Multidimensional Skills, Sorting, and Human Capital Accumulation*

Jeremy Lise
University of Minnesota

Fabien Postel-Vinay
UCL and IFS

August 2017 under revision

Abstract

We construct a structural model of on-the-job search in which workers differ in skills along several dimensions (cognitive, manual, interpersonal...) and sort themselves into jobs with heterogeneous skill requirements along those same dimensions. We further allow for skills to be accumulated when used, and eroded away when not used. We estimate the model using occupation-level measures of skill requirements based on O*NET data, combined with a worker-level panel from the NLSY79. We use the estimated model to shed light on the origins and costs of mismatch along the cognitive, manual, and interpersonal skill dimensions.

*We would like to thank Fatih Guvenen, Jim Heckman, Derek Neal, Carl Sanders, David Wiczer, and seminar audiences at ASU, Bonn, Cambridge, Chicago, Collegio Carlo Alberto (Turin), Cornell, CREST (Paris), Ecole Polytechnique, Helsinki, Hitotsubashi University Tokyo, Lausanne, Minnesota, NYU, Penn State, Vienna, Washington St. Louis/St. Louis Fed and Wisconsin Madison, as well as participants in the NBER Summer Institute, RES meetings, SED meetings, SaM annual conference, Philadelphia Workshop on Macroeconomics, and Barcelona GSE Summer Forum for useful feedback on earlier versions of this paper. We are additionally grateful to Carl Sanders for kindly providing his data and helping us get started with O*NET. All errors and shortcomings are ours.

1 Introduction

The traditional approach to studying wage and employment inequality, as emphasised by Acemoglu and Autor (2011), relies on a view of labor markets where each worker is endowed with a certain level of "human capital" that rigidly dictates the type of job they are able to hold. This view has gradually evolved into one of labor markets as institutions mediating the endogenous allocation of workers with heterogeneous skills into tasks with heterogeneous skill requirements: any worker can now potentially perform any job, with their skills determining how good they are at any given job, while the market determines the assignment of skills to tasks. This more general view of labor markets has afforded great progress in our understanding of wage and employment inequality.[1]

A subsisting limitation of this approach, however, is that it routinely models skill and task heterogeneity as one-dimensional: workers have more or less of one catch-all "skill", and jobs differ in their requirements for that skill. This representation is at odds with intuition: if one worker is very good at abstract problem-solving but inept at manual work, while another excels in manual tasks but struggles with abstract reasoning, how does one decide which worker is more "skilled"? It is also at odds with the perception of statistical agencies and practitioners of human resources, which maintain and use data describing workers and occupations along many different and imperfectly correlated dimensions such as years and field of education, training, health, aptitude and psychometric test scores, *etc.*, or the occupational skill requirements descriptors available from the O*NET program discussed below. Moreover, it is likely that workers improve the skills that they use regularly and tend to lose some of those they do not use so much, a pattern that a scalar representation of human capital is bound to miss altogether.

The alternative view that workers are endowed with bundles of different skills used in different proportions depending on the task they perform has some history in labor economics (Sanders and Taber, 2012). But at present, few quantitative modeling tools

[1]It shed light on secular trends such as the "polarization" of wages and employment (the simultaneous growth in wages and employment shares of both high-and low-skill workers, at the expense of the middle part of the skill distribution - Acemoglu and Autor, 2011). It contributed to explaining business-cycle fluctuations in aggregate output and employment (Lise and Robin, 2016). It gave substance to the intuitive notion of "skill mismatch" and helped make sense of patterns of worker turnover (Lise, Meghir, and Robin, 2016). It helped clarify the informational content of wage data (Eeckhout and Kircher, 2011).

exist that fully exploit the wealth of information on heterogeneous, multidimensional worker skills and job skill requirements available in the data in a description of labor market equilibrium.

In this paper, we contribute to filling this gap: we extend an otherwise standard and well-tested search-theoretic model of individual careers to allow for multidimensional skills and on-the-job learning. We estimate the model using occupation-level measures of skill requirements based on O*NET data, combined with a worker-level panel (NLSY79). We use the estimated model to shed light on the origins and costs of mismatch along three dimensions of skills: cognitive, manual, and interpersonal[2]. We then proceed to showing that the equilibrium allocation of workers into jobs generically differs from the allocation that a Planner would choose, and investigate the nature and magnitude of the resulting inefficiencies based on our estimated structural model.

Our main findings are the following. The model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have moderate returns and adjust quickly (*i.e.*, they are easily accumulated on the job, and relatively easily lost when left unused). Cognitive skills have much higher returns, but are much slower to adjust. Interpersonal skills have moderate returns, and are essentially fixed over a worker's lifetime. Next, the cost of skill mismatch (modeled as the combination of an output loss and a loss of worker utility caused by skill mismatch) is very high for cognitive skills, an order of magnitude greater than for manual or interpersonal skills. Moreover, this cost is asymmetric: employing a worker who is under-qualified in cognitive skills (*i.e.* has a level of skills that falls short of the job's skill requirements) is several orders of magnitude more costly than employing an over-qualified worker. Those important differences between various skill dimensions are missed when subsuming worker productive heterogeneity into one single scalar index.

The paper is organized as follows. Section 2 provides a brief discussion of some of the related literature. Section 3 lays out the formal model, Section 4 describes the data used for estimation, with some emphasis on O*NET, Section 5 explains the simulation/estimation protocol, Section 6 presents the estimation results and discusses some of the model's predictions on skill mismatch and sorting, and Section 7 shows results from counterfactual experiments aiming to quantify the distance be-

[2]What we term interpersonal skills are sometimes referred to as non-cognitive or personality traits (Heckman, Stixrud, and Urzua, 2006; Borghans, Duckworth, Heckman, and ter Weel, 2008)

tween the model's decentralized equilibrium and the allocation that a Planner would select, as well as the social cost of various aspects of skill mismatch. Finally, Section 8 concludes.

2 Related Literature

This paper is related to the vast empirical literature on the returns to firm and occupation tenure and to more recent work on task-specific human capital. Those literatures, and the connections between them are covered in the excellent survey paper by Sanders and Taber (2012).[3] As a preamble to their review of the empirical literature, Sanders and Taber (2012) offer an elegant theoretical model of job search and investment in multi-dimensional skills which, on many aspects, can be seen as a special case of the model in this paper.[4] However, they only use their model to provide intuition and highlight key qualitative predictions of the theory, and do not bring it to the data.

In a more structural vein, attempts to model the allocation and pricing of heterogeneous supply and demand of indivisible and multi-dimensional bundles dates back at least to Tinbergen (1956) and the hedonic model of Rosen (1974). Generic non-parametric identification of the hedonic model is established in Ekeland, Heckman, and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2010). Recently Lindenlaub (2014) estimates the quadratic-normal assignment model of Tinbergen (1956) along two dimensions of skills (manual and cognitive) for two different cohorts using the same combination of O*NET and NLSY data as we do in this paper. She finds an interesting pattern of technological change: the complementarity between her measures of cognitive worker skills and cognitive job skill requirements increased substantially during the 1990s, while the complementarity between manual job and worker attributes decreased. She then analyzes the consequences of that technological shift for sorting and wage inequality.

While Lindenlaub's analysis brings about many valuable new insights, it assumes away market imperfections which limits its applicability to empirical and quantitative

[3]We thank John Kennan for bringing this particular reference to our attention.

[4]Sanders and Taber (2012) model individual skill accumulation as the outcome of endogenous investment decisions (in the spirit of Ben-Porath (1967)), whereas we consider (occupation-specific) learning-by-doing. While the conceptual differences between those two models are important, they are notoriously difficult to tell apart empirically.

policy analysis. First, it is difficult to define a meaningful notion of unemployment or of skill mismatch in a Walrasian (frictionless) world where, given the economy's primitives (*i.e.* the production technology and the distributions of job and worker attributes), equilibrium is by construction efficient. By contrast, allowing for market imperfections creates scope for welfare-improving policy intervention. Second, frictionless matching and assignment models, including Rosen's hedonic model, are static.[5] As such, they are largely silent on questions relating to a worker's life cycle, such as the cost of skill mismatch throughout a worker's career, the way in which individual skills evolve over a career, how this skill accumulation is priced in the market, or the reasons why workers switch occupations as often as they do.[6]

Next, following in the tradition of Heckman and Sedlacek (1985), Keane and Wolpin (1997), and Lee and Wolpin (2006), the important contribution by Yamaguchi (2012) provides the first estimation of a Roy-type model of task-specific human capital accumulation and occupation choices over the life cycle based on the combination of the NLSY with data on occupation-level attributes, interpreted as "task complexity", from the Dictionary of Occupational Titles (DOT, the predecessor of O*NET). The broad approach is the same as in the present paper: each occupation is characterized by the vector of weights (the degree of task complexity) it places on a limited number of different skill dimensions, as in Lazear's (2009) skill-weights approach. Worker skills are not directly observed, but their accumulation is modeled as a hidden Markov chain, the parameters of which are identified from observed choices of occupations with different task contents (observed from the DOT data), using the model's structure. Yamaguchi's findings suggest that higher task complexity is associated with higher wage returns to, and faster growth of the skills relevant to the task. A wage variance decomposition further suggests that both cognitive and motor skills (the two skill dimensions considered by Yamaguchi) are important determinants of cross-sectional log wage variance. A decomposition of wage growth shows that cog-

[5]See Chiappori and Salanié (2016) for a recent survey of the econometrics of static, frictionless matching models.

[6]Models of experimentation and learning following on from Jovanovic (1979) such as Neal (1999); Pavan (2011); Golan and Antonovics (2012) have been useful to make sense of the patterns of between and within occupation switches in the NLSY. They cannot, however, easily rationalize the frequent transitions in and out of unemployment observed in the same data. Lindenlaub and Postel-Vinay (2016) extend Lindelaub's frictionless model to a frictional environment using a basic framework that has much in common with our model. However, Lindenlaub and Postel-Vinay (2016) is a theoretical exercise focusing on conditions under which specific sorting patterns emerge in steady-state equilibrium, and their model does not feature human capital accumulation.

nitive skills account for all of the wage growth of high-school and college graduates, while motor skills account for about half of the wage growth of high-school dropouts.

While clearly related in spirit, our model differs from Yamaguchi (2012) in several important ways. First, Yamaguchi (2012) is a frictionless model in which occupation mobility is largely governed by unobserved shocks to an exogenously posited wage function, to workers' skills, and to workers' preferences for any given type of job.[7] We propose a more parsimonious random search model, in which the only shocks are the receipt of job offers by workers. Wages and mobility decisions are then endogenously determined through between-employer competition for labor services. Our less flexible, but more transparent and readily interpretable specification offers a remarkably good fit to the data. Second, the only engine of wage growth in Yamaguchi's model is skill accumulation. Other sources of wage growth, such as job-shopping or learning, are therefore partly picked up by skills in that model, which may lead to an overstatement of the role of skills. Our model also ignores learning, but explicitly models job-shopping as an additional source of wage dynamics.[8] Adding search frictions allows us to address issues related to unemployment and skill mismatch.[9]

Two recent papers are particularly related. Taber and Vejlin (2016) estimate a model which allows for search, Roy-type selection, human capital accumulation and non wage amenities. Workers are modeled as having a time invariant relative ability at each job-type in the economy. In the absence of frictions they would choose a single job-type and remain indefinitely. Human capital is assumed to be general and accumulated while working. Job mobility is informative about the degree of search frictions, and wage cuts are informative about non wage amenities. Taber and Vejlin (2016) model relative ability between jobs/occupations as an unobserved vector with dimension equal to the number of job-types in the economy. We take a substantially more parsimonious approach: a worker's relative productivity across jobs/occupations depends on the amount of skills (cognitive, manual, interpersonal...) they currently possess and whether or not they are a good fit for the demands of a particular job.

[7]One immediately apparent drawback of this frictionless approach is that, taken literally, it predicts that workers should change occupations continuously (or in every period, in Yamaguchi's discrete-time model), which is obviously at odds with observation.

[8]Sanders (2012) considers learning in a model otherwise similar to Yamaguchi's.

[9]Moscarini (2001) combines a two-sector Roy model into an equilibrium matching model and analyzes the partially directed search patterns arising in equilibrium and governing equilibrium selection of skill bundles into sectors. His setup has great descriptive appeal, but remains far too stylized to be taken directly to the data.

Rather than treating these skills as unobserved, we use a large set of premarket measurements to estimate a worker's initial endowment, and a similar large set of measurements on occupations to estimate the skill requirements of jobs. A second notable modeling difference is that we allow these skills to evolve differentially depending on the extent to which they are being used in a particular job/occupation. We are particularly interested in the the differential returns to these skills, and the extent to which each type of skill can be learned on the job.

From an empirical perspective, perhaps the closest paper is Guvenen, Kuruscu, Tanaka, and Wiczer (2016). They use the same combination of NLSY and O*NET data as we do to construct a summary index of multidimensional skill mismatch which they use to assess the impact of skill mismatch on wages and patterns of occupational switching. They produce a rich set of empirical results, a rough summary of which is that both current and past mismatch strongly impact wages (negatively), the probability of switching occupations (positively), and the direction of said switching.

Their index of skill mismatch is derived from a model of occupation choice with workers learning about their own ability. Mismatch arises in this model, not because of search frictions, but because workers have imperfect information about their own skills and sort into occupations that are optimal for their *perceived* skill bundle (which differs from their true one). As they gradually learn about their true skills (about which they observe a sequence of noisy signals over time), workers switch occupations. This model gives rise to an intuitive summary mismatch measure that is based on the distance between a worker's skill bundle and the set of skills required by their occupation, which the authors then use as a regressor in Mincer-type wage equations and in statistical models of occupation switching.

While our paper shares some of its basic objectives with Guvenen *et al.* (2016) (chiefly, an assessment of the production/wage cost of skill mismatch in various dimensions), the two contributions differ in terms of both approach and focus. Aside from substantive differences in modeling choices, Guvenen *et al.* use their theory as a guide for intuition and specification of reduced-form statistical models rather than as an actual structure for estimation. More importantly, they provide detailed results on the impact of mismatch on the probability and direction of occupational switching, whereas we focus (1) on differences between skill categories in the speed of human capital accumulation or decay and (2) on the social cost of various forms of mismatch. Our structural approach is especially useful to address the latter broad question, which we do by means of counterfactual simulations.

3 Job Search with Multi-dimensional Job and Worker Attributes

3.1 The Model

The Environment. Workers are characterized by general and specialized skills. The market productivity of specialized skills will depend on the technology of a particular firm, while general skills have a common effect on output, independent of the particular firm technology a worker is currently matched with. Match output is $f(\mathbf{x},\mathbf{y})$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^K$ describes the worker's set of skills, and $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^L$ describes the firm's technology, with $L \leq K$. The first L worker skills are specialized with the remaining $K-L$ being general skills. The firm's technology is fixed, but the worker's skills gradually adjust to the firm's technology as follows:

$$\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x},\mathbf{y}),$$

where $\mathbf{g} : \mathbb{R}^K \times \mathbb{R}^L \to \mathbb{R}^K$ is a continuous function. Just as in production, the adjustment of specialized skills differs depending on the firm technology, while the adjustment of general skills depends only on experience.

Time is continuous. Upon entering the labor market, workers draw their initial skill vectors from an exogenous distribution $N(\cdot)$ [with density $\nu(\cdot)$]. Workers can be matched to a firm or unemployed. If matched, they lose their job at rate δ, and they sample alternate job offers from the fixed sampling distribution $\Upsilon(\mathbf{y})$ [with density $\upsilon(\mathbf{y})$] at rate λ_1. Unemployed workers sample job offers from the same sampling distribution at rate λ_0. Workers exit the market at rate μ. All four transition rates $(\lambda_0, \lambda_1, \delta, \mu)$ are exogenous.

All agents have linear preferences over income and discount the future at rate r. A type-$\mathbf{x}$ worker's flow utility from working in a type-$\mathbf{y}$ job for a wage w is $w - c(\mathbf{x},\mathbf{y})$, where $c(\mathbf{x},\mathbf{y})$ is disutility from work, which depends on the type of the match, $(\mathbf{x},\mathbf{y})$. A type-$\mathbf{x}$ unemployed worker receives a flow income $b(\mathbf{x})$ and has no disutility of being unemployed.

Firm, worker, and match values. We denote the total private value (*i.e.* the value to the firm-worker pair) of a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ firm by $P(\mathbf{x},\mathbf{y})$. Under linear preferences over wages, this value is independent of the way in which it is shared between the two parties in the match, and only depends on match attributes $(\mathbf{x},\mathbf{y})$. We further denote the value of unemployment by $U(\mathbf{x})$, and the worker's value of his current wage contract by W, where $W \geq U(\mathbf{x})$ (otherwise the worker would quit into unemployment), and $W \leq P(\mathbf{x},\mathbf{y})$ (otherwise the firm would fire the worker). Assuming that the employer's value of a job vacancy is zero (which would arise under free entry and exit of vacancies on the search market), the total *surplus* generated by a type-$(\mathbf{x},\mathbf{y})$ match is $P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$, and the worker's share of that surplus is $(W - U(\mathbf{x}))\,/\,(P(\mathbf{x},\mathbf{y}) - U(\mathbf{x}))$.

Rent sharing and wages. Wage contracts are renegotiated sequentially by mutual agreement, as in the sequential auction model of Postel-Vinay and Robin (2002). Workers have the possibility of playing off their current employer against any firm from which they receive an outside offer. If they do so, the current and outside employers Bertrand-compete over the worker's services.

Consider a type-$\mathbf{x}$ worker employed at a type-$\mathbf{y}$ firm and assume that the worker receives an outside offer from a firm of type $\mathbf{y}'$. Bertrand competition between the type-$\mathbf{y}$ and type-$\mathbf{y}'$ employers implies that the worker ends up in the match that has higher total value — that is, he stays in his initial job if $P(\mathbf{x},\mathbf{y}) \geq P(\mathbf{x},\mathbf{y}')$ and moves to the type-$\mathbf{y}'$ job otherwise — with a new wage contract worth $W' = \min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\}$.

Suppose, for the sake of argument, that $P(\mathbf{x},\mathbf{y}) \geq P(\mathbf{x},\mathbf{y}') > W$. In this case, the outcome of the renegotiation is such that the worker stays with his initial type-$\mathbf{y}$ employer under a new contract with value $W' = P(\mathbf{x},\mathbf{y}')$.[10] The worker's renegotiated share of the match surplus, denoted $\sigma(\mathbf{x},\mathbf{y},\mathbf{y}')$, is therefore:

$$\sigma(\mathbf{x},\mathbf{y},\mathbf{y}') = \frac{P(\mathbf{x},\mathbf{y}') - U(\mathbf{x})}{P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})} \in [0,1]. \tag{1}$$

To pin down the way in which the value $W' = P(\mathbf{x},\mathbf{y}') = U(\mathbf{x}) + \sigma(\mathbf{x},\mathbf{y},\mathbf{y}')\,[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})]$ is delivered over time by the firm to the worker, we assume that the surplus share

[10] Obviously, renegotiation only takes place if $P(\mathbf{x},\mathbf{y}') > W$, as otherwise the type-$\mathbf{y}'$ employer is unable to make a (profitable) offer that improves on the worker's initial value W, and the worker's threat of accepting an offer from that employer is not credible.

σ, negotiated at the time the worker receives an outside offer from the type-$\mathbf{y}'$ job, stays constant until the following renegotiation. Put differently, while the worker's skill bundle $\mathbf{x}$ and, as a consequence, the match surplus $P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$ evolve over the course of his tenure in the type-$\mathbf{y}$ job, the share of that surplus transferred to the worker stays constant between negotiations and is determined as per equation (1) by the best outside offer previously received by the worker. The particular way in which the type-$\mathbf{y}$ employer delivers the value $P(\mathbf{x},\mathbf{y}')$ to the worker only affects the time profile of wage payments and the timing of renegotiation. It makes no difference to the allocation of workers into jobs, as mobility decisions are only based on comparisons of total match values, which, under linear preferences, are independent of the time profile of wage payments.[11]

Value functions and wage equation. The total private value of a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ firm, $P(\mathbf{x},\mathbf{y})$, solves:[12]

$$(r+\mu+\delta)P(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - c\,(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} P(\mathbf{x},\mathbf{y}). \qquad (2)$$

Note that the frequency at which the worker collects offers, λ_1, does not affect $P(\mathbf{x},\mathbf{y})$. Upon receiving an outside offer, the worker either stays in his initial match, in which case the continuation value for that match is $P(\mathbf{x},\mathbf{y})$, or he accepts the offer, in which case he extracts a value of $P(\mathbf{x},\mathbf{y})$ from the poacher (as a result of Bertrand competition) and leaves his initial employer with a vacant job worth 0. Either way, the joint continuation value for the partners in the initial match equals $P(\mathbf{x},\mathbf{y})$. This is a key implication of Bertrand competition between employers: from a social perspective, cases where the worker accepts the outside offer and moves to a match with higher value $P(\mathbf{x},\mathbf{y}')$ are associated with a net surplus gain of $P(\mathbf{x},\mathbf{y}') - P(\mathbf{x},\mathbf{y})$. Yet none of the social gains associated with future job mobility are internalized by private agents, as those gains accrue to a third party (the worker's future employer). We discuss some of the consequences of this property in Section 7.1.

[11] Common alternative assumptions about the way in which firms deliver value to workers include a constant wage or a constant share of match output (a piece rate). Our assumption of a constant surplus share has the merit of simplifying computations considerably. Note that the wages produced by the constant wage, constant piece rate or constant surplus share assumptions are exactly identical if the worker's skills stay constant over time ($\dot{\mathbf{x}} \equiv 0$).

[12] The dot ("$\cdot$") denotes the outer product and ∇ denotes the gradient.

The value of unemployment, $U(\mathbf{x})$, solves:

$$(r+\mu)U(\mathbf{x}) = b(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{0}) \cdot \nabla U(\mathbf{x}), \tag{3}$$

where by convention the employer type is set to $\mathbf{y} = \mathbf{0}_L$ for an unemployed worker. For reasons similar to those just discussed about $P(\mathbf{x},\mathbf{y})$, the worker fails to internalize the gain in surplus associated with him accepting a job offer, and the private value of unemployment is independent of the frequency at which those offers arrive.

The worker receives an endogenous share σ of the match surplus $P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$, which he values at $W(\mathbf{x},\mathbf{y},\sigma) = (1-\sigma)U(\mathbf{x}) + \sigma P(\mathbf{x},\mathbf{y})$. The wage $w(\mathbf{x},\mathbf{y},\sigma)$ implementing that value solves:

$$\begin{aligned}(r+\delta+\mu)W(\mathbf{x},\mathbf{y},\sigma) &= w(\mathbf{x},\mathbf{y},\sigma) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) \\ &+ \lambda_1 \mathbf{E} \max\{0, \min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\} - W(\mathbf{x},\mathbf{y},\sigma)\} + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} W(\mathbf{x},\mathbf{y},\sigma),\end{aligned} \tag{4}$$

where the expectation is taken over the sampling distribution, $\mathbf{y}' \sim \Upsilon$.

Combining 2, (3) and (4) (using $W(\mathbf{x},\mathbf{y},\sigma) = (1-\sigma)U(\mathbf{x}) + \sigma P(\mathbf{x},\mathbf{y})$) further yields the following wage equation:

$$\begin{aligned}w(\mathbf{x},\mathbf{y},\sigma) &= \sigma f(\mathbf{x},\mathbf{y}) + (1-\sigma)b(\mathbf{x}) + (1-\sigma)c(\mathbf{x},\mathbf{y}) \\ &- \lambda_1 \mathbf{E} \max\{0, \min\{P(\mathbf{x},\mathbf{y}') - P(\mathbf{x},\mathbf{y}), 0\} + (1-\sigma)\left(P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})\right)\} \\ &- (1-\sigma)\left(\mathbf{g}(\mathbf{x},\mathbf{y}) - \mathbf{g}(\mathbf{x},\mathbf{0})\right) \cdot \nabla U(\mathbf{x}).\end{aligned} \tag{5}$$

The first term $\sigma f(\mathbf{x},\mathbf{y}) + (1-\sigma)b(\mathbf{x}) + (1-\sigma)c(\mathbf{x},\mathbf{y})$ reflects static sharing of the match surplus flow, in shares $(\sigma, 1-\sigma)$ resulting from the worker's history of outside job offers. Note that the worker always has to be compensated for a share $(1-\sigma)$ of his disutility of work $c(\mathbf{x},\mathbf{y})$. The next (expectation) term reflects value of future outside offers, which the worker pays for by accepting a lower starting wage. The final term reflects the fact that an employed worker's skill bundle evolves towards the job's skill requirements $\mathbf{y}$, whereas those skills would erode towards $\mathbf{0}_L$ if the worker was unemployed. This, in general, benefits the worker in the event he becomes unemployed, and therefore affects the wage negatively.[13]

[13] As mentioned in Section 2, the model in Sanders and Taber (2012) is close to a special case of our model where $f(\mathbf{x},\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ and where workers always receive a fixed share of the match surplus (*i.e.* σ is a fixed constant). Predicted wages differ between our model and theirs, but the

3.2 Model Analysis

A fully closed-form case. Full closed-form solutions can be obtained under specific functional form assumptions. We now give an example, which we will use in our empirical specification below.

We first restrict the dimensionality of worker and job attributes, both for simplicity of exposition and because those restrictions are relevant to the empirical application below (nothing in the theory depends on those particular restrictions). We think of a typical worker's skill bundle $\mathbf{x} = (x_C, x_M, x_I, x_T)$ as capturing *(i)* the worker's cognitive skills x_C, *(ii)* the worker's manual skills x_M, *(ii)* the worker's interpersonal skills x_I, and *(iv)* capture the worker's "general efficiency" x_T. Jobs are likewise characterized by a three-dimensional bundle $\mathbf{y} = (y_C, y_M, y_I)$ capturing measures of the job's requirements in cognitive, manual, and interpersonal skills. All three job attributes are fixed over time, whereas a worker's cognitive, manual, and interpersonal skills (x_C, x_M, x_I) are allowed to adjust over time to the requirements of the particular job the worker holds (learning by doing).

The key functional form assumption is to assume a linear adjustment for skills. In particular, we assume that a worker's specialized (*i.e.* cognitive, manual, and interpersonal) skills adjust linearly to his/her job's skill requirements, *i.e.* we specify the function $\mathbf{g}(\mathbf{x}, \mathbf{y})$ as:

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \dot{x}_C \\ \dot{x}_M \\ \dot{x}_I \\ \dot{x}_T \end{pmatrix} = \begin{pmatrix} \gamma_C^u \max\{y_C - x_C, 0\} + \gamma_C^o \min\{y_C - x_C, 0\} \\ \gamma_M^u \max\{y_M - x_M, 0\} + \gamma_M^o \min\{y_M - x_M, 0\} \\ \gamma_I^u \max\{y_I - x_I, 0\} + \gamma_I^o \min\{y_I - x_I, 0\} \\ g x_T \end{pmatrix}, \tag{6}$$

where the $\gamma_k^{u/o}$'s are all positive constants governing the speed at which worker skills adjust to a job's requirements. Note that we allow that speed to differ between upward and downward adjustments (γ_k^u vs γ_k^o for $k = C, M, I$, where "u" stands for "under-qualified" and "o" stands for "over-qualified"), and between skill types ($\gamma_C^{u/o}$ vs $\gamma_M^{u/o}$ vs $\gamma_I^{u/o}$). In this case, (13) solves as:

$$x_k(s) = y_k - e^{-\gamma_k^{u/o}(s-t)} (y_k - x_k(t)), \tag{7}$$

worker-job allocation for given distributions of $\mathbf{x}$ and $\mathbf{y}$ is identical. As already mentioned, the two models further differ in the specific assumption regarding skill accumulation (endogenous investment decisions vs. learning-by-doing).

where the adjustment speed $\gamma_k^{u/o}$ that applies depends on whether $k = C, M$ or I and whether $x_k(t) \gtrless y_k$. Over time a worker's specialized skills will adjust to the requirements of the job. Finally, a worker's general efficiency simply grows at a constant rate: $x_T(t) = x_T(0) \times e^{gt}$, independently of the worker's cognitive/manual skills or of the worker's employment status. This very simple specification will help the model capture the wage/experience trend observed in the data.

The production function is then specified as follows:

$$f(\mathbf{x}, \mathbf{y}) = x_T \times \left[\varphi(\mathbf{y}) - \sum_{k=C,M,I} \kappa_k^u \min\left\{x_k - y_k, 0\right\}^2 \right]. \tag{8}$$

The function $\varphi(\mathbf{y})$ is assumed increasing in all components of $\mathbf{y}$, implying that jobs with higher requirements in any of cognitive, manual or interpersonal skills are inherently more productive, regardless of the worker they are matched with. The terms $-\kappa_k^u \min\left\{x_k - y_k, 0\right\}^2$, $k = C, M, I$ capture the idea that a worker with a shortage of skills $\mathbf{x}$ compared to the job's skill requirement level $\mathbf{y}$ in any dimension (cognitive, manual, or interpersonal) causes a loss of output (assuming that all κ_k's are non-negative). We allow for the output loss caused by skill mismatch to differ depending on which skills the worker is deficient in. Note that specification (8) carries the implicit assumption that an *over*-qualified worker (such that $x_k > y_k$) produces the same output as a worker whose skills are a perfect match for the job: over-qualification causes neither a direct gain, nor a loss of output. We will return to this shortly. General efficiency, x_T, acts to scale up or down output, conditional on $\mathbf{x}$ and $\mathbf{y}$. In estimation we will allow x_T to be correlated with $\mathbf{x}_0$, allowing for the possibility that workers with, for example, high cognitive skills also have high general skills. Finally, we simply specify unemployment income as depending on general skill only, $b(\mathbf{x}) = bx_T$, with b a positive constant, so that $U(\mathbf{x}) = bx_T/(r + \mu - g)$ is independent of the specialized skills (x_C, x_M, x_I).

The last object we need to specify is the flow disutility of work:

$$c(\mathbf{x}, \mathbf{y}) = x_T \times \sum_{k=C,M,I} \kappa_k^o \max\left\{x_k - y_k, 0\right\}^2. \tag{9}$$

According to this specification, disutility of work is only positive if the worker is over-qualified for her/his job in some skill dimension. We interpret this as a utility cost of being under-matched. This assumption brings about some comments. As is arguably intuitively natural, the production function (8) only allows for a *shortage* of worker cognitive, manual, or interpersonal skills compared to the job's requirements to cause a loss of output (and hence of match value). Yet it is important that we also allow for an *excess* of skills to cause a loss of match value. We assume that this loss takes the form of a utility cost of being under-matched. This utility cost is the only cost of over-qualification that is internalized by the match parties. To see this, consider a match in which the worker is over-qualified in all dimensions, *i.e.* $x_k > y_k$ for all $k = C, M, I$. If all of the κ_k^o's were equal to zero, then the value of that match would be the same (given equal general skills x_T) regardless of the amount by which the worker is over qualified, even though a "more" over-qualified worker stands to lose more skills than a "less" over-qualified one when taking up this job. So long as the worker is over-qualified, a marginal change in the worker's cognitive or manual skills affects neither match output (8) nor the value of unemployment (as $b(\mathbf{x})$ is independent of (x_C, x_M, x_I)). In that case, any loss of skills only changes the surplus in the worker's *future* matches, which, because of Bertrand competition, do not affect the surplus from the current match (as discussed in Subsection 3.1). Conversely, with positive κ_k^o's, a match with an over-qualified worker has lower value, the further the worker's skills are above the job's skill requirements. Finally, the specific functional form (9) of $c(\mathbf{x}, \mathbf{y})$ echoes the cost of mismatch (of under-qualification) in the production function (8), mostly for analytical convenience and simplicity.[14]

With those specifications, equations (2) and (3) imply (see Appendix A.1):

$$P(\mathbf{x}(t), \mathbf{y}) - U(\mathbf{x}) = x_T(t) \times \left\{ \frac{\varphi(\mathbf{y}) - b}{r + \delta + \mu - g} - \sum_{k=C,M,I} \left(\frac{\kappa_k^u \min\{x_k(t) - y_k, 0\}^2}{r + \delta + \mu - g + 2\gamma_k^u} + \frac{\kappa_k^o \max\{x_k(t) - y_k, 0\}^2}{r + \delta + \mu - g + 2\gamma_k^o} \right) \right\}. \quad (10)$$

The first term in the equation defining match surplus $P(\mathbf{x}(t), \mathbf{y}) - U(\mathbf{x})$ is the (maximum, given $\mathbf{y}$) surplus achieved if the worker's skills are perfectly matched to the job's requirements - *i.e.* if $(x_C(t), x_M(t), x_I(t)) = (y_C, y_M, y_I)$. The remaining terms reflect the (private surplus) cost of initial cognitive, manual, and interpersonal skill mismatch. This cost obviously depends on the weights of cognitive, manual and in-

[14]See Appendix A.2 for an alternative specification.

terpersonal mismatch in the technology and utility function, but also on the speed of skill adjustment: if adjustment is instantaneous ($\gamma_k^{u/o} \to +\infty$), the cost of mismatch becomes negligible.

4 Data

Our estimation sample is a panel of worker-level data from the 1979 *National Longitudinal Survey of Youth* (NLSY79) combined with occupation-level data on skill requirements from the O*NET program (`www.onetcenter.org`). We describe both data sets and the way we combine them before turning to a description of the estimation sample itself.[15]

4.1 Construction of the Estimation Sample

Data sources. The NLSY79 is well known and requires little description. Our extract from that data set is a weekly unbalanced panel of workers whom we follow from first entry into the labor market. For each worker in the panel, time is set to zero at the first week they cease to be in full-time education. We focus on males from the main sample who were never in the military,[16] and retain all individual histories until the first occurrence of a non-employment spell of 18 months or more: we consider individuals experiencing such a long spell of non-employment as losing their attachment to the labor force, which we treat as attrition from the sample. We retain information on labor force status and transitions, weekly earnings, occupation of current job (Census codes), education (highest grade completed), performance in a battery of ten aptitude tests called the *Armed Services Vocational Aptitude Battery* (ASVAB), measures of anti-social behavior, measures of health, and scores in two psychometric tests measuring social skills. Education, ASVAB scores, and measures of social skills and health will be used as measures of the initial skill bundles $\mathbf{x}$ of those workers (more below).

[15] We are not the first authors to combine these data sources. A non-exhaustive list includes Autor, Levy, and Murnane (2003), Acemoglu and Autor (2011), Autor and Dorn (2013), Yamaguchi (2012), Sanders (2012), Lindenlaub (2014), and Guvenen et al. (2016), who all use combinations of the NLSY with occupation data from O*NET or from its predecessor, the *Dictionary of Occupational Titles*.

[16] The NLSY over-samples ethnic minorities, people in the military, and the poor. We drop all such over-sampled observations.

To obtain measures of the skill requirements $\mathbf{y}$ attached to the occupations observed in the NLSY sample, we combine the latter with data from the O*NET program. O*NET, a.k.a. the *Occupational Information Network*, is a database describing occupations in terms of skill and knowledge requirements, work practices, and work settings.[17] It comes as a list of 277 descriptors, with ratings of importance, level, relevance or extent, for over 970 different occupations. O*NET descriptors are organized into nine broad categories: skills, abilities, knowledge, work activities, work context, experience/education levels required, job interests, work values, and work styles. O*NET ratings come from two different sources: a survey of workers, who are asked to rate their own occupation in terms of a subset of the O*NET descriptors, and a survey of "occupation analysts" who are asked to rate other descriptors in the O*NET set.

We retain descriptors from the skills, abilities, knowledge, work activities, and work context O*NET files, as descriptors contained in the other files (job interests, work values, and work styles) are less directly interpretable in terms of skill requirements, and merge those files with our NLSY sample, based on occupation codes.[18]

Job skill requirements. Our selection from the O*NET database leaves us with over 200 different descriptors, which we take as measures of the underlying skill requirements. We reduce this large set of descriptors to three dimensions, which we interpret as "cognitive", "manual", and "interpersonal" skill requirements, using the following procedure.[19] First, we run Principal Component Analysis (PCA) on our large set of O*NET measures and keep the first three principal components. We then recover our cognitive, manual, and interpersonal skill requirement indices by recombining those three principal components (which by default are constructed to be orthonormal) in such a way that they satisfy the following three exclusion restrictions: (1) the *mathematics* score only reflects cognitive skill requirements; (2) the *mechanical*

[17] O*NET is developed by the North Carolina Department of Commerce and sponsored by the US Department of Labor. Its initial purpose was to replace the old *Dictionary of Occupational Titles*. More information is available on `www.onetcenter.org`, or on the related Department of Labor site `www.doleta.gov/programs/onet/eta_default.cfm`.

[18] The NLSY79 uses 1970, 1980 and 2000 Census codes for occupation, whereas O*NET uses 2009 SOC codes. Crosswalks exists between those different nomenclatures. The crosswalks we use were kindly provided to us by Carl Sanders, whose help is gratefully acknowledged. Using those, over 92% of occupation codes records in the NLSY sample have a match in the O*NET data.

[19] Technical details of the construction of our job skill requirement and worker skill bundles are given in Appendix A.5.

knowledge score only reflects manual skill requirements; (3) the *social perceptiveness* score only reflects interpersonal skill requirements). Interpretation of the three skill requirement indices thus obtained as cognitive, manual and interpersonal therefore relies on those exclusion restrictions. Finally, we rescale our skill requirement indices so that they lie in $[0, 1]$.[20]

Worker skill bundles. Finally, we need to construct a distribution of initial worker skill bundles, *i.e.* the distribution $N(\mathbf{x})$ of cognitive, interpersonal and manual skills among labor market entrants. For this we follow a similar procedure as for the distribution of skill requirements, using PCA and exclusion restrictions. We use the following sets of measures: the ten ASVAB scores that are directly available from the NLSY sample, individual scores on the Rotter locus-of-control scale and the Rosenberg self-esteem scale tests,[21] three measures of criminal and anti-social behavior, two measures of health (BMI and weight), and an O*NET-based measure of cognitive, manual, and interpersonal skills attached to the level of education attained by each NLSY sample member. The latter is constructed using the "experience/education requirements" file from O*NET, which informs about the education requirements of each occupation in O*NET, and from which we take the average value, for each education level, of the cognitive, manual and interpersonal scores constructed above. As exclusion restriction we assume that (1) the ASVAB *mathematics knowledge* score only reflects cognitive skills; (2) the ASVAB *automotive and shop information* score only reflects manual skills; (3) the Rosenberg *self-esteem* score only reflects interpersonal skills. Those particular exclusion restrictions were chosen for their intuitive consistency with the exclusion restrictions used in the construction of job skill requirements, so as to ensure that worker skill indices are reasonably well "aligned" with the corresponding skill requirement indices in all three dimensions. Yet in the estimation, we will allow for the possibility of less-than-perfect alignment between worker skill and job skill requirement scores (see below).

Our final estimation sample consists of an initial cross-section of 1,770 males whom

[20]We do this using linear transforms (rather than by converting the initial indices to ranks, as has been done elsewhere in the literature), because we expect there to be useful information in the distance between two different occupations in terms of cognitive, manual, or interpersonal skill requirements. Linear rescaling preserves relative distances, whereas conversion into ranks renders all occupations equidistant.

[21]See `https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/attitudes` for a description of those two tests.

we follow over up to 30 years. This sample is described in detail in Appendix A.4.

5 Estimation

We estimate the model by indirect inference. To this end, the first step is to simulate a panel that mimics our estimation sample. We first describe the simulation protocol, then discuss the moments we choose to match in the estimation as well as identification of the model.

5.1 Simulation

Solution method. The model has a convenient recursive structure. Equations (2) and (3) can be solved jointly for $U(\mathbf{x})$ and $P(\mathbf{x}, \mathbf{y})$ in a first step. Wages are then obtained from the combination of (4) and the assumption of Bertrand competition: the surplus share $\sigma(\mathbf{x}, \mathbf{y}, \mathbf{y}')$ obtained by a type-$\mathbf{x}$ worker playing off employers $\mathbf{y}$ and $\mathbf{y}'$ (with $P(\mathbf{x}, \mathbf{y}) > P(\mathbf{x}, \mathbf{y}'))$ against each other solves (1), and the wages that follow from that renegotiation solve (5). Finally, given those value functions, a cohort of workers can be simulated as we now describe.

Simulation protocol. We simulate a cohort of N workers (indexed by $i = 1, \cdots, N$) over $T = 300$ months (indexed $t = 0, \cdots, T-1$) using a discrete-time approximation of our model. All workers start out in period $t = 0$ endowed with an initial skill bundle $\mathbf{x}_{i0} = (x_{C,i0}, x_{M,i0}, x_{I,i0}, x_{T,i0})$ drawn from the distribution $\nu(\cdot)$, and in an initial labor market state (unemployed or employed in a job with attributes $\mathbf{y}_{i1}$ under some initial labor contract giving him a share σ_{i1} of the surplus associated with his job) determined as described below. In each subsequent period $t = 1, \cdots, T-1$, we update each worker's skill bundle iteratively using the solution to $\dot{\mathbf{x}}_{is} = \mathbf{g}(\mathbf{x}_{is}, \mathbf{y}_{i,t-1})$ over $s \in [t-1, t]$ given the initial condition $\mathbf{x}_{i,t-1}$, and where $\mathbf{y}_{i,t-1}$ is the skill requirement vector of the worker's current job (normalized to zero for unemployed workers). We then let any employed worker be randomly hit by a job destruction shock (probability δ) or an outside offer (probability λ_1). Any employed worker hit by a job destruction shock starts the following period as unemployed. Any employed worker receiving an outside offer draws job attributes $\mathbf{y}'$ from the sampling distribution $\Upsilon(\cdot)$ and, depending on the comparison between the value of his current job $P(\mathbf{x}_{it}, \mathbf{y}_{i,t-1})$

and that of his outside offer $P(\mathbf{x}_{it}, \mathbf{y}')$, either accepts the offer (in which case their job attribute vector gets updated to $\mathbf{y}_{it} = \mathbf{y}'$), or stays in his job, with or without a contract renegotiation. In each case, the worker's period-t wage $w_{i,t}$ is updated according to equation (5). Symmetrically, we let any unemployed worker draw a job offer (probability λ_0) with job attributes $\mathbf{y}' \sim \Upsilon(\cdot)$, which the worker accepts if and only if $P(\mathbf{x}_{it}, \mathbf{y}') \geq U(\mathbf{x}_{it})$. Again, the worker's wage is updated.[22]

To set the initial ($t = 0$) condition, we simulate the model over a "pre-sampling" period, starting from a situation where all workers are unemployed. We then run the simulation as described above, shutting down skill updating and layoffs. We stop the pre-sampling simulation when the simulated nonemployment rate reaches a value of 35% (the observed nonemployment rate in our NLSY sample), and take the current state of the sample at that point as the initial condition.

Each simulation thus produces an $N \times T$ (balanced) panel of worker data with the same format as our estimation sample. The simulated sample keeps track of each worker's employment status, labor market transitions, wages w_{it}, skill bundle $\mathbf{x}_{it}$, and job attributes $\mathbf{y}_{it}$.

Model parameterization. We use the specification introduced in Subsection (3.2) which, because it affords closed-form solutions, considerably reduces the computational burden. The skill adjustment, production, and disutility of work functions are specified as in (6), (8), and (9) respectively, with the following additional parameterization:

$$\varphi(\mathbf{y}) = \alpha_T + \alpha_C y_C + \alpha_M y_M + \alpha_I y_I.$$

We further impose $\alpha_k > 0$, $k = C, M, I$ to ensure that $\varphi(\cdot)$ is an increasing function.

We interpret (x_C, x_M, x_I) and (y_C, y_M, y_I) as the model counterparts of the cognitive and manual skill indices we constructed from out combination of O*NET and NLSY data as explained in the previous section. The joint distribution of initial cognitive, manual, and interpersonal worker skills $(x_C(0), x_M(0), x_I(0))$ is fully observed in the data, and requires no parameterization. General worker efficiency grows along with potential experience t at a constant rate g. In addition, we allow it to be correlated in an unrestricted way with initial cognitive, manual and interpersonal skills

[22] Note that, in the simulation, we shut down sample attrition (which in the model occurs at rate μ). Attrition is random in the model, the only impact would be to reduce the simulated sample size, which we can usefully avoid.

$(x_C(0), x_M(0), x_I(0))$, as well as education:

$$x_T(t) = \exp\left(g \cdot t + \zeta_S \cdot \text{YEARS\_OF\_SCHOOLING} + \zeta_C x_C(0) + \zeta_M x_M(0) + \zeta_I x_I(0) + \varepsilon_0\right),$$

where the ζ's are coefficients and ε_0 is an uncorrelated unobserved heterogeneity term such that the mean of e^{ε_0} is normalized to 1. Given the model's structure, this makes e^{ε_0} an uncorrelated mixing variable that multiplies all individual wages and values. In particular, observed log-wages $\ln w$ are such that $\ln w \stackrel{d}{=} \ln w|_{\varepsilon_0=0} + \varepsilon_0$, where $\stackrel{d}{=}$ denotes equality in distributions, and $w|_{\varepsilon_0=0}$ denote simulated wages under the assumption that all workers have $\varepsilon_0 = 0$. We can thus estimate the model abstracting from this particular heterogeneity (*i.e.* assuming $\varepsilon_0 = 0$ for all workers), then retrieve the distribution of ε_0 by deconvolution.

Finally, we specify the skill requirements (y_C, y_M, y_I) as simple transforms of the skill requirement indices $(\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ constructed from the O*NET data as described in Section 4. This is to allow for the possibility that our constructed $\widetilde{\mathbf{y}}$'s might not be exactly aligned with our measured worker skills $\mathbf{x}$ (see the discussion in Section 4). Specifically, we assume that $y_k = \widetilde{y}_k^{\xi_k}$, with $\xi_k > 0$, thus ensuring that y_k is an increasing transformation of $\widetilde{y}_k$ that stays in the unit interval. We then approximate the joint sampling distributions of job attributes $\Upsilon(\mathbf{y})$ using a Gaussian copula and Pareto marginals (this means assuming that the marginal sampling density of skill requirement y_k, $k = C, M, I$ is proportional to $(1 - y_k)^{\beta_k - 1}$). The rank correlation parameters $(\rho_{CM}, \rho_{CI}, \rho_{MI})$ of the Gaussian copula are to be estimated, together with the parameters $(\beta_C, \beta_M, \beta_I)$ of the three marginals. This specification proves flexible enough to offer a good fit.

5.2 Targeted Moments

The specification of our model laid out in Subsection 5.1 involves the parameter vector described earlier in this paper and summarized in Appendix A.3. Among those parameters, we fix the discount rate r and the sample attrition rate μ to "standard" values (the monthly equivalent of 10% per annum for r, and 0.002 for μ, implying an average working life of 42 years). As explained before, the joint distribution of initial cognitive, manual, and interpersonal worker skills $(x_C(0), x_M(0), x_I(0))$ is observed in the initial cross-section of our estimation panel. Finally, the job destruction rate δ

has a direct empirical counterpart, namely the sample average job loss ("E2U") rate.[23] With this subset of parameters estimated - or calibrated - in a preliminary step, we are left with a 32-dimensional parameter vector to estimate (summarized in Appendix A.3). We estimate these parameters by matching the following set of moments: *(i)* sample mean U2E rate, *(ii)* mean E2E rate profile (summarized by average E2E rates over six consecutive equal-length subsets of the observation window), *(iii)* mean and standard deviation of the marginal cross-sectional distributions of current job attributes $\widetilde{\mathbf{y}}_{it}$ among employed workers at a selection of sampling dates,[24] *(iv)* pairwise correlations of skill requirements, $\text{corr}(\widetilde{y}_{k,it}, \widetilde{y}_{k',it})$, $k' \neq k$, $(k, k') \in \{C, M, I\}^2$ among jobs held by employed workers at a selection of sampling dates *(v)* correlations of initial worker cognitive, manual and interpersonal skills and the skill requirements of jobs held, $\text{corr}(x_{k,i0}, \widetilde{y}_{k,it})$, $k = C, M, I$ at a selection of sampling dates *(vi)* coefficients of a regression of log wages $\ln w_{it}$ on initial skills $\mathbf{x}_{i0}$, current job attributes $\widetilde{\mathbf{y}}_{it}$, tenure, experience, and years of schooling (*i.e.* the regression in Table 5). We drop the first simulated wage out of unemployment for the wage regression.[25] The model-based moments are computed from simulated samples of $N = 35,400$ workers - twenty replicas of the initial NLSY cross-section.

5.3 Identification

Appendix A.6 formally discusses identification of the model laid out in Section 3 (given the parameterization described in 3.2) from a data set with the structure and contents described in Section 4. In this Subsection we summarize the main sources of information that identify the various components of our model.

The levels of wages conditional on education, experience, initial skills and (observed) job skill requirements identify the returns to education and initial skills (the parameters ζ_S, ζ_C, ζ_M and ζ_I), the wage trend (the parameter g), and the baseline returns to job skill requirements (the function $\varphi(\mathbf{y})$). The (production/utility) costs

[23]Figure 6 suggests that the job loss rate is not exactly constant over a worker's life cycle. We abstract from this feature of the data.

[24]In practice, we compute those moments at six dates corresponding to 2.5, 5, 7.5, 10, 12.5 and 15 years into the sample.

[25]In this version of the sequential auction model, in which workers are risk-neutral and have no bargaining power, workers tend to accept very low wages upon exiting unemployment, to "buy their way" onto the job ladder. As soon as a worker receives her first outside offer the wage will jump. We drop the initial wage out of unemployment so as not to bias our estimate of human capital accumulation due to the large wage change at the very beginning of an employment spell.

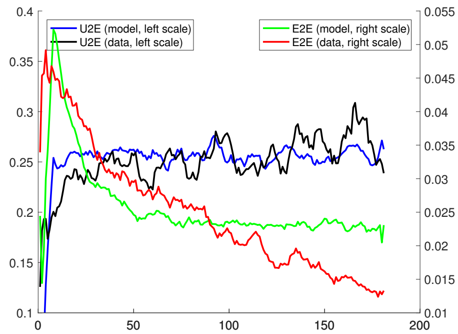
of mismatch and the speed of human capital accumulation or decay (parameters $\kappa_k^{u/o}$ and $\gamma_k^{u/o}$, $k = C, M, I$) are identified from comparisons of the sets of job types $\mathbf{y}$ that are acceptable to workers with equal initial skills $\mathbf{x}(0)$, but have experienced different employment histories. Knowledge of any worker's initial skill bundle $\mathbf{x}(0)$ and full labor market spell history, combined with the knowledge of the skill adjustment process (parameters $\gamma_k^{u/o}$) then enables us to construct the full path of skill bundles $\mathbf{x}(t)$ for all workers in the sample. The set of job offers accepted by unemployed workers with given skill bundle $\mathbf{x}$ then identifies the sampling distribution $\Upsilon(\mathbf{y})$ over the set $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) \geq U\}$, so that $\Upsilon(\mathbf{y})$ is identified over the union of all such sets for all skill bundles $\mathbf{x}$ observed in the sample (that is, $\Upsilon(\mathbf{y})$ is identified at all skill requirement levels $\mathbf{y}$ that are acceptable by at least some worker types). Finally, the offer arrival rates λ_0 and λ_1 are identified, conditionally on the rest of the model, from sample U2E and E2E transition probabilities.

Although the exact arguments used in Appendix A.6 to establish identification are not literally taken up in the practical estimation protocol, the information contained in the moments we use for estimation (listed in Subsection 5.2) does echo those arguments. In particular, the cross-section wage regression coefficients that we seek to replicate contain the information needed to identify the parameters of $\varphi(\mathbf{y})$, ζ, and g. Moreover, the various moments of the joint distribution of initial worker skills and current job skill requirements convey information about the set of matches that are acceptable to a given worker type, which is used to identify $\kappa_k^{u/o}$, $\gamma_k^{u/o}$, and ultimately the sampling distribution $\Upsilon(\cdot)$. In practice, our chosen moments ensure precise "local" identification of the model's parameters, in the sense that the distance between data-based model-predicted moments has a clear local minimum at the estimated parameter value.

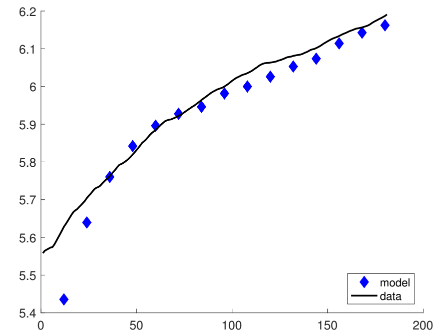
6 Results

6.1 Model Fit

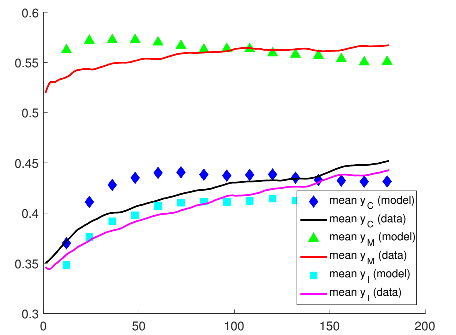
Figure 1 illustrates various aspects of the fit. All time series on Figure 1 are plotted over a period of 15 years (180 months, *i.e.* the sample window used for estimation). Table 1 further shows the fit in terms of the descriptive wage regression discussed in Section 4.



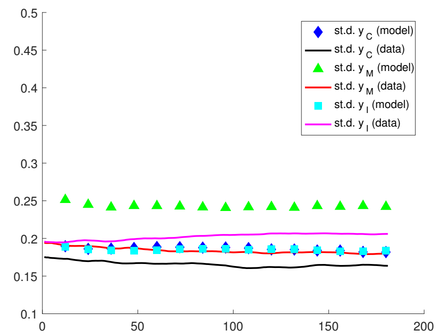
(a) Transition rates



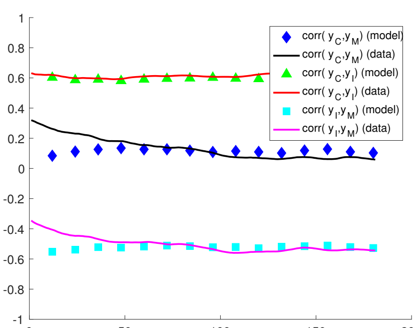
(b) Wage/experience profile



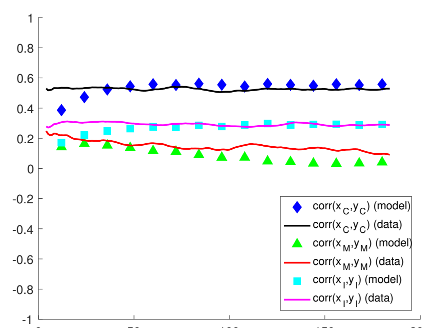
(c) Cross-sectional mean job attributes



(d) Cross-sectional st.d. of job attributes



(e) Correlation of job attributes



(f) Corr. of job and worker attributes

Figure 1: Model fit

Table 1: Fit to wage regression coefficients

	data	model		data	model
$\widetilde{y}_C$	0.664	0.646	x_{I0}	0.270	0.270
$\widetilde{y}_M$	0.246	0.240	years of schooling	0.026	0.0256
$\widetilde{y}_I$	0.378	0.371	experience	2.24e − 3	2.16e − 3
x_{C0}	0.363	0.363	tenure	1.98e − 3	4.99e − 3
x_{M0}	−0.101	−0.101	constant	4.455	4.455

The model fits both the nonemployment exit rate (Figure 1a, left scale) and the job-to-job (E2E) transition rate (Figure 1a, right scale) reasonably well. The decline of E2E rates with experience is correctly captured by the model (it occurs as a consequence of workers gradually settling into jobs to which their skills are better suited, both because they sort into better matches over time and because their skills adjust to whatever job they are in at any given time), even though it overstates both the initial speed of that decline and the level of the E2E rate at high levels of experience. The fit to the U2E rate is good, although the model does not capture the mild upward trend in that rate. All of the discrepancies between data and model in Figure 1a are largely due to our restriction to experience-invariant contact rates, λ_0 and λ_1.

The sample average wage/experience profile is shown on Figure 1b and is reasonably well captured by the model, despite a tendency to overstate its concavity.

Figures 1c through f show the time-profiles of various fitted cross-sectional moments of the joint distribution of workers' initial skills $(x_{C,i0}, x_{M,i0}, x_{I,i0})$ and current job attributes $(\widetilde{y}_{C,it}, \widetilde{y}_{M,it}, \widetilde{y}_{I,it})$ in the population of employed workers, at a selection of experience levels. The model offers a good fit to all targeted moments: it captures the rise in average cognitive and interpersonal job attributes (although is does not fully replicate the magnitude of the changes over the first four years), as well as the near constancy of average manual job attributes. The average levels of manual and interpersonal skill requirements are mildly understated, and the standard deviation of manual skill requirements is slightly overstated by the model.

We next turn to the model's ability to replicate the pooled cross-section wage regression shown in Table 1. The model correctly predicts a much stronger cross-section correlation of log wages with the cognitive skill requirement index $\widetilde{y}_C$ than with the manual ($\widetilde{y}_M$) or interpersonal ($\widetilde{y}_I$) skill requirement indices. Coefficients on "conventional" regressors (schooling, experience, and tenure) are well captured by the model (despite a tendency to overstate the returns to job tenure), and so is the conditional correlation between wages and initial worker skills.[26]

6.2 Parameter Estimates

Table 2 shows point estimates of the model parameters. Asymptotic standard errors are reported in parentheses, below each point estimate.[27] There is little to say about the offer arrival and job destruction rates, apart maybe from the fact that the estimated relative search intensity of employed workers, λ_1/λ_0, is in the region of 0.45, which is on the high side, although not completely outside of the set of standard estimates on US data. Overall job productivity $\varphi(\mathbf{y})$ is increasing in all cognitive, manual and interpersonal skill requirements, with the loading on cognitive skills about twice as large as the ones on manual and interpersonal skills. This is consistent with the lower coefficients on $\widetilde{y}_M$ and $\widetilde{y}_I$ than on $\widetilde{y}_C$ in the wage regression shown in Table 1.

Overall worker efficiency x_T is positively associated with a high initial endowment in cognitive and (to a lesser extent) in interpersonal skills ($\zeta_C > \zeta_I > 0$), while initial manual skills are negatively correlated with x_T ($\zeta_M < 0$). One additional year of education increases efficiency by 2.4 percent (ζ_S). However one should bear in mind that education is positively correlated with initial cognitive and interpersonal skills and (weakly) negatively correlated with initial manual skills in the sample. The value of ζ_S taken in isolation therefore understates the overall returns to education.

The employment of an under-qualified worker in any skill dimension is costly in terms of output, yet the output loss caused by this type of mismatch by far most severe in the cognitive dimension and least severe in the interpersonal dimension. The utility cost of being under-matched - *i.e.* the surplus cost of the worker being *over*-qualified worker - is positive in all dimensions, but generally much smaller than

[26]See Appendix A.7 for additional details on the cross-sectional fit to wages with and with and without unobserved heterogeneity ε_0.

[27]The covariance matrix of the parameter vector is estimated as $\left(G^\top G\right)^{-1} G^\top \Omega G \left(G^\top G\right)^{-1}$, where $G = \mathbb{E}\left[\partial \mathbf{m}/\partial \Theta^\top\right]$, the (expectation of the) Jacobian matrix of the moment function $\mathbf{m}(\Theta)$, is obtained by numerical differentiation and where Ω, the covariance matrix of the moment function, is estimated by resampling the data 500 times, repeating the construction of job and worker attributes each time. Reported standard errors therefore account for the fact that $\mathbf{x}$ and $\widetilde{\mathbf{y}}$ are estimated in a preliminary step.

Table 2: Parameter estimates

production function*							disutility of work*			un. inc.
α_T	α_C	α_M	α_I	κ_C^u	κ_M^u	κ_I^u	κ_C^o	κ_M^o	κ_I^o	b
108.4	119.6	54.0	54.2	3,188.5	474.2	178.8	42.7	207.5	66.5	108.4
(21.0)	(15.1)	(4.50)	(8.56)	(492.9)	(109.3)	(47.9)	(8.33)	(58.1)	(13.7)	(21.7)
				(128.8)	*(14.9)*	*(8.3)*	*(2.1)*	*(7.0)*	*(3.3)*	

skill accumulation function**							general efficiency			
γ_C^u	γ_C^o	γ_M^u	γ_M^o	γ_I^u	γ_I^o	g	ζ_S	ζ_C	ζ_M	ζ_I
0.008	0.004	0.034	0.031	0.001	5.8e − 7	0.002	0.029	0.59	−0.12	0.34
(.001)	(.001)	(.008)	(.004)	(.006)	(.002)	(9e−4)	(.023)	(.238)	(.201)	(.208)
(7.54)	*(16.6)*	*(1.70)*	*(1.88)*	*(55.8)*	*(99,424)*					

sampling distribution***									transition rates		
ξ_C	ξ_M	ξ_I	ρ_{CM}	ρ_{CI}	ρ_{IM}	β_C	β_M	β_I	λ_0	λ_1	δ^{****}
1.00	0.79	0.86	0.13	0.70	−0.47	2.55	1.21	3.04	0.33	0.17	0.02
(.044)	(.072)	(.056)	(.021)	(.015)	(.019)	(.109)	(.117)	(.261)	(.014)	(.003)	(.001)
			(0.16)	*(0.67)*	*(−0.42)*	*(0.29)*	*(0.46)*	*(0.25)*			

*percent surplus loss caused by deviating from ideal match by 1 SD of Υ at mean $\mathbf{y}$ in italics;

** half-life in years in italics ; *** implied correlations and means in italics ; **** estimated in first step the corresponding surplus (production) cost of *under*-qualification. To give a sense of the orders of magnitude involved, the numbers in italics below the estimates of the various κ parameters give the percentage flow-surplus cost of deviating from the ideal match at the mean sampled $\mathbf{y}$ by one standard deviation of the sampling distribution Υ.[28]

The correlation patterns between skill requirements in the sampling distribution ($\rho_{CM} > 0$, $\rho_{CI} > 0$, $\rho_{IM} < 0$) suggests that jobs requiring high levels of cognitive skills also tend to require high levels of skills in at least one of the other two dimensions, particularly interpersonal. Jobs with a high manual content, however, tend to have low interpersonal requirements. The next section offers a more complete interpretation of the sampling distribution.

The pattern of skill adjustment differs vastly between cognitive and manual skills. Manual skills adjust much faster than cognitive skills. Cognitive skills are very persistent (*i.e.* not easily accumulated or lost) with a half-life of 7.6 years to learn and 17.5 years to forget. The half-life of manual skills is much shorter, about 16 months to acquire and two years to lose. Interpersonal skills essentially do not adjust over a worker's typical horizon and can, to a good approximation, be treated as fixed worker traits.

Perhaps the clearest message from those estimates is that the model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have relatively low returns and adjust quickly in both directions, cognitive skills have much higher returns, but are much slower to adjust. Interpersonal skills have lower returns than cognitive skills, but higher than manual skills (especially through their effect on overall worker efficiency), and are essentially fixed over a worker's lifetime. Finally, skill mismatch is most costly in the cognitive dimension and in the "underskilled" direction (*i.e.* when the worker has lower skills than the job requires).

6.3 Skill Mismatch, Skill Changes, and Sorting

Distributions of skills and skill requirements. The broad question of skill mismatch can be understood in many different ways. One aspect of that question is the alignment (or lack thereof) between the skills that workers are equipped with

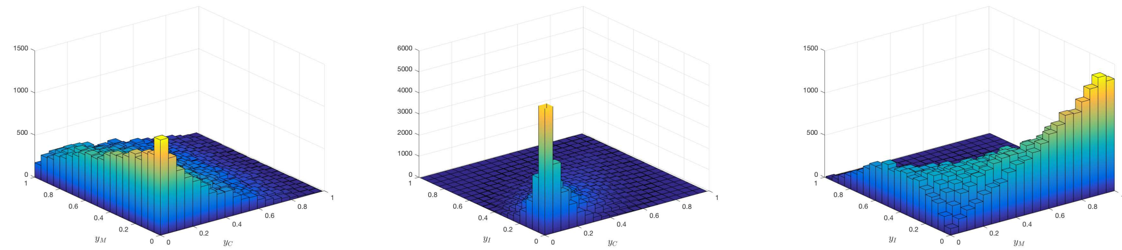
[28]For example, denoting the mean of Υ as $\mathbf{y}^m = (y_C^m, y_M^m, y_I^m)$ and the standard deviations of the marginals of Υ as $(\sigma_C, \sigma_M, \sigma_I)$, the percentage reported in italics below the estimate of κ_C^u in Table 2 is $100 \times \left[1 - \frac{P(\mathbf{y}^m + (\sigma_C,0,0), \mathbf{y}^m) - U(\mathbf{y}^m + (\sigma_C,0,0))}{P(\mathbf{y}^m, \mathbf{y}^m) - U(\mathbf{y}^m)}\right]$.

when they leave education - the distribution $N(\cdot)$ of initial worker skill bundles $\mathbf{x}(0)$, in the parlance of the model - and the firms' skill requirements - the model counterpart of which is the sampling distribution $\Upsilon(\mathbf{y})$.

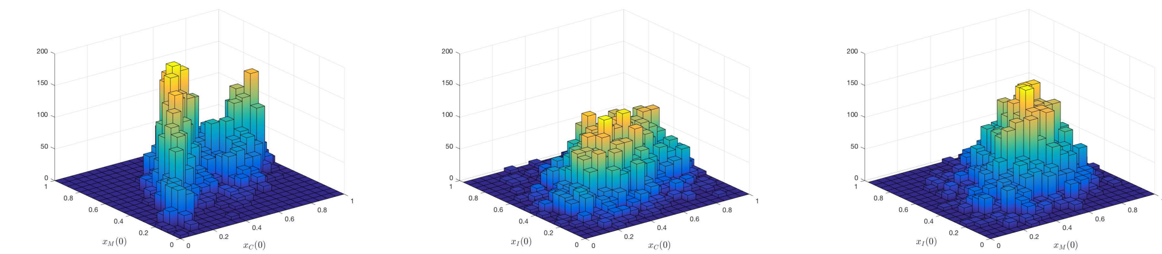
The obvious common features of those two distributions are the negative correlation between manual skills and both other skill dimensions, and the concentration around the "corners" of the skill spectrum, particularly around (x_C, x_M) or $(y_C, y_M) = (1, 0)$ on one hand, and (x_C, x_M) or $(y_C, y_M) = (0, 1)$ on the other (see Figures 2a, b and c). This strongly suggests that employers are looking for "specialist" workers, endowed with a high amount of either cognitive or manual skills, rather than "generalists" who would have average skills in both dimensions. Figures 2d, e and f also suggest that the distribution of skills among labor market entrants reflects this demand for specialists to an extent. However, those skill distributions look somewhat different from the sampling distribution of job skill requirements. In particular, the distribution of initial worker skills has more mass toward the high cognitive skill end of the skill spectrum than the distribution of job offers: employers seem to demand fewer cognitive skills than are available in the population of labor market entrants.[29] The same is true, to a somewhat lesser extent, of the other two skill types: the distribution of skills among labor market entrants is more concentrated around "middling" skills than the sampling distribution is.

Figures 2g-l next shows how the distribution of worker skills changes as the cohort of workers accumulates experience. The evolution is clearly towards workers gaining cognitive skills and losing manual skills on average (while, as we saw earlier, interpersonal skills hardly change over a worker's lifetime). This can be explained by the fact that jobs with high cognitive skill requirements are intrinsically more productive (the estimated weight on y_C in the production function, α_C, is an order of magnitude larger than the weights on y_M and y_I), inducing workers to accept jobs with the highest cognitive content compatible with their level of cognitive skills, even if it means ending up severely overskilled in the manual dimension. Because jobs with higher cognitive skill requirements tend to have relatively low manual skill requirements (as y_C and y_M are strongly negatively correlated in the job offer sampling distribution),

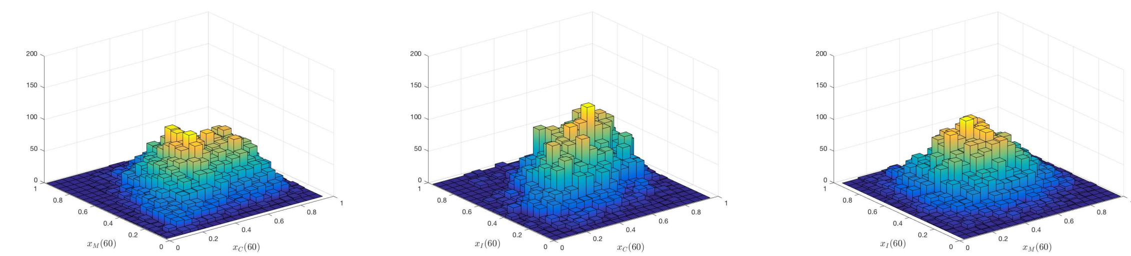
[29] Those differences are not entirely surprising: the population of workers whose skills are represented on Figures 2d, e and f is a cohort of relatively young workers. As such, their skills may not be representative of those in the entire active workforce. By contrast, at least under random search, the sampling distribution addresses all workers, the majority of which are from older cohorts among which the skill distribution may be quite different.



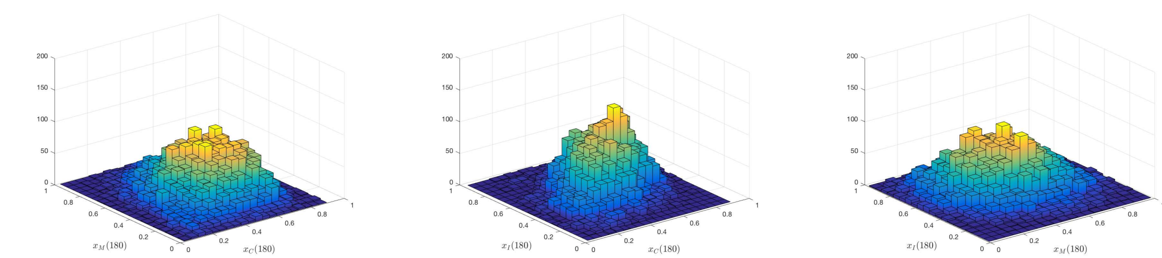
(a) Sampling density, CM (b) Sampling density, CI (c) Sampling density, MI



(d) Initial worker skill dist., CM (e) Initial worker skill dist., CI (f) Initial worker skill dist., MI



(g) 5 years of experience, CM (h) 5 years of experience, CI (i) 5 years of experience, MI



(j) 15 years of experience, CM (k) 15 years of experience, CI (l) 15 years of experience, MI

Figure 2: Distribution of skill requirements and evolution of worker skills with experience

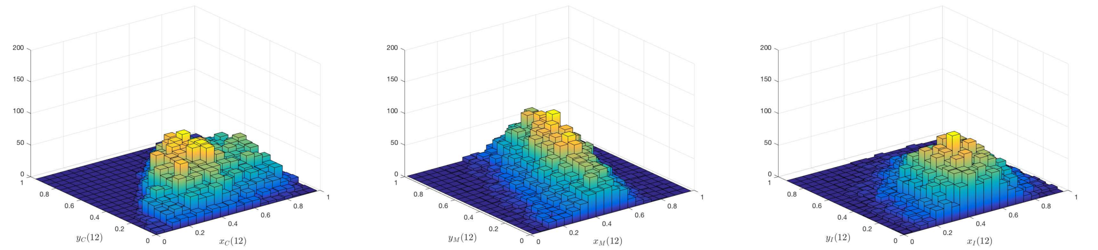
by following this strategy workers tend to maintain or gradually acquire cognitive skills and lose manual skills (which they do relatively fast, given the high estimated adjustment speed of manual skills, $\gamma_M^{u/o}$). A second striking feature of Figures 2 is that, although the typical worker's skills tend to migrate toward more cognitive skills and less manual skills on average, the already limited degree of specialization apparent in the initial skill distribution (Figure 2a) regresses even further as workers gain experience.

Skill sorting and mismatch. We next examine the joint distribution of worker skill bundles and job skill requirements among ongoing matches. Figure 3 shows two examples of those joint distributions, among workers who are one year into their careers (panels a, c and e), and among workers with fifteen years of experience (panels b, d and f). Simply eyeballing these histograms gives a distinct impression of positive sorting in all skill dimensions, even at early stages of the working life. Moreover, the "strength" of this positive sorting - as measured by the (inverse of the) conditional dispersion in worker skills for a given level of skill requirement - clearly increases as workers accumulate experience. This results from the combination of workers gradually sorting themselves into jobs for which their skills are better suited, and adjusting their initial skills to their job's requirements: as can be seen from Figure 3, sorting at 15 years of experience is strongest in the manual dimension (as manual skills adjust quickly), and weakest in the interpersonal dimension (as interpersonal skills do not adjust).

Concerning early-career skill mismatch, inspection of Figures 3a and 3c suggests that it is stronger in the manual than in the cognitive dimension.[30] This echoes a remark made before, that workers tend to prioritize a good match in the cognitive dimension, sometimes to the detriment of match quality in the other skill dimensions.

A final feature of Figure 3 is that, while there is largely positive sorting in all skill dimensions, a substantial mass of workers appear "under-matched" in the cognitive dimension, in the sense that their job's cognitive skill requirement is lower than their own cognitive skill level. By contrast, very few workers are "over-matched" in the cognitive dimension (and those who are are so by a small margin). The tradeoff from the perspective of a worker contemplating a job is between the job's overall produc-

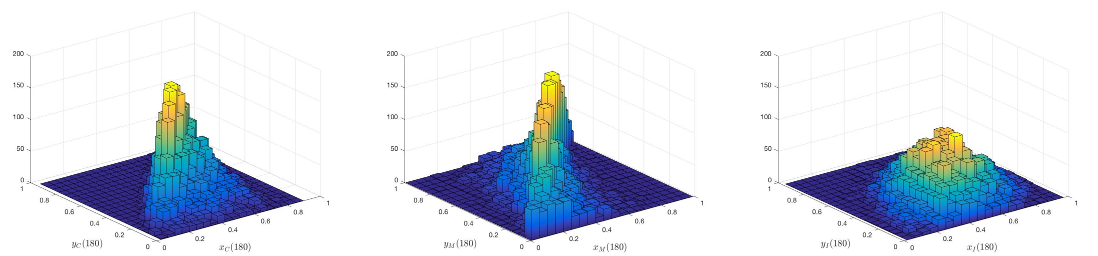
[30]This corroborates the descriptive evidence in Table 4 which showed that the correlation between a worker's skill level and the corresponding requirement in their first job is much stronger in the cognitive than in the manual or interpersonal dimension.



(a) Cognitive skills, one year of exp.

(b) Manual skills, one year of exp.

(c) Interpersonal skills, one year of exp.



(d) Cognitive skills, 15 years of exp.

(e) Manual skills, 15 years of exp.

(f) Interpersonal skills, 15 years of exp.

Figure 3: Sorting

tivity (the $\varphi(\mathbf{y})$ term in the production function), and any cost of being mismatched. In the case of cognitive skills, the cost of being "over-matched" (or "under-skilled"), measured by κ_C^u, is prohibitively high, even accounting for the fact that $\varphi(\mathbf{y})$ increases much more steeply with y_C than with y_M or y_I (see Table 2).

7 Counterfactual Experiments

We now use our structural model to run three different counterfactual experiments. In the first experiment, we ask how distant the estimated market allocation is from the constrained efficient allocation, by which we mean the allocation that a benevolent Planner would select for given job contact and job destruction rates, $(\lambda_0, \lambda_1, \delta)$. In other words, we are investigating the efficiency of job acceptance and rejection decisions, given the estimated rates at which workers receive offers. In the second experiment, we offer a decomposition of the aggregate output cost of frictions. In the third experiment, we focus on the cost of early-career mismatch.

7.1 Comparison with the Constrained Efficient Allocation

As discussed in Section 3, a consequence of Bertrand competition as a rent-sharing mechanism is that private firm-worker collectives do not internalize the extra surplus that is created when the worker moves to a competing job for which her/his skills are better suited, as that extra surplus is captured by a third party, namely the worker's future employer. Similarly, unemployed workers fail to internalize the surplus associated with them finding a job, as that surplus is entirely captured by their future employer. As a consequence, neither the private value of a match (2) nor the private value of unemployment (3) depend on the job contact rates (λ_0, λ_1).

By contrast, the *social* value of a match, or that of unemployed search (both denoted with stars in what follows) do internalize the surplus generated in future matches. Those values solve:

$$(r+\delta+\mu)P^*(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta U^*(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} P^*(\mathbf{x},\mathbf{y}) \\ + \lambda_1 \mathbf{E} \max\left\{P^*(\mathbf{x},\mathbf{y}') - P^*(\mathbf{x},\mathbf{y}), 0\right\} \quad (11)$$

and:

$$(r+\mu)U^*(\mathbf{x}) = b(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{0}) \cdot \nabla U^*(\mathbf{x}) + \lambda_0 \mathbf{E} \max\left\{P^*(\mathbf{x},\mathbf{y}') - U^*(\mathbf{x}), 0\right\} \quad (12)$$

where, in both cases, the last (expectation) term captures the expected surplus gains from future matches. Those social values $[P^*(\mathbf{x},\mathbf{y}), U^*(\mathbf{x})]$ differ from their private counterparts $[P(\mathbf{x},\mathbf{y}), U(\mathbf{x})]$ (given by (2) and (3)) precisely because of those expectation terms. We now look at the nature of those discrepancies, starting with the simpler case without human capital accumulation ($\mathbf{g}(\mathbf{x},\mathbf{y}) \equiv 0$).

The no-human-capital-accumulation benchmark: $\mathbf{g}(\mathbf{x},\mathbf{y}) \equiv 0$. In the case without human capital accumulation, job-to-job reallocation is efficient (for given job contact rates) but nonemployment-to-employment reallocation is not. To see this, observe that a combination of equations (2) and (11) when $\mathbf{g}(\mathbf{x},\mathbf{y}) \equiv 0$ produces, for any $(\mathbf{y}_1, \mathbf{y}_2)$:

$$(r+\delta+\mu)\left[P^*(\mathbf{x},\mathbf{y}_2) - P^*(\mathbf{x},\mathbf{y}_1)\right] = (r+\delta+\mu)\left[P(\mathbf{x},\mathbf{y}_2) - P(\mathbf{x},\mathbf{y}_1)\right] \\ + \lambda_1 \mathbf{E}\left[\max\left\{P^*(\mathbf{x},\mathbf{y}') - P^*(\mathbf{x},\mathbf{y}_2), 0\right\} - \max\left\{P^*(\mathbf{x},\mathbf{y}') - P^*(\mathbf{x},\mathbf{y}_1), 0\right\}\right].$$

This latter equation implies that $P(\mathbf{x},\mathbf{y}_2) > P(\mathbf{x},\mathbf{y}_1) \iff P^*(\mathbf{x},\mathbf{y}_2) > P^*(\mathbf{x},\mathbf{y}_1)$, which in turn implies that job-to-job reallocation is efficient in the decentralized economy: workers move from job $\mathbf{y}_1$ to job $\mathbf{y}_2$ iff. $P(\mathbf{x},\mathbf{y}_2) > P(\mathbf{x},\mathbf{y}_1)$ *i.e.* iff $P^*(\mathbf{x},\mathbf{y}_2) > P^*(\mathbf{x},\mathbf{y}_1)$, which would be the Planner's criterion for reallocation.

Things are different at the employment/nonemployment margin. Workers in the decentralized economy move from nonemployment to job $\mathbf{y}$ iff. $P(\mathbf{x},\mathbf{y}) > U(\mathbf{x})$. Combining (2)-(3) and (11)-(12):

$$\begin{aligned}(r+\delta+\mu)\left[P^*(\mathbf{x},\mathbf{y}) - U^*(\mathbf{x})\right] &= (r+\delta+\mu)\left[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})\right] \\ &- \lambda_0 \mathbf{E}\max\left\{P^*(\mathbf{x},\mathbf{y}') - U^*(\mathbf{x}), 0\right\} + \lambda_1 \mathbf{E}\max\left\{P^*(\mathbf{x},\mathbf{y}') - P^*(\mathbf{x},\mathbf{y}), 0\right\}.\end{aligned}$$

Assuming that $\lambda_0 > \lambda_1$ as found in the estimation (the converse assumption is easily investigated), the latter equation implies:

$$\begin{aligned}(r+\delta+\mu)\left[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})\right] &> (r+\delta+\mu)\left[P^*(\mathbf{x},\mathbf{y}) - U^*(\mathbf{x})\right] \\ &- \lambda_0 \left[\mathbf{E}\max\left\{P^*(\mathbf{x},\mathbf{y}') - U^*(\mathbf{x}), 0\right\} - \mathbf{E}\max\left\{P^*(\mathbf{x},\mathbf{y}') - P^*(\mathbf{x},\mathbf{y}), 0\right\}\right].\end{aligned}$$

In this situation, $P^*(\mathbf{x},\mathbf{y}) \geq U^*(\mathbf{x}) \Longrightarrow P(\mathbf{x},\mathbf{y}) > U(\mathbf{x})$, but the converse implication is not true. Thus, the decentralized economy has too high an unemployment exit rate: some matches are accepted which would be better rejected from a social point of view. The reason is that private agents do not internalize the loss of search efficiency entailed by a move from non-employment into employment.[31]

Reintroducing human capital accumulation. Job-specific human capital accumulation adds a layer of complexity to the problem: human capital is now accumulated at different speeds in different matches, which affects the relative social value of potential matches, in a way that is not fully internalized by private agents. Once again, private agents do not internalize the expected surplus created in future matches, part of which arises from the fact that workers will accumulate more skills, and therefore gain access to even better jobs, if they move to a job with higher skill requirements. This phenomenon is particularly glaring at the employment/nonemployment margin, where a nonemployed worker contemplating a job fails to internalize the fact

[31] It is easy to show along the same lines that unemployment exit is inefficiently low in the decentralized economy if $\lambda_0 < \lambda_1$, and is efficient if $\lambda_0 = \lambda_1$.

(a) Transition rates

(b) Percent difference in net aggregate output

Figure 4: Comparison of market and constrained-efficient allocations

that, should she take the job, her skills will stop atrophying (or at least stop atrophying as fast) as they would if she stayed unemployed. The failure to internalize this gain in terms of skill accumulation tends to induce nonemployed individuals to *reject too many job offers*, thus balancing their tendency (highlighted in the previous paragraph) to accept too many such jobs as a result of their failure to internalize the loss of job search efficiency associated with a move into employment.

Unfortunately, the combined net effect of heterogeneity in skill accumulation across potential matches and a different search technology between employment and nonemployment is too complex to make any generic theoretical statement about the direction of the inefficiencies arising in equilibrium. Is the nonemployment rate too high or too low in equilibrium? Is the turnover rate too high or too low? We now address these questions *at the estimated parameter values*, by comparing our model's simulated equilibrium to a simulated "Planner's solution". We obtain the latter by simulating the model using the exact same protocol as described in Sub-section (5.1), except that the criterion used for job acceptance hinges on the comparison of social values $[P^*(\mathbf{x},\mathbf{y}), U^*(\mathbf{x})]$ rather than private ones $[P(\mathbf{x},\mathbf{y}), U(\mathbf{x})]$.[32]

Figure 4 shows, by experience level, comparisons of the Planner's average U2E and E2E transition rates (Panel a), as well as the percentage difference in aggregate net output between the Planner's and the market allocation (Panel b). Aggregate net output is defined as the sum of total output produced by employed workers net of possible utility costs of under-qualification, plus total output produced by nonemployed workers.

Figure 4a shows that the Planner's hiring rate from unemployment is consistently and markedly *lower* than the market's. In other words, similar to what the model

[32]Unlike (2) and (3), equations (11) and (12) cannot be solved in closed form even under the functional forms used in this paper. We thus solve (11) and (12) for $P^*(\mathbf{x},\mathbf{y})$ and $U^*(\mathbf{x})$ numerically using projections of $P^*(\mathbf{x},\mathbf{y})$ and $U^*(\mathbf{x})$ on complete sets of polynomials.

without human capital accumulation would predict, the market moves too many workers out of nonemployment even in the presence of human capital accumulation: the cost in terms of human capital loss of keeping a worker in unemployment, which the market fails to internalize, mitigates but does not dominate the higher efficiency of nonemployed over employed job search.

Interestingly, the market's average E2E transition rate is also higher than the Planner's at all experience levels. One way to see this is that the private agents' failure to internalize the impact of human capital accumulation in their current match on the surplus they will generate in future matches makes them too eager to take up jobs with high short-term gain but limited scope for further learning. As a result, the decentralized economy features excess worker turnover and too little unemployment.

Finally, Figure 4b suggests that, while the Planner's solution produces higher surplus for the cohort under study, the improvement brought about by the Planner is very limited (in the order of one to two percentage points), and concentrated around early stages of workers' careers (which is when worker reallocation is mostly from nonemployment to employment). This improvement can be viewed as an upper bound on the potential gain achievable through policy intervention since we are comparing the case where worker-firm matches do not internalize any of the effect of the current match on future matches with the Planner's solution that fully internalizes these effects. In a model where workers have some bargaining power (Cahuc, Postel-Vinay, and Robin, 2006) they would internalize part of this externality. Indeed, our decentralized economy would correspond to the case in which firms have all the bargaining power and the Planner's solution corresponds to the case in which workers have all the bargaining power. In a model where firms do not have all the bargaining power, the decentralized solution would be closer to our Planner's solution, leaving even less scope for policies which improve aggregate output. However it is important to recall at this point that this efficiency analysis is only partial in that it is conducted under constant job offer arrival rates. As such, it ignores any distortion in labor demand induced by the discrepancies between private and social match values.

7.2 The Cost of Frictions

In the previous subsection we looked at the difference between the market outcome and the allocations chosen by a Planner who is constrained by the same frictions as the market. We now examine the costs of those frictions. The frictions in the model restrict both the frequency and direction of worker mobility. We consider the costs of those frictions in two steps.

First, we keep the frequency of job contacts constant, but allow the worker to be allocated to the most appropriate job when an opportunity arrives. In this experiment, a worker given an opportunity to pick a new job (which happens, as in the baseline case, at flow rates λ_0 and λ_1) will choose a job whose attributes yield maximum surplus $P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})$ given her current skills $\mathbf{x}$.[33]

This first exercise eliminates mismatch upon job change, but keeps workers in a job longer than they would like if they were not constrained by search frictions. In a second exercise, we eliminate frictions altogether and solve the problem of optimally moving each worker between jobs to maximize the present value of output less the disutility of work. In this case we are trading off maximizing current output and human capital accumulation of workers. The friction-free problem is presented in detail in Appendix A.8.

Figure 5a shows, as a function of the cohort's labor market experience, the percentage difference in aggregate output between the decentralized benchmark and, respectively, a "mismatch-free" economy (first exercise, blue line) and a frictionless economy (second exercise, red line). The model estimates the overall cost of frictions to be very large: removing all frictions from the economy would increase the cohort's long-run output flow by about 40% (red line on Figure 5a). What the blue line on Figure 5a further reveals is that most of this cost arises from frictions restricting the *direction* (rather than the frequency) of reallocation: allowing workers to reallocate freely, although infrequently, would already increase long-run output by about 35%.[34]

[33]This chosen job type (say, $\mathbf{y}^\star$) is obtained by maximization of (10) w.r.t. $\mathbf{y}$: $y_k^\star = \min\left\{x_k + \frac{\alpha_k}{2\kappa_k^u}\frac{r+\delta+\mu-g+2\gamma_k^u}{r+\delta+\mu-g}, 1\right\}$. Note that the worker will systematically choose a job for which s/he is slightly *underskilled*, due to the intrinsic returns of job attributes ($\alpha_k > 0$). Also note that we consider the private worker's surplus in this exercise, as opposed to the Planner's surplus. The latter would internalize the surplus gain in future matches from human capital accumulation in the current match, as explained in the previous subsection. Implementing the Planner's rather than the private worker's solution complicates computations considerably and makes little quantitative difference.

[34]"Infrequently" in the context of this analysis should be understood as "not continuously", but still much more frequently than in the decentralized equilibrium. Indeed, this counterfactual exercise assumes that the job contact rate λ_1 is equal to its estimated value from Table 2, the same as in the decentralized equilibrium. Yet in the counterfactual, workers switch jobs each time they receive an opportunity to do so (since they are allowed to freely choose their preferred job upon being hit by a λ_1-shock), whereas they reject a substantial fraction of the randomly drawn offers they receive

../figures/fig_cf_output_cost.eps ../figures/fig_cf_early_career_mis.eps

(a) The output cost of frictions

(b) The cost of early-career mismatch

Figure 5

7.3 The Cost of Early Career Mismatch

In this final counterfactual experiment, we assess the social gain that would be brought about if workers could *initially* (*i.e.*, as they enter the labor market) be placed in their preferred job, after which they are left to behave as they do in the decentralized equilibrium. We measure this gain in terms of the expected present discounted sum of future output produced by labor market entrants (the "career output" of entrants), namely

$$Q(\mathbf{x}_0) = \mathbb{E}\left\{\int_0^{+\infty} \left(\ell_t\left[f(\mathbf{x}(t),\mathbf{y}(t)) - c(\mathbf{x}(t),\mathbf{y}(t))\right] + (1-\ell_t)b\right) e^{-rt}dt \mid \mathbf{x}(0) = \mathbf{x}_0\right\}$$

where $\ell_t \in \{0,1\}$ indicates the worker being employed at date t. Specifically, Figure 5b plots the ratio $Q^{CF}(\mathbf{x}_0)/Q^{DC}(\mathbf{x}_0) - 1$, where $Q^{DC}(\mathbf{x}_0)$ is the average career output of type-$\mathbf{x}_0$ entrants in the decentralized equilibrium and $Q^{CF}(\mathbf{x}_0)$ is the career output of a type-$\mathbf{x}_0$ entrant who is allocated to her surplus-maximizing job upon entering the labor market.[35] The ratio is plotted on Figure 5b against deciles of (x_C, x_M), for three different values of x_I, the blue, green, and red surface corresponding, respectively, to the 1st, 5th, and 9th deciles of initial interpersonal skills.

The cost of early career mismatch thus measured is large: depending on initial worker skills, career output would be 8 to 22% higher if workers could be assigned to their preferred job upon entering the labor market. The mean cost $\mathbb{E}\left[Q^{CF}(\mathbf{x}_0)/Q^{DC}(\mathbf{x}_0) - 1\right]$ where the expectation is taken over the observed distribution of $\mathbf{x}_0$ is 13.8%. A small share of this cost can be put down to initial unemployment (arguably the worst form

in the decentralized equilibrium. As a consequence, the conterfactual E2E rate is much higher than the observed (decentralized-equilibrium) one.

[35] CF stands for "counterfactual". The decentralized $Q^{DC}(\mathbf{x}_0)$ is computed under the assumption that workers start their career (at $t = 0$) in the same initial conditions as are estimated from the data. Note that the aggregate nonemployment rate amongst labor market entrants is around 30% in the data (whereas it is zero by design in the counterfactual).

of mismatch): recall that around 30% of labor market entrants are nonemployed in the decentralized equilibrium, whereas none are, by construction, in the counterfactual. The cost increases in the initial level of manual and interpersonal skills and, surprisingly, does not vary monotonically with initial cognitive skills. A useful comparison for our estimated costs come from the literature measuring the long-term wage costs associated with graduating from college during bad *vs* good times. Kahn (2010) finds substantial variation in wage losses, ranging from 1 to 20% each year for workers graduating in the worst relative to best economic conditions in the US. Oreopoulos, von Wachter, and Heisz (2012), using Canadian data, estimate an average cumulative wage loss of 5% after 10 years for individuals graduating from college in a typical recession. Our counterfactual should be viewed as an upper bound on the cost of early career mismatch since it measures the expected difference in output between an average career and a career beginning with the ideal starting job. Even when the economy is booming the ideal starting job will be substantially better than the average starting job.

8 Conclusion

In this paper we extend an otherwise standard and well-tested search-theoretic model of individual careers to allow for multidimensional skills and on-the-job learning. We estimate the model using occupation-level measures of skill requirements based on O*NET data, combined with a worker-level panel (NLSY79). We use the estimated model to shed light on the origins and costs of mismatch along three dimensions of skills: cognitive, manual, and interpersonal. We then proceed to show that the equilibrium allocation of workers into jobs generically differs from the allocation that a Planner would choose, and investigate the nature and magnitude of the resulting inefficiencies based on our estimated structural model.

Our main findings are the following. The model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have moderate returns and adjust quickly (*i.e.*, they are easily accumulated on the job, and relatively easily lost when left unused). Cognitive skills have much higher returns, but are much slower to adjust. Interpersonal skills have moderate returns, and are very slow to adjust over a worker's lifetime. Next, the cost of skill mismatch (modeled as the combination of an output loss and a loss of worker utility caused by skill mismatch) is very high for cognitive skills, an order of magnitude greater than for manual or interpersonal skills. Moreover, this cost is asymmetric: employing a worker who is under-qualified in cognitive skills (*i.e.* has a level of skills that falls short of the job's skill requirements) is several orders of magnitude more costly than employing an over-qualified worker. Those important differences between various skill dimensions are missed when subsuming worker productive heterogeneity into one single scalar index.

References

Acemoglu, D., and D. Autor (2011): "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," vol. 4, Part B of *Handbook of Labor Economics*, pp. 1043 – 1171. Elsevier.

Autor, D. H., and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103(5), 1553–97.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), pp. 1279–1333.

Ben-Porath, Y. (1967): "The Production of Human Capital and Life Cycle of Earnings," *Journal of Political Economy*, 75, 352–65.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

Chiappori, P.-A., and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54(3), 832–861.

Eeckhout, J., and P. Kircher (2011): "Identifying Sorting — In Theory," *The Review of Economic Studies*, 78, 872–906.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–S109.

Golan, L., and K. Antonovics (2012): "Experimentation and Job Choice," *Journal of Labor Economics*.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2016): "Multidimensional Skill Mismatch," Manuscript, University of Minnesota.

Heckman, J., J. Stixrud, and S. Urzua (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 24, 411–82.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591.

Heckman, J. J., and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market," *Journal of Political Economy*, 93(6), 1077–1125.

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

Kahn, L. (2010): "The long-term labor market consequences of graduating from college in a bad economy," *Labor Economics*, 17, 303–316.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

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

Lee, D., and K. I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74(1), 1–46.

Lindenlaub, I. (2014): "Sorting Multi-dimensional Types: Theory and Application," Manuscript, European University Institute.

Lindenlaub, I., and F. Postel-Vinay (2016): "Multi-dimensional Sorting Under Random Search," Manuscript, University College London.

Lise, J., C. Meghir, and J.-M. Robin (2016): "Matching, Sorting, and Wages," *Review of Economic Dynamics*, 19(1), 63–87.

Lise, J., and J.-M. Robin (2016): "The Macro-dynamics of Sorting between Workers and Firms," Forthcoming, *American Economic Review*.

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

Neal, D. (1999): "The Complexity of Job Mobility among Young Men," *Journal of Labor Economics*.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4, 1–29.

Pavan, R. (2011): "Career Choice and Wage Growth," *Journal of Labor Economics*.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55.

Sanders, C. (2012): "Skill Uncertainty, Skill Accumulation, and Occupational Choice," Manuscript, Washington University in St. Louis.

Sanders, C., and C. Taber (2012): "Life-Cycle Wage Growth and Heterogeneous Human Capital," *Annual Review of Economics*, 4, 399–425.

Taber, C., and R. Vejlin (2016): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," unpublished manuscript.

Tinbergen, J. (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77, 155–175.

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

A Web Appendix For Online Publication

A.1 Solving for the value functions.

The value functions can be solved for in quasi-closed form. We first focus on the match value $P(\mathbf{x},\mathbf{y})$, taking the value of unemployment $U(\mathbf{x})$ as given. To solve for $P(\mathbf{x},\mathbf{y})$, it is convenient to parameterize P and $\mathbf{x}$ as a function of the worker's tenure, say t, in the job under consideration. The solution to the first-order linear PDE (2) is then characterized by the following system of $K+1$ ODEs:

$$\frac{dx_k}{dt} = g_k\left(\mathbf{x}(t),\mathbf{y}\right) \qquad k = 1,\cdots,K \tag{13}$$

$$\frac{dz}{dt} = (r+\mu+\delta)z - \left[f\left(\mathbf{x}(t),\mathbf{y}\right) - c\left(\mathbf{x}(t),\mathbf{y}\right)\right] - \delta U\left(\mathbf{x}(t)\right) \tag{14}$$

which are indeed the characteristic equations of (2). Match value is then the solution to $P\left(\mathbf{x}(t),\mathbf{y}\right) = z(t)$. Initial conditions for the first K equations (13) are given by the worker's skill vector $\mathbf{x}(0)$ at the point of hire. The last initial condition, $z(0)$, is unknown, but we can impose the boundary condition $z(t)\exp\left[-\left(r+\mu+\delta\right)t\right] \to 0$ as $t \to +\infty$ to pin down a unique solution to (14).

To be more explicit, let us denote by $\mathbf{X}\left(t;\mathbf{y},\mathbf{x}_0\right)$ the solution to (13) given initial condition $\mathbf{x}_0$ and job type $\mathbf{y}$ (possibly equal to $\mathbf{0}_L$ if the worker is unemployed). The date-t value of a match between a job with attributes $\mathbf{y}$ and a worker with current skill bundle $\mathbf{x}(t)$ is then given by the solution to (14):

$$\begin{aligned} &P\left(\mathbf{x}(t),\mathbf{y}\right) \\ &= \int_t^{+\infty} \left[f\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right),\mathbf{y}\right) - c\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right),\mathbf{y}\right) + \delta U\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right)\right)\right] e^{-(r+\mu+\delta)(s-t)} ds. \end{aligned}$$

The value of unemployment $U\left(\mathbf{x}(t)\right) = \int_t^{+\infty} b\left(\mathbf{X}\left(s;\mathbf{0},\mathbf{x}(t)\right)\right) e^{-(r+\mu)(s-t)} ds$ is solved for in a similar fashion, and the surplus associated with a typical match is obtained by subtraction:

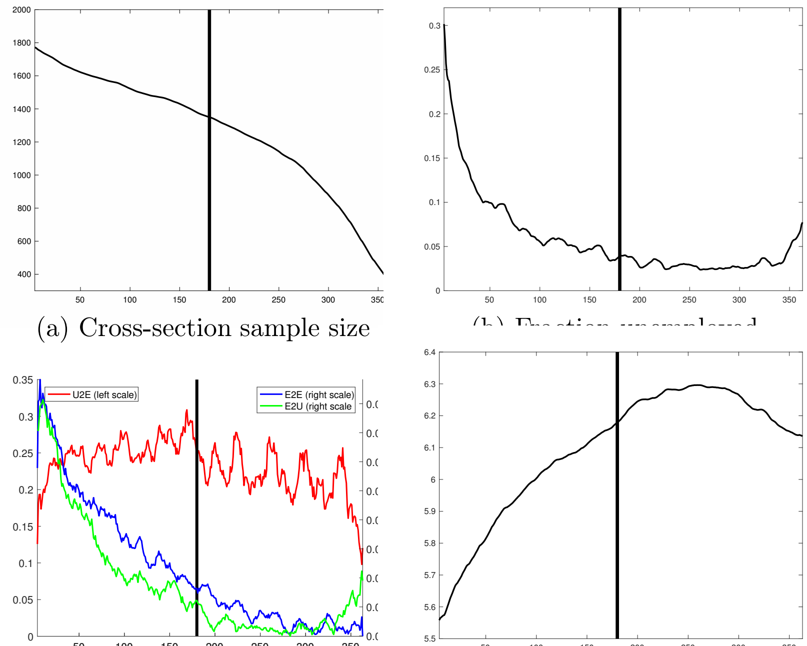
$$\begin{aligned} &P\left(\mathbf{x}(t),\mathbf{y}\right) - U\left(\mathbf{x}(t)\right) = \\ &= \int_t^{+\infty} \left[f\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right),\mathbf{y}\right) - c\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right),\mathbf{y}\right) - b\left(\mathbf{X}\left(s;\mathbf{y},\mathbf{x}(t)\right)\right)\right] e^{-(r+\mu+\delta)(s-t)} ds. \end{aligned} \tag{15}$$

A.2 Alternative Specification for Over-qualification

A possible alternative to our assumption that over-qualification entails a utility cost that would also allow for over-qualification to be costly in terms of match surplus would be to assume that over-qualification causes a loss of output (and does not cause any disutility of work). Formally, this would mean specifying the production function as $f^{\text{alt}}(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})$ and the worker's flow utility function as simply equal to the wage w. This alternative specification would yield exactly the same match values (and therefore the same worker-job allocation pattern) as our utility cost of being under-matched version of the model. Where the two models would differ, though, would be in terms of predicted wages. Under our utility-cost assumption, an over qualified worker produces the same output as an ideally-qualified worker, but suffers an extra cost of working in that match. The over-qualified worker will therefore receive a higher wage in compensation for that cost. Under the alternative production-cost assumption, the over-qualified worker just produces less output than the ideally-suited worker, and will therefore earn a lower wage, even though s/he has more skills.

A.3 Parameter Summary

Offer arrival rates:	(λ_0, λ_1)	
Job destruction rate:	δ	*Estimated as sample mean E2U rate*
Unemployment income:	b	
Production function f:	$(\alpha_T, \alpha_C, \alpha_M, \alpha_I$ $\kappa_C^u, \kappa_M^u, \kappa_I^u)$	
Utility cost of over-qualification c:	$(\kappa_C^o, \kappa_M^o, \kappa_I^o)$	
Skill accumulation function $\mathbf{g}$:	$(\gamma_C^u, \gamma_M^u, \gamma_I^u$ $\gamma_C^o, \gamma_M^o, \gamma_I^o, g)$	
Joint distribution of initial worker skills:		*Observed from initial sample cross section*
General worker efficiency x_T:	$(\zeta_S, \zeta_C, \zeta_M, \zeta_I, \varepsilon_0)$	*Distribution of ε_0 estimated by deconvolution in final step*
Sampling distribution of job attributes Υ:	$(\xi_C, \xi_M, \xi_I, \rho_{CM},$ $\rho_{CI}, \rho_{MI}, \beta_C, \beta_M, \beta_I)$	*Gaussian copula with Pareto marginals*
Attrition and discount rates:	(r, μ)	*Calibrated*



(a) Cross-section sample size

(b) Fraction unemployed

(c) Transition rates

(d) Mean log wages

Figure 6: Sample description

A.4 Data Details

Our final estimation sample consists of an initial cross-section of 1,840 males whom we follow over up to 30 years. There is, however, a substantial amount of attrition, which we comment on in the next paragraph.

Figure 6 describes our sample in terms of a set of times series about worker stocks, labor market transition rates, and average wages over the full 30-year sample window. The horizontal-axis variable is time, measured in months since labor market entry.

Figure 6a shows the pattern of attrition from our sample. Attrition is initially very gradual, with the sample cross-section size declining by about 30 percent over the initial twenty years. Past that point, attrition accelerates considerably. This is partly a consequence of the fact that we follow a cohort of individuals from the date they leave full-time education, resetting time to zero on the week they enter the labor market. Individuals having spent more time at school enter the labor market later, and are therefore observed for fewer years than less educated individuals. This causes the composition of the sample to shift toward less educated individuals as one approaches the end of the observation window. To circumvent this problem, we restrict our estimation sample to the first 15 years (180 months) of the initial sample. This 180-month cutoff is materialized by a thick vertical black line on all panels of Figure 6.

Figure 6b shows the nonemployment rate among sample members. As one would

Table 3: Examples of skill requirement scores

Occupation title	Skill requirements: Cognitive	Manual	Interpersonal
Physicists	1	0.755	0.692
Graders and Sorters, Agricultural Products	0	0.138	0.058
Aircraft Mechanics and Service Technicians	0.613	1	0.318
Telemarketers	0.147	0	0.330
Preventive Medicine Physicians	0.658	0.410	1
Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic	0.302	0.641	0

Source: O*NET and authors' calculations

expect, this rate declines monotonically over time, until it reaches a steady level slightly under 5 percent. It rises again slightly after about 20/25 years, likely as a result of the compositional shift discussed above. Perhaps slightly more surprising is the long time it takes for the nonemployment rate to reach this steady state (roughly ten years). Figure 6c shows the rates of transition between labor market states. The nonemployment exit rate is roughly stable at around 25 percent per month, while the transition rates from job to job and into nonemployment decline smoothly over the sample window. Finally, Figure 6d plots average log wages among employed sample members which, again as one would expect, increase monotonically over time until they reach a point where, mirroring the nonemployment rate, they start declining, again a likely consequence of non-random attrition from the sample.

Table 3 lists some examples of the cognitive, manual and interpersonal skill requirement scores we constructed for a few occupations. We denote those scores by $\widetilde{\mathbf{y}} = (\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ and will use them as empirical measures of the model's job attributes $\mathbf{y}$. Examples in Table 3 include the occupations with the highest cognitive (Physicist), manual (Aircraft Mechanics and Service Technicians), and interpersonal (Preventive Medicine Physicians) skill requirements in the sample, and the occupations with the lowest cognitive (Graders and Sorters, Agricultural Products), manual (Telemarketers), and interpersonal (Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic) skill requirements.

The correlation pattern of workers' *initial* skills and the skill requirements of the *first jobs they are observed in* is described in Table 4, where workers' initial cognitive, manual, and interpersonal skill indices are denoted by (x_{C0}, x_{M0}, x_{I0}) , while $(\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ refer to the empirical measures of job skill requirements in a worker's first job. This correlation pattern reveals several features of the data. First, (x_{C0}, x_{M0}, x_{I0}) are positively correlated in our cross-section of workers. Even though those correlation

Table 4: Correlation pattern of initial skills and skill requirements in first job

	x_{C0}	x_{M0}	x_{I0}	$\widetilde{y}_C$	$\widetilde{y}_M$	$\widetilde{y}_I$
x_{C0}	1					
x_{M0}	0.46	1				
x_{I0}	0.39	0.31	1			
$\widetilde{y}_C$	0.48	0.15	0.23	1		
$\widetilde{y}_M$	−0.18	0.18	0.01	0.32	1	
$\widetilde{y}_I$	0.55	−0.01	0.27	0.64	−0.34	1

coefficients are far below one, they suggest that workers with high skills in one dimension tend to have high skills in the other two. Cognitive and manual skills appear slightly more strongly associated with each other than either is with interpersonal skills. Second, $\widetilde{y}_C$ is positively correlated with both $\widetilde{y}_M$ and $\widetilde{y}_I$ in the cross section of workers' first jobs. Even though there is obviously some selection here (as the set of jobs a worker will take up depends on their own skill bundle $\mathbf{x}$), this suggests that jobs requiring high levels of cognitive skills also tend to require high skill levels in one of the manual or interpersonal dimensions. While manual and interpersonal skill requirements are both positively correlated with cognitive skill requirements, they are negatively correlated with each other. Third, $(x_{C0}, \widetilde{y}_C)$, $(x_{M0}, \widetilde{y}_M)$, and $(x_{I0}, \widetilde{y}_I)$ are positively correlated (as expected), and so are $(x_{C0}, \widetilde{y}_I)$, $(x_{M0}, \widetilde{y}_C)$, $(x_{I0}, \widetilde{y}_M)$, $(x_{I0}, \widetilde{y}_C)$. By contrast, $(x_{C0}, \widetilde{y}_M)$ and $(x_{M0}, \widetilde{y}_I)$ are negatively correlated, suggesting that workers select themselves into either manual or non-manual jobs, as fits their skill bundles.

Table 5 shows results from an OLS regression of log weekly earnings on worker initial skill and job skill requirement indices, with additional controls for experience, tenure, and schooling. All skill requirements are positively associated with wages in the cross-section, the effect of cognitive skills being 3.25 times larger than that of manual skills and 1.66 times larger than interpersonal skills. Workers' initial levels of cognitive and interpersonal skills are positively correlated with wages (interpersonal skills less so), while initial manual skill levels come out with a negative sign. This latter result may be caused by an extra, unobserved worker productive attribute (x_T in the model) which is negatively associated with initial manual skills.

A.5 Construction of skill measures

The data sets from which we construct worker skill and job skill requirement scores both consist of a set of P different measures observed for N individuals (workers in

Table 5: Descriptive earnings regression

	coeff.	std. err.		coeff.	std. err.
$\widetilde{y}_C$	0.664	(.010)	x_{I0}	0.270	(.006)
$\widetilde{y}_M$	0.246	(.008)	years of schooling	0.026	(.001)
$\widetilde{y}_I$	0.378	(.009)	experience	2.24e − 3	(.000)
x_{C0}	0.363	(.014)	tenure	1.98e − 3	(.000)
x_{M0}	−0.101	(.011)	constant	4.455	(.014)
$\mathbf{R^2}$		0.371			
Observations		232,303			

the case of the NLSY, and occupations in the case of O*NET). We denote the $N \times P$ matrix of all observations by $\mathbf{M}$. PCA decomposes the matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{FL}$, where $\mathbf{F}$ is the orthonormal $N \times P$ matrix of principal eigenvectors of $\mathbf{M}^\top \mathbf{M}$ and $\mathbf{L}$ is a $P \times P$ matrix of factor loadings. We consider the first 3 principal components only, *i.e.* we consider the decomposition $\mathbf{M} = \mathbf{F}_3\mathbf{L}_3 + \mathbf{U}$, where $\mathbf{F}_3$ is the $N \times 3$ matrix formed by taking the first 3 columns of $\mathbf{F}$ and $\mathbf{L}_3$ is the $3 \times P$ matrix formed by taking the first 3 rows of $\mathbf{L}$.

For any invertible 3×3 matrix $\mathbf{T}$, the above decomposition of $\mathbf{M}$ can be rewritten as $\mathbf{M} = (\mathbf{F}_3\mathbf{T})(\mathbf{T}^{-1}\mathbf{L}_3) + \mathbf{U}$, which is an alternative decomposition of $\mathbf{M}$ into new (linearly recombined) factors $\mathbf{F}_3\mathbf{T}$ with loadings $\mathbf{T}^{-1}\mathbf{L}_3$. We choose $\mathbf{T}$ such that our decomposition of $\mathbf{M}$ satisfies our chosen exclusion restrictions. Taking the case of O*NET as an example, we order the measures such that measure 1 (the first column of $\mathbf{M}$) is the score on *mathematics knowledge*, measure 2 is the score on *mechanical knowledge*, and measure 3 is the score on *social perceptiveness*, then define $\mathbf{T} = \mathbf{L}_{3,3}$ where $\mathbf{L}_{3,3}$ is the 3×3 matrix made up of the first three columns of $\mathbf{L}_3$.

It should be emphasized that our method of constructing worker skill and job skill requirement scores differs slightly from the approach usually taken in the related literature. The conventional approach consists of assigning each of the P data measures (of skills or skill requirements, as the case may be) to one of K different bins, where K is the number of skill dimensions relevant to the model (three, in our case), and set the score in skill dimension k as the average of all measures in bin k.

The conventional method therefore assumes that any given measure is only relevant to one single skill dimension. Which skill dimension a measure is relevant to must be decided *a priori*. In our case, this would mean deciding for every NLSY or O*NET descriptor whether it relates to cognitive, manual, or interpersonal skills. While this decision may seem relatively straightforward, at least on an intuitive level, for some measures (for instance the six measures on which we impose exclusion restrictions), it is far from clear-cut for most measures, which can easily be argued to be relevant for two or more skill dimensions. We therefore choose to minimize the number of exclusion restrictions we impose on the data. We believe that our approach offers a good compromise between interpretability, parsimony, and ability to capture the covariance patterns between skills and skill requirement dimensions.

A.6 Identification

This appendix contains a formal discussion of identification. Identification is, in large part, parametric, in that many of the arguments below make use of the specific functional forms assumed in the main text.

The job loss rate δ is directly observed in the data. We assume that so is the population distribution of initial skill bundles. Moreover, we discuss identification conditional on knowledge of the discount rate r and the sample attrition rate μ.

The wage equation (5) can be written as:

$$\begin{aligned} w(\mathbf{x},\mathbf{y},\sigma) = \sigma f(\mathbf{x},\mathbf{y}) + (1-\sigma)b(\mathbf{x}) + (1-\sigma)c(\mathbf{x},\mathbf{y}) \\ - \lambda_1(1-\sigma)\left[P(\mathbf{x},\mathbf{y})-U\right]\int_{\mathcal{Y}}\mathbf{1}\left\{P(\mathbf{x},\mathbf{y}')\geq P(\mathbf{x},\mathbf{y})\right\}d\Upsilon(\mathbf{y}') \\ - \lambda_1\int_{\mathcal{Y}}\left[\mathbf{1}\left\{P(\mathbf{x},\mathbf{y}')\geq \sigma P(\mathbf{x},\mathbf{y})+(1-\sigma)U\right\}-\mathbf{1}\left\{P(\mathbf{x},\mathbf{y}')\geq P(\mathbf{x},\mathbf{y})\right\}\right]\times \\ \left[P(\mathbf{x},\mathbf{y}')-\sigma P(\mathbf{x},\mathbf{y})-(1-\sigma)U\right]d\Upsilon(\mathbf{y}'). \end{aligned} \quad (16)$$

A first important implication of (16) is that the maximum wage given $(\mathbf{x},\mathbf{y})$ is $f(\mathbf{x},\mathbf{y})$, implying in turn that the maximum wage given $\mathbf{y}$ is $f(\mathbf{y},\mathbf{y}) = x_T\varphi(\mathbf{y})$. Because $\mathbf{y}$ is observed for all employed workers, the function $\varphi(\cdot)$ is (non-parametrically) identified, up to x_T.[36] But x_T is itself a function of observables (up to the uncorrelated heterogeneity term ε_0), namely the worker's education, initial skill bundle and experience, which is therefore also identified. This proves identification of the parameters α_T, α_C, α_M, α_I, g, ζ_S, ζ_C, ζ_M, ζ_I.

Next, consider the set of workers with initial skill bundle $\mathbf{x}$ exiting nonemployment at any experience level. The (observed) set of job types $\mathbf{y}$ that those workers accept is the set $\{\mathbf{y} : P(\mathbf{x},\mathbf{y}) \geq U\}$, and its boundary is the set $\{\mathbf{y} : P(\mathbf{x},\mathbf{y}) = U\}$. This latter set is therefore identified, conditional on knowledge of $\mathbf{x}$. We now show that this latter fact allows identification of the parameters of the match value function $P(\mathbf{x},\mathbf{y}) = U$.

First, from the expression of the match surplus (10), one can show that joint observation of $\mathbf{x}$ and the set $\{\mathbf{y} : P(\mathbf{x},\mathbf{y}) = U\}$ allows separate identification of the parameters of $P(\mathbf{x},\mathbf{y})$, *i.e.* the composite parameters $\kappa_k^{u/o}/\left(r+\delta+\mu-g+2\gamma_k^{u/o}\right)$, $k =$

[36]What is, in fact, observed, is not directly $\mathbf{y}$ but rather its empirical counterpart $\widetilde{\mathbf{y}}$. With our functional form assumptions $\varphi(\mathbf{y}) = \alpha_T + \alpha_C y_C + \alpha_M y_M + \alpha_I y_I$ and $y_k = \widetilde{y}_k^{\xi_k}$, $k = C, M, I$, the maximum wage given $\mathbf{y}$ jointly identifies the α's and the ξ's.

C, M, I.[37] Now, the issue is that we do not directly observe worker skills at all levels of experience: rather, we only observe workers' *initial* skill bundles. However, consider a worker with (observed) initial skill bundle $\mathbf{x}(0)$ starting his working life in unemployment, and who finds a job after an initial unemployment spell of duration $d^{(1)}$. From the human capital accumulation function (6), we know that this worker's skill bundle by the time s/he finds a job is $\mathbf{x}\left(d^{(1)}\right) = \left(x_C(0)e^{-\gamma_C^o d^{(1)}}, x_M(0)e^{-\gamma_M^o d^{(1)}}, x_I(0)e^{-\gamma_I^o d^{(1)}}\right)$. Identification of the parameters of $P(\mathbf{x}, \mathbf{y})$, $\kappa_k^{u/o}/\left(r + \delta + \mu - g + 2\gamma_k^{u/o}\right)$, is thus obtained from the set of initially unemployed workers whose initial unemployment spell duration $d^{(1)} \to 0$. Furthermore, once the parameters of $P(\mathbf{x}, \mathbf{y})$ are known, observation of the set $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) = U\}$ given $\mathbf{x} = \left(x_C(0)e^{-\gamma_C^o d^{(1)}}, x_M(0)e^{-\gamma_M^o d^{(1)}}, x_I(0)e^{-\gamma_I^o d^{(1)}}\right)$ with $\mathbf{x}(0)$ observed identifies γ_k^o for $k = C, M, I$. Combining those results, we now have separate identification of γ_k^o, κ_k^o, and $\kappa_k^u/\left(r + \delta + \mu - g + 2\gamma_k^u\right)$, and still need to separate κ_k^u from γ_k^u in the latter composite parameter. This can be done by repeating the latter argument for workers who are initially employed in matches with skill requirements $\mathbf{y}^{(1)}$ for which they are *under*-qualified, *i.e.* such that $x_k(0) < y_k^{(1)}$ for $k = C, M, I$, become unemployed after an initial spell duration of $d^{(1)}$, then find a job again after an unemployment spell of duration $d^{(2)}$. From the human capital accumulation function (6), those workers' skill bundles when they find their second job (at experience $d^{(1)}+d^{(2)}$ is given by $x_k\left(d^{(1)} + d^{(2)}\right) = e^{-\gamma_k^o d^{(2)}}\left[y_k^{(1)} - e^{-\gamma_k^u d^{(1)}}\left(y_k^{(1)} - x_k(0)\right)\right]$. The only unknown parameter in this expression is γ_k^u, which is then again identified from the set $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) = U\}$.

The full set of production, utility, and human capital accumulation parameters is thus identified. Note that, while the arguments laid out above rely on the specific functional forms assumed in the main text, the background source of identification for the cost of mismatch and the speed of human capital accumulation (or decay) is a comparison of the set of job types $\mathbf{y}$ that are acceptable to workers with equal initial skills $\mathbf{x}(0)$, but have experienced different employment histories.

Once the parameters of the human capital accumulation function $\mathbf{g}(\mathbf{x}, \mathbf{y})$ are known, we can construct any worker's full path of skill bundles $\mathbf{x}$: consider a worker in his n^{th} spell (which could be a spell of unemployment). Denote the skill requirements in that spell by $\mathbf{y}^{(n)} = \left(y_C^{(n)}, y_M^{(n)}, y_I^{(n)}\right)$ (both equal to 0 if the spell is one of unemployment), the worker's skill bundle at the beginning of that spell by $\mathbf{x}^{(n)} = \left(x_C^{(n)}, x_M^{(n)}, x_I^{(n)}\right)$, and the duration of that spell by $d^{(n)}$. Spell duration $d^{(n)}$ and the vector $\mathbf{y}^{(n)}$ are observed in all spells, while $\mathbf{x}^{(n)}$ is only observed in the initial spell, $n = 1$, where it equals $\mathbf{x}(0)$. Then, using the skill accumulation equation

[37] One way to see this is to realize from (10) that the set $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) = U\}$ is the union of four quarter-ellipses, the centers and axes of which can be expressed as simple functions of $\mathbf{x}$ and the parameter combinations $\kappa_k^{u/o}/\left(r + \delta + \mu - g + 2\gamma_k^{u/o}\right)$. Observation of $\mathbf{x}$ and $\mathbf{y}$ for this set identifies these centers and axes.

$\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{y})$, we have that $\mathbf{x}^{(n+1)} = \mathbf{X}\left(d^{(n)}; \mathbf{y}^{(n+1)}, \mathbf{x}^{(n)}\right)$, where $\mathbf{X}(\cdot)$ denotes the solution to (13) as explained in the main text. Using backward substitution, we can then construct $\mathbf{x}^{(n)}$ for any spell as a function of the history of durations and skill requirements of past spells and the worker's initial skill level $\mathbf{x}^{(1)} = \mathbf{x}(0)$.

Next, the set of job offers accepted by unemployed workers with skills $\mathbf{x}$ identifies the sampling distribution $\Upsilon(\mathbf{y})$ over the set $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) \geq U\}$. $\Upsilon(\mathbf{y})$ is thus (non-parametrically, conditionally on the rest of the model) identified over the union of all such sets for all skill bundles $\mathbf{x}$ observed in the sample. That is, $\Upsilon(\mathbf{y})$ is identified at all skill requirement levels $\mathbf{y}$ that are acceptable by at least some worker types.

Finally, the offer arrival rates λ_0 and λ_1 are identified, conditionally on the rest of the model, from sample U2E and E2E transition probabilities, and the flow value of nonemployment, $b(\mathbf{x})$, is identified from the wage of workers exiting nonemployment: applying (16) to workers just exiting nonemployment ($\sigma = 0$) yields:

$$w(\mathbf{x}, \mathbf{y}, 0) = b(\mathbf{x}) + c(\mathbf{x}, \mathbf{y}) + \lambda_1 \int_{\mathcal{Y}} \mathbf{1}\left\{P(\mathbf{x}, \mathbf{y}') \geq P(\mathbf{x}, \mathbf{y})\right\} \left[P(\mathbf{x}, \mathbf{y}') - P(\mathbf{x}, \mathbf{y})\right] d\Upsilon(\mathbf{y}') - \lambda_1 \int_{\mathcal{Y}} \mathbf{1}\left\{P(\mathbf{x}, \mathbf{y}') \geq U\right\} \left[P(\mathbf{x}, \mathbf{y}') - U\right] d\Upsilon(\mathbf{y}'),$$

which equals $b(\mathbf{x}) + c(\mathbf{x}, \mathbf{y})$ on the (known) set of $\mathbf{y}$'s such that $P(\mathbf{x}, \mathbf{y}) = U$.

A.7 Unobserved Heterogeneity and the Fit to theWage Distribution

The black line in Figure 7 shows a kernel density estimate of log wages in the final period of the simulation *still assuming away unobserved worker heterogeneity*, *i.e.* that $\varepsilon_0 = 0$ for all workers. The underlying histogram shows the corresponding empirical distribution. The simulated distribution is slightly more concentrated than the empirical one, and has a long left tail and a short right tail. Those are again standard predictions of the sequential auction model with linear preferences and no unobserved worker heterogeneity: the long left tail reflects the low wages accepted by workers hired out of unemployment. Adding in permanent worker heterogeneity (heterogeneity in ε_0) results in the unconditional wage distribution being a mixture of distributions like the one represented by the black line on Figure 7. To illustrate this, the red line on 7 shows a kernel density estimate of the sum of simulated wages and an uncorrelated normal random variable (which can be interpreted as the sum of measurement error and permanent worker heterogeneity ε_0) with a variance set to match the empirical wage variance. Even this very simple form of unobserved heterogeneity improves the fit to the tails of the wage distribution.

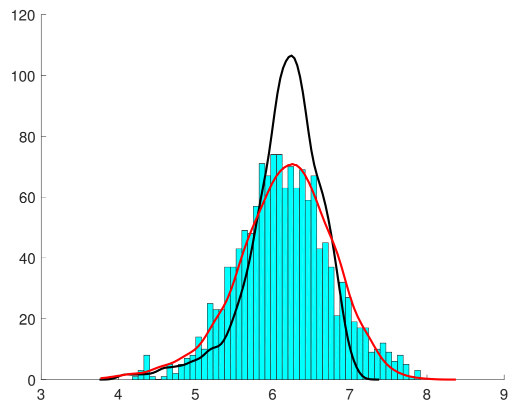


Figure 7: Log wage density

A.8 The friction-free Planner's problem

We first define the following convenient notation: $s(\mathbf{x}, \mathbf{y}) = [f(\mathbf{x}, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})] / x_T$. Absent any frictions, the worker's problem can be expressed as follows:

$$\max_{\{\mathbf{y}_t, \ell_t\}} \int_0^{+\infty} [\ell_t s(\mathbf{x}_t, \mathbf{y}_t) + (1 - \ell_t) b] \, e^{-(r+\mu-g)t} dt$$
$$\text{subject to } \dot{\mathbf{x}}_t = \ell_t \mathbf{g}(\mathbf{x}_t, \mathbf{y}_t) + (1 - \ell_t)\mathbf{g}(\mathbf{x}_t, 0), \quad \ell_t \in [0, 1], \quad \mathbf{x}_0 \text{ given}$$

where ℓ_t stands for within-period labor supply of the worker under consideration. Denoting this problem's Hamiltonian by $\mathcal{H}$ and the costate vector by $\boldsymbol{\xi}$, the first-order and Euler conditions are, for $k = C, M, I$:

$$\frac{\partial \mathcal{H}}{\partial \ell_t} = s(\mathbf{x}_t, \mathbf{y}_t) - b + \boldsymbol{\xi} \cdot [\mathbf{g}(\mathbf{x}_t, \mathbf{y}_t) - \mathbf{g}(\mathbf{x}_t, 0)]$$
$$\frac{\partial \mathcal{H}}{\partial y_{kt}} = \ell_t \left\{ \frac{\partial s(\mathbf{x}_t, \mathbf{y}_t)}{\partial y_{kt}} + \boldsymbol{\xi} \cdot \frac{\partial \mathbf{g}(\mathbf{x}_t, \mathbf{y}_t)}{\partial y_{kt}} \right\}$$
$$\dot{\xi}_{kt} = (r + \mu - g)\xi_{kt} - \frac{\partial s(\mathbf{x}_t, \mathbf{y}_t)}{\partial x_{kt}} - \boldsymbol{\xi} \cdot \left[\ell_t \frac{\partial \mathbf{g}(\mathbf{x}_t, \mathbf{y}_t)}{\partial x_{kt}} + (1 - \ell_t) \frac{\partial \mathbf{g}(\mathbf{x}_t, \mathbf{y}_t)}{\partial x_{kt}} \right],$$

to which the transversality condition $\boldsymbol{\xi}_t e^{-(r+\mu-g)t} \to \mathbf{0}$ as $t \to +\infty$ must be added. Assuming $\ell_t = 1$, and using the functional forms in the main text, the latter two sets of conditions solve as $\xi_{kt} \equiv \alpha_k / (r + \mu - g)$ and:

$$y_{kt} = \min \left\{ x_{kt} + \frac{\alpha_k}{2\kappa_k^u} \left(1 + \frac{\gamma_k^u}{r + \mu - g} \right), 1 \right\},$$

implying that it is always optimal to assign workers to jobs for which they are slightly under qualified in all skill dimensions to benefit from on-the-job learning. One can then check that $\ell_t = 1$ is indeed the optimal choice for all $\mathbf{x}$ and t if the condition

$b \leq \sum_{k=C,M,I} \left(\frac{\alpha_k \left(r+\mu-g+\gamma_k^u \right)}{2(r+\mu-g)\sqrt{\kappa_k^u}} \right)^2$ holds, which is the case at our parameter estimates.