Multidimensional Skills, Sorting, and Human Capital Accumulation[*]

Jeremy Lise

UCL and IFS

Fabien Postel-Vinay

UCL, IFS and Sciences Po

February 2015- IN PROGRESS

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. Our results clearly suggest that those three types of skills are very different productive attributes.

1   Introduction

Productive heterogeneity among workers and jobs is routinely modeled in the theoretical/structural literature using scalar indices. Yet statistical agencies, as well as a growing empirical literature have recognized that workers differ in skills along several dimensions (e.g. manual, cognitive, interpersonal. . .) and that jobs (occupations) require different mixes of various types of skills. 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.

In this paper, we generalize the sequential auction model of Postel-Vinay and Robin (2002) to allow for multi-dimensional 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 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 [this part is TBC].

Our main findings are the following. The model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have low returns and adjust quickly (i.e., they are easily accumulated on the job, and easily lost when left unused). Cognitive skills have much higher returns, but are much slower to adjust, both up and down. Interpersonal skills have very modest 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, substantial for manual skills, and negligible for interpersonal skills. Moreover, this cost is asymmetric: employing a worker who is under-qualified in either cognitive or manual 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. Finally, Section 7 concludes.

2   Related Literature

This paper is obviously related to the vast empirical literature on the returns to firm and occupation tenure and to its more recent successor focusing on task-specific human capital. Those literatures, and the connections between them are covered in the excellent survey paper by Sanders and Taber (2012), to which we refer the reader.[1] As a preamble to their review of the empirical literature, Sanders and Taber (2012) offer an elegant theoretical model of job search

---

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

and investment in multi-dimensional skills which, on many aspects, can be seen as a special case of the model in this paper (we briefly return to this comparison below).[2] 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, Lindenlaub (2014) estimates a model of frictionless assignment along two dimensions of skills (manual and cognitive), using the same combination of O*NET and NLSY data as we do in this paper. Specifically, she estimates her model on two different cross-sections of worker data, one from the NLSY79 and the other from the NLSY97, and 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 over the same period. She then analyzes the consequences of that technological shift for sorting and wage inequality. While Lindenlaub's frictionless assignment model brings about many valuable new insights, the fact that it assumes away market imperfections limits its applicability to empirical and quantitative 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 models are static descriptions of the "long run". As such, they are useful to analyze the long-run consequences of various forms of skill- or task-biased technological change (as Lindenlaub does), but are largely silent on any question relating to a worker's life cycle, such as the cost of skill mismatch at various points of a worker's career, or the way in which individual skills evolve over a career, or the reasons why workers switch occupations as often as they do in the data.

Finally, 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 (to our knowledge) 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 Occupa-

---

[2]The main conceptual difference between out model and the one in Sanders and Taber (2012) is that they 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.

tional Titles (DOT, the predecessor of O*NET - see Section 4 for details). 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, and a decomposition of wage growth shows that cognitive 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 closely related in spirit, our model differs from Yamaguchi (2012) in several important ways. First, Yamaguchi (2012) is a frictionless (Roy) 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.[3] We propose an arguably more parsimonious random search model, in which the only unobserved shocks are layoff shocks and the receipt of job offers by workers (both modeled as constant-intensity Poisson processes). Wages and mobility decisions are then jointly and endogenously determined through explicit between-employer competition for labor services. Our less flexible, but more transparent and more 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,[4] but explicitly models job-shopping as an additional source of wage dynamics (Postel-Vinay and Robin, 2002). Finally, like Lindenlaub's (2014) paper, Yamaguchi's frictionless model is silent on issues relating to unemployment or skill mismatch. Adding search frictions, as we do in this paper, allows us to make progress in this direction.[5]

---

[3]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.

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

[5]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.

3 Job Search with Multi-dimensional Job and Worker Attributes

3.1 The Model

The Environment. 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 $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^K$ describes the firm's technology. 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 $\mathbf{g} : \mathbb{R}^K \to \mathbb{R}^K$ is a continuous function with $\mathbf{g}(\mathbf{y}, \mathbf{y}) = \mathbf{0}_{K \times 1}$.

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 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:

$$v(w, \mathbf{x}, \mathbf{y}) = w - c(\mathbf{x}, \mathbf{y}),$$

where $c(\mathbf{x}, \mathbf{y})$ is disutility from work, which depends on the type of the match, $(\mathbf{x}, \mathbf{y})$. We normalize this disutility term assuming $c(\mathbf{y}, \mathbf{y}) = 0$. 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 $P(\mathbf{x}, \mathbf{y}) \geq P(\mathbf{x}, \mathbf{y}') > W$.[6] 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 $W' = P(\mathbf{x}, \mathbf{y}')$. The worker's renegotiated share of the match surplus is therefore:

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

We assume that this share stays constant until the following renegotiation. 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.[7]

---

[6]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.

[7]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 ($\dot{\mathbf{x}} \equiv 0$).

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:[8]

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

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 her/his initial match, in which case the continuation value for that match is $P(\mathbf{x}, \mathbf{y})$, or s/he accepts the offer, in which case s/he extracts a value of $P(\mathbf{x}, \mathbf{y})$ from the poacher (as a result of Bertrand competition) and leaves her/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 in this economy.

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 $\mathbf{y} = \mathbf{0}_K$ 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 her/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 s/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:

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

where the expectation is taken over the sampling distribution Υ. Substitution of (3) and (4)

---

[8]The dot ("·") denotes the outer product and ∇ denotes the gradient.

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

$$w(\mathbf{x}, \mathbf{y}, \sigma) = \sigma f(\mathbf{x}, \mathbf{y}) + (1 - \sigma)b(\mathbf{x}) + (1 - \sigma)c(\mathbf{x}, \mathbf{y})$$

$$- \lambda_1 \mathbf{E} \max \left\{ 0, \min \left\{ P(\mathbf{x}, \mathbf{y}') - P(\mathbf{x}, \mathbf{y}), 0 \right\} + (1 - \sigma) \left(P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x}) \right) \right\}$$

$$- (1 - \sigma) \left(\mathbf{g}(\mathbf{x}, \mathbf{y}) - \mathbf{g}(\mathbf{x}, \mathbf{0}) \right) \cdot \nabla U(\mathbf{x}). \quad (5)$$

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 her/his disutility of work $c(\mathbf{x}, \mathbf{y})$. The next (expectation) term is the option value of future outside offers. 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 $\mathbf{0}$ if the worker was unemployed. This, in general, benefits the worker in the event s/he becomes unemployed, and therefore affects the surplus negatively.[9]

3.2 Model Analysis

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:

$$\frac{dx_k}{dt} = g_k \left(\mathbf{x}(t), \mathbf{y} \right) \qquad k = 1, \cdots, K \quad (6)$$

$$\frac{dz}{dt} = (r + \mu + \delta)z - [f \left(\mathbf{x}(t), \mathbf{y} \right) - c \left(\mathbf{x}(t), \mathbf{y} \right)] - \delta U \left(\mathbf{x}(t) \right) \quad (7)$$

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 (6) 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

---

[9]As mentioned in Section 2, the model in Sanders and Taber (2012) can 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 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).

impose the boundary condition $z(t) \exp\left[-(r + \mu + \delta) t\right] \to 0$ as $t \to +\infty$ to pin down a unique solution to (7).

To be more explicit, let us denote by $\mathbf{X}(t; \mathbf{y}, \mathbf{x}_0)$ the solution to (6) given initial condition $\mathbf{x}_0$ and job type $\mathbf{y}$ (possibly equal to $\mathbf{0}$ 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 (7):

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

The value of unemployment is solved for in a similar fashion:

$$U\left(\mathbf{x}(t)\right) = \int_t^{+\infty} b\left(\mathbf{X}\left(s; \mathbf{0}, \mathbf{x}(t)\right)\right) e^{-(r+\mu)(s-t)} ds. \tag{8}$$

Combining those last two equations, one obtains the surplus associated with a typical match:

$$P\left(\mathbf{x}(t), \mathbf{y}\right) - U\left(\mathbf{x}(t)\right) =$$
$$= \int_t^{+\infty} \left[f\left(\mathbf{X}\left(s; \mathbf{y}, \mathbf{x}(t)\right), \mathbf{y}\right) - c\left(\mathbf{X}\left(s; \mathbf{y}, \mathbf{x}(t)\right), \mathbf{y}\right) - b\left(\mathbf{X}\left(s; \mathbf{y}, \mathbf{x}(t)\right)\right)\right] e^{-(r+\mu+\delta)(s-t)} ds. \tag{9}$$

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)* a measure of the worker's cognitive skills x_C, *(ii)* a measure of the worker's manual skills x_M, *(ii)* a measure of the worker's interpersonal skills x_I, and *(iv)* a measure of 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.

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

$$
\mathbf{g}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \dot{x}_C \\ \dot{x}_M \\ \dot{x}_I \\ \dot{x}_T \end{pmatrix} = \begin{pmatrix} \gamma_C^u \max\{y_C - x_C, 0\} + \gamma_C^o \min\{y_C - x_C, 0\} \\ \gamma_M^u \max\{y_M - x_M, 0\} + \gamma_M^o \min\{y_M - x_M, 0\} \\ \gamma_I^u \max\{y_I - x_I, 0\} + \gamma_I^o \min\{y_I - x_I, 0\} \\ g x_T \end{pmatrix},
\tag{10}
$$

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 form "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, the solution to (6) is:

$$
x_k(s) = y_k - e^{-\gamma_k^{u/o}(s-t)}\left(y_k - x_k(t)\right),
\tag{11}
$$

where the adjustment speed $\gamma_k^{u/o}$ that applies depends on whether $k = C, M$ or I and whether $x_k(t) \gtrless y_k$. Finally, a worker's general efficiency simply grows with experience 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:

$$
f(\mathbf{x}, \mathbf{y}) = x_T \times \left[\varphi(\mathbf{y}) - \sum_{k=C,M,I} \kappa_k^u \min\{x_k - y_k, 0\}^2\right].
\tag{12}
$$

The function $\varphi(\mathbf{y})$ is assumed increasing in both components of $\mathbf{y}$, implying that jobs with higher requirements in either cognitive or manual skills are inherently more productive, regardless of the worker they are matched with. The remaining terms $-\kappa_k^u \min\{x_k - y_k, 0\}^2$, $k = C, M, I$ 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 whether the worker is under-qualified in any skill dimension. Note that specification (12) 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 gain, nor a loss of output. We will return to this shortly. Finally, we simply specify unemployment income as $b(\mathbf{x}) = bx_T$, with b a positive constant, so that $U(\mathbf{x}) = bx_T/(r + \mu - g)$ is independent of (x_C, x_M, x_I).

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

$$c(\mathbf{x}, \mathbf{y}) = x_T \times \sum_{k=C,M,I} \kappa_k^o \max\{x_k - y_k, 0\}^2. \tag{13}$$

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 (12) 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 (12) 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 (13) of $c(\mathbf{x}, \mathbf{y})$ echoes the cost of mismatch (of under-qualification) in the production function (12), mostly for analytical convenience and simplicity.[10]

---

[10] 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 $v^{\text{alt}}(w, \mathbf{x}, \mathbf{y}) = w$. 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,

With those specifications, equations (8) and (9) become:

$$P\left(\mathbf{x}(t),\mathbf{y}\right)-U(\mathbf{x})=x_T(t)\times\left\{\frac{\varphi\left(\mathbf{y}\right)-b}{r+\delta+\mu-g}\right.$$

$$\left.-\sum_{k=C,M,I}\left(\frac{\kappa_k^u\min\left\{x_k(t)-y_k,0\right\}^2}{r+\delta+\mu-g+2\gamma_k^u}+\frac{\kappa_k^o\max\left\{x_k(t)-y_k,0\right\}^2}{r+\delta+\mu-g+2\gamma_k^o}\right)\right\}.\quad(14)$$

The first term in the equation defining match surplus $P\left(\mathbf{x}(t),\mathbf{y}\right)-U(\mathbf{x})$ 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 interpersonal 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 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.[11]

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 on 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,[12] and retain all individual histories until the first occurrence of a non-employment spell

---

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.

[11]We are not the first authors to combine data sources in this way. 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, Kuruscu, Tanaka, and Wiczer (2014), who all use combinations of the NLSY with occupation data from O*NET or from its predecessor, the *Dictionary of Occupational Titles.*

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

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), and performance in a battery of seven aptitude tests called the *Armed Services Vocational Aptitude Battery* (ASVAB). Education and ASVAB scores will be used as measures of the initial skill bundles $\mathbf{x}$ of those workers (more below).

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.[13] 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. 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 are 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.[14]

Job skill requirements. Our selection from the O*NET database leaves us with over 200 different descriptors. We reduce those to three dimensions, which we interpret as "cognitive", "manual", and "interpersonal" skill requirements, using the following procedure. First, we classify each descriptor (based on its textual definition) as cognitive, manual, interpersonal, or "other", and discard the latter category.[15] This leaves us with two sets of 100 "cognitive", 65 "manual"

---

[13]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`.

[14]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.

[15]This classification is straightforward for some descriptors (e.g. "written expression", "mathematical reasoning", "processing information", "performing general physical activities", "spend time bending or twisting the

and 33 "interpersonal" descriptors, respectively, which we each collapse to one dimension using Principal Component Analysis (PCA) in the pooled sample of employed workers in all periods.[16]

Worker skill bundles. Finally, we need to construct a distribution of initial worker skill bundles, i.e. the distribution $N(\mathbf{x})$ of cognitive and manual skills among labor market entrants. For this follow a similar procedure as for the distribution of skill requirements: we classify a set of skill measures as either cognitive, manual, or interpersonal and collapse each set of measures to one dimension using PCA in the initial cross-section of our NLSY panel. We use the following sets of measures: the seven ASVAB scores that are directly available from the NLSY sample,[17] , individual scores on the Rotter locus-of-control scale and the Rosenberg self-esteem scale tests[18] (those are used as measures of interpersonal skills), three measures of criminal behavior (again used as measures of interpersonal skills), 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.

4.2 Sample Description

Our final estimation sample consists of an initial cross-section of 1,773 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.

---

body","establishing and maintaining interpersonal relationships","selling or influencing others". . .), and slightly less obvious for others, such as "controlling machines and processes" (classified as cognitive) or "repairing and maintaining mechanical equipment" (classified as manual). The "other" category, which we discard, mostly contains items that relate to the general work environment (e.g. "consequence of error", "freedom to make decisions", or "pace determined by speed of equipment").

[16] An alternative to this *a priori* classification of descriptors would be to run PCA on the whole set of descriptors, and retain the two components associated with the largest two eigenvalues (those two components would then be uncorrelated by construction). The main advantage of this alternative method is that it does not rely on *a priori* judgment of which descriptors are measures of cognitive *vs.* manual skills. The flip side of this advantage is also the main drawback of this alternative approach: the two components produced by PCA on the full sample cannot be interpreted individually as cognitive, manual, or indeed any other type of skill. The only thing that can be interpreted is the space spanned by the family of components retained, not each individual component.

[17] Among the ASVAB battery, only one test can clearly be interpreted as a test of manual skills (the "automotive and shop information" test). The other six tests are clearly of a cognitive nature.

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

(a) Cross-section sample size   (b) Fraction unemployed
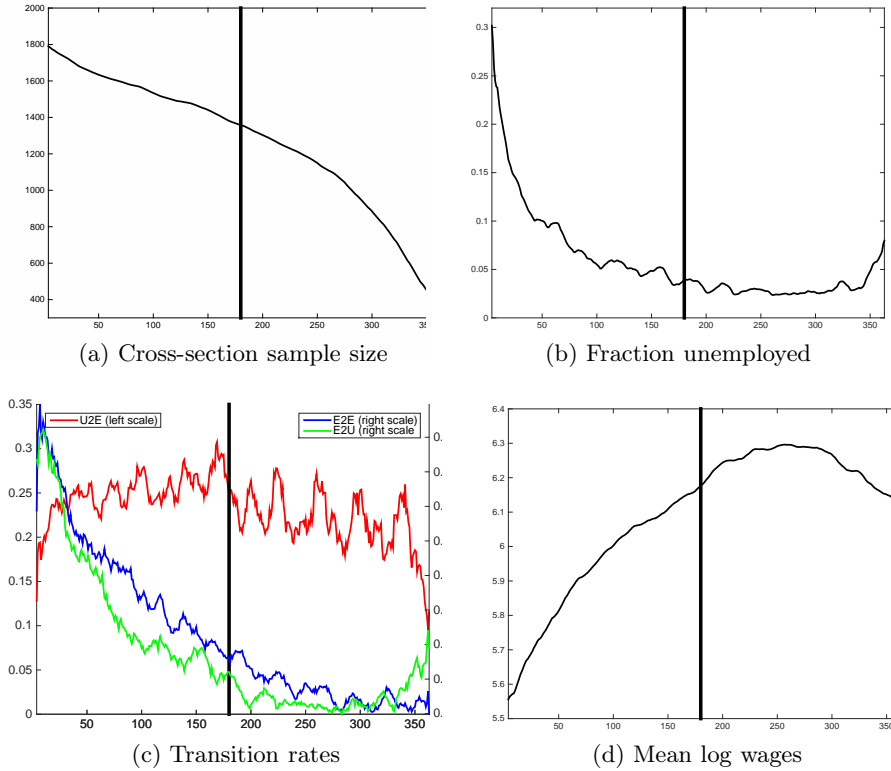
(c) Transition rates   (d) Mean log wages

Figure 1: Sample description

Flows, stocks, and wages over time. Figure 1 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 1a 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 be observed for fewer years than less educated individuals. This causes the composition of the sample to shift toward less educated individuals toward 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 1.

Figure 1b shows the nonemployment rate among sample members. As one would 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 1c 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 1d 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.

Worker skills and job skill requirements. Table 1 lists some examples of the cognitive, manual and interpersonal skill requirement indices we constructed for a few occupations. Because the skill requirement scores produced by our PCA do not have a natural scale, we normalize them to $[0, 1]$ by using the ranks of those scores in the pooled sample of employed workers. We denote those ranks 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 1 include the occupations with the highest cognitive (Physicist), manual (Elevator Installers and Repairers), and interpersonal (Preventive Medicine Physicians) skill requirements in the sample, and the occupations with the lowest cognitive (Graders and Sorters, Agricultural Products), manual (Anthropologists), and interpersonal (Pressers, Textile, Garment, and Related Materials) skill requirements.

Occupation title	Skill requirements:		
	Cognitive	Manual	Interpersonal
Physicists	1	0.207	0.741
Graders and Sorters, Agricultural Products	0	0.343	0.001
Elevator Installers and Repairers	0.497	1	0.485
Anthropologists	0.940	0	0.752
Preventive Medicine Physicians	0.999	0.095	1
Pressers, Textile, Garment, and Related Materials	0.005	0.578	0
Economists	0.891	0.009	0.697

Source: O*NET and authors' calculations

Table 1: Examples of skill requirement scores

The correlation pattern of workers' *initial* skills and the skill requirements of the jobs they are observed in is described in Table 2, where workers' initial cognitive, manual, and interpersonal skill indices are denoted by (x_{C0}, x_{M0}, x_{I0}), while $(\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ denote the empirical measures of a worker's current job skill requirements (those obviously evolve over time as the worker

changes jobs). This covariance pattern reveals three main features of the data. First, (x_{C0}, x_{M0}), (x_{I0}, x_{M0}), $(\widetilde{y}_C, \widetilde{y}_M)$, and $(\widetilde{y}_I, \widetilde{y}_M)$ are negatively correlated among workers, while (x_{C0}, x_{I0}) and $(\widetilde{y}_C, \widetilde{y}_I)$ are positively correlated. This suggests that workers specialize to an extent, in either manual or non-manual (meaning cognitive and interpersonal) skills, and that firms tend to require such specialization. Second, $(x_{C0}, \widetilde{y}_C)$, $(x_{M0}, \widetilde{y}_M)$ and $(x_{I0}, \widetilde{y}_I)$ are positively correlated, and so are $(x_{C0}, \widetilde{y}_I)$, $(x_{I0}, \widetilde{y}_C)$. By contrast, $(x_{C0}, \widetilde{y}_M)$, $(x_{M0}, \widetilde{y}_C)$, $(x_{I0}, \widetilde{y}_M)$, $(x_{M0}, \widetilde{y}_I)$ are all negatively correlated, suggesting that workers select themselves into either manual or non-manual jobs, as fits their skill bundles. Third, the correlations between workers' initial skill measures and job skill requirements remain relatively stable over the observation window. Our model will interpret this as resulting from the combination of two forces: workers gradually selecting themselves into jobs for which their skills are better suited, while at the same time seeing their skills evolve away from their initial skill bundle and toward the skill requirements of the jobs they happen to hold.

Wages. Table 3 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. Cognitive and manual skill requirements are positively associated with wages in the cross-section, the effect of cognitive skills being 10 times larger than that of manual skills. Oddly, interpersonal skill requirements are negatively associated with wages, suggesting some negative selection into jobs that are intensive in interpersonal skills. Workers' initial levels of cognitive and interpersonal skills are positively correlated with wages (interpersonal skills much more weakly so), while initial manual skill levels come out with a very slightly negative, marginally significant coefficient. Overall, those results suggest that, after controlling for education, cognitive job and worker attributes are strongly positively associated with wages, while the effects of manual and interpersonal job and worker on wages are much weaker.

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.

1-5 years in sample

	x_{C0}	x_{M0}	x_{I0}	$\widetilde{y}_C$	$\widetilde{y}_M$	$\widetilde{y}_I$
x_{C0}	1					
x_{M0}	-0.79	1				
x_{I0}	0.52	-0.52	1			
$\widetilde{y}_C$	0.50	-0.49	0.37	1		
$\widetilde{y}_M$	-0.47	0.39	-0.36	-0.52	1	
$\widetilde{y}_I$	0.44	-0.41	0.37	0.85	-0.56	1

5-10 years in sample

	x_{C0}	x_{M0}	x_{I0}	$\widetilde{y}_C$	$\widetilde{y}_M$	$\widetilde{y}_I$
x_{C0}	1					
x_{M0}	-0.79	1				
x_{I0}	0.51	-0.53	1			
$\widetilde{y}_C$	0.50	-0.47	0.35	1		
$\widetilde{y}_M$	-0.51	0.43	-0.38	-0.58	1	
$\widetilde{y}_I$	0.44	-0.40	0.33	0.85	-0.61	1

10-15 years in sample

	x_{C0}	x_{M0}	x_{I0}	$\widetilde{y}_C$	$\widetilde{y}_M$	$\widetilde{y}_I$
x_{C0}	1					
x_{M0}	-0.79	1				
x_{I0}	0.51	-0.53	1			
$\widetilde{y}_C$	0.52	-0.50	0.37	1		
$\widetilde{y}_M$	-0.51	0.44	-0.40	-0.61	1	
$\widetilde{y}_I$	0.46	-0.42	0.36	0.87	-0.63	1

15-20 years in sample

	x_{C0}	x_{M0}	x_{I0}	$\widetilde{y}_C$	$\widetilde{y}_M$	$\widetilde{y}_I$
x_{C0}	1					
x_{M0}	-0.79	1				
x_{I0}	0.51	-0.52	1			
$\widetilde{y}_C$	0.47	-0.45	0.34	1		
$\widetilde{y}_M$	-0.49	0.41	-0.41	-0.62	1	
$\widetilde{y}_I$	0.43	-0.38	0.35	0.86	-0.65	1

Table 2: Correlation pattern of initial skills and skill requirements

	coeff.	std. err.		coeff.	std. err.
$\widetilde{y}_C$	0.738	(.008)	x_{I0}	0.047	(.005)
$\widetilde{y}_M$	0.046	(.006)	years of schooling	0.042	(.001)
$\widetilde{y}_I$	-0.168	(.008)	experience	$2.2e-3$	(.000)
x_{C0}	0.224	(.007)	tenure	$1.9e-3$	(.000)
x_{M0}	-0.019	(.007)	constant	4.639	(.014)
R^2		0.371			
Observations		168,318			

Table 3: Descriptive wage regression

5.1 Simulation

Solution method. The model has a convenient recursive structure. Equations (3) and (2) 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

19

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 iff. $P(\mathbf{x}_{it}, \mathbf{y}') \geq U(\mathbf{x}_{it})$. Again, the worker's wage is updated.[19]

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 (10), (12), and (13) 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

---

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

interpersonal skills $(x_C(0), x_M(0), x_I(0))$, 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 $\ln w \stackrel{d}{=} \ln w|_{\varepsilon_0=1} + \varepsilon_0$, where $\stackrel{d}{=}$ denotes equality in distributions, and $w|_{\varepsilon_0=1}$ 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. 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, the rank correlation parameters $(\rho_{CM}, \rho_{CI}, \rho_{MI})$ of which are to be estimated.

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. 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.[20] With this subset of parameters estimated - or calibrated - in a preliminary step,

---

[20]Figure 1 suggests that the job loss rate is not exactly constant over a worker's life cycle. We abstract from this feature of the data.

we are left with the following 29-dimensional parameter vector to estimate:

$$\Theta = [\lambda_0, \lambda_1, b, \alpha_T, \alpha_C, \alpha_M, \alpha_I, \kappa_C^u, \kappa_C^o, \kappa_M^u, \kappa_M^o, \kappa_I^u, \kappa_i^o,$$

$$\xi_C, \xi_M, \xi_I, \gamma_C^u, \gamma_C^o, \gamma_M^u, \gamma_M^o, \gamma_I^u, \gamma_I^o, g, \zeta_S, \zeta_C, \zeta_M, \rho_{CM}, \rho_{CI}, \rho_{MI}]$$

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)* cross-sectional mean log wage at a selection of sampling dates,[21] *(iv)* 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, *(v)* 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 *(vi)* 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 *(vii)* coefficients of a regression of log wages $\ln w_{it}$ on initial skills $\mathbf{x}_{i1}$, current job attributes $\widetilde{\mathbf{y}}_{it}$, tenure, experience, and years of schooling (i.e. the regression in the first column of Table (3)). The model-based moments are computed from simulated samples of 36,800 workers - twenty replicas of the initial NLSY cross-section.

5.3 Identification

Appendix B 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. Leaving out the details of the formal identification arguments, in this Subsection we summarize the main sources of information that identify the various components of our theoretical 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 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 $\mathbf{x}(0)$

---

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

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 B 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 $\Upsilon(\cdot)$. The second subsection of Appendix B shows that, in practice, our chosen set of moments ensures precise "local" identification of the model's parameters, in the sense that the objective function (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 2 gives a visual idea of some aspects of the fit. All time series on Figure 2 are plotted over a period of 15 years (180 months, i.e. the sample window used for estimation). Table 4 further shows the fit in terms of the descriptive age regression discussed in Section 4.

The model fits both the nonemployment exit rate (Figure 2a, left scale) and the job-to-job transition rate (Figure 2a, right scale) reasonably well. The decline of E2E rates with experience is correctly captured by the model (it occurs as a consequence of workers climbing the job ladder and settling into jobs to which their skills are better suited), 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 very good, in spite of the model's failure to capture the mild upward trend
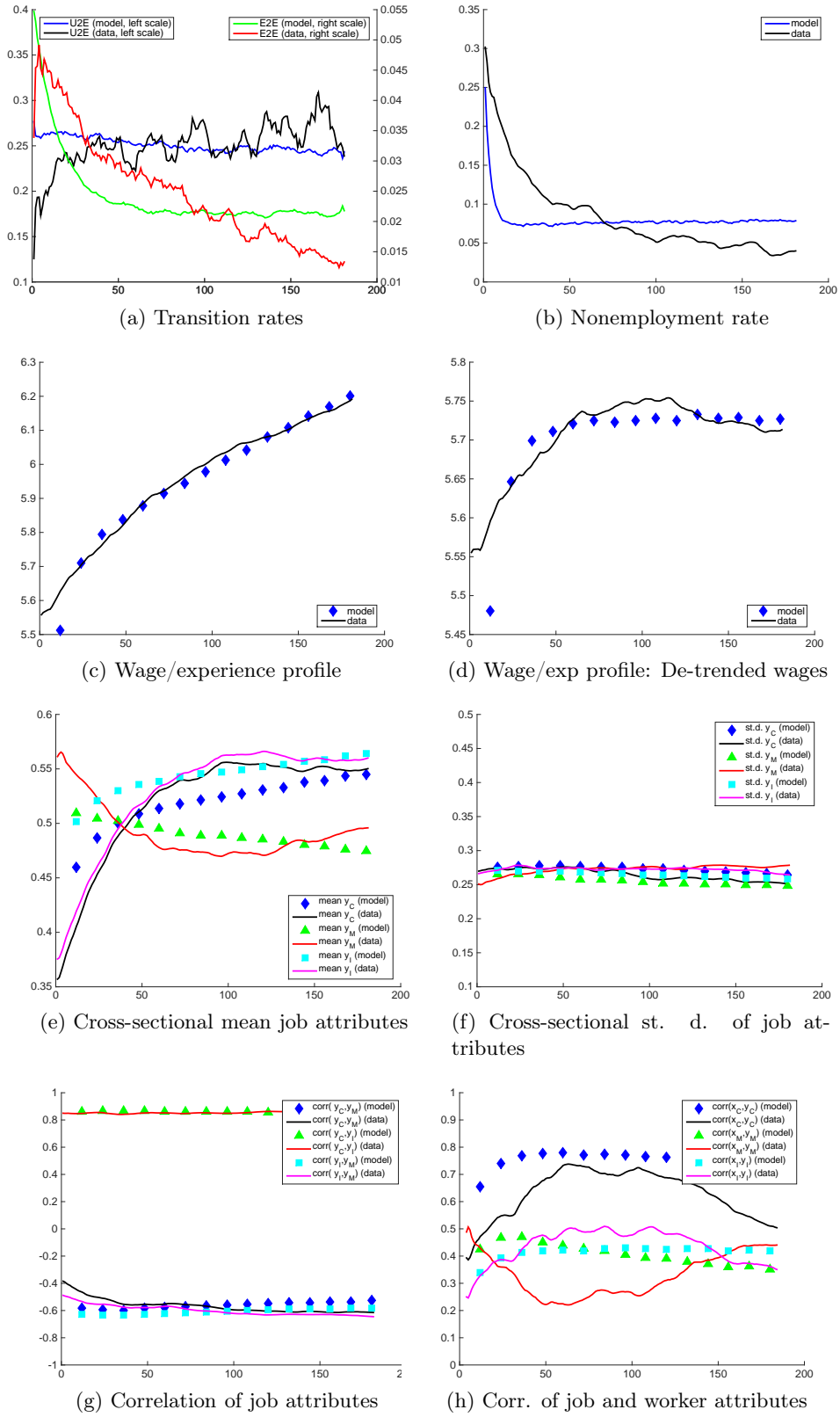
23

(a) Transition rates

(b) Nonemployment rate

(c) Wage/experience profile

(d) Wage/exp profile: De-trended wages

(e) Cross-sectional mean job attributes

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

(g) Correlation of job attributes

(h) Corr. of job and worker attributes

Figure 2: Model fit

24

	data	model		data	model
$\widetilde{y}_C$	0.738	0.559	x_{I0}	0.047	0.047
$\widetilde{y}_M$	0.046	0.011	years of schooling	0.042	0.048
$\widetilde{y}_I$	−0.168	−0.097	experience	2.2e − 3	2.2e − 3
x_{C0}	0.224	0.224	tenure	1.9e − 3	5.8e − 3
x_{M0}	−0.019	−0.019	constant	4.639	4.577

Table 4: Fit to wage regression coefficients

in that rate. The discrepancies between the observed and model-based nonemployment rate profiles (Figure 2b) are mainly caused by our restriction to a constant job loss rate which, as discussed in Section 4, is at odds with observation.

The sample average wage/experience profile is shown on Figure 2c, while Figure 2d shows the fit to de-trended (i.e. wages multiplied by e^{-gt}) to separate the impact of the trend in individual efficiency from that of on-the-job search. Both profiles are reasonably well captured by the model, despite a tendency to understate starting wages. This is a recurring issue with 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. This issue is partly compensated in this model by the utility cost of being under-matched, which over-qualified workers have to be compensated for.

Figures 2e through h 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 reasonably good fit to all targeted moments moments: looking at the scale of Figures 2e-h suggests that the magnitude of the discrepancies between the data and model-based series is modest.

We next turn to the model's ability to replicate the pooled cross-section wage regression shown in Tables 3 and 4. 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 (although the magnitude of all of those coefficients are slightly understated by the model). Coefficients on "conventional" regressors (schooling, experience, and tenure) are also reasonably well captured by the model (despite a tendency
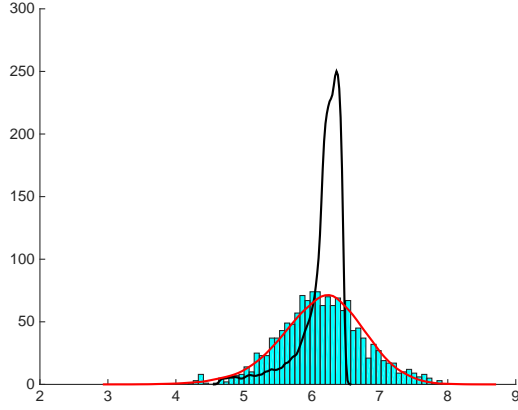
Figure 3: Log wage density

to overstate the returns to schooling and to job tenure), and so is the conditional correlation between wages and initial worker skills.

The black line on Figure 3 shows a kernel density estimate of log wages in the final period of the simulation *still assuming away unobserved worker heterogeneity*, i.e. under that assumption 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 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 3. To illustrate this, the red line on 3 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.

6.2   Parameter Estimates

Table 5 shows point estimates of the model parameters. 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 0.4, 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, although its

26

production function							un. inc.
α_T	α_C	α_M	α_I	κ_C^u	κ_M^u	κ_I^u	b
112.5	66.6	8.58	0.673	242.7	146.1	11.6	120.3

disu. of work			general efficiency			
κ_C^o	κ_M^o	κ_I^o	ζ_S	ζ_C	ζ_M	ζ_I
3.31	0.091	1.35	0.045	0.621	-0.029	0.075

skill accumulation function*						
γ_C^u	γ_C^o	γ_M^u	γ_M^o	γ_I^u	γ_I^o	g
4.0e $-$ 3	4.1e $-$ 3	0.039	0.039	5.6e $-$ 7	5.6e $-$ 7	2.6e $-$ 3
(14.5)	*(14.0)*	*(1.48)*	*(1.48)*	*(>100,000)*	*(>100,000)*	

sampling distribution						transition rates		
ξ_C	ξ_M	ξ_I	ρ_{CM}	ρ_{CI}	ρ_{IM}	λ_0	λ_1	$\delta^{\star\star}$
1.45	0.87	1.19	-0.628	0.897	-0.672	0.409	0.155	0.021

*: half-life in years in parentheses

**: estimated in first step

Table 5: Parameter estimates

dependence on manual skill requirements is small, and its dependence on interpersonal skill requirements is negligible compared to the effect of cognitive skill requirements. This is consistent with the much lower coefficients on $\widetilde{y}_M$ and $\widetilde{y}_I$ than on $\widetilde{y}_C$ in the wage regression shown in Table 4.

Overall worker efficiency x_T is strongly positively associated with a high initial endowment in cognitive skills (ζ_C), while initial other skill types are less strongly correlated with efficiency (especially so for initial manual skills, which are slightly negatively correlated with x_T: $\zeta_M < 0$). One additional year of education increases efficiency by 4.5 percent (ζ_S). However one should bear in mind that education is positively correlated with initial cognitive and interpersonal skills and 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 either cognitive or manual skills is (very) costly in terms of output, yet the output loss caused by this type of mismatch in the cognitive dimension is about almost twice as severe as that caused by manual skill mismatch. The output loss caused by under-qualification in the interpersonal dimension is comparatively very small. The utility cost of being under-matched - i.e. the surplus cost of the worker being *over*-qualified worker - is positive in the all dimensions, but considerably smaller than the corresponding surplus (production) cost of *under*-qualification. The utility cost of being under-matched in the manual dimension is negligible.

The correlation patterns between skill requirements in the sampling distribution ($\rho_{CM} < 0$,$\rho_{CI} > 0$,$\rho_{IM} < 0$), suggesting that employers are looking for "specialist" workers rather than generalists, where specialization occurs along the manual/non-manual dimension. The pattern of skill adjustment differs vastly between cognitive and manual skills. All skill types adjust at remarkably similar speeds in both directions. Yet 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 13 to 16 years. The half-life of manual skills is much shorter, about 18 months in both directions. Interpersonal skills essentially never adjust 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 low returns and adjust quickly in both directions, cognitive skills have much higher returns, but are much slower to adjust. Interpersonal skills essentially affect overall worker efficiency but have very low

"direct" returns in production, and are essentially fixed over a worker's lifetime.

6.3 Skill Mismatch, Skill Changes, and Sorting

Distributions of skills and skill requirements. The broad question of skill mismatch, which we examine in this sub-section, 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 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 $\Upsilon(\mathbf{y})$. Figure 4 shows histogram plots of both distributions, with a common scale for comparability. Specifically, Figure 4 shows the three *marginal* sampling distributions of (y_C, y_M), (y_C, y_I), and (y_I, y_M) (panels a, c, and e) and the corresponding three marginal distributions of pairs of initial worker skills (panels b, d, and f).[22]

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. 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 4b, d and f also suggests that the distribution of skills among labor market entrants reflects this demand for specialists to an extent. However, a closer look reveals that the distribution of initial worker skills is much more skewed towards the low cognitive skill end of the skill spectrum than the distribution of job offers: employers seem to demand fewer manual skills than are available in the population of labor market entrants.

Figure 5 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 explained in previous paragraphs, 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 such jobs in which they gradually acquire cognitive skills. Since those jobs generally have low manual skill requirements (as y_C and y_M are strongly negatively
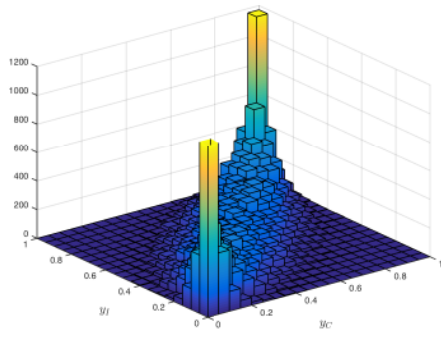
---

[22]For example, the marginal sampling distribution of (y_C, y_M) plotted on panel a is $\int_0^1 \gamma(y_C, y_M, y_I) dy_I$.
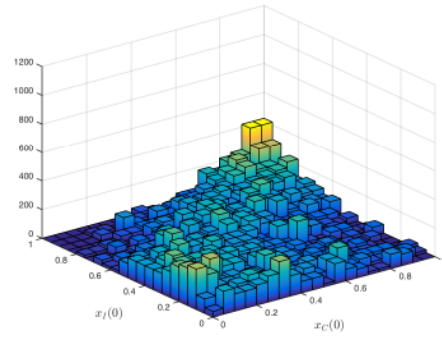
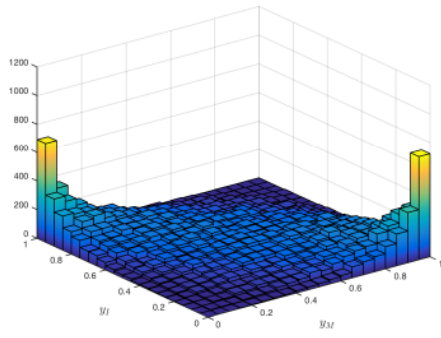29
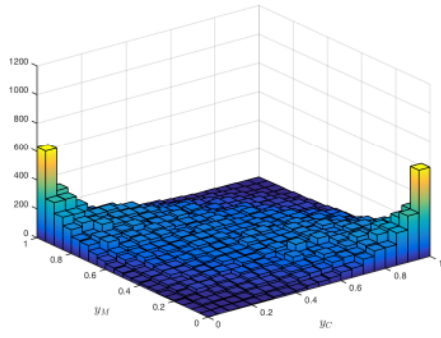
(a) Sampling density (CM)
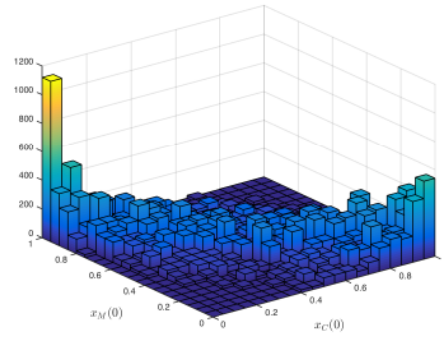
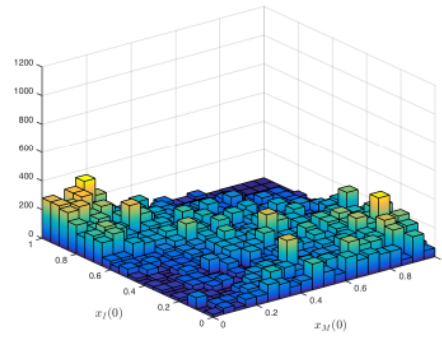(b) Initial worker skill dist. (CM)

(c) Sampling density (CI)
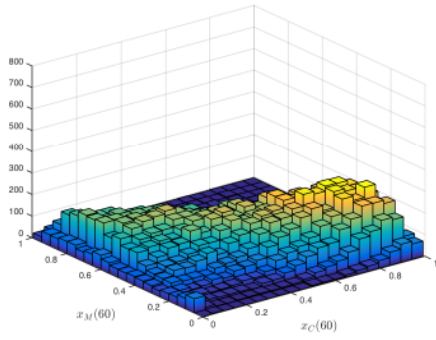
(d) Initial worker skill dist. (CI)
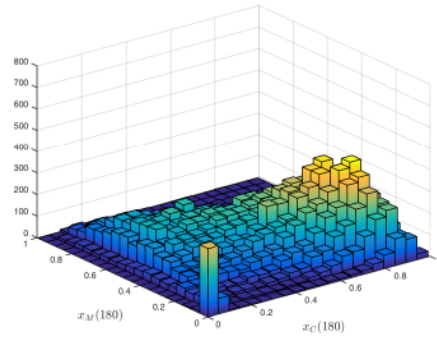
(e) Sampling density (MI)
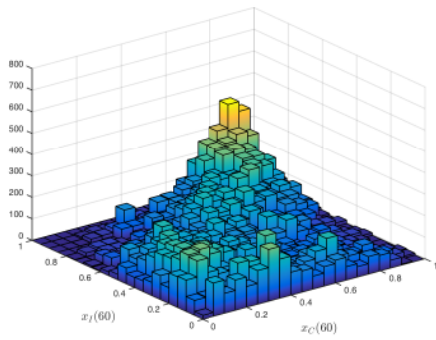
(f) Initial worker skill dist. (MI)

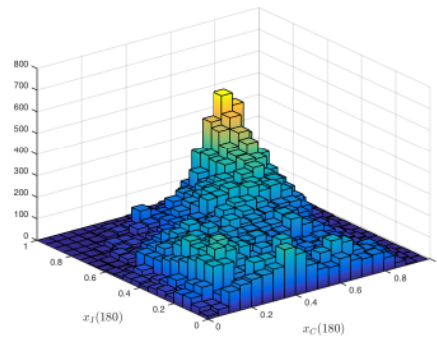Figure 4: Distribution of initial skills and skill requirements
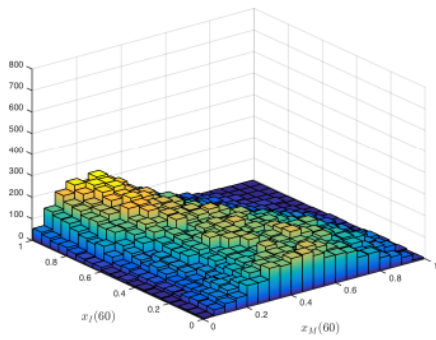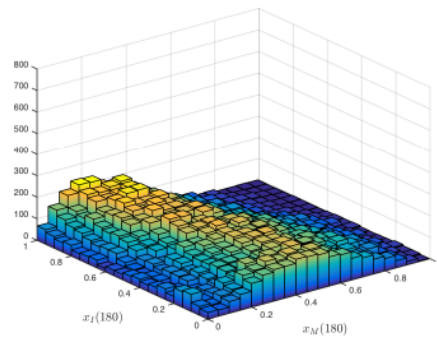
(a) CM, 5 years of exp.

(b) CM, 15 years of exp.

(c) CI, 5 years of exp.

(d) CI, 15 years of exp.

(e) MI, 5 years of exp.

(f) MI, 15 years of exp.

Figure 5: Evolution of worker skills with experience

correlated in the job offer sampling distribution), workers also tend to lose their manual skills by working in those jobs. They do so relatively fast, given the high estimated adjustment speed of manual skills, $\gamma_M^{u/o}$. A second striking feature of Figures 5 is that, although the typical worker's skills tend to migrate toward more cognitive skills and less manual skills on average, the extreme degree of specialization apparent in the initial skill distribution (Figure 4b) regresses substantially as workers gain experience: the initially bimodal distribution of skills changes to a single-mode distribution relatively fast.

Finally, Figures 4c-d show the emergence of a small group of workers with very low levels of both types of skills. Those are the long-term unemployed, who have stayed long enough out of a job to lose all of their initial skills and become unemployable.[23] Closer examination of the simulated dynamics reveals that those unemployable workers are predominantly workers who entered the labor market with very low initial cognitive skills and failed to find a suitably manual job soon enough to prevent their manual skills from eroding. Because cognitive skills atrophy at a much slower pace than manual skills when they are not used (see Table 5), workers entering the market with a reasonable initial level of cognitive skills are, to a large extent, immune from the risk of becoming unemployable.

Skill sorting and mismatch. We next examine the joint distribution of worker skill bundles and job skill requirements among ongoing matches. Figure 6 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.

Another, perhaps less immediately obvious feature of Figure 6 is that workers tend to be slightly "over-matched" in the cognitive dimension, in the sense that their job's cognitive skill requirement is typically slightly higher than their own cognitive skill level. This is most easily
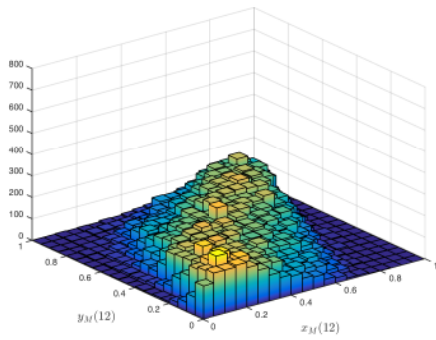
---

[23]We can define "unemployability" formally within the model: a worker with skill bundle $\mathbf{x}$ such that $\max_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{x}, \mathbf{y}) < U(\mathbf{x})$ generates negative net private surplus in all potential job types, and is therefore unemployable.
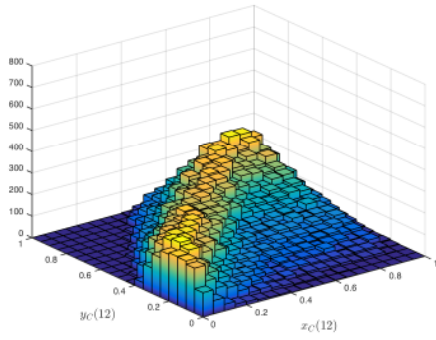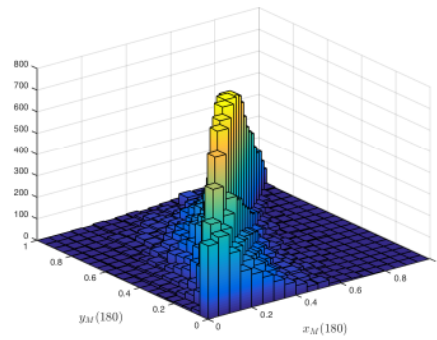
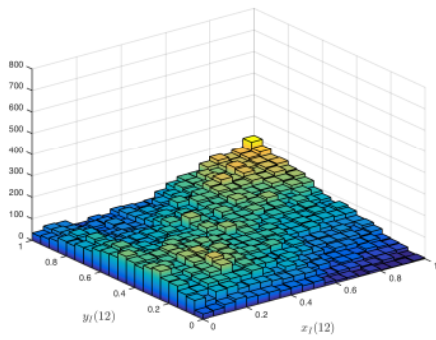(a) Cognitive skills, one year of exp.
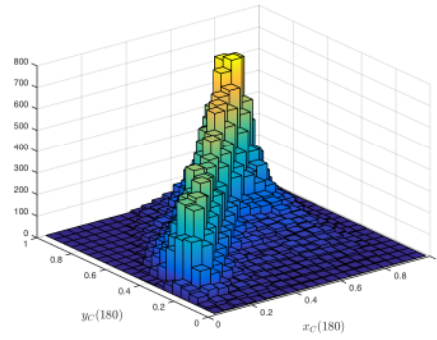
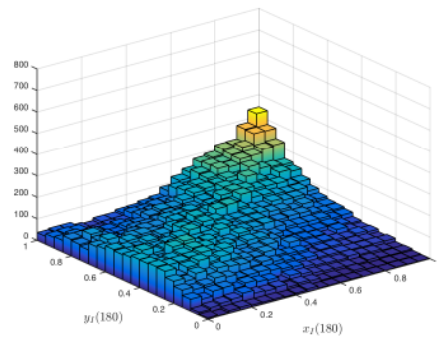(b) Cognitive skills, 15 years of exp.

(c) Manual skills, one year of exp.

(d) Manual skills, 15 years of exp.

(e) Interpersonal skills, one year of exp.

(f) Interpersonal skills, 15 years of exp.

Figure 6: Sorting

seen on Figure 6b, where the main mass of the distribution is slightly on the skill requirement side of the main diagonal. This is due to the fact that jobs with higher cognitive skill requirements are intrinsically more productive for any level of worker skills, and are therefore more attractive to all workers. The tradeoff from the perspective of a worker contemplating a job is between the job's overall productivity (the $\varphi(\mathbf{y})$ term in the production function), and any cost from being mismatched. The optimal point in that tradeoff is a level of skill requirement slightly above the current worker's skills. This tradeoff applies to all skill types, but the relative weight on overall job productivity is much higher in the cognitive than in the manual or interpersonal dimensions, because $\varphi(\mathbf{y})$ increases much more steeply with y_C than with y_M or y_I (see Table 5).

7 Conclusion [TBC]

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.

GUVENEN, F., B. KURUSCU, S. TANAKA, AND D. WICZER (2014): "Occupational Switching and Self-Discovery in the Labor Market," Manuscript, Federal Reserve Bank of St. Louis.

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.

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.

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

POSTEL-VINAY, F., AND J.-M. ROBIN (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

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.

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

A   Parameter Summary

Offer arrival rates: (λ_0, λ_1)

Job destruction rate: δ *Estimated in first step as*
the 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 $\mathbf{g}$: $(\gamma_C^u, \gamma_C^o, \gamma_M^u, \gamma_M^o, \gamma_I^u, \gamma_I^o, g)$

Joint distribution of initial *Observed from initial*
cognitive and manual skills: *cross section in the sample*

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 $(\xi_C, \xi_M, \xi_I, \rho_{CM}, \rho_{CI}, \rho_{MI})$ *Gaussian copula with implied*
of job attributes Υ: *correlation $(\rho_{CM}, \rho_{CI}, \rho_{MICM})$*

Attrition and discount rates: (r, μ) *Calibrated*

B   Identification

B.1   In theory

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:

$$w(\mathbf{x}, \mathbf{y}, \sigma) = \sigma f(\mathbf{x}, \mathbf{y}) + (1 - \sigma)b(\mathbf{x}) + (1 - \sigma)c(\mathbf{x}, \mathbf{y})$$

$$- \lambda_1 (1 - \sigma) \left[P(\mathbf{x}, \mathbf{y}) - U \right] \int_{\mathcal{Y}} \mathbf{1} \left\{ P(\mathbf{x}, \mathbf{y}') \geq P(\mathbf{x}, \mathbf{y}) \right\} d\Upsilon(\mathbf{y}')$$

$$- \lambda_1 \int_{\mathcal{Y}} \left[\mathbf{1} \left\{ P(\mathbf{x}, \mathbf{y}') \geq \sigma P(\mathbf{x}, \mathbf{y}) + (1 - \sigma)U \right\} - \mathbf{1} \left\{ P(\mathbf{x}, \mathbf{y}') \geq P(\mathbf{x}, \mathbf{y}) \right\} \right] \times$$

$$\left[P(\mathbf{x}, \mathbf{y}') - \sigma P(\mathbf{x}, \mathbf{y}) - (1 - \sigma)U \right] d\Upsilon(\mathbf{y}'). \quad (15)$$

A first important implication of (15) 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.[24] 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 (14), 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 $\kappa_k^{u/o} / \left(r + \delta + \mu - g + 2\gamma_k^{u/o} \right)$, $k = C, M, I$.[25] 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 10, we know that this worker's skill bundle by the time 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 $\kappa_k^u / (r + \delta + \mu - g + 2\gamma_k^u)$, and still need to separate κ_k^u from γ_k^u in the latter

---

[24]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.

[25]One way to see this is to realize from (14) 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.

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 10, 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 $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{y})$, 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 (6) 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 (15) 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 $P(\mathbf{x}, \mathbf{y}) = U$.

B.2   In practice

Figure 7 shows how the objective function (the distance between empirical and simulated moments) responds to changes in each parameter individually around the estimated value (indicated by a red dot on the figure). This gives a heuristic idea of the strength of local parametric identification of our model, based on the moments we use for estimation. The figure generally suggests that the objective function has a reasonably clear local minimum at the parameter estimates with our simulated sample size, despite the presence of a fair amount of simulation noise. We suspect that most of this noise arises from the lack of precision with which we are able to compute the three-dimensional integral $\mathbf{E} \max \{0, \min \{P(\mathbf{x}, \mathbf{y}') - P(\mathbf{x}, \mathbf{y}), 0\} + (1 - \sigma)(P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x}))\}$ appearing in the wage equation (5). To limit compute time and memory requirements, we approximate this integral using the trapezoid rule over a grid of $10 \times 10 \times 10 = 1,000$ integration nodes for $\mathbf{y}' = (y'_C, y'_M, y'_I)$.
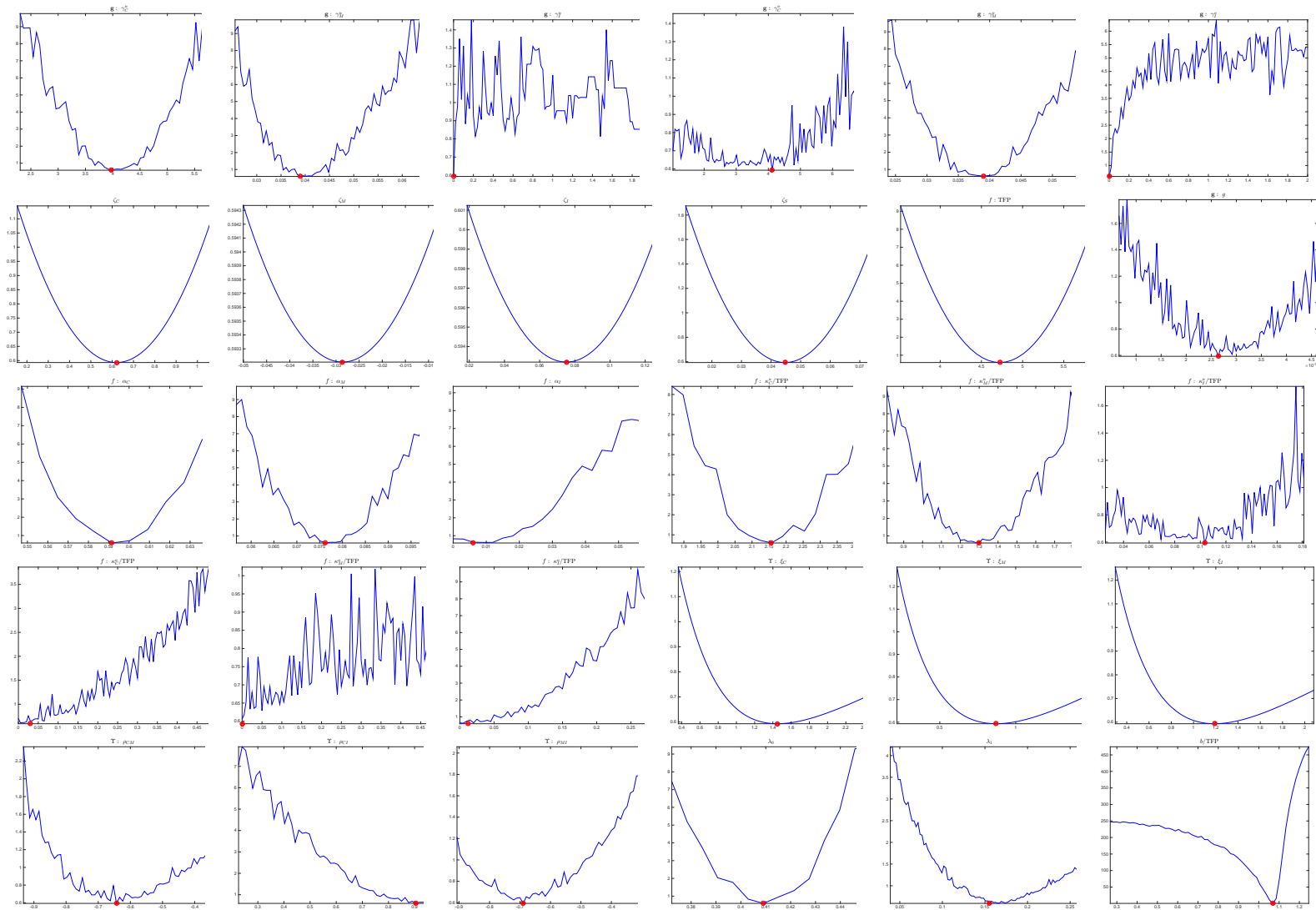
Figure 7: Identification