of change across specializations. Second, to look for evidence of experimentation I adapt an argument from Neal (1999) to my context and argue that optimal search behavior should lead to workers making “complex” changes, that is changes between occupations that perform very different tasks, before making “simple” changes, which are job switches to occupations that perform similar tasks.

Overall sample summary statistics for the values of the specialization and magnitude measures can be found in Table 3. This table shows that even though the sample average values of τ_c and τ_m (also in the table) are similar, they tend to be used in bundles where manual tasks make up a higher proportion of the overall task bundle. Looking at the aggregate levels does not give much information about individual-level dynamics, however, which are the focus of this section.

Table 10 reports the means, medians, and standard deviations for the variables $\Delta\theta$ and $|\Delta\theta|$, the signed value and absolute value of a worker’s change in specialization since the last period respectively. The mean of $\Delta\theta$ is nearly 0, which indicates that workers tend to remain using a similar mix of tasks between periods. This is true even conditioning on switching occupations. The standard deviation of $\Delta\theta$ decreases slightly over time, which shows that there is a decrease in unusually large or small moves over time. The size of moves, regardless of direction, is also decreasing over time looking at the median of $|\Delta\theta|$.

The dispersion of workers across θ is not simply given by an initial distribution of workers across jobs, but worker mobility within a career plays a significant role. The values in Table 10 help to summarize the amount of switching within individuals over time, as opposed to just a fixed distribution of workers across skills. Looking at the standard deviation of $\Delta\theta$, the value is around half of the total sample standard deviation of θ . Movement across θ within workers over time is a significant aspect of their careers, not just their initial level of θ .

Just showing that workers move across different θ is not direct evidence that workers move systemically across occupations since observed movement could reflect random shocks or measurement error. I present two related pieces of evidence to support the claim that this movement is consistent with experimentation. First, I create a variable every year for whether there was a “simple change” or a “complex change” from the previous year. A “simple change” is defined as one where the change of the value of θ from the previous year is less than the sample median of all the year-to-year changes, and a “complex change” is when the change is greater than that. Neal (1999) first used these terms as definitions of different types of worker moves. He defines simple changes as ones that involve a change of firm but not of career, and complex changes to be changes of firm and career. While our models are quite different and there are no firm-level effects in my model, the intuition that experimentation involves making complex moves before simple moves largely carries over here.

In my case, if the change in θ every period was drawn from the empirical distribution in the sample, I can calculate the probability that complex changes would never follow simple changes given any total number of observations. I then look at the number of workers in my sample where they actually follow that pattern. The results are shown in Table 5. They indicate that across different total numbers of possible changes, workers tend to follow the “complex change first” / “simple change second” pattern more often than random chance would suggest.

There are a couple things to note about the results from Table 5. First, the results are similar even conditioning on firm switches like Neal does. Second, the magnitude of the gaps between the data and random movement are about the same as Neal’s for a small number of possible moves, but not quite as large as his effects for the numbers of possible moves from 7 and up. Still, the values are uniformly higher than the predicted probabilities from random movement. This is especially true for individuals with a high number of moves. If movement was random almost no one in the data set would follow the pattern that is observed for approximately 5% of them.

There could be a few problems with the above analysis that do not come directly from the model, but may still be important. For one thing, perhaps the declining number of complex changes is just because of decreasing occupation to occupation or firm to firm mobility. Since my skills data is only at the occupational level, if there was some other reason that occupational

mobility declined, it would look the number of complex changes were declining even if workers would still switch across skill sets if they were to switch occupations. Similarly, it could just reflect a decrease in firm-to-firm movements. With fixed occupation-specific or firm-specific matches, which I do not include in my model, if the worker finds a good match and stays there it adds a significant number of 0's (which simple changes) to the changes in θ , which would raise the chances of the pattern “complex change” / “simple change” even though it has nothing to do with skills. To try to check whether this is what is happening, I repeat the analysis above for two different conditions. First, I condition on there being a change in primary occupation between the years. This raises the median size of the change in θ (because of the elimination of all the 0's), so I redefine a complex change to be the median change conditional on the change being positive and recreate the table. The second check I do is condition on firm changes, calculate the new median size of the θ move, and repeat the analysis again. While the tables are not included here, the results are basically the same as the unconditional version of the analysis, although the sample sizes in some of the cells get quite small. For each total number of changes, the pattern “complex changes first” / “simple changes later” is systematically overrepresented in the data compared to random movement. Since the results for the conditional analysis are so similar, I use the unconditional data for the regressions I run below.

The second set of stylized facts come from controlling for individual-level heterogeneity. One way the results above could be spurious is if individual workers differed according to their probability of making complex changes, perhaps just because they had preferences for a higher or lower amount of occupational mobility than others. If there are some types of workers who are less likely to make simple changes and others who are less likely to make complex changes, the pattern would tend to look like the results above, since there would tend to be clustering of one type of changes within a career, and any pattern of just one type of change is consistent with the two-stage search argument. To try to control for this individual heterogeneity, I run a fixed effects regression of the size of current changes in specialization on the total amount of past changes in specialization. With only skill accumulation, changes in θ should be correlated with time since entry but not necessarily the previous number of simple changes. With uncertainty, the expected sign of the coefficient β_1 is negative, which is to say that as the number of times the worker does not move to an occupation that uses significantly different skills increases, he is less likely to make that move in the current period. Table 6 shows the coefficients from this estimation. Formally,

$$|\Delta\theta|_{it} = \beta_0 + \beta_1 \sum_{k=0}^{t-1} |\Delta\theta|_{ik} + \beta_2 \times TimeSinceEntry_{it} + \beta_3 \times TimeSinceEntry_{it}^2 + u_i + \varepsilon_{it}$$

While these coefficients are difficult to interpret, the negative estimate for the β_1 coefficient in-

indicates that the more the worker moved earlier in his career, the less likely he is to move in the current period, as long as the individual-specific heterogeneity in moving type is controlled for with the fixed effects. If the regression is run without the individual-specific effect, the estimate for β_1 is positive and significant. Adding the fixed effects helps to control for the fact that individuals who are more likely to make larger moves in each period are more likely to have a large sum of past moves, which would make it look like past moves caused current moves.

In summary, looking at changes in the θ measure over a worker's lifetime is consistent with a model where workers need to experiment not just across different occupations or firm but also across different tasks in order to learn their skills. In the next section, I will use the model explicitly to determine the importance of skill uncertainty for mobility across θ .

6 Model Estimation

While the discussion above shows that there are aspects of the data that are consistent with a model of worker learning, it does not help to quantify the effects of learning compared to skill accumulation. If the quantitative implications of endogenous learning are small it may not be worthwhile to consider it further. To quantify the effects of skill uncertainty on task mobility, I take the model above and estimate it directly from the data. Using the estimates, I can simulate counterfactual worker careers, changing the amount of initial skill uncertainty or rate of skill accumulation.

The full model can be written

$$\begin{aligned}
V_t(\mu_c, \sigma_c, \mu_m, \sigma_m) &= \max_{\tau \in [0,1]^2} w_j(\mu_c, \sigma_c, \mu_m, \sigma_m) + \beta E[V_{t+1}(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m) | \tau] \\
V_{T+j}(\mu'_c, \sigma'_c, \mu'_m, \sigma'_m) &= 0, \quad j > 0 \\
w_j &= \psi_j + \tau_{cj}(\mu_{ci} + \eta_c \sigma_{ci}) + \tau_{mj}(\mu_{mi} + \eta_m \sigma_{mi}) \\
\psi_j &= a_0 + a_1 \tau_c + a_2 \tau_m + a_3 \tau_c \tau_m + a_4 \tau_c^2 + a_5 \tau_m^2 \\
\mu'_k &= \omega_k(\tau_k, \sigma_k) \cdot X_k + (1 - \omega_k(\tau_k, \sigma_k)) \cdot \mu_k + \tau_k^{\alpha_k} R_k(\gamma_{0k} + \gamma_{1k}t + \gamma_{2k}t^2)^{-1} - \delta_k, \quad k = \{c, m\} \\
\sigma'_k &= \sigma_k \cdot \sqrt{1 - \omega_k(\tau_k, \sigma_k)}, \quad k = \{c, m\} \\
\omega_k(\tau_k, \sigma_k) &\equiv \frac{\tau_k^2 \sigma_k^2}{\tau_k^2 \sigma_k^2 + \sigma_{\varepsilon_k}^2}, \quad k = \{c, m\}
\end{aligned}$$

The first equation is the value function in period t , which says that the worker chooses an optimal occupation to maximize their log wage in the current period plus the expected discounted continuation value. The second equation says that the worker's career ends at time

T , which I set to 40, since I use yearly observations. The third equation is the wage function that determines what workers of given beliefs matched with each occupation will receive. The occupation-specific prices ψ_j that enter the log wage function are unobserved. The prices could be estimated as occupation-specific effects, but there would be one additional parameter to estimate for each occupation in the data set. There are 440 occupations in the data and at this point I cannot estimate this number of parameters. Instead, I estimate the parameters of pricing function that is quadratic in the task intensities τ . The fourth equation shows the form of the function, and the coefficient vector a is estimated within the model. The fifth equation is the transition of mean beliefs about the workers skills, the sixth the transition of the precision of those beliefs, and the last equation is the definition of the weighting function workers use to update their beliefs. I assume that I observe the worker's choice of tasks τ , their log wage w , and their time since labor market entry t .

I do not observe any of the worker's state variables, since I do not have any data on their beliefs about their skills levels. This is the opposite of a standard utility-maximization problem where the state (e.g. the worker's income) is observed but the level of utility is not. From observation of the age, choice, and wage, I try to back out the unobserved state. These types of problems are referred to as Hidden Markov Models in the inference literature. I cannot use typical solution algorithms from the HMM literature since my problem has endogenous transitions between states, and typical HMM algorithms assume a fixed exogenous Markov transition matrix.

The parameters of interest are the initial distribution of worker skills, the distribution of their initial beliefs about their skills, the variance of the noise ε terms that dictates the speed of learning, and skill accumulation rates R_1 and R_2 , the skill accumulation shape parameters α_c , α_m and the γ , and the depreciation rates δ . I will discuss the identification of these parameters below in the "Indirect Inference Moments" section.

Since there is no closed-form solution to the problem, I use numerical methods to solve it. I use backwards induction on the value function with Chebyshev interpolation of the value function in each period and numerical integration of the continuation values.

Because of the computational complexity of the problem, even solving the individual's problem is difficult, since the value function has 4 continuous states and there is a double integral with no closed form solution inside the continuation value. Estimating the model is even more computationally demanding, since I do not observe the states, but only the controls and wages. Instead of directly solving the model, I will estimate the model using Indirect Inference, which has had some success in estimating learning models such as Guvenen and Smith (2010). The advantage is that this model is easier to simulate from than estimate: simulating the data for one individual is orders of magnitude easier than calculating the likelihood for that individual. The

essence of indirect inference is to simulate data from the model such that certain aspects of the simulated data compare well with the same aspects of the true data set. Included in this general method are Simulated Method of Moments and Simulated Maximum Likelihood estimators. I will estimate parameters by matching a series of moments from the data with the simulated data. Another important part of indirect inference is that the ways the model matches the data should be driven by the identification of the model. Here that leads me to choose moments related to wage growth, occupational switching rates, and the relationship between changes in θ and time in the labor force, all of which I used in claiming evidence for the model in the first estimation section.

6.1 Indirect Inference Moments

6.1.1 Wages

The model has clear implications for wage growth. Workers become more productive through skill accumulation over time, while also gaining more information about their true skills levels. The first set of moments I use related to wages is a cross-sectional regression of wages onto time since labor market entry and the task choices of individuals,

$$E[w_{it}|e_{it}, c_{it}, m_{it}] = \beta_0 + \beta_1 e_{it} + \beta_2 e_{it}^2 + \beta_3 c_{it} + \beta_4 m_{it}$$

where w is log wages, e is years since labor market entry, c is choice of cognitive task intensity, and m is choice of manual task intensity. Effectively, matching this regression from the true data will mean that the means of wages are similar within experience/task choice cells. Since wages are closely related to the unobserved underlying skill distribution, matching these coefficients will help to recover workers' underlying skills.

The sample values for the mean wage regressions from the NLSY are shown in Table 7. While my sample is restricted slightly differently these results are consistent with Topel and Ward (1992), who find that wages approximately double over the first 20 years of workers' careers. Considering the effects of model parameters of the simulated moments, the effects of R_c and R_m , the skill accumulation rates, are obvious: higher levels lead to more wage growth because workers become more productive. The effects of other parameters on this moment are more subtle. For example, changes any parameters that lead to workers being more uncertain about their skills at labor market entry (such as increasing the variance of the initial signal) lead to higher levels of wage growth, whereas changes to parameters that lead to less learning during the career (such as increasing the variance of the noise σ_ϵ) lead to lower wage growth.

Since skill heterogeneity across workers plays an important part in the model as well, I at-

tempt to match within experience/task choice cell heterogeneity as well as the overall averages. To do this, consider the true residuals from the mean wage regression:

$$\varepsilon_w = w - \beta_0 - \beta_1 e_{it} - \beta_2 e_{it}^2 - \beta_3 c_{it} - \beta_4 m_{it}$$

$$E[\varepsilon_w | e_{it}, e_{it}^2, c_{it}, m_{it}] = 0$$

I linearly project the conditional variance of the unobserved heterogeneity onto the independent variables to get

$$E[\varepsilon_w^2 | e_{it}, c_{it}, m_{it}] = \gamma_0 + \gamma_1 e_{it} + \gamma_2 e_{it}^2 + \gamma_3 c_{it} + \gamma_4 m_{it}$$

which can be estimated by least squares. In practice, I form the residuals from the estimates of the cross-sectional wage equation and regress the squares of those observed residuals onto the independent variables. As above, this is an estimate of the amount of unobserved wage heterogeneity within experience and task choice cells. The estimates for this regression are shown in Table 8.

In addition to the cross-sectional regression, matching the relationship between individual mobility and wage changes captures the dynamic aspect of the model. As one measure of this relationship, I match the coefficients of the regression

$$E[\Delta w_{it} | \Delta \tau_{it}, \Delta \theta_{it}] = \nu_0 + \nu_1 \Delta \tau_{it} + \nu_2 \Delta \theta_{it}$$

where Δ is the first difference operator within individuals. This regression says that the average change in wages individuals experience is systematically related to their shifts in overall task magnitude and their changes in specialization. The panel nature of this regression is robust to the initial selection of workers into occupations based on unobservables.

The estimates for this regression are shown in Table 16.

6.1.2 Task Intensity

The model also has predictions about the levels and changes for task intensity in the economy. Because of the higher occupation-specific prices of low τ occupations, when workers begin their career with low skill levels the wage-maximizing occupations are the ones with lower task intensities. As they accumulate skills, they move up the occupation ladder to higher τ occupations. This is only exactly true in a repeated static model, however. With skill accumulation and learning, higher τ_c and τ_m jobs are more attractive because of the opportunity to accumulate additional skills and learn about skill levels, both of which will have a positive value in the fu-

ture. It is possible that these effects could completely offset each other, such that the increasing τ_c and τ_m that come from productivity gains could be offset by decreasing τ_c and τ_m motives once skill accumulation and learning become less important. The variance of the τ 's also has information about the distribution of s_c and s_m across the economy.

Similar to the wage regressions, I use two regressions to recover the conditional mean and variances of task choices, although here I only condition on worker experience. This will capture the overall task choices in the economy, the transition in the average choices over time, as well as the unobserved heterogeneity of task choices across time. The 12 additional moments I will match are the coefficients on the regressions

$$E[c_{it}|e_{it}] = \beta_0^c + \beta_1^c e_{it} + \beta_2^c e_{it}^2$$

$$E[m_{it}|e_{it}] = \beta_0^m + \beta_1^m e_{it} + \beta_2^m e_{it}^2$$

$$E[\varepsilon_c^2|e_{it}] = \gamma_0^c + \gamma_1^c e_{it} + \gamma_2^c e_{it}^2$$

$$E[\varepsilon_m^2|e_{it}] = \gamma_0^m + \gamma_1^m e_{it} + \gamma_2^m e_{it}^2$$

where as before the ε are the true regression errors from the respective conditional means of the task choices.

The values for these regressions are shown in Tables 9 and 11-14.

6.1.3 Distribution of Changes in θ

The last set of moments use changes within individuals across time to try to differentiate skill accumulation from learning. My argument in the model focused on changes in θ to determine when a move is because of skill accumulation and when it is because of learning. Large moves in θ imply that individuals chose to look for something different in terms of specialization, since moves generated by skill accumulation will tend to make the next occupation have a similar mix of tasks. I use two different regressions to consider the relationship between moves across specializations.

The first regression I match is looking at the conditional mean of the size of changes in specializations across time,

$$E[|\Delta\theta|_{it}|e_{it}, \|\tau\|_{it}] = \eta_0 + \eta_1 e_{it} + \eta_2 e_{it}^2 + \eta_3 \|\tau\|_{it}$$

Matching these coefficients requires matching the mean level of mobility in the economy as well as its change over time, all conditional on the total magnitude of the worker's tasks. Estimates from this regression are included in Table 15.

In order to match more of the individual-level mobility dynamics from the data, I also use the fixed-effects regressions from the first part of the estimation section. The next set of moments are the estimated parameters from the fixed-effects regression

$$|\Delta\theta|_{it} = \beta_0 + \beta_1 \sum_{k=0}^{t-1} |\Delta\theta|_{ik} + \beta_2 \times TimeSinceEntry_{it} + \beta_3 \times TimeSinceEntry_{it}^2 + u_i + \varepsilon_{it}$$

In the data, the coefficient on β_1 was negative, which meant that individuals who had more mobility across θ were less likely to move in the current period. If the data generated by the model matched these effects, it would be because the information value of a different type of job had decreased due to earlier information. Estimates of these coefficients are reported in Table 7.

6.2 Estimation Procedure

To estimate the model, I choose model parameters to minimize the objective function

$$\hat{\gamma} = \arg \min_{\tilde{\gamma}} g(\tilde{\gamma}) = \arg \min_{\tilde{\gamma}} (h(\tilde{\gamma}) - m)' W_n (h(\tilde{\gamma}) - m)$$

where m is a $k \times 1$ vector of the moments from the data, $h(\gamma)$ is the function that maps any $l \times 1$ vector of the parameters into a $k \times 1$ vector that contains the same moments from the simulation, and W_n is a positive definite weighting matrix. For each possible value of γ , to get $h(\gamma)$ I solve the individual's problem to generate the optimal policy for those parameter choices, and then I simulate 10,000 worker careers and calculate the indirect inference moments summarized above. I then search over values of γ until I have reached a local minimum, and repeat from different starting values to try to ensure I can find a global minimum to the objective function. For the weighting function, I follow Blundell et al. (2008) and use the diagonal matrix that contains the inverse of the standard errors of the moments.

To solve the individual problem for each set of parameter values, I use standard backwards induction of the value function in each period. Beginning in the last period, I use Chebyshev interpolation to generate coefficients for a polynomial that approximates the value function in that period, then in the previous period I do the same thing, using the interpolated continuation value. Getting the expected continuation value requires a numerical integration, which I do using Gauss-Hermite integration. Because of the structure of the problem, when it comes time to simulate the careers of the workers, I can just use the interpolated expected continuation value instead of having to do the integral of the value function again. Because of this, simulating worker careers is very fast relative to getting the Chebyshev coefficients the first time through.

Calculating the Chebyshev coefficients takes around 2 seconds per period, which simulating 10,000 workers takes around the time to interpolate one period. The largest use of time for calculation comes from the numerical integration in the expected continuation value.

7 Results

In this section, I summarize the estimation results and the model fit. Tables 14, 15, and 16 show the values of the data moments and the corresponding moments the model generates at the estimated parameters.

The estimated parameters are shown in Table 17. While the parameter values themselves do not have clear interpretations outside of simulations of the model, a few things can be seen directly. I estimate that cognitive skills are more widely dispersed across the population than manual skills, which reflects the fact that the wage gap between high and low cognitive jobs is larger than the gap between high and low manual task jobs. The skill accumulation rate is estimated to be almost 0 for manual type jobs. This is because workers do not move to higher manual intensity tasks on average, so in the model they must not be accumulating additional manual skills. This result is consistent with the estimates of Yamaguchi (2010), whose estimates are from a model with only skill accumulation.

The other parameters reported in Table 17 are the estimates of the pricing function coefficients. When workers are deciding between occupations the differences in prices are all they consider, so I show the partial derivatives of the price function in the table along with the estimated parameters. The results indicate that higher cognitive and manual skill occupations are more desired by workers in the model and the lower prices for those occupations are used to reduce the amount of workers in that occupation.

The overall model fit for the aggregate parameters varies. For wages, shown in Table 14, the time trend is consistent with the data, but the model estimates that the effects of cognitive and manual tasks levels are too high relative to the data. With other forces acting on wages such as frictions or contracting, some of the wage growth experienced by workers would be attributed to those instead of the growth in task choices. The variance of wages is predicted to start too low and grow too quickly to make up for the difference.

The model fit for tasks shown in Table 14 is generally good, with the aggregate trends of both cognitive and manual task growth replicated by the model. The growth in the dispersion of both types of tasks are too high relative to the data, particularly for manual tasks. The results for the mobility moments in Table 16 show that the model can capture the general trends of mobility with some exceptions. Of particular note is the fact the model predicts the wrong sign for the relationship between magnitude and the size of moves; workers in higher overall intensity jobs

tend to move more across specializations. The model can fit the qualitative relationship between past moves across specializations and current moves, though the magnitude of some of the estimations are off by around half.

My estimates of the amount of information that workers possess at labor market entry are quite low: using the estimates of the variances of the initial signal from Table 17 implies that workers' precisions about their beliefs are only 5% below the population precision. Looking at the model fit for wages in Table 14, the model fit for wage means and variances is good. The variance of wages grows slightly faster in the model than the data. Similarly, the model fit for cognitive tasks is overall quite close. The model overestimates the average manual task levels in the economy, and they do not decrease quite as quickly as they do in the data. The main area where the model has trouble fitting the aggregates is in the variances of task choices. The simulated standard deviations are around half as large as those observed. The reason for this is that the model needs to have small variances of skills in order to match the standard deviations of wages, but then at this distribution of skills workers will all take very similar jobs. One potential reason for this is that I do not allow for measurement error in tasks.

Along with aggregate moments, I use individual-specific moments related to task mobility. Table 16 shows the estimates from those moments. Similar to the aggregate variances, the model predicts less mobility across different specializations than is observed in the data since the constant in the regressions are similar but the coefficient on experience is much more negative in the model, meaning mobility decreases too quickly. In Table 16 I also show that the model accurately reproduces the pattern of the effect of the size of past moves on current moves that I discussed in Section 5 above.

Overall, the model can do a good job reproducing the aggregate moments of the data as well as the general trends in the individual-level time series. The largest amount of room for improving the fit lies in the fact that workers' task choices are too similar to each other and they do not move enough across different tasks in the simulation. Allowing for more general substitution patterns between the skills could lead to workers choosing a wider variety of tasks even with similar skill levels.

7.1 Counterfactuals

I have emphasized the role that skill uncertainty can play in explaining the fact that workers move across jobs of different specializations. In this section I will show that in the estimated model, skill uncertainty and experimentation play a major role in this phenomenon. To do this, I use the estimated model parameters and change those that govern the amount of skill uncertainty so that workers have perfect information about their skills, keeping all other parameters

the same. I then simulate the model under both sets of parameters and look at the changes induced in the movement of workers across jobs with different task mixes. If a large amount of the movement across specializations would be eliminated with perfect worker knowledge, that is evidence that experimentation can play an important role in using models of skill accumulation to explain task mobility.

From the model, there are three primary ways that workers can move across different specializations. First, if the skill accumulation rates are different, e.g. if cognitive skills accumulate faster than manual skills the increase in the level of cognitive tasks workers use is faster than manual tasks, and workers would not stay at the same angle from the x-axis as before. Second, the equilibrium pricing function could potentially drive worker to be more specialized at higher task levels than they would be at lower task levels. Third, workers could experiment across different specializations in order to learn their skill levels early on. The first two reasons are ways the standard skill accumulation model can explain mobility across different specializations, and the third is a contribution to the model. From the parameter estimates it is clear that the skill accumulation rates are different across skills and that the pricing function is not symmetric over skills. To determine how much mobility skill uncertainty explains, it is necessary to simulate the model.

The results from the counterfactual simulations are shown in Table 16. The first column shows the values of three measures of the magnitude of moves in the simulated data at the true parameter values. The first measure is the mean of $|\Delta\theta|$, which is the average value of a worker's move between the last period and the current one in the sample. The second measure is the median of the size of moves, and the third is the standard deviation of the size of moves. The mean and median are both measures of the overall tendency for workers to move to jobs where they perform tasks with different specializations than they use in their current job, and the standard deviation summarizes the fact that the size of moves differs across workers.

The second column in Table 16 shows the results of a simulation where the variance of the worker's initial signal is set to 0 and holding the rest of the parameters fixed. The thought experiment this corresponds to would be each worker being told their true skill levels as soon as they enter the labor market, and then they solve their lifetime problem where the only dynamic motive is skill accumulation. The results indicate skill uncertainty decreases the mean size of moves by 45%, the median of the size of moves by 22%, and also decreases the standard deviation of the size of moves by 26%. Intuitively, since workers no longer need to use skills they are uncertain about to learn about them, there is less of a need to move across different specializations. While there is not one statistic that summarizes the amount the distribution of moves changes by, it seems reasonable to summarize these results by saying that skill uncertainty can explain about 30% of movement across different task specializations.

8 Conclusion

Workers entering the labor market are uncertain about their skill set. In this paper, I argue that worker skill uncertainty can help explain worker mobility across jobs where they perform different types of tasks, a fact that standard models of skill accumulation typically ignore. To do this, I combine a model of worker skill uncertainty and skill accumulation with panel data on workers' tasks over their career. Intuition from the model suggests that workers will move across jobs where they use tasks in different mixes when they are uncertain about their skills. Mobility patterns across different specializations in the data are consistent with the model. I estimate the parameters of the model and find that skill uncertainty accounts for approximately 30% of mobility across different specializations.

The model extends previously used models of skill-specific human capital accumulation to allow for skill uncertainty and worker experimentation in an equilibrium framework. Workers choose what tasks to perform each period, but instead of basing their decisions on their true skill levels they act on beliefs about those skill levels and experiment in order to gain information. Using the model, I show that the larger the skill uncertainty and value of experimentation, the more workers will move across jobs that use significantly different tasks. In the model, taking jobs where workers perform different tasks than they did before has the highest value early in the career, when they have not worked in similar jobs before. I also allow worker skills to change over time through skill accumulation, which comes from a learning-by-doing mechanism. The incentives for investment are stronger in occupations where the worker believes he is already high skilled.

To look at whether the data is consistent with the implications of the model, I combine panel data on worker careers from the National Longitudinal Survey of Youth 1979 with data on the tasks performed in occupations from the US Department on Labor O*NET database. From this data I decompose the worker's chosen bundle of tasks into measures of "specialization" and "magnitude," and document that moves across different specializations are common in the data.

I then estimate the parameters of the model and simulate counterfactuals to determine the empirical importance of skill uncertainty in explaining moves across different specializations. Results from the simulations indicate that if workers knew their true skill levels at labor market entry, the size of changes across specializations would be reduced by approximately 30%. This suggests that worker experimentation plays an important role in worker mobility across tasks.

For future research, I will look at counterfactuals that allow the wage structure to shift in equilibrium. In this paper, the counterfactuals are only partial equilibrium and hold wage contracts fixed, but if it were possible to estimate the parameters of the labor demand side of the

market I could use this framework to analyze the effects of shifts on the amount of worker skill uncertainty on the equilibrium distribution of wages.

References

- Abraham, K. and Spletzer, J. (2009). New evidence on the returns to job skills. *American Economic Review*, 99(2):52–57.
- Antonovics, K. and Golan, L. (2011). Experimentation and job choice. *Working Paper*.
- Autor, D. and Handel, M. (2009). Putting tasks to the test: human capital, job tasks and wages. *NBER Working Paper 15116*.
- Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption Inequality and Partial Insurance. *American Economic Review*, 98(5):1887–1921.
- Borghans, L., Ter Weel, B., and Weinberg, B. (2010). Interpersonal styles and labor market outcomes. *Journal of Human Resources*, Forthcoming.
- Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.
- Farber, H. (1994). The analysis of interfirm worker mobility. *Journal of Labor Economics*, 12(4):554–593.
- Farber, H. (1999). Mobility and stability: the dynamics of job change in labor markets. *Handbook of Labor Economics*, 3:2439–2483.
- Groes, F., Kircher, P., and Manovskii, I. (2009). The U-Shapes of occupational mobility. *Working Paper*.
- Guvenen, F. and Smith, A. (2010). Inferring labor income risk from economic choices: an indirect inference approach. *NBER Working Paper*.
- Heckman, J. and Sedlacek, G. (1985). Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *The Journal of Political Economy*, 93(6):1077–1125.
- Heckman, J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3).

- Jovanovic, B. (1979). Job matching and the theory of turnover. *The Journal of Political Economy*, 87(5).
- Jovanovic, B. and Nyarko, Y. (1996). Learning by doing and the choice of technology. *Econometrica: Journal of the Econometric Society*, 64(6):1299–1310.
- Kambourov, G. and Manovskii, I. (2009). Occupational specificity of human capital. *International Economic Review*, 50(1):63–115.
- Keane, M. and Wolpin, K. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.
- McCall, B. (1990). Occupational matching: a test of sorts. *Journal of Political Economy*, pages 45–69.
- Miller, A. (1980). *Work, jobs, and occupations: A critical review of the Dictionary of Occupational Titles*. Natl Academy Pr.
- Miller, R. (1984). Job matching and occupational choice. *The Journal of Political Economy*, 92(6):1086–1120.
- Moscarini, G. (2005). Job matching and the wage distribution. *Econometrica*, 73(2):481–516.
- Nagypal, E. (2007). Learning-by-doing versus learning about match quality: can we tell them apart? *Review of Economic Studies*, 74(2):537–566.
- Neal, D. (1999). The complexity of job mobility among young men. *Journal of Labor Economics*, 17(2):237–261.
- Papageorgiou, T. (2009). Learning your comparative advantages. *Working Paper*.
- Poletaev, M. and Robinson, C. (2008). Human capital specificity: evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000. *Journal of Labor Economics*, 26(3).
- Pries, M. and Rogerson, R. (2005). Hiring policies, labor market institutions, and labor market flows. *Journal of Political Economy*, 113(4):811–839.
- Robinson, C. (2010). Occupational mobility, occupation distance and basic skills: evidence from job-based skill measures. *Working Paper*.
- Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135.

- Topel, R. and Ward, M. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479.
- Woodcock, S. (2010). Heterogeneity and learning in labor markets. *The BE Journal of Economic Analysis & Policy, Advances*, 10(1):85.
- Yamaguchi, S. (2010). Tasks and heterogeneous human capital. *Working Paper*.

A Tables and Figures

Figure 1: Occupations Characterized in Cartesian Coordinates

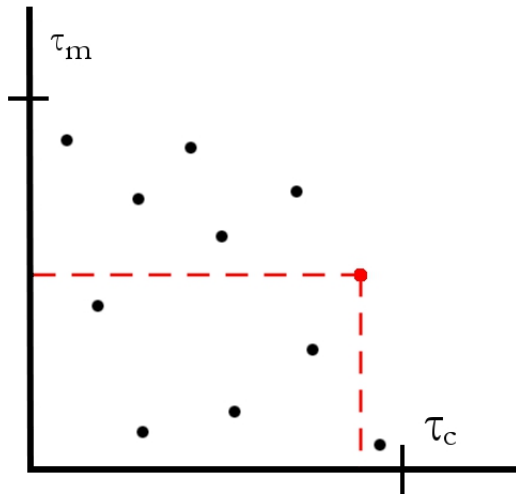


Figure 2: Occupations Characterized in Polar Coordinates

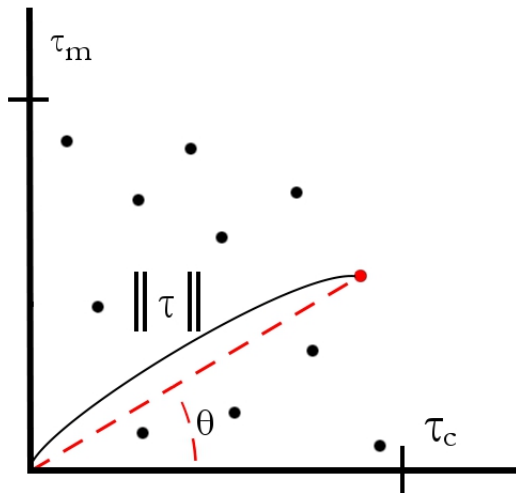


Table 1: O*NET Summary Statistics, Occupation Weighted			
	Mean	Std. Dev.	N
Cognitive Tasks, τ_c	.49	.21	440
Manual Tasks, τ_m	.45	.234	440
$\text{Corr}(\tau_c, \tau_m) = -.21$			
The minimum for both task measures is 0 and the maximum is 1.			

Figure 3: Example Question from O*NET Survey

5. Mathematics **Using mathematics to solve problems.**

A. How important is MATHEMATICS to the performance of *your current job*?

Not Important*	Somewhat Important	Important	Very Important	Extremely Important
①	②	③	④	⑤

* If you marked Not Important, skip LEVEL below and go on to the next skill.

B. What level of MATHEMATICS is needed to perform *your current job*?

Count the amount of change to be given to a customer	Calculate the square footage of a new home under construction	Develop a mathematical model to simulate and resolve an engineering problem
↓	↓	↓
①	②	③
④	⑤	⑥
⑦	Highest Level	

Figure 4: Histogram of the Density of Cognitive Skills across Occupations

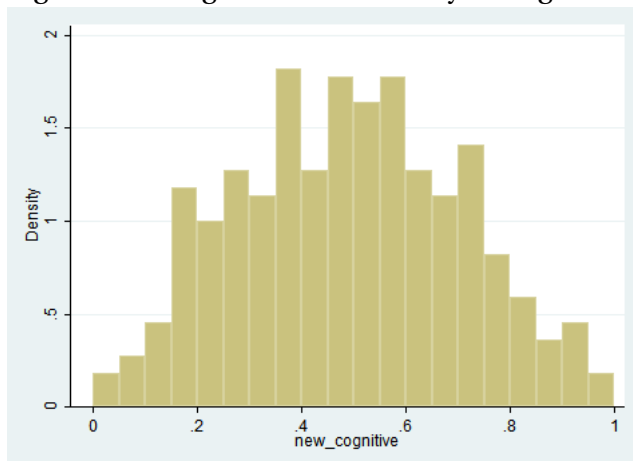


Figure 5: Histogram of the Density of Manual Skills across Occupations

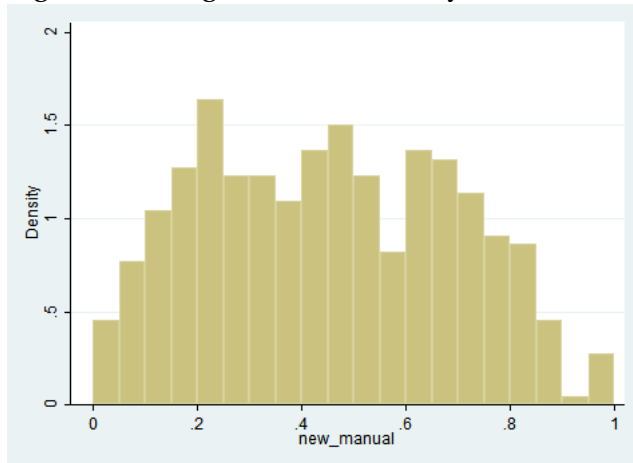


Table 2: Example Occupations and Constructed O*NET Scores

	Cognitive	Manual
Dining Room and Cafeteria Attendants	0.00	0.30
Telemarketers	0.10	0.08
Mechanical Door Repairers	0.33	0.78
Supervisors of Mechanics and Repairers	0.59	0.68
Librarians	0.59	0.26
Economists	0.76	0.02
Hazardous Materials Removal Workers	0.83	0.81
Actuaries	0.94	0.00

Table 3: Summary Statistics for the NLSY 1979 White Male High School Graduates

	Mean	Std. Dev.	Min.	Max.	N
AFQT	41	24	0	100	840
Age at Labor Market Entry	18	2.5	15	37	840
Hourly Wage	16.7	9.4	2.5	70	20121
Cognitive Task Score	.41	.19	0	1	17101
Manual Task Score	.56	.20	0	1	17101
θ	1.2	.27	.19	1.57	17101
$\ \tau\ $	.7	.17	.14	1.26	17101

Wages are in 2010 dollars. See the text for additional sample selection criteria.

Table 4: Measures of θ and $\|\tau\|$ for Example Occupations

	θ	$\ \tau\ $
Dining Room and Cafeteria Attendants	1.57	0.30
Telemarketers	0.67	0.13
Mechanical Door Repairers	1.17	0.84
Supervisors of Mechanics and Repairers	0.85	0.90
Librarians	0.42	0.64
Economists	0.03	0.72
Hazardous Materials Removal Workers	0.77	1.16
Actuaries	0.00	0.94

Table 5: Pattern of Changes that Satisfy “Complex” then “Simple” Moves

# of Total Changes	% that Satisfy $C \Rightarrow S$	% under random movement
2	43	34
3	20	25
4	22	16
5	13	9
6	12	6
7	5	3
8	6	2
9	4	1
10	2	.5
11	3	.1
12	3	.01

Table 6: Fixed Effects Regression for $\Delta\theta$			
Dependent Variable is $ \Delta\theta $			
Variable	Coefficient	Std. Err.	p-val
$\sum_{k=0}^{t-1} \Delta\theta _{ik}$	-.08	.002	.000
Time Since Entry	.004	.007	.000
Time Since Entry ²	-.00008	.00012	.000
Constant	.05	.0021	.000
$N_{total} = 9492$	$\sigma_i = .008$	R^2 (between) = .08	R^2 (within) = .003
$N_i = 840$	$\sigma_\varepsilon = .05$		R^2 (overall) = .004
σ_i is the estimated standard deviation of the distribution of fixed effects. σ_ε is the estimated standard deviation of the error term.			

Table 7: Indirect Inference Moments, Part 1				
Dependent Variable is Log Wages				
Coefficient	Estimate	Std. Err.	t-stat	p-val
c	.61	.003	67	.000
m	.28	.0059	9.5	.000
Time Since Entry	0.018	.000045	46	.000
Time Since Entry ²	-.000146	.00004	-18	.000
$N = 13678, R^2 = .109$				
The constant was estimated but is not reported since it is not used as a moment.				

Table 8: Indirect Inference Moments, Part 2				
Dependent Variable is Squared Residuals from Log Wage Regression				
Coefficient	Estimate	Std. Err.	t-stat	p-val
Constant	.179	.01	17	.000
Time Since Entry	0.009	.0004	22	.000
Time Since Entry ²	$-3.43 \times e - 4$	$1.26 \times e - 5$	-20	.000
$N = 13678, R^2 = .01$				

Table 9: Indirect Inference Moments, Part 3				
Dependent Variable is Cognitive Task Choices				
Coefficient	Estimate	Std. Err.	t-stat	p-val
Constant	0.331	.03	11	.000
Time Since Entry	.007	.003	-2	.05
Time Since Entry ²	-.00012	.00001	12	.000
$N = 13678, R^2 = .31$				

Table 10: Mobility Across θ				
	Year 1	5	10	21
Mean of $\Delta\theta$	-0.01	-0.01	-0.01	.0004
Std. Dev of $\Delta\theta$	.142	.146	.141	.126
Mean of $ \Delta\theta $	.08	.08	.07	.08
Median of $ \Delta\theta $	.03	.025	.007	.006
Std. Dev of $ \Delta\theta $	.12	.12	.12	.13
The sample standard deviation of θ is .27				

Table 11: Indirect Inference Moments, Part 4				
Dependent Variable is Manual Task Choices				
Coefficient	Estimate	Std. Err.	t-stat	p-val
Constant	0.45	.04	11	.000
Time Since Entry	.003	.001	-3	.001
Time Since Entry ²	-.00016	.001	16	.000
$N = 13678, R^2 = .24$				

Table 12: Indirect Inference Moments, Part 5				
Dependent Variable is Squared Residuals from Cog. Task Choices				
Coefficient	Estimate	Std. Err.	t-stat	p-val
Constant	.026	.0005	50	.000
Time Since Entry	0.0016	.00004	40	.000
Time Since Entry ²	$-4.43 \times e - 5$	$1.26 \times e - 6$	-12	.000
$N = 13678, R^2 = .01$				

Table 13: Indirect Inference Moments, Part 6				
Dependent Variable is Squared Residuals from Man. Task Choices				
Coefficient	Estimate	Std. Err.	t-stat	p-val
Constant	0.42	.06	8	.000
Time Since Entry	0.00055	.00007	9	.000
Time Since Entry ²	$-9.91 \times e - 6$	$3.31 \times e - 6$	-3	.001
$N = 13678, R^2 = .005$				

Table 14: Data Moments and Model Moments, Wages			
	Moment	Data	Model
$w = \alpha_0 + \alpha_1 e + \alpha_2 e^2 + \alpha_3 c + \alpha_4 m$	α_1	0.0181	0.0182
	α_2	-0.0001	-0.0004
	α_3	0.6102	1.2344
	α_4	0.2792	0.6046
$\varepsilon_w^2 = \gamma_0 + \gamma_1 e + \gamma_2 e$	γ_0	0.1797	0.1458
	γ_1	0.0096	0.0232
	γ_2	-0.0003	0.0024

Table 15: Data Moments and Model Moments, Tasks			
	Moment	Data	Model
$c = \beta_0^c + \beta_1^c e + \beta_2^c e^2$	β_0^c	0.3310	0.1877
	β_1^c	0.0070	0.0786
	β_2^c	-0.0001	-0.0008
$m = \beta_0^m + \beta_1^m e + \beta_2^m e^2$	β_0^m	0.4510	0.3276
	β_0^m	0.0032	0.0040
	β_0^m	-0.0002	-0.0016
$\varepsilon_c^2 = \gamma_0^c + \gamma_1^c e$	γ_0^c	0.0269	0.0331
	γ_1^c	0.0016	0.0325
$\varepsilon_m^2 = \gamma_0^m + \gamma_1^m e$	γ_0^m	0.0419	0.1075
	γ_1^m	0.0006	0.0024

Table 16: Data Moments and Model Moments, Mobility			
	Moment	Data	Model
$ \Delta\theta = \eta_0 + \eta_1 e + \eta_2 \ \tau\ $	η_0	0.0160	0.0229
	η_1	-0.0003	-0.0437
	η_2	-0.0051	0.0231
$\Delta w = \nu_0 + \nu_1 \Delta \ \tau\ + \nu_2 \Delta\theta$	ν_0	0.0014	0.0014
	ν_1	0.1028	0.1667
	ν_2	-0.0229	-0.0326
$ \Delta\theta _{it} = \zeta_0 + \zeta_1 e + \zeta_2 e^2 + \zeta_3 \sum_{k=0}^{t-1} \Delta\theta _{ik}$	ζ_0	.05	.02
	ζ_1	.004	.003
	ζ_2	-.00008	-.00004
	ζ_3	-.08	-.16

Table 17: Parameter Estimates		
Parameter	Meaning	Value
(μ_c, σ_c)	Initial Distribution of Cognitive Skills	2.03, .40 [.43, .12]
(μ_m, σ_m)	Initial Distribution of Manual Skills	2.43, .44 [.12, .08]
(R_c, R_m)	Skill Accumulation Rates	4.47, .01 [.21, .01]
(α_c, α_m)	Skill Accumulation Shape Params	.869, .4 [.21, .5]
γ	Decrease in SA Ability	10.05 [1.2]
σ_v	Initial Information About Skills	3.1 [.91]
σ_ε	Noise in the Learning Process	.27 [.13]
(δ_c, δ_m)	Skill Depreciation Rates	.001, .03 [.001, .01]
Pricing Function		
$p_j = a_1\tau_c + a_2\tau_m + a_3\tau_c\tau_m + a_4\tau_c^2 + a_5\tau_m^2$		
$\frac{\partial p_j}{\partial \tau_c}$	$.003\tau_m - 1.89\tau_c$	
$\frac{\partial p_j}{\partial \tau_m}$	$-.003\tau_c - 1.95\tau_m$	
a_1	-1.89 [.12]	
a_2	-1.95 [.15]	
a_3	$-.003$ [.002]	
a_4	0.01 [.01]	
a_5	-0.02 [.02]	
Asymptotic Std. Err. in brackets.		

Table 18: Counterfactual Simulations			
Moment	Baseline Case	No Skill Uncertainty	% Change
Mean of $ \Delta\theta $	.042	.023	-45%
Median of $ \Delta\theta $	.04	.031	-22%
Std. Dev of $ \Delta\theta $	.06	.044	-26%