Multidimensional Skills, Sorting, and Human Capital Accumulation*

Jeremy Lise
University of Minnesota
and FRB Minneapolis

Fabien Postel-Vinay
UCL and IFS

December 2019

Abstract

We construct a structural model of on-the-job search in which workers differ in skills along several dimensions and sort themselves into jobs with heterogeneous skill requirements along those same dimensions. Skills are accumulated when used, and depreciate when not used. We estimate the model combining data from O*NET with the NLSY79. We use the model to shed light on the origins and costs of mismatch along heterogeneous skill dimensions. We highlight the deficiencies of relying on a unidimensional model of skill when decomposing the sources of variation in the value of lifetime output between initial conditions and career shocks.

*We would like to thank three anonymous referees and two associate editors, plus Fatih Guvenen, Jim Heckman, John Kennan, Derek Neal, Carl Sanders, David Wiczer, Ken Wolpin and seminar audiences at ASU, Bonn, Cambridge, Chicago, Collegio Carlo Alberto (Turin), Cornell, CREST (Paris), Ecole Polytechnique, Helsinki, Hitotsubashi University Tokyo, Lausanne, Minnesota, NYU, U Penn, Penn State, Pitt/CMU, USC, Rice, Richmond Fed, Toronto, Vienna, Washington St. Louis/St. Louis Fed and Wisconsin Madison, as well as participants in the NBER Summer Institute, RES meetings, SED meetings, SaM annual conference, Philadelphia Workshop on Macroeconomics, and Barcelona GSE Summer Forum for useful feedback on earlier versions of this paper. We are additionally grateful to Carl Sanders for kindly providing his data and helping us get started with O*NET. All errors and shortcomings are ours. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

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

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

The alternative view that workers are endowed with bundles of different skills used in different proportions depending on the task they perform has some history in labor economics (see Sanders and Taber, 2012, for a review). But at present, few quantitative modeling tools exist that fully exploit the wealth of information on heterogeneous, multidimensional worker skills and job skill requirements available in the data in a description of labor market equilibrium.

In this paper, we contribute to filling this gap: we extend an otherwise standard and

[1]It shed light on secular trends such as the "polarization" of wages and employment (the simultaneous growth in wages and employment shares of both high-and low-skill workers, at the expense of the middle part of the skill distribution - Acemoglu and Autor, 2011). It contributed to explaining business-cycle fluctuations in aggregate output and employment (Lise and Robin, 2017). It gave substance to the intuitive notion of "skill mismatch" and helped make sense of patterns of worker turnover (Shimer and Smith, 2000; Lise, Meghir, and Robin, 2016). It helped clarify the informational content of wage data (Eeckhout and Kircher, 2011; Hagedorn, Law, and Manovskii, 2017).

well-tested search-theoretic model of individual careers to allow for multidimensional skills and on-the-job learning. We estimate the model using occupation-level measures of skill requirements based on O*NET data, combined with a worker-level panel (NLSY79). We use the estimated model to shed light on the origins and costs of mismatch along three dimensions of skills: cognitive, manual, and interpersonal,[2] and the sources of variation in lifetime output.

Our main findings are the following. The model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have moderate returns and adjust quickly (*i.e.*, they are easily accumulated on the job, and relatively easily lost when left unused). Cognitive skills have much higher returns, but are much slower to adjust. Interpersonal skills have only slightly higher returns than manual skills, and are essentially fixed over a worker's lifetime. Next, the cost of skill mismatch (modeled as the sum of an output loss and a loss of worker utility caused by skill mismatch) is very high for cognitive skills, an order of magnitude greater than for manual or interpersonal skills. Moreover, this cost is asymmetric: employing a worker who is under-qualified in cognitive skills (*i.e.* has a level of skills that falls short of the job's skill requirements) is more than twice as costly in terms of lost surplus as employing an over-qualified worker. Those important differences between various skill dimensions are missed when subsuming worker productive heterogeneity into one single scalar index. Indeed, when we consider a decomposition of lifetime output, and compare our multidimensional model to a version of the model with a only a single skill, the single skill version overestimates the importance of unobserved heterogeneity (relative to observed initial skills) by a factor of two, and underestimates the contribution of career shocks relative to initial observed skills by one half.

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. Section 7 presents some results on the determinants of social output, and decomposes the variance of expected lifetime output into components due to initial endowments of worker skills and randomness during workers' careers. Finally, Section 8 concludes.

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

2 Related Literature

This paper is related to the vast empirical literature on the returns to firm and occupation tenure and to more recent work on task-specific human capital (see, for example, Poletaev and Robinson, 2008; Gathmann and Schonberg, 2010, among others). Those literatures, and the connections between them are covered in the excellent survey paper by Sanders and Taber (2012). As a preamble to their review of the empirical literature, Sanders and Taber (2012) offer an elegant theoretical model of job search and investment in multi-dimensional skills which, on many aspects, can be seen as a special case of the model in this paper.[3] However, they only use their model to provide intuition and highlight key qualitative predictions of the theory, and do not bring it to the data.

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

While Lindenlaub's analysis brings about many valuable new insights, it assumes away market imperfections which limits its applicability to empirical and quantitative 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, the extent of the imperfections is an empirical question. Second, frictionless matching and assignment models, including Rosen's hedonic model, are static.[4] As such, they are largely silent on questions relating to a worker's life cycle, such as the cost of

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

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

skill mismatch throughout a worker's career, the way in which individual skills evolve over a career, how this skill accumulation is priced in the market, or the reasons why workers switch occupations as often as they do.[5]

Next, following in the tradition of Heckman and Sedlacek (1985), Keane and Wolpin (1997), and Lee and Wolpin (2006), the important contribution by Yamaguchi (2012) provides the first estimation of a Roy-type model of task-specific human capital accumulation and occupation choices over the life cycle based on the combination of the NLSY with data on occupation-level attributes, interpreted as "task complexity", from the Dictionary of Occupational Titles (DOT, the predecessor of O*NET). The broad approach is the same as in the present paper: each occupation is characterized by the vector of weights (the degree of task complexity) it places on a limited number of different skill dimensions, as in Lazear's (2009) skill-weights approach. Worker skills are not directly observed, but their accumulation is modeled as a hidden Markov chain, the parameters of which are identified from observed choices of occupations with different task contents (observed from the DOT data), using the model's structure. Yamaguchi's findings suggest that higher task complexity is associated with higher wage returns to, and faster growth of the skills relevant to the task. A wage variance decomposition further suggests that both cognitive and motor skills (the two skill dimensions considered by Yamaguchi) are important determinants of cross-sectional log wage variance. A decomposition of wage growth shows that 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 clearly related in spirit, our model differs from Yamaguchi (2012) in several important ways. First, Yamaguchi (2012) is a frictionless model in which occupation mobility is largely governed by unobserved shocks to an exogenously posited wage function, unobserved shocks to workers' skills, and unobserved shocks to workers' preferences for any given type of job.[6] We propose a more parsimonious random search model, in which the only shocks

[5]Models of experimentation and learning following on from Jovanovic (1979) such as Neal (1999); Pavan (2011); Golan and Antonovics (2012) have been useful to make sense of the patterns of between and within occupation switches in the NLSY. They cannot, however, easily rationalize the frequent transitions in and out of unemployment observed in the same data. Lindenlaub and Postel-Vinay (2016) extend Lindenlaub's frictionless model to a frictional environment using a basic framework that has much in common with our model. However, Lindenlaub and Postel-Vinay (2016) is a theoretical exercise focusing on conditions under which specific sorting patterns emerge in steady-state equilibrium, and their model does not feature human capital accumulation. Anderson and Smith (2010) develop a matching model with evolving one-dimensional types in a frictionless environment. They characterize the conditions for assortative matching in this context, as well as implications for life-cycle dynamics of wages.

[6]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. Note that any dynamic model that aims to replicate the empirical job mobility patterns will need to incorporate some type of mobility friction (for example a mobility

are the receipt of job offers by workers. Wages and mobility decisions are then endogenously determined through between-employer competition for labor services. Our less flexible, but more transparent and readily interpretable specification offers a remarkably good fit to the data. Second, the only engine of wage growth in Yamaguchi's model is skill accumulation. Other sources of wage growth, such as job-shopping or learning, are therefore partly picked up by skills in that model, which may lead to an overstatement of the role of skills. Our model also ignores learning, but explicitly models job-shopping as an additional source of wage dynamics.[7] Adding search frictions allows us to address issues related to unemployment and skill mismatch.[8]

Two recent papers are particularly related. Taber and Vejlin (2016) estimate a model which allows for search, Roy-type selection, human capital accumulation and non wage amenities. Workers are modeled as having a time invariant relative ability at each job-type in the economy. In the absence of frictions they would choose a single job-type and remain indefinitely. Human capital is assumed to be general and accumulated while working. Job mobility is informative about the degree of search frictions, and wage cuts are informative about non wage amenities. Taber and Vejlin (2016) model relative ability between jobs/occupations as an unobserved vector with dimension equal to the number of job-types in the economy. We take a substantially more parsimonious approach: a worker's relative productivity across jobs/occupations depends on the amount of skills (cognitive, manual, interpersonal. . .) they currently possess and whether or not they are a good fit for the demands of a particular job. Rather than treating these skills as unobserved, we use a large set of premarket measurements to estimate a worker's initial endowment, and a similar large set of measurements on occupations to estimate the skill requirements of jobs. A second notable modeling difference is that we allow these skills to evolve differentially depending on the extent to which they are being used in a particular job/occupation. We are particularly interested in the the differential returns to these skills, and the extent to which each type of skill can be learned on the job.

From an empirical perspective, perhaps the closest paper is Guvenen, Kuruscu, Tanaka, and Wiczer (2018). They use the same combination of NLSY and O*NET data as we do to construct a summary index of multidimensional skill mismatch which they use to

cost) to prevent workers from continuously reallocating in the face of new shocks or new information. For example, Guvenen, Kuruscu, Tanaka, and Wiczer (2018), which we discuss below, rely on a combination of information and mobility frictions to generate the empirical implications of their learning model.

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

[8] 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.

assess the impact of skill mismatch on wages and patterns of occupational switching. They produce a rich set of empirical results, a rough summary of which is that both current and past mismatch strongly impact wages (negatively), the probability of switching occupations (positively), and the direction of said switching.

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

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

3 Job Search with Multi-dimensional Job and Worker Attributes

3.1 The Model

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

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

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

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

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

Firm, worker, and match values. We denote the total private value (*i.e.* the value to the firm-worker pair) of a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ firm by $P(\mathbf{x}, \mathbf{y})$. Under linear preferences over wages, this value is independent of the way in which it is shared between the two parties, 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 their current wage contract by W, which we discuss in detail below. Admissible worker values imply $W \geq U(\mathbf{x})$ (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.[9]

[9]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

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, they stay in their 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 $W' = \min\{P(\mathbf{x},\mathbf{y}), \max\{P(\mathbf{x},\mathbf{y}'), W\}\}$.

Suppose, for the sake of argument, that $P(\mathbf{x},\mathbf{y}) \geq P(\mathbf{x},\mathbf{y}') > W$. In this case, the outcome of the renegotiation is such that the worker stays with their 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, denoted $\sigma(\mathbf{x},\mathbf{y},\mathbf{y}')$, is therefore:

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

To pin down the way in which the value $W' = P(\mathbf{x},\mathbf{y}') = U(\mathbf{x}) + \sigma(\mathbf{x},\mathbf{y},\mathbf{y}')\,[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})]$ is delivered over time by the firm to the worker, we assume that the surplus share σ, negotiated at the time the worker receives an outside offer from the type-$\mathbf{y}'$ job, stays constant until the following renegotiation. Put differently, while the worker's skill bundle $\mathbf{x}$ and, as a consequence, the match surplus $P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$ evolve over the course of their tenure in the type-$\mathbf{y}$ job, the share of that surplus transferred to the worker stays constant between negotiations and is determined as per equation (1) by the best outside offer previously received by the worker. The particular way in which the type-$\mathbf{y}$ employer delivers the value $P(\mathbf{x},\mathbf{y}')$ to the worker only affects the time profile of wage payments and the timing of renegotiation. It makes no difference to the allocation of workers into jobs, as mobility decisions are only based on comparisons of total match values, which, under linear preferences, are independent of the time profile of wage payments.[10]

an offer from that employer is not credible.

[10]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:[11]

$$\begin{aligned}(r+\mu+\delta)P(\mathbf{x},\mathbf{y}) &= f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} P(\mathbf{x},\mathbf{y}) \\ &\quad + \lambda_1 \mathbf{E}\left[W(\mathbf{x},\mathbf{y}',\sigma) - P(\mathbf{x},\mathbf{y}) | P(\mathbf{x},\mathbf{y}') > P(\mathbf{x},\mathbf{y})\right] \\ &= f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} P(\mathbf{x},\mathbf{y}). \qquad (2)\end{aligned}$$

Note that Bertrand competition implies that the frequency at which the worker collects offers, λ_1, does not affect $P(\mathbf{x}, \mathbf{y})$. Upon receiving an outside offer, the worker either stays in his initial match, in which case the continuation value for that match is $P(\mathbf{x}, \mathbf{y})$, or they accept the offer, in which case they extract a value of $P(\mathbf{x}, \mathbf{y})$ from the poacher (as a result of Bertrand competition) and leave their initial employer with a vacant job worth 0. Either way, the joint continuation value for the partners in the initial match equals $P(\mathbf{x}, \mathbf{y})$. This is a key implication of Bertrand competition between employers. From a social perspective, cases where the worker accepts the outside offer and moves to a match with higher value $P(\mathbf{x}, \mathbf{y}')$ are associated with a net surplus gain of $P(\mathbf{x}, \mathbf{y}') - P(\mathbf{x}, \mathbf{y})$. Yet none of the social gains associated with future job mobility are internalized by private agents, as those gains accrue to a third party (the worker's future employer).[12]

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

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

where the employer type is set to $\mathbf{y} = \mathbf{0}_L$ for an unemployed worker. For reasons similar to those just discussed about $P(\mathbf{x}, \mathbf{y})$, the worker fails to internalize the gain in surplus associated with 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 they value at $W(\mathbf{x}, \mathbf{y}, \sigma) = (1-\sigma)U(\mathbf{x}) + \sigma P(\mathbf{x}, \mathbf{y})$. The wage $w(\mathbf{x}, \mathbf{y}, \sigma)$ implementing that value solves:

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

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

[11] The dot ("$\cdot$") denotes the outer product, ∇ denotes the gradient, and $\nabla_{\mathbf{x}}$ denotes the gradient with respect to the subset $\mathbf{x}$ of the function's arguments.

[12] We discuss some of the consequences of this property in Section 6.4 and Web Appendix B.1.

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

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

The first term $\sigma f(\mathbf{x},\mathbf{y}) + (1-\sigma)\left[b(\mathbf{x}) + c(\mathbf{x},\mathbf{y})\right]$ 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 their forgone home production and disutility of work $c(\mathbf{x},\mathbf{y})$. The next (expectation) term reflects value of future outside offers, which the worker pays for by accepting a lower starting wage. The final term reflects the fact that an employed worker's skill bundle evolves towards the job's skill requirements $\mathbf{y}$, whereas those skills would erode towards $\mathbf{0}_L$ if the worker was unemployed. This, in general, benefits the worker in the event they become unemployed, and therefore affects the wage negatively.[13]

3.2 Model Analysis

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

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

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

The key functional form assumption is to assume a linear adjustment for skills. In particular, we assume that a worker's specialized (*i.e.* cognitive, manual, and interpersonal) skills adjust linearly to their job's skill requirements:

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

where the $\gamma_k^{u/o}$'s are all positive constants governing the speed at which worker skills adjust to a job's requirements. Note that we allow that speed to differ between upward and downward adjustments (γ_k^u vs γ_k^o for $k = C, M, I$, where "u" stands for "under-qualified" and "o" stands for "over-qualified"), and between skill types ($\gamma_C^{u/o}$ vs $\gamma_M^{u/o}$ vs $\gamma_I^{u/o}$). In this case a worker's skills relate to job tenure $s - t$ as follows:

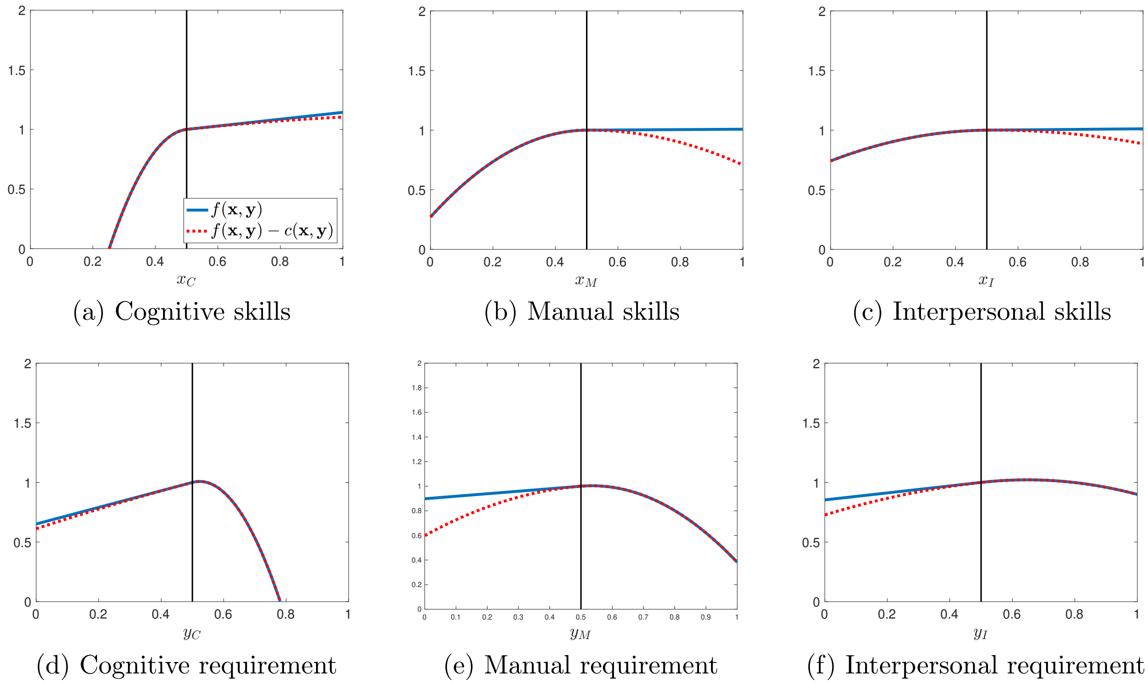
$$x_k(s) = y_k - e^{-\gamma_k^{u/o}(s-t)} \left(y_k - x_k(t)\right), \tag{7}$$

where the adjustment speed $\gamma_k^{u/o}$ that applies depends on whether $k = C, M$ or I and whether $x_k(t) \gtrless y_k$ (see Appendix A.1 for details). Over time a worker's specialized skills will adjust to the requirements of the job. Finally, a worker's general efficiency simply grows at a constant rate: $x_T(t) = x_T(0) \times e^{gt}$, independently of the worker's cognitive/manual skills or of the worker's employment status. This simple specification will help the model capture the wage/experience trend observed in the data.

The production function is then specified as follows:

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

In this specification of $f(\mathbf{x},\mathbf{y})$, the linear terms $\alpha_k y_k$ capture the possibility that jobs with different requirements in any of cognitive, manual or interpersonal skills have inherently different productivity levels, regardless of the worker they are matched with. The next three 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^u's are non-negative). We allow for the output loss caused by skill mismatch to differ depending on which skills the worker is deficient in. The last three terms, $\alpha_{kk} x_k y_k$, $k = C, M, I$, are added to provide additional flexibility in the modeling of complementarity between worker skills and job skill



(a) Cognitive skills (b) Manual skills (c) Interpersonal skills

(d) Cognitive requirement (e) Manual requirement (f) Interpersonal requirement

Figure 1: The production function

Note: $f(x, y)$ and $c(x, y)$ are plotted using the estimated parameter values from Table 4 below.

requirements within all three skill categories. In particular, assuming that $\alpha_{kk} \geq 0$, those last terms imply that an over-qualified worker (with $x_k > y_k$) produces more output than a worker with $x_k = y_k$, whose skills are an exact match for the job's requirements in skill dimension k. Finally, general efficiency, x_T, merely scales output up or down, conditional on $\mathbf{x}$ and $\mathbf{y}$. In the estimation we will allow x_T to be correlated with initial skill bundles $\mathbf{x}_0$, allowing for the possibility that workers with, for example, high cognitive skills also have high general skills.

Figure 1 provides a visual impression of the way in which the production function $f(\mathbf{x}, \mathbf{y})$ varies with its various inputs. The solid lines on the six panels on Figure 1 show the response of $f(\mathbf{x}, \mathbf{y})$ to variations in each of the three components of $\mathbf{x}$ (top row) and of $\mathbf{y}$ (bottom row), holding all other components of $\mathbf{x}$ and $\mathbf{y}$ fixed at 0.5. All graphs are put on a common scale, with $f(\mathbf{x}, \mathbf{y})$ normalized to 1 at the point where $\mathbf{x} = \mathbf{y}$, and constructed using our parameter estimates, presented below.

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

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

According to this specification, disutility of work is only positive if the worker is over-qualified for their job in some skill dimension. We interpret this as a utility cost of being under-matched. Note that our specification of $c(\mathbf{x}, \mathbf{y})$ is formally identical to the one we use to capture the output loss caused by a *shortage* of worker cognitive, manual, or interpersonal skills compared to the job's requirements in the production function (8). The utility cost of being under-matched further allows for an *excess* of skills to cause a loss of match value, albeit without causing a loss of output. An appealing implication of this specification is that over-qualified workers will have to be compensated for that utility cost, and will therefore have to be paid more in a given job than workers whose skills exactly match the job's requirements. Again, a visual impression of the flow surplus from a match, $f(\mathbf{x}, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})$, is is given by the dotted lines on Figure 1, which are constructed in the same way as the corresponding production function lines discussed above. The relatively small vertical distance between the solid and dotted lines in regions of the graphs where the worker is overqualified suggests that the utility cost of over-qualification is quantitatively relatively small - a point to which we will return when we discuss estimation results.

Finally, for simplicity, we specify unemployment income as depending on general skill only, $b(\mathbf{x}) = bx_T$, with b a positive constant, so that $U(\mathbf{x}) = bx_T/(r + \mu - g)$ is independent of the specialized skills (x_C, x_M, x_I).

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

$$
\begin{aligned}
P(\mathbf{x}(t), \mathbf{y}) - U(\mathbf{x}(t)) = x_T(t) \times \Bigg\{ & \frac{\alpha_T + \sum_{k=C,M,I} (\alpha_k y_k + \alpha_{kk} y_k^2) - b}{r + \delta + \mu - g} \\
& - \sum_{k=C,M,I} \left(\frac{\kappa_k^u \min\{x_k(t) - y_k, 0\}^2}{r + \delta + \mu - g + 2\gamma_k^u} + \frac{\kappa_k^o \max\{x_k(t) - y_k, 0\}^2}{r + \delta + \mu - g + 2\gamma_k^o} \right) \\
& + \sum_{k=C,M,I} \alpha_{kk} y_k \times \left(\frac{\min\{x_k(t) - y_k, 0\}}{r + \delta + \mu - g + \gamma_k^u} + \frac{\max\{x_k(t) - y_k, 0\}}{r + \delta + \mu - g + \gamma_k^o} \right) \Bigg\}. \qquad (10)
\end{aligned}
$$

The first term in the equation defining match surplus $P(\mathbf{x}(t), \mathbf{y}) - U(\mathbf{x}(t))$ is the 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 net 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, and on their degree of complementarity between worker skills and job skill requirements. But it also depends on the speed of skill adjustment: if adjustment is instantaneous ($\gamma_k^{u/o} \to +\infty$), the cost of mismatch becomes negligible.

4 Data

Our estimation sample is a panel of worker-level data from the 1979 *National Longitudinal Survey of Youth* (NLSY79) combined with occupation-level data on skill requirements from the O*NET program (`www.onetcenter.org`). We describe both data sets and the way we combine them before turning to a description of the estimation sample itself.[14] Additional details can be found in Appendix A.2.

4.1 Construction of the Estimation Sample

Data sources. The NLSY79 is well known and requires little description. Our extract from that data set is a weekly unbalanced panel of workers whom we follow from first entry into the labor market. For each worker in the panel, time is set to zero at the first week they cease to be in full-time education. We focus on males from the main sample who were never in the military,[15] and retain all individual histories until the first occurrence of a non-employment spell of 18 months or more: we consider individuals experiencing such a long spell of non-employment as losing their attachment to the labor force, which we treat as attrition from the sample. We retain information on labor force status and transitions, weekly earnings, occupation of current job (Census codes), education (highest grade completed), performance in a battery of ten aptitude tests called the *Armed Services Vocational Aptitude Battery* (ASVAB), measures of anti-social behavior, measures of health, and scores in two psychometric tests measuring social skills. Education, ASVAB scores, and measures of social skills and health will be used as measures of the initial skill bundles $\mathbf{x}_0$ 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. 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.[16] 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,

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

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

[16] 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`.

knowledge, work activities, work context, experience/education levels required, job interests, work values, and work styles. O*NET ratings come from two different sources: a survey of workers, who are asked to rate their own occupation in terms of a subset of the O*NET descriptors, and a survey of "occupation analysts" who are asked to rate other descriptors in the O*NET set.

We retain descriptors from the skills, abilities, knowledge, work activities, and work context O*NET files, as descriptors contained in the other files (job interests, work values, and work styles) are less directly interpretable in terms of skill requirements, and merge those files with our NLSY sample, based on occupation codes.[17,18]

Job skill requirements. Our selection from the O*NET database leaves us with over 200 different descriptors, which we take as measures of the underlying skill requirements. We reduce this large set of descriptors to three dimensions, which we interpret as "cognitive", "manual", and "interpersonal" skill requirements, using the following procedure.[19] First, we run Principal Component Analysis (PCA) on our large set of O*NET measures and keep the first three principal components. We then recover our cognitive, manual, and interpersonal skill requirement indices by recombining those three principal components (which by default are constructed to be orthonormal) in such a way that they satisfy the following three exclusion restrictions: (1) the *mathematics* score only reflects cognitive skill requirements; (2) the *mechanical knowledge* score only reflects manual skill requirements; (3) the *social perceptiveness* score only reflects interpersonal skill requirements). Interpretation of the three skill requirement indices thus obtained as cognitive, manual and interpersonal therefore relies on those exclusion restrictions. Finally, we rescale our skill requirement indices so that they lie in $[0, 1]$.[20]

[17]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.

[18]In our data sample we define a job change as a change of employer. In some cases, individuals are observed changing occupations *within an employer spell*. The typical pattern in those cases is that workers are oscillating between two or three "similar-looking" occupations (for example, an individual in our sample has a spell of employment starting with 10 weeks as an Electrical Engineer, continuing with 132 weeks as an Aerospace Engineer, followed by 50 weeks as a Mechanical Engineer, and ending with 9 weeks as an Aerospace Engineer again). This is strongly suggestive of classification error. To address that, we average the skill requirements of those occupations within an employer spell.

[19]Technical details of the construction of our job skill requirement and worker skill bundles are given in Appendix A.3.

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

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

4.2 Empirical content of skill and skill requirement bundles

In this sub-section we present reduced form evidence that our measures of worker skills and job skill requirements have predictive power for wages in the current job. We also present evidence that our measure of a worker's skills, deviation from their job's skill requirements, and duration in their first job have predictive power for the skill requirements of the job they move to next. We also compare our measure to some common alternatives from the literature.

Skills, skill requirements, and wages. Our measures of occupational skill requirements and initial worker skill bundles are not particularly sensitive to our choice of exclusion restrictions (see Appendix A.4 for details on robustness). Additionally, because our skill requirement measures make maximal use of the variation contained in the O*NET, they out perform the abstract-, manual- and routine-task measures proposed by Autor and Dorn

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

Table 1: Empirical content of skill and task measures

log wage	(1)	(2)	(3)	(4)	(5)	(6)
x_{C0}	0.339	0.483	0.316	-0.054	0.318	-0.076
	(0.115)	(0.117)	(0.115)	(0.152)	(0.129)	(0.153)
x_{M0}	-0.088	-0.093	-0.070	0.064	-0.122	0.024
	(0.087)	(0.089)	(0.087)	(0.168)	(0.099)	(0.169)
x_{I0}	0.268	0.311	0.262	0.234	0.394	0.305
	(0.053)	(0.054)	(0.053)	(0.101)	(0.095)	(0.141)
$\tilde{y}_C$	0.657		0.704	0.048		0.197
	(0.071)		(0.082)	(0.165)		(0.185)
$\tilde{y}_M$	0.259		0.170	0.418		0.402
	(0.058)		(0.066)	(0.165)		(0.178)
$\tilde{y}_I$	0.389		0.285	0.411		0.361
	(0.063)		(0.066)	(0.142)		(0.147)
$\tau^{AD}_{abstract}$		0.483	0.121		0.191	-0.052
		(0.032)	(0.034)		(0.104)	(0.115)
τ^{AD}_{manual}		0.251	0.194		-0.017	-0.291
		(0.048)	(0.050)		(0.177)	(0.179)
$\tau^{AD}_{routine}$		0.171	-0.012		0.273	0.071
		(0.028)	(0.030)		(0.088)	(0.089)
$x_{C0} \times \tilde{y}_C$				0.902		0.691
				(0.221)		(0.250)
$x_{M0} \times \tilde{y}_M$				-0.172		-0.267
				(0.249)		(0.266)
$x_{I0} \times \tilde{y}_I$				0.084		0.026
				(0.233)		(0.239)
$x_{C0} \times \tau^{AD}_{abstract}$					0.466	0.264
					(0.158)	(0.173)
$x_{M0} \times \tau^{AD}_{manual}$					0.419	0.730
					(0.282)	(0.285)
$x_{I0} \times \tau^{AD}_{routine}$					-0.199	-0.130
					(0.162)	(0.163)
tenure	0.238	0.239	0.234	0.235	0.240	0.233
	(0.025)	(0.026)	(0.025)	(0.025)	(0.026)	(0.025)
experience	0.268	0.299	0.266	0.269	0.300	0.267
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
years of education	0.270	0.336	0.274	0.271	0.329	0.266
	(0.082)	(0.084)	(0.081)	(0.081)	(0.083)	(0.080)
constant	4.436	4.524	4.448	4.552	4.600	4.562
	(0.094)	(0.095)	(0.094)	(0.145)	(0.110)	(0.149)
N	224,417	224,417	224,417	224,417	224,417	224,417
adjusted R^2	0.373	0.339	0.375	0.376	0.340	0.378

Standard errors clustered at the individual level.

τ^{AD} are the task measures from Autor and Dorn (2013).

(2013, AD) in terms of explanatory power in a descriptive wage regression. This can be seen in Table 1, which presents the results from regressions of log wages on various sets of skill and skill requirement measures. The regressors include our measures of initial cognitive, manual and interpersonal skills (x_{C0}, x_{M0}, x_{I0}), our measures of cognitive-, manual- and interpersonal-skill requirements for the current occupation ($\tilde{y}_C$, $\tilde{y}_M$, $\tilde{y}_I$), and the AD measures of the abstract-, manual- and routine-task intensity of the current occupation ($\tau^{AD}_{abstract}$,τ^{AD}_{manual}, $\tau^{AD}_{routine}$). The adjusted R^2 when including our skill requirement measures is higher than when including the AD task measures (0.37 compared to 0.34). When both sets of measures are included in the regression, the adjusted R^2 barely increases relative to the specification using only our measures, and many of the coefficient estimates on the AD task measures become small and statistically insignificant, while our measures continue to have the same statistically significant coefficients as when entered alone. These results hold both when we include only level effects and interactions between initial worker skills and job skill requirement (columns 1–3 compared to 4–6). We interpret the results from this descriptive regression as indicating that our measures contain predictive power for wages and that they contain all the information in the AD task measures, and more.[22]

In our model we represent an occupation by a vector of three skill requirements: cognitive, manual and interpersonal. This is a parsimonious representation with a natural measure of how similar occupations are to each other. We represent how good an individual is at an occupation by how close their own skill vector is to the occupation skill vector. This representation provides, for each worker, a complete ranking of how good they are at each occupation, and assumes they will be similarly good at occupations that have similar skill requirements. An alternative would be to directly represent how good each worker is at each occupation, without imposing any structure on similar occupations (see, for example, Kambourov and Manovskii, 2009a,b). This is very much in the spirit of a generalized Roy model. This would be more flexible from a purely empirical vantage point, but substantially more demanding to model, as it would imply that the worker's state space has at least the dimension of the number of occupations. Additionally, it would require the identification and estimation of the latent stochastic processes for each occupation specific skill, and the correlation structure. Authors who have pursued this modeling strategy have typically restricted attention to a small number (3 to 5) of very coarse occupation groups such as, white, blue and pink collar, service, professional, etc, and assumed independence of ability across occupations (see, for example, Keane and Wolpin, 1997; Lee and Wolpin, 2006; Sullivan, 2010).

[22]We are able to create consistent occupational classifications for both our measures and the AD measures for 97% of our observations.

Table 2: Occupation and Individual Fixed Effects

log wage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
x_{C0}	-0.036	0.567	-0.130	0.449	-0.144					
	(0.153)	(0.116)	(0.127)	(0.105)	(0.121)					
x_{M0}	0.014	-0.153	-0.150	-0.124	-0.065					
	(0.169)	(0.090)	(0.107)	(0.082)	(0.110)					
x_{I0}	0.232	0.311	0.033	0.276	0.105					
	(0.101)	(0.055)	(0.067)	(0.049)	(0.069)					
$\tilde{y}_C$	0.041					-0.532				
	(0.164)					(0.154)				
$\tilde{y}_M$	0.365					0.561				
	(0.171)					(0.154)				
$\tilde{y}_I$	0.395					0.388				
	(0.143)					(0.148)				
$x_{C0} \times \tilde{y}_C$	0.921		1.161		1.114	1.356		0.731		0.752
	(0.221)		(0.102)		(0.123)	(0.228)		(0.117)		(0.116)
$x_{M0} \times \tilde{y}_M$	-0.109		0.202		0.076	-0.279		0.279		0.170
	(0.254)		(0.091)		(0.110)	(0.237)		(0.085)		(0.088)
$x_{I0} \times \tilde{y}_I$	0.095		0.556		0.350	-0.144		0.304		0.183
	(0.233)		(0.101)		(0.124)	(0.257)		(0.109)		(0.112)
tenure	0.234	0.261	0.242	0.232	0.232	0.142	0.121	0.138	0.115	0.134
	(0.025)	(0.025)	(0.024)	(0.023)	(0.023)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)
experience	0.269	0.289	0.264	0.257	0.244	0.335	0.363	0.334	0.343	0.322
	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
years of education	0.256	0.321	0.294	0.306	0.289					
	(0.081)	(0.085)	(0.080)	(0.075)	(0.073)					
constant	4.603	4.237	4.440	4.579	4.751	5.297	5.303	4.991	5.548	5.332
	(0.148)	(0.194)	(0.200)	(0.221)	(0.248)	(0.058)	(0.173)	(0.185)	(0.130)	(0.151)
occupation FE 1 digit		✓	✓				✓	✓		
occupation FE 3 digit				✓	✓				✓	✓
worker FE						✓	✓	✓	✓	✓
N	232,303	232,303	232,303	232,303	232,303	232,303	232,303	232,303	232,303	232,303
adjusted R^2	0.374	0.347	0.388	0.430	0.448	0.682	0.677	0.684	0.697	0.701

Standard errors clustered at the individual level.

In Table 2 we consider the empirical content of our skill measures relative to occupation fixed effects, again in terms of the descriptive wage regression. Column 1 regresses log wages on our vector of skill measures for workers, the skill requirements of their occupation and the interactions. These coefficients are all used as moments in our estimation.[23] In column 2 we drop our occupation skill demand measures and replace them with occupation fixed effects at the one-digit level[24]. In column 3 we include the occupation fixed effects and the interactions between our worker skill measures and our occupation skill requirement measures. There are several things to note. First, the adjusted R^2 is higher when we use our measures than when we use the one-digit occupation fixed effects. Second, adding both the interaction of worker skills with occupation skill requirements and occupation fixed effect improves the fit substantially over the fixed effects regression and only marginally over the skill requirement regression. Additionally, the coefficients on the interaction terms are all significantly different from zero, even when occupation fixed effects are included.

In terms of explaining wage variation, our representation with three worker skills and three skill requirements provides a better fit than including one-digit occupation fixed effects. One-digit occupations are arguably too coarse. Columns 4 and 5 replicate the regressions in columns 2 and 3 with three-digit occupation fixed effects. In this case the adjusted R^2 from the fixed effects regression is higher than using only our skill measures, but the coefficient on the interaction of workers cognitive skills and the cognitive skill requirements of an occupation are still highly statistically significant. There is clearly additional content in our measure of skills and skills requirements. That said, a Roy model of occupation choice at the three digit level would require over 700 additional state variables, each with a latent stochastic process to be estimated. This is well beyond what we consider feasible. An alternative would be to consider a Lucas and Prescott (1974) type islands model (see Kambourov and Manovskii, 2009a). This has some attraction, but comes at the cost of giving up any notion of workers being more likely to move to similar occupations.

In columns 6–10 we repeat this exercise including worker fixed effects. Our structural model has unobserved worker heterogeneity that affects the wage multiplicatively, so the fixed effect regression is coherent here too. The main takeaway from these regressions is that even after we include worker fixed effects and three-digit occupation fixed effects, the coefficients on the interaction of worker skills and occupation skill requirements are statisti-

[23] The coefficient estimates differ slightly from 1 column 4 due to the fact that we loose some observations in 1 when matching our occupations to the AD task measures.

[24] These are the 1970 Census codes used by the NLSY79, corresponding to the 12 categories: Professional, Technical, and Kindred Workers; Managers and Administrators, except Farm; Sales Workers; Clerical and Unskilled Workers; Craftsmen and Kindred Workers; Operatives, except Transport; Transport Equipment Operatives; Laborers, except Farm; Farmers and Farm Managers; Farm Laborers and Farm Foremen; Service Workers, except Private Household; Private Household Workers.

cally significant, with the exception of interpersonal skills (we will return to this point when interpreting the structural estimates in Section 6.2).

Skill gaps and worker mobility. In Table 3 we present reduced form evidence on how skill gaps affect mobility between a worker's first and second jobs. What we mean by "skill gap" is the gap between measured worker skills and measured job requirements in a worker's first job. We examine how the first-job skill gap, interacted with the amount of time the worker spends in that first job, affects the type of job the worker moves to next. Columns 1–3 present results from a set of regressions where the dependent variable is the cognitive, manual, or interpersonal skill requirement at a worker's second job (denoted $\tilde{y}_k^+$). The right hand side variables are the same in each regression and consist of the worker's initial skill bundle; the squared difference between worker's initial skills and the skill requirements in the worker's first job, interacted with an indicator for whether the worker is under- or over-qualified (*i.e.* the skill gaps in all three skill dimensions); the duration of the first job spell; and the duration of the first job spell interacted with each skill gap. We exclude from this sample workers who have a period of unemployment between the measurement of $\mathbf{x}_0$ and their first job.

The coefficients of interest in each regression are on the interactions between the skill gaps in all three skill dimensions interacted with the duration of the first job spell. This tells us how the average skill-k requirement at the second job is affected by the duration of exposure to the same skill requirement at the first job. Take, for example, two workers who have identical initial skill vectors $\mathbf{x}_0$, who both have a first job with identical skill demands $\tilde{\mathbf{y}}$. Assume that these workers are under-qualified for their initial job, $\tilde{y}_C - x_{C0} > 0$. The positive coefficient on duration$\times$ $(\tilde{y}_C - x_{C0})$ in column 1 implies that between two identical workers who have identical first jobs for which they are under-qualified on the cognitive skill dimension, the worker who is exposed to the first job for a longer duration will, on average, move to a second job with a higher cognitive skill demand than the shorter duration worker. Similarly, if the workers are over-qualified in the cognitive dimension in their first job, the one who is exposed to this job for a longer duration will, on average, move to a second job with a lower cognitive skill requirement than the short duration worker. Under- or over-qualification on the manual or interpersonal skills in the first job has no effect on the expected cognitive skill demand in the second job. We see the same pattern for manual and interpersonal skills in columns 2 and 3, with the exception that the cognitive and manual gaps also influence the interpersonal skill requirements of the second job, although with opposite signs.[25]

[25] Groes, Kircher, and Manovskii (2015) find that the probability of occupational mobility is U-shaped in terms of a worker's occupation-specific wage ranking. In columns 4, 5 and 6 we control for a workers' occupation-specific wage decile (see Appendix A.2.3 for details). This has essentially no effect on the results

Table 3: Effect of quality and duration of first job on quality of second job

	(1) $\tilde{y}_C^+$	(2) $\tilde{y}_M^+$	(3) $\tilde{y}_I^+$	(4) $\tilde{y}_C^+$	(5) $\tilde{y}_M^+$	(6) $\tilde{y}_I^+$
x_{C0}	0.650	-0.300	0.460	0.659	-0.303	0.472
	(0.062)	(0.074)	(0.061)	(0.062)	(0.074)	(0.061)
x_{M0}	-0.117	0.687	-0.409	-0.124	0.677	-0.401
	(0.062)	(0.074)	(0.061)	(0.063)	(0.075)	(0.062)
x_{I0}	0.054	0.013	0.395	0.062	0.032	0.385
	(0.065)	(0.077)	(0.064)	(0.065)	(0.078)	(0.064)
$\max\{\tilde{y}_C - x_{C0}, 0\}^2$	3.044	0.998	1.102	3.321	0.932	1.379
	(0.694)	(0.827)	(0.686)	(0.696)	(0.836)	(0.690)
$\min\{\tilde{y}_C - x_{C0}, 0\}^2$	-0.677	-0.164	-0.096	-0.678	-0.168	-0.098
	(0.106)	(0.126)	(0.104)	(0.105)	(0.126)	(0.104)
$\max\{\tilde{y}_M - x_{M0}, 0\}^2$	-0.171	0.682	-0.450	-0.230	0.630	-0.484
	(0.227)	(0.270)	(0.224)	(0.228)	(0.274)	(0.226)
$\min\{\tilde{y}_M - x_{M0}, 0\}^2$	0.226	-0.420	0.190	0.213	-0.431	0.178
	(0.123)	(0.146)	(0.121)	(0.123)	(0.148)	(0.122)
$\max\{\tilde{y}_I - x_{I0}, 0\}^2$	-0.049	0.011	0.980	-0.058	0.008	0.981
	(0.312)	(0.371)	(0.308)	(0.312)	(0.375)	(0.309)
$\min\{\tilde{y}_I - x_{I0}, 0\}^2$	0.104	0.026	-0.399	0.121	0.019	-0.381
	(0.109)	(0.129)	(0.107)	(0.109)	(0.130)	(0.108)
duration	0.014	-0.001	0.017	0.016	-0.001	0.018
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
duration$\times(\tilde{y}_C - x_{C0})$	0.050	-0.038	0.035	0.050	-0.036	0.036
	(0.020)	(0.023)	(0.019)	(0.020)	(0.024)	(0.020)
duration$\times(\tilde{y}_M - x_{M0})$	-0.003	0.078	-0.025	-0.006	0.078	-0.028
	(0.016)	(0.019)	(0.016)	(0.016)	(0.019)	(0.016)
duration$\times(\tilde{y}_I - x_{I0})$	0.002	-0.001	0.031	0.002	0.001	0.029
	(0.013)	(0.015)	(0.012)	(0.013)	(0.015)	(0.012)
constant	0.091	0.327	0.161	0.083	0.332	0.159
	(0.040)	(0.047)	(0.039)	(0.043)	(0.052)	(0.043)
controls for occupation-specific wage decile				✓	✓	✓
N	528	528	528	528	528	528
adjusted R^2	0.376	0.276	0.497	0.385	0.274	0.502

Standard errors in parentheses

Taking stock. There are clear patterns in the raw data: the interaction between worker skills and skill requirements affect wages, and how long a worker spends at an initial job where she/he is under- or over-qualified affects the set of jobs she will move to. We next turn to estimation of the structural model to provide a coherent interpretation of these patterns.

5 Estimation

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

5.1 Simulation

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

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

for the effect of the quality and duration of a match and mobility.

and that of their outside offer $P(\mathbf{x}_{it}, \mathbf{y}')$, either accepts the offer (in which case their job attribute vector gets updated to $\mathbf{y}_{it} = \mathbf{y}'$), or stays in their job, with or without a contract renegotiation. In each case, the worker's period-t wage $w_{i,t}$ is updated according to equation (5). Symmetrically, we let any unemployed worker draw a job offer (probability λ_0) with job attributes $\mathbf{y}' \sim \Upsilon(\cdot)$, which the worker accepts if and only if $P(\mathbf{x}_{it}, \mathbf{y}') \geq U(\mathbf{x}_{it})$. Again, the worker's wage is updated.[26]

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 non-employment rate reaches a value of 35% (the observed non-employment rate on labor market entry in our NLSY sample), and take the current state of the sample at that point as the initial condition. This pre-sampling period is necessary to generate a latent outside option for each worker who is already employed at the time of the survey.

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

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

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 our combination of O*NET and NLSY data as explained in the previous section. The joint distribution of initial cognitive, manual, and interpersonal worker skills $(x_C(0), x_M(0), x_I(0))$ is fully observed in the data, and requires no parameterization. General worker efficiency grows along with potential experience t at a constant rate g. In addition, we allow it to be correlated in an unrestricted way with initial cognitive, manual and interpersonal skills $(x_C(0), x_M(0), x_I(0))$, as well as education:

$$x_T(t) = \exp\left(g \cdot t + \zeta_S \cdot \text{YEARS\_OF\_EDUCATION} + \zeta_C x_C(0) + \zeta_M x_M(0) + \zeta_I x_I(0) + \varepsilon_0\right), \tag{11}$$

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

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

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

Finally, we specify the skill requirements (y_C, y_M, y_I) as simple transforms of the skill requirement indices $(\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ constructed from the O*NET data as described in Section 4. This is to allow for the possibility that our constructed $\widetilde{\mathbf{y}}$'s might not be exactly aligned with our measured worker skills $\mathbf{x}$ (see the discussion in Section 4). Specifically, we assume that $y_k = \widetilde{y}_k^{\xi_k}$, with $\xi_k > 0$, thus ensuring that y_k is an increasing transformation of $\widetilde{y}_k$ that stays in the unit interval. We then approximate the joint sampling distributions of job attributes $\Upsilon(\mathbf{y})$ using a Gaussian copula and Beta marginals with skill-specific parameters (η_k^1, η_k^2), $k = C, M, I$. The rank correlation parameters $(\rho_{CM}, \rho_{CI}, \rho_{MI})$ of the Gaussian copula are to be estimated, together with the parameters of the three marginals.

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 Table A.1 in the Appendix. 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.[27] With this subset of parameters estimated – or calibrated – in a preliminary step, we are left with a 39-dimensional parameter vector to estimate (summarized in Table A.1 in the Appendix). We estimate these parameters by matching the following set of moments: *(i)* sample mean U2E rate, *(ii)* mean E2E rate profile (summarized by average E2E rates over six consecutive equal-length subsets of the observation window), *(iii)* mean and standard deviation of the marginal cross-sectional distributions of current job attributes $\widetilde{\mathbf{y}}_{it}$ among employed workers at a selection of sampling dates,[28] *(iv)* pairwise correlations of skill requirements, $\text{corr}(\widetilde{y}_{k,it}, \widetilde{y}_{k',it})$, $k' \neq k$, $(k, k') \in \{C, M, I\}^2$ among jobs held by employed workers at a selection of sampling dates

[27]Figure A.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.

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

(v) correlations of initial worker cognitive, manual and interpersonal skills and the skill requirements of jobs held, $\text{corr}(x_{k,i0}, \widetilde{y}_{k,it})$, $k = C, M, I$ at a selection of sampling dates *(vi)* coefficients of a regression of log wages $\ln w_{it}$ on initial skills $\mathbf{x}_{i0}$, current job attributes $\widetilde{\mathbf{y}}_{it}$, interactions $x_{k,i0} \times y_{k,it}$ for $k = C, M, I$, tenure, experience, and years of education (*i.e.* the regression in Table 2 column 1). We drop the first simulated wage out of unemployment for the wage regression.[29] The model-based moments are computed from simulated samples of $N = 18,400$ workers – ten replicas of the initial NLSY cross-section.

5.3 Identification

Appendix A.5 formally discusses identification of the model laid out in Section 3 (given the parameterization described in 3.2) from a data set with the structure and contents described in Section 4. In this Subsection we summarize the main sources of information that identify the various components of our model.

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

Although the exact arguments used in Appendix A.5 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

[29]In this version of the sequential auction model, in which workers are risk-neutral and have no bargaining power, workers tend to accept very low wages upon exiting unemployment, to "buy their way" onto the job ladder. As soon as a worker receives her/his first outside offer the wage will jump. We drop the initial wage out of unemployment so as not to bias our estimate of human capital accumulation due to the large wage change at the very beginning of an employment spell. We return to this issue in Section B.1.

particular, the cross-section wage regression coefficients that we seek to replicate contain the information needed to identify the parameters α_k, α_{kk}, ζ_k, 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 descriptive wage regression in Table 1 provides some reduced form evidence of which skills requirements are most important in production, and illustrates where some of the identification of the production function is coming from. In column 1, log wage is regressed on our measures of initial worker skill bundles and the skill requirements of the job they are currently in. When entered linearly, both cognitive and interpersonal skills of the workers affect the wage positively, as do the cognitive, manual and interpersonal skill requirements of the job. A worker's initial level of manual skills does not have a statistically significant effect on wages. In column 4 we add to this regression interaction terms between a worker's measured skill and the skill requirement of the job ($x_{k0} \times \tilde{y}_k$), for $k \in \{C, M, I\}$. An informative pattern emerges. The coefficient on the interaction of a worker's initial cognitive skills and the cognitive skill requirement of the job is strongly positive, while the coefficients on x_{C0} and $\tilde{y}_C$ entered separately become small and statistically insignificant. Interestingly, the coefficient on the interaction of worker skills and job skill requirements for manual and for interpersonal dimensions are small and statistically insignificant, and the coefficient on these skills and skill requirements entered individually remain essentially unchanged from column 1. In terms of identification of our model, these moments will imply a production function in which employing a worker who meets the cognitive skill requirements for a job is very important, much more so than meeting the manual or interpersonal skill requirements.

Identification of the human capital accumulation process for each skill type comes from how past jobs affect future opportunities. We presented some direct evidence on this in Table 3. Workers who are employed at a job in which they are initially under-qualified in a particular skill dimension move, on average, to jobs that require even higher levels of this skill requirement, and this is increasing in duration of the first match. We see clearly that the initial gap between a worker's skill and the job skill requirements, interacted with duration of the first job affects the set of jobs the worker accepts in the future. This dependence of accepted jobs in the future on early career skill gaps is informative on the extent to which workers learn different skills depending on the skill requirements of the job they are in, and how long they stay at the job.

In practice, our chosen moments ensure precise "local" identification of the model's parameters, in the sense that the distance between data-based model-predicted moments has a clear local minimum at the estimated parameter value.[30]

Several caveats on identification are worth mentioning here. First, our model implies that unobserved heterogeneity ε_0 affects wages, but does not affect a worker's decision to accept or reject a job offer. To the extent that workers do select jobs based on unobserved characteristics, our model will tend to overestimate the extent of mismatch and overstate the implied role of search frictions. Second, identification is parametric. For example, it requires that the shape of the production function estimated at the boundary of the acceptance set is informative about the shape of the production function in the interior.

6 Results

6.1 Model Fit

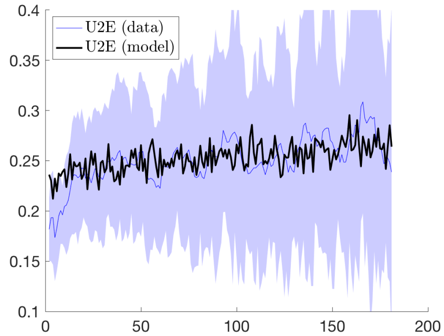
Figure 2 illustrates various aspects of the fit. All time series in Figure 2 are plotted over a period of 15 years (180 months, *i.e.* the sample window used for estimation), together with 95% confidence bands (based on 1,500 bootstrap replications) around the data series. Figure 2h further shows the fit in terms of the descriptive wage regression discussed in Section 4.

The model fits both the non-employment exit rate (U2E, Figure 2a) and the job-to-job transition rate (E2E, Figure 2b) reasonably well, considering the restriction to constant job contact rates λ_0 and λ_1 over the life cycle. The decline of E2E rates with experience is correctly captured by the model (it occurs as a consequence of workers gradually settling into jobs to which their skills are better suited, both because they sort into better matches over time and because their skills adjust to whatever job they are in at any given time), even though the model overstates both the initial speed of that decline and the level of the E2E rate at high levels of experience. All of the discrepancies between data and model in Figures 2a,b are largely due to our restriction to experience-invariant contact rates, λ_0 and λ_1.

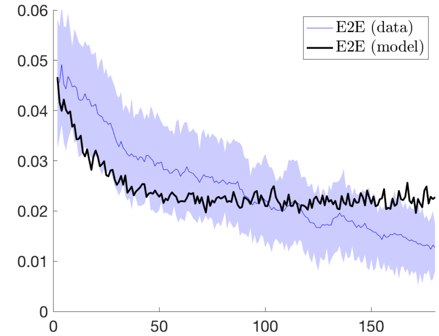
The sample average wage/experience profile is shown in Figure 2c. We do not directly target that particular profile in the estimation, yet the model captures it reasonably well, despite a tendency to overstate its concavity.

Figures 2d through g 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 gets the levels and experience profiles of those cross-sectional moments roughly right. In particular, the model captures the rise in average cognitive and interpersonal

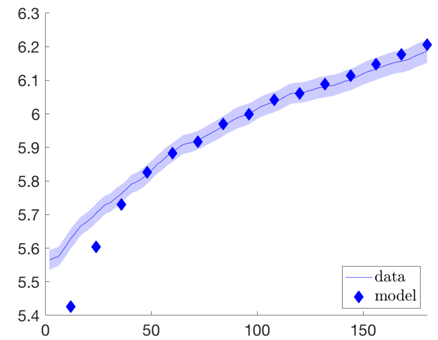
[30] We present additional measures of the local sensitivity of the parameter estimates to the data moment in Web Appendix B.2.



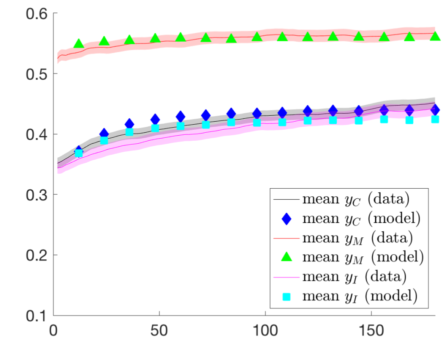
(a) U2E rate



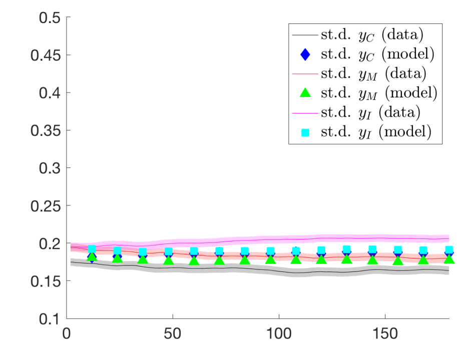
(b) E2E rate



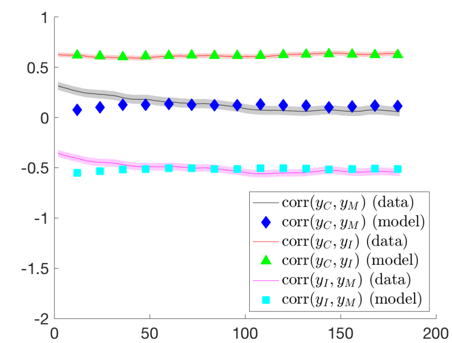
(c) Log wage/experience profile



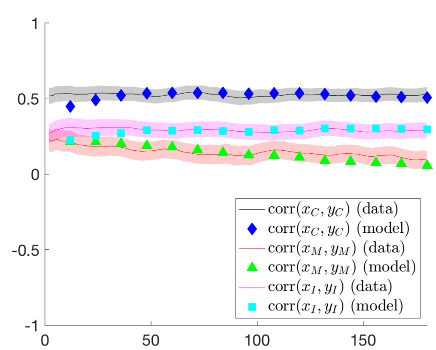
(d) Cross-sectional mean job attributes



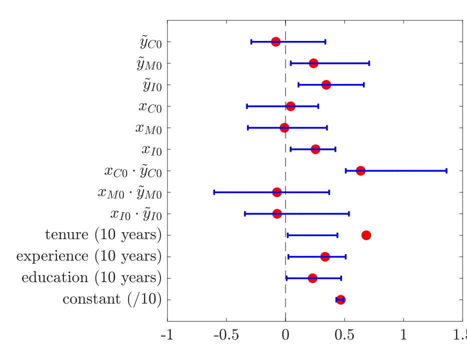
(e) Cross-sectional st.d. of job attributes



(f) Correlation of job attributes



(g) Corr. of job and worker attributes



(h) Descriptive (log) wage regression

Figure 2: Model fit

job attributes, although it tends to overstate the concavity of the experience profiles of those two job attributes at low levels of experience. It also replicates the near constancy of average manual job attributes. The model further captures the experience profiles of correlations between the various job attributes (Figure 2f), and between job requirements in each skill category and the corresponding initial worker skill (Figures 2g), again despite slight discrepancies between data and simulation at low levels of experience. Finally, while the model correctly predicts a flat experience profile for the standard deviations of all three job attributes, it does a little less well fitting the levels of those standard deviations, overstating the standard deviations of cognitive skill requirements and understating that of interpersonal skill requirements by about 10% at all levels of experience.

We finally turn to the model's ability to replicate the pooled cross-section wage regression shown in Table 2 column 1. Figure 2h shows, for each coefficient of the regression, the bootstrap 95% confidence interval of the empirical estimate, together with the structural model-based point estimate (the large dots). The structural estimates are all well within the empirical confidence bounds, with the exception of the returns to job tenure which the model has a tendency to overstate.

6.2 Parameter Estimates

Table 4 shows point estimates of the model parameters with asymptotic standard errors in parentheses below each estimate.[31] There is little to say about the offer arrival and job destruction rates, which are within the range of standard estimates on US data, even though the ratio $\lambda_1/\lambda_0 \simeq 0.42$ is on the high end of that range. Overall job productivity is increasing in all cognitive, manual and interpersonal skill requirements, with the loading on cognitive skills between 1.5 and two times as large as the ones on manual and interpersonal skills. Job skill requirements are complementary to the corresponding worker skills (α_{CC}, α_{MM} and α_{II} are all positive), although complementarity is an order of magnitude stronger in the cognitive than in the other two skill dimensions.

Overall worker efficiency x_T is positively associated with a high initial endowment in cognitive and interpersonal skills ($\zeta_C, \zeta_I > 0$), while initial manual skills are negatively correlated with x_T ($\zeta_M < 0$). One additional year of education increases efficiency by 2.4 percent (ζ_S). However, one should bear in mind that education is positively correlated with initial cognitive and interpersonal skills and (weakly) negatively correlated with initial

[31] The covariance matrix of the parameter vector Θ is estimated as $\left(G^\top G\right)^{-1} G^\top \Omega G \left(G^\top G\right)^{-1}$, where $G = \mathbb{E}\left[\partial \mathbf{m}/\partial \Theta^\top\right]$, the (expectation of the) Jacobian matrix of the moment function $\mathbf{m}(\Theta)$, is obtained by numerical differentiation and where Ω, the covariance matrix of the moment function, is estimated by resampling the data 1,500 times.

Table 4: Parameter estimates

production function*										disutility of work*			un. inc.
α_T	α_C	α_M	α_I	α_{CC}	α_{MM}	α_{II}	κ_C^u	κ_M^u	κ_I^u	κ_C^o	κ_M^o	κ_I^o	b
137.5	140.3	64.4	92.4	195.6	10.7	15.4	5,165.3	984.5	337.6	54.1	409.6	171.9	137.5
(17.0)	(24.8)	(10.1)	(11.1)	(24.2)	(2.40)	(4.29)	(533.5)	(265.4)	(97.4)	(7.14)	(71.9)	(23.9)	(17.0)
							(45.7)	*(3.9)*	*(4.1)*	*(20.9)*	*(8.8)*	*(5.1)*	

skill accumulation function**							general efficiency			
γ_C^u	γ_C^o	γ_M^u	γ_M^o	γ_I^u	γ_I^o	g	ζ_S	ζ_C	ζ_M	ζ_I
7.7e − 3	2.1e − 3	3.4e − 2	7.7e − 3	1.0e − 3	5.8e − 7	2.3e − 3	2.4e − 2	0.18	−0.17	0.20
(8.4e−4)	(5.3e−4)	(3.1e−3)	(6.4e−4)	(3.8e−3)	(1.4e−3)	(5.0e−4)	(.031)	(.501)	(.521)	(.261)
(7.54)	*(27.3)*	*(1.70)*	*(7.51)*	*(55.8)*	*(99,407)*					

sampling distribution***												transition rates		
ξ_C	ξ_M	ξ_I	ρ_{CM}	ρ_{CI}	ρ_{IM}	η_C^1	η_C^2	η_M^1	η_M^2	η_I^1	η_I^2	λ_0	λ_1	δ^{****}
1.21	0.79	0.88	0.14	0.73	−0.44	1.22	2.86	2.15	2.76	0.93	2.96	0.39	0.16	2.1e − 2
(.038)	(.038)	(.040)	(.019)	(.011)	(.019)	(.066)	(.095)	(.143)	(.117)	(.085)	(.124)	(.011)	(1.2e−3)	(3.3e−7)
			(0.13)	*(0.71)*	*(−0.42)*	*(0.30, 0.20)*		*(0.44, 0.20)*		*(0.24, 0.19)*				

*percent surplus loss caused by deviating from output-maximizing match by 1 SD of Υ at mean $\mathbf{x}$ in italics;

** half-life in years in italics ; *** implied correlations and (means, standard deviations) in italics ; **** estimated in first step manual skills in the sample. The value of ζ_S taken in isolation therefore understates the overall returns to education. Also note that the parameters $(\zeta_S, \zeta_C, \zeta_M, \zeta_I)$ are imprecisely estimated.

The employment of an under-qualified worker in any skill dimension is costly in terms of output, yet the output loss caused by this type of mismatch is by far most severe in the cognitive dimension and least severe in the interpersonal dimension. The utility cost of being under-matched – *i.e.* the surplus cost of the worker being *over*-qualified – is positive in all dimensions, but generally much smaller than the corresponding surplus (production) cost of *under*-qualification. To give a sense of the orders of magnitude involved, the numbers in italics below the estimates of the various κ's give the percentage flow-surplus cost of deviating by one standard deviation of the sampling distribution Υ from the output-maximizing match for a worker with the mean skill level in the distribution of initial skills $\mathbf{x}_0$.[32]

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

All three types of skills are accumulated faster (by an under-qualified worker) than they are lost (by an over-qualified worker), i.e. $\gamma_k^u > \gamma_k^o$ for $k = C, M, I$. Yet apart from that common property, patterns of skill adjustment differ vastly between skill dimensions. Manual skills adjust much faster than cognitive skills. Cognitive skills are very persistent (*i.e.* not easily accumulated or lost) with a half-life of 7.5 years to learn and 27.3 years to forget. The half-life of manual skills is much shorter, about 20 months to acquire and 7.5 years to lose. Interpersonal skills essentially do not adjust over a worker's typical horizon and can, to a good approximation, be treated as fixed worker traits, determined prior to labor market entry.

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

[32] Denoting the mean of N as $\mathbf{x}^m = (x_C^m, x_M^m, x_I^m)$, the output-maximizing match is with $\mathbf{y}^\star = \text{argmax}_{\mathbf{y}} f(\mathbf{x}^m, \mathbf{y})$. Then, denoting the standard deviations of the marginals of Υ as $(\sigma_C, \sigma_M, \sigma_I)$, the percentage reported in italics below the estimate of κ_C^u in Table 4 is $100 \times \left[1 - \frac{P(\mathbf{x}^m, \mathbf{y}^\star + (\sigma_C, 0, 0)) - U(\mathbf{x}^m)}{P(\mathbf{x}^m, \mathbf{y}^\star) - U(\mathbf{x}^m)}\right]$.

when the worker has lower skills than the job requires).

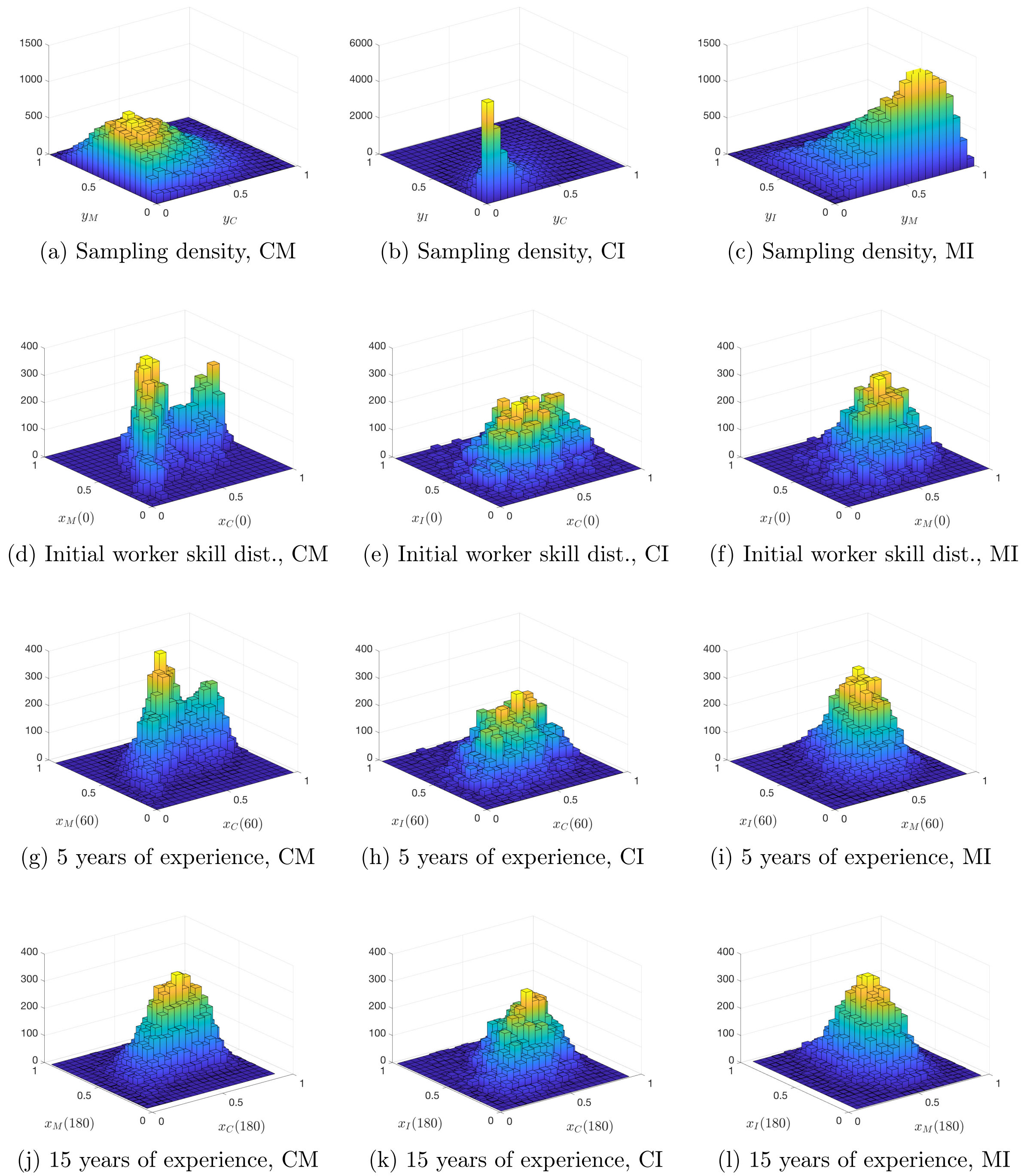
6.3 Skill Mismatch, Skill Changes, and Sorting

Distributions of skills and skill requirements. The broad question of skill mismatch can be understood in many different ways. One aspect of that question is the alignment (or lack thereof) between the skills that workers are equipped with when they leave education – the distribution $N(\cdot)$ of initial worker skill bundles $\mathbf{x}(0)$, in the parlance of the model – and the firms' skill requirements – the model counterpart of which is the sampling distribution $\Upsilon(\mathbf{y})$.

The top row of Figure 3 (Panels a, b and c) show the marginal sampling distributions of pairs of job attributes, integrating out one skill dimension at a time.[33] The second row of Figure 3 (Panels d, e and f) do the same for the distribution of initial skills among labor market entrants, $N(\cdot)$. Plots of the sampling distribution suggest that labor demand is concentrated around jobs with intermediate to high manual skill requirements (y_M around 0.5 to 0.6), and modest levels of cognitive and interpersonal skill requirements (y_C and y_I around 0.2). A visual comparison of the top two rows of Figure 3 further suggests that labor market entrants are, on average, endowed with levels of manual skills that roughly coincide with what the sampling distribution suggests employers are looking for, but also seem to have much higher levels of cognitive and interpersonal skills than is required in most jobs.

Of course, there are narrow limits to the amount of information conveyed by a visual comparison of the sampling and initial skill distributions. First, discrepancies between those two distributions are not entirely surprising: the population of workers whose skills are represented in Figures 3d-f is a cohort of relatively young workers. As such, their skills may not be representative of those in the entire active workforce. By contrast, at least under random search, the sampling distribution addresses all workers, the majority of which are from older cohorts among which the skill distribution may be quite different. Second, work skills and job attributes are combined into the surplus function (10), and what relative shapes of Υ and N constitute a "good fit" in the sense of surplus maximization is not visually obvious. For example, as discussed above, parameter estimates show that being over-qualified in the manual dimension is costly relative to being over-qualified in any of the other two skill dimensions, while being under-qualified in the cognitive dimension is very costly. The apparent excess of cognitive skills in the population of labor market entrants may be a sensible response to those features of the technology. We consider specific surplus-based measures of mismatch in the next section. Third, matching occurs gradually over time, with

[33]For example, Figure 3a is a plot of $\int v(y_C, y_M, y_I)dy_I$ against (y_C, y_M).



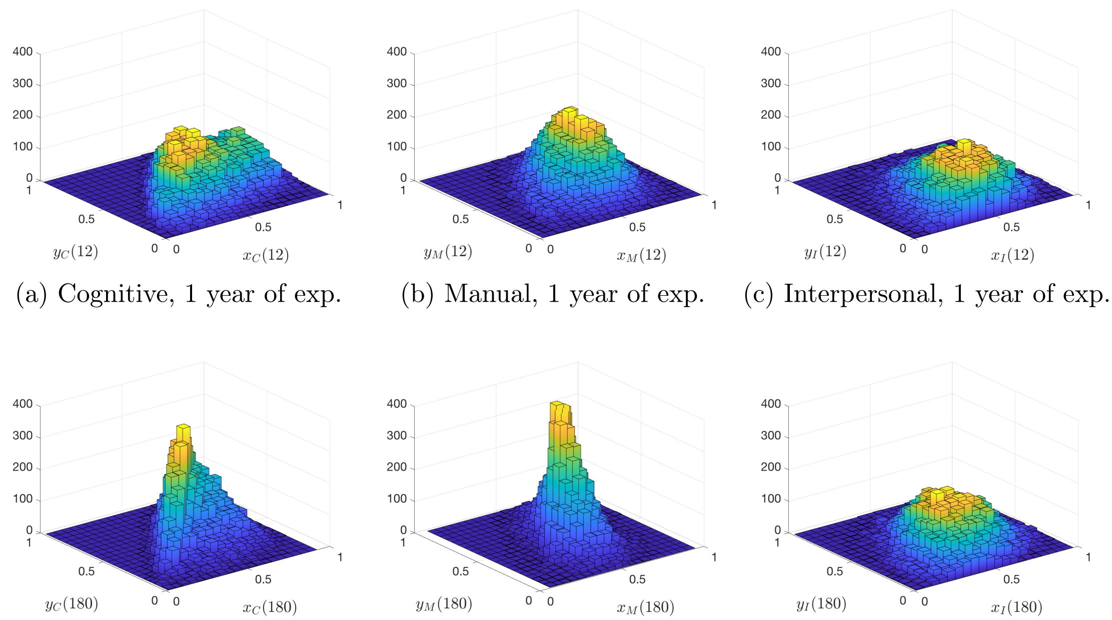
(a) Sampling density, CM (b) Sampling density, CI (c) Sampling density, MI

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

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

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

Figure 3: Distribution of skill requirements and evolution of worker skills with experience



(a) Cognitive, 1 year of exp. (b) Manual, 1 year of exp. (c) Interpersonal, 1 year of exp.

(d) Cognitive, 15 years of exp. (e) Manual, 15 years of exp. (f) Interpersonal, 15 years of exp.

Figure 4: Sorting

workers sampling repeatedly from Υ, while their skills evolve over time.

We now investigate this last point. Figures 3g–l show how the distribution of worker skills in the model changes as the cohort of workers accumulates experience. The evolution is clearly towards workers gaining cognitive skills and maintaining their manual skills on average (while, as we saw earlier, interpersonal skills hardly adjust at all 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 weights on y_C and $x_C \cdot y_C$ in the production function, α_C and α_{CC}, are an order of magnitude larger than their manual and interpersonal counterparts), inducing workers to accept jobs with the highest cognitive content compatible with their level of cognitive skills, while at the same time avoiding costly over-qualification in manual skills. By following this strategy workers tend to maintain or gradually acquire cognitive skills. A second striking feature of Figures 3 is that the already limited degree of specialization apparent in the initial skill distribution (Figure 3a) regresses even further as workers gain experience: the skill distribution becomes increasingly unimodal.

Skill sorting and mismatch. We next examine the joint distribution of worker skill bundles and job skill requirements among ongoing matches. Figure 4 shows two examples of those joint distributions, among workers who are one year into their careers (Panels a, b and c), and among workers with fifteen years of experience (Panels d, e and f). Simply eyeballing these histograms gives a distinct impression of positive sorting in all skill dimensions, even at early stages of the working life. Moreover, the "strength" of this positive sorting - as measured by the (inverse of the) conditional dispersion in worker skills for a given level of skill requirement - clearly increases as workers accumulate experience. This results from the combination of workers gradually sorting themselves into jobs for which their skills are better suited, and adjusting their initial skills to their job's requirements: as can be seen from Figure 4, sorting at 15 years of experience is strongest in the manual dimension (as manual skills adjust quickly), and weakest in the interpersonal dimension (as interpersonal skills do not adjust).

A final feature of Figure 4 is that, while there is largely positive sorting in all skill dimensions, a substantial mass of workers appear "under-matched" in the cognitive dimension, in the sense that their job's cognitive skill requirement is lower than their own cognitive skill level. By contrast, very few workers are "over-matched" in cognitive skills (and those who are are so by a small margin). The tradeoff from the perspective of a worker contemplating a job is between the job's overall productivity (the $\alpha_C y_C$ and $\alpha_{CC} x_C \cdot y_C$ terms in the production function), and any cost of being mismatched. In the case of cognitive skills, the cost of being "over-matched" (or "under-skilled"), measured by κ_C^u, is prohibitively high, even accounting for the fact that output increases much more steeply with y_C than with y_M or y_I (Table 4).

6.4 Worker Bargaining Power

Our baseline model relies on the sequential auction rent-sharing protocol of Postel-Vinay and Robin (2002), in which the only source of worker rent is Bertrand competition between employers. Upon receiving a job offer, a worker compares the total private values of his/her current labor market state (employment in a given type-$\mathbf{y}$ job or unemployment) with that of accepting the job offer (employment in an alternative type-$\mathbf{y}'$ job), and never extracts more than the minimum of those two values. This assumption has the implication that workers tend to accept very low wages upon exiting unemployment, to "buy their way" onto the job ladder, after which wages tend to increase very steeply as soon as they receive their first outside offer on their new job. The model thus tends to overstate returns to tenure, as well as returns to experience early on in a worker's career. The sequential auction protocol further implies that private firm-worker pairs fail to internalize the extra surplus that is created when the worker moves to a competing job for which her/his skills are better suited, as that extra surplus is entirely captured by a third party, namely the worker's future employer. For the same reason, unemployed workers fail to internalize the surplus associated with them finding a job. Both of those features of the sequential auction model can be mitigated by extending the model to give workers some additional bargaining power that enables them to capture a share $\beta \in [0,1]$ of the match rent over and above what they receive from sheer Bertrand competition between employers (Cahuc, Postel-Vinay, and Robin, 2006; Dey and Flinn, 2005). In Web Appendix B.1 we investigate this extension and show that, while endowing workers with positive bargaining power is theoretically straightforward in the context of our model, it has a very high cost in terms of computational tractability.

7 The Determinants of Social Output

We now analyze the determinants of the social value of output in our model economy. Specifically, we focus on the expected present discounted sum of future output produced by a worker from experience t onwards (the "experience-t expected career output"). Consider a worker i with experience t, who is either unemployed (denoted by $\ell_{it} = 0$) or employed ($\ell_{it} = 1$) in a job with attributes $\mathbf{y}_{it}$. The worker has education (years of education) ed_i, initial skill bundle $\mathbf{x}_{i0}$, unobserved ability ε_{0i} and current skills $\mathbf{x}_{it}$. Their experience-t expected career output is then defined as:

$$Q_{it} = \mathbf{E}\left[\int_t^{+\infty} \left(\ell_{is}\left[f\left(\mathbf{x}_{is}, \mathbf{y}_{is}\right) - c\left(\mathbf{x}_{is}, \mathbf{y}_{is}\right)\right] + \left(1-\ell_{is}\right) b\left(\mathbf{x}_{is}\right)\right) e^{-(r+\mu)(s-t)} ds \;\middle|\; \mathbf{x}_{i0}, \mathrm{ed}_i, \varepsilon_{0i}, \mathbf{x}_{it}, \ell_{it}, \mathbf{y}_{it}\right]. \quad (12)$$

Dispersion in initial skill bundles $\mathbf{x}_{i0}$, years of education ed_i, and unobserved ability ε_{0i}, along with the random career shocks faced by workers (job mobility, job destruction and attrition) up to experience t in our cohort of workers induces a cross-section distribution of experience-t career output within the cohort, for any experience level t. We now seek to assess the contributions of those different sources of randomness to the variance of (the natural logarithm of) Q_{it} for a fixed experience level t.[34]

To that end, note that, with our specification assumptions, all three functions $f(\cdot)$, $c(\cdot)$ and $b(\cdot)$ are multiplicatively separable in the worker's initial general efficiency, $x_{T,i0}$. This implies that Q_{it} can be decomposed as $\ln Q_{it} = \ln x_{T,i0} + \ln \widetilde{Q}_{it}$. Recalling the definition (11) of $x_{T,i0}$, we see that dispersion in $\ln x_{T,i0}$ is induced by dispersion in $\mathbf{x}_{i0}$, years of education ed_i, and ε_{0i}, while dispersion in $\ln \widetilde{Q}_{it}$ is induced by dispersion in $\mathbf{x}_{i0}$ and by random career shocks. Indeed, $\ln \widetilde{Q}_{it}$ is independent of years of education and ε_{0i} conditional on $\mathbf{x}_{i0}$. The variance of

[34]See Appendix A.7 for details on the computation of Q_{it}.

Table 5: Decomposition of $\mathbf{Var}\ \ln Q_{it}$

	Share of $\mathbf{Var}\ \ln Q_{it}$ due to...			
	initial skills $\mathbf{x}_0$ (term 1)	shocks (term 2)	heterogeneity ε_0 (term 3)	education $\mid \mathbf{x}_0$ (term 4)
Whole sample	65.0%	16.4%	18.9%	0.0%
College +	17.2%	48.3%	35.5%	0.0%
Some college	27.5%	34.2%	38.9%	0.0%
Non-college	37.9%	22.4%	40.1%	0.0%

Level of experience: $t = 10$ years.

career output $\mathbf{Var}\ \ln Q_{it}$ can therefore be decomposed in the following interpretable manner:

$$\begin{aligned}
\mathbf{Var}\ \ln Q_{it} &= \mathbf{Var}[\mathbf{E}(\ln Q_{it} \mid \mathbf{x}_{i0})] + \mathbf{E}\left[\mathbf{Var}(\ln Q_{it} \mid \mathbf{x}_{i0})\right] \\
&= \mathbf{Var}[\mathbf{E}(\ln Q_{it} \mid \mathbf{x}_{i0})] + \mathbf{E}[\mathbf{Var}(\ln \widetilde{Q}_{it} \mid \mathbf{x}_{i0})] + \mathbf{E}\left[\mathbf{Var}(\ln x_{T,i0} \mid \mathbf{x}_{i0})\right] \\
&= \underbrace{\mathbf{Var}[\mathbf{E}(\ln Q_{it} \mid \mathbf{x}_{i0})]}_{\text{1: between } \mathbf{x}_0} + \underbrace{\mathbf{E}[\mathbf{Var}(\ln \widetilde{Q}_{it} \mid \mathbf{x}_{i0})]}_{\text{2: shocks}} + \underbrace{\mathbf{Var}\ \varepsilon_{i0}}_{\text{3: heterogeneity}} + \underbrace{\zeta_S^2 \mathbf{E}[\mathbf{Var}(\text{ed}_i \mid \mathbf{x}_{i0})]}_{\text{4: schooling given } \mathbf{x}_0}
\end{aligned}$$

where the last equality uses (11) to further decompose the conditional variance of $\ln x_{T,i0}$. The first term is the between-$\mathbf{x}_0$ variance, i.e. the part of the cross-section variance in experience-t career output that is explained by dispersion in initial skill endowments, $\mathbf{x}_{i0}$. The second term is the part of the cross-section variance in Q_{it} that is attributable to the randomness in a worker's employment history. The third term is the variance attributable to dispersion in unobserved ability ε_{0i} (which, by construction of the model, is uncorrelated to any other source of dispersion). Finally, the fourth term is the residual variance, due to the dispersion in education that is not explained by dispersion in initial skills $\mathbf{x}_{i0}$. Results from this variance decomposition are gathered in Table 5 for $t = 10$ years of experience. For practical interpretability, we perform our variance decomposition on three sub-samples, stratified by education: College and above (16+ years of education, 30.7% of the sample), Some college (13-15 years of education. 19.3% of the sample), and Non-college (12 years of education or less, 50.0% of the sample).

For our entire cohort, almost two thirds (65%) of the variance is accounted for by the vector of initial skill endowments $\mathbf{x}_0$, and an additional 19% is explained by unobserved heterogeneity ε_0. The additional information brought about by education given $\mathbf{x}_0$ is negligible. Taken together, those numbers mean that about 84% of the variance in expected career output at 10 years of experience is explained by dispersion in initial conditions. The rest – a little over 16% – is due to the randomness associated with job offers and job displacement during the first 10 years in the labor market, and the differential skill accumulation that

Table 6: Further decomposition of $\mathbf{Var}\ \ln Q_{it}$

	Share of $\mathbf{Var}\ \ln Q_{it}$ due to ...			
	$\mathbf{x}_0$	x_{0C}	x_{0M}	x_{0I}
Whole sample	65.0%	58.9%	11.9%	19.3%
College +	17.2%	9.4%	2.8%	3.6%
Some college	27.5%	22.2%	10.6%	8.0%
Non-college	37.9%	33.9%	27.7%	10.5%

Note: Level of experience: $t = 10$ years. The share of the variance explained by the individual skills x_{0k} do not sum to the share explained by $\mathbf{x}_0$ due to the fact that (x_{0C}, x_{0M}, x_{0I}) are not mutually independent, and to non-linearities.

results from this randomness.[35] Conditioning on broad levels of education reduces the share of variance explained by initial skill bundles, as those are correlated with education.[36] Yet comparison of the bottom three rows of Table 5 suggests that the importance of labor market shocks increases with education, while, vice versa, the importance of initial conditions (the sum of $\mathbf{x}_0$ and ε_0) decreases with education.

To get a sense of which type(s) of skills generate the most variance in career output, Table 6 reports the shares of $\mathbf{Var}\ \ln Q_{it}$ explained by each separate skill dimension of $\mathbf{x}_0$ taken in isolation, $\mathbf{Var}\left[\mathbf{E}\left(\ln Q_{it} \mid x_{0k}\right)\right]$ for $k \in \{C, M, I\}$. We see that cognitive skills alone can account for almost 60% of the variance, much more than what either manual skills (12%) or interpersonal skills (19%) can account for on their own. Cognitive skills are the most important dimension of initial skills for all education groups, however their relative importance is much smaller for the low-educated: initial cognitive skills explain more than three times as much variance as initial manual skills for the top education group, but only account for 22% more variance than initial manual skills for the Non-college group.

Given the dominant role of cognitive skills illustrated in Table 6, a natural question to ask is: What would a model with only a single dimension for skill miss? Table 7 reports

[35]Keane and Wolpin (1997) provide a similar decomposition in which they conclude that 90% of the variation is due to initial conditions and 10% is due to shocks. The main difference with our approach is that they assume workers are always employed at their best job. From the point of view of our model, this would be equivalent to labelling all of the variation due to mismatch as due to initial skills. Cunha, Heckman, and Navarro (2005) and Cunha and Heckman (2016) use observations on individuals' decision of whether or not to attend college to help disentangle heterogeneity (known to individuals at the time of making the college attendance choice) from uncertainty (realized and observed later on in the life cycle). They estimate that 50% of the variation in lifetime earnings is due to heterogeneity. We consider lifetime earnings post schooling choice and estimate that between 65% and 84% is due to heterogeneity, depending on when in the life cycle we assume ε_0 to be observed by agents. Indeed, even though so far we have implicitly interpreted ε_0 as heterogeneity realized upon or before entry into the labor market, since ε_0 affects wages but does not affect any labor market decisions in our model, the exact timing of the realization of ε_0 is largely a matter of assumption. We would need to model additional choices, such as consumption/savings decisions, in order to parse out how much is heterogeneity known ex-ante *vs* how much is uncertainty realized later in the life cycle.

[36]See Web Appendix B.3 for a formal presentation of this statement.

Table 7: Decomposition of $\mathbf{Var}\ \ln Q_{it}$: one-dimensional model

	Share of $\mathbf{Var}\ \ln Q_{it}$ due to...			
	initial skills x_0 (term 1)	shocks (term 2)	heterogeneity ε_0 (term 3)	education $\mid x_0$ (term 4)
Whole sample	32.5%	3.94%	60.4%	3.16%
College +	10.6%	6.01%	81.3%	2.08%
Some college	28.0%	3.94%	67.6%	0.43%
Non-college	24.6%	3.73%	71.6%	0.13%

Note: Level of experience: $t = 10$ years.

the decomposition of $\mathbf{Var}\ \ln Q_{it}$ as obtained from a version of our model with a scalar (one-dimensional) measure of job skill requirements. Specifically, instead of allowing for a three-dimensional job attribute $\mathbf{y} = (y_C, y_M, y_I)$, we only allow for a scalar y, constructed as the first principal component of our initial large set of O*NET measures.[37] Symmetrically, we only allow for one dimension of specific worker skills x instead of the three-dimensional $\mathbf{x} = (x_C, x_M, x_I)$ considered so far. We estimate the resulting "one-dimensional" version of our model and produce the decomposition of $\mathbf{Var}\ \ln Q_{it}$ in exactly the same way as we did for our full, three-dimensional model.[38]

A comparison of Tables 7 and 5 immediately shows that, as might have been expected, the single-dimensional skill index model is capable of explaining substantially less of the variance of career output based on observables than our original model, which allows for three-dimensional skill bundles. This reduced explanatory power of the single-dimensional skill index manifests itself in two different ways. First, and most directly, the share of variance explained by observed initial worker skills (first column in Tables 7 and 5) tends to be smaller in the scalar case than in the three-dimensional case. Second, the share explained by labor market shocks (second column in Tables 7 and 5) is markedly smaller in the scalar case than in the three-dimensional case. Intuitively, two occupations that look similar in terms of the single-dimensional attribute y (which is constructed as the first principal component of our set of O*NET descriptors) may look rather different in terms of the three-dimensional bundle $\mathbf{y}$ (which is based on the first three principal components), along dimensions of $\mathbf{y}$ that are not captured by the single-dimensional index y.

[37]Note that, unlike in the three-dimensional case, where we further transformed the first three principal components based on exclusion restrictions for interpretability, here we keep the first principal component as it is produced by the PCA, and remain agnostic as to the label of the "skill" that this captured by this scalar measure. Details of this construction, and detailed estimation results for this one-dimensional version of our model are available on request.

[38]In the notation of Section 3, we re-do the entire estimation exercise replacing $L = 3$ by $L = 1$. This results in a reduction in the number of data moments to be calculated (for example, the covariance terms involving y_C and y_M disappear), as well as a reduction in the number of parameters to be estimated (for example, there are fewer parameters describing the production function and the distribution of y).

Table 8: Elasticities of Q_{it}

Elasticity of Q_{it} with respect to:	**Whole sample**	**College +**	**Some college**	**Non-college**
x_{0C}	0.31	0.33	0.32	0.28
x_{0M}	−0.11	−0.10	−0.13	−0.11
x_{0I}	0.07	0.07	0.06	0.06
$\mathbf{E}_\Upsilon y_C$	0.08	0.13	0.09	0.04
$\mathbf{E}_\Upsilon y_M$	0.07	0.06	0.07	0.07
$\mathbf{E}_\Upsilon y_I$	0.06	0.05	0.06	0.06
λ_0	0.02	0.03	0.02	0.02
λ_1	0.01	0.01	0.04	0.02
mismatch	0.26	0.31	0.26	0.22

Level of experience: $t = 10$ years.

What labor market shocks do is move workers between occupations with different productive attributes. If two occupations have a similar single-dimensional attribute y, but different three-dimensional bundles $\mathbf{y}$, then moving a worker between those two occupations will not cause a lot of variation in that worker's career output in the single-dimensional model, which fails to pick up the differences between the two occupations, but will do so in the three-dimensional world where the two occupations do look different. Similarly, two workers may look similar in terms on the single-dimensional attribute x, but very different along dimensions of $\mathbf{x}$ that are not captured by x. When these workers are confronted with the opportunity of a job with skill requirements summarized by y, the scalar model will not generate any wage variation. The multidimensional model will generate substantial wage variation between these workers if one is much better suited than the other in terms of the vector of requirements $\mathbf{y}$.

While the model with unidimensional skill underestimates the share of the variance explained by both initial skills and shocks during working life, it underestimates the effect due to shocks more. Comparing the first rows of Tables 5 and 7, we see that the model with multidimensional skills implies initial skills explain 3.96 times as much of the variance as career shocks when looking at the entire sample. The corresponding ratio for the unidimensional model is 8.25, reducing the relative role of shocks by more than half. This same relative reduction of the role for shocks occurs within each education group.

Next we compute the cohort-wide average elasticities of expected career output (at ten years of experience) with respect to the subset of model parameters listed in the first column of Table 8. The way in which we construct some of those elasticities requires some explanation. The first three rows of Table 8 aim to measure the impact on average career output of a marginal proportional change in initial workforce skills. What we mean by that is the following. For each separate skill dimension k, we can multiply the initial endowment x_{0k} of every worker in the sample by a common scale $1+\Delta$ and compute their career output $Q_{it}(\Delta)$ under this modified distribution of initial skills. The numbers reported in the first three rows of Table 8 are $\mathbf{E}\left[\partial \ln Q_{it}(\Delta)/\partial\Delta\right]$, evaluated at $\Delta=0$.

The following three rows turn to the labor demand side and assess the impact on average career output of a marginal change in the skill requirements demanded by employers. We measure that by the average elasticity of Q_{it} with respect to the mean skill requirement in each separate dimension.[39]

The next three rows are about the matching technology, starting with the (self-explanatory) elasticities of Q_{it} with respect to the frequencies of job contacts, λ_0 and λ_1, followed by a row that we label "mismatch". Numbers in this last row are based on counterfactual simulations where we reassign employed workers as follows. If a worker with current skills $\mathbf{x}$ is employed in a job with attributes $\mathbf{y}$ by the time they reach ten years of experience, we reassign them to a job of type $\mathbf{y}'=(1-\Delta)\mathbf{y}+\Delta\mathbf{y}^\star(\mathbf{x})$, where Δ is a small, positive number and where $\mathbf{y}^\star(\mathbf{x})$ is the output-maximizing job type for a worker with skills $\mathbf{x}$. We can compute that worker's career output $Q_{it}(\Delta)$ following this reassignment for any value of Δ. The numbers reported in the "mismatch" row of Table 8 $\mathbf{E}\left[\partial \ln Q_{it}(\Delta)/\partial\Delta\right]$, evaluated at $\Delta=0$.

Comparing the elasticities for the whole sample we see that the largest effect on output is associated with a marginal improvement in the cognitive skills of the entering workforce. Interestingly, a marginal increase in the interpersonal skill endowment of entrants only has a comparatively small effect and an increase in their manual skills would have a negative effect, the latter being a result of the small effect of manual skills on output and the large effect on disutility of work.

The (quantitatively close) second largest effect on output is associated with a marginal reduction in the mismatch between worker skill and job requirements. This one-time reduction in mismatch has a direct effect on increasing current output, but also an effect on future output through improved skill accumulation of workers. The output loss due to mismatch dwarves the losses associated with the sheer frequency of worker-firm contacts, λ_0 and λ_1, which our model estimates to be an order of magnitude smaller in comparison.

Finally, increasing the mean skill requirement in the sampling distribution has moderately positive effects on output, which are roughly equal across skill types.

Looking across the education groups the same broad pattern holds as in the full sample. Reductions in mismatch and increases in the cognitive ability of the entering workforce have

[39] The marginal sampling distribution of skill requirement k is parameterized as a Beta distribution with parameters $\left(\eta_k^1,\eta_k^2\right)$. Those parameters obviously map 1:1 into the mean and variance of that marginal distribution. To compute the elasticity of Q_{it} with respect to the mean sampled skill requirement in dimension k, $\mathbf{E}_\Upsilon y_k$, we vary $\left(\eta_k^1,\eta_k^2\right)$ to induce a marginal change in the mean, keeping the variance constant.

the largest effects on output. The magnitudes of both effects increase somewhat in education.

8 Conclusion

We extend a standard and well-tested search-theoretic model of individual careers to allow for multidimensional skills and on-the-job learning. We estimate the model using occupation-level measures of skill requirements based on O*NET data, combined with a worker-level panel (NLSY79). We use the estimated model to shed light on the origins and costs of mismatch along three dimensions of skills: cognitive, manual, and interpersonal.

Our main findings are the following. The model sees cognitive, manual and interpersonal skills as very different productive attributes. Manual skills have moderate returns and adjust quickly (*i.e.*, they are easily accumulated on the job, and relatively easily lost when left unused). Cognitive skills have much higher returns, but are much slower to adjust. The returns on interpersonal skills are similar to (slightly higher than) those on manual skills. Yet interpersonal skills are all but fixed over a worker's lifetime. Next, the cost of skill mismatch (modeled as the combination of an output loss and a loss of worker utility caused by skill mismatch) is very high for cognitive skills, an order of magnitude greater than for manual or interpersonal skills. Moreover, this cost is asymmetric: employing a worker who is under-qualified in cognitive skills (*i.e.* has a level of skills that falls short of the job's skill requirements) is several orders of magnitude more costly than employing an over-qualified worker. Those important differences between various skill dimensions are missed when subsuming worker productive heterogeneity into one single scalar index. This is highlighted when we decompose the variance of lifetime output between initial skills (observed and unobserved) and shocks over the career. The use of a unidimensional model of skill rather than a multidimensional model underestimates the explained share of the variance (of the present value of output) by half and, additionally, underestimates the contribution of shocks relative to initial skills by half.

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.

Anderson, A., and L. Smith (2010): "Dynamic matching and evolving reputations," *The Review of Economic Studies*, 77(1), 3–29.

Andrews, I., M. Gentzkow, and J. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132(4), 1553–1592.

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

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

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

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

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

Census Bureau (1989): "The Relationship Between the 1970 and the 1980 Industry and Occupation Classification Systems," Technical Paper 59.

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

Cunha, F., and J. Heckman (2016): "Decomposing Trends in Inequality in Earnings into Forecastable and Uncertain Components," *Journal of Labor Economics*, 34(2), S31–S65.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57, 191–261.

Dey, M. S., and C. J. Flinn (2005): "An Equilibrium Model of Health Insurance Provision and Wage Determination," *Econometrica*, 73(2), 571–627.

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

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

Gathmann, C., and U. Schonberg (2010): "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 28(1), 1–49.

Golan, L., and K. Antonovics (2012): "Experimentation and Job Choice," *Journal of Labor Economics*, 30(2), 333–66.

Groes, F., P. Kircher, and I. Manovskii (2015): "The U-Shapes of Occupational Mobility," *The Review of Economic Studies*, 82(2), 659–692.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2018): "Multidimensional Skill Mismatch," Discussion paper, forthcoming *American Economic Journal: Macroeconomics*.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85(1), 29–65.

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

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

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

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

Kambourov, G., and I. Manovskii (2009a): "Occupational Mobility and Wage Inequality," *The Review of Economic Studies*, 76(2), 731–759.

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

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

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

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

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

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

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

Lise, J., and J.-M. Robin (2017): "The Macrodynamics of Sorting between Workers and Firms," *American Economic Review*, 107(4), 1104–1135.

Lucas, R. E., and E. C. Prescott (1974): "Equilibrium Search and Unemployment," 7(2), 188–209.

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

Neal, D. (1999): "The Complexity of Job Mobility among Young Men," *Journal of Labor Economics*, 17(2), 237–61.

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

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

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

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

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

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

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

Sullivan, P. (2010): "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search," 51(1), 289–317.

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

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

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

A Appendix

Table A.1: Summary of Model Parameters

Structural object	Symbol	How is is estimated
Offer arrival rates:	(λ_0, λ_1)	*Estimated within the model*
Job destruction rate:	δ	*Pre-estimated as sample mean E2U rate*
Unemployment income:	b	*Estimated within the model*
Production function f:	$(\alpha_T, \alpha_C, \alpha_M, \alpha_I, \alpha_{CC}, \alpha_{MM}, \alpha_{II}, \kappa_C^u, \kappa_M^u, \kappa_I^u)$	*Estimated within the model*
Utility cost of over-qualification c:	$(\kappa_C^o, \kappa_M^o, \kappa_I^o)$	*Estimated within the model*
Skill accumulation function $\mathbf{g}$:	$(\gamma_C^u, \gamma_M^u, \gamma_I^u$ $\gamma_C^o, \gamma_M^o, \gamma_I^o, g)$	*Estimated within the model*
Joint distribution of initial worker skills:		*Directly observed from initial sample cross section*
General worker efficiency x_T:	$(\zeta_S, \zeta_C, \zeta_M, \zeta_I, \varepsilon_0)$	*Distribution of ε_0 estimated by deconvolution in final step*
Sampling distribution of job attributes Υ:	$(\xi_C, \xi_M, \xi_I, \rho_{CM}, \rho_{CI}, \rho_{MI}, \eta_C^1, \eta_C^2, \eta_M^1, \eta_M^2, \eta_I^1, \eta_I^2)$	*Estimated within the model, specified as Gaussian Copula with Beta marginals*
Attrition and discount rates:	(r, μ)	*Calibrated outside model*

A.1 Solving for the Value Functions.

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

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

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

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

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

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

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

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

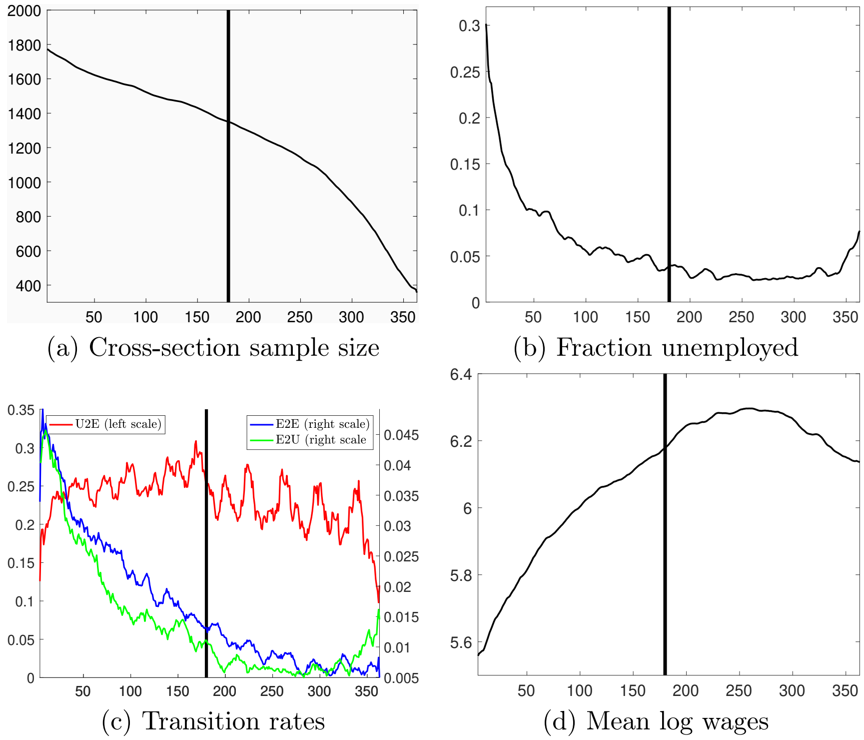
Using the functional forms assumptions (8) and (9) in the above formula produces (10).

A.2 Data Details

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

A.2.1 Flows, stocks, and wages over time.

Figure A.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.



(a) Cross-section sample size (b) Fraction unemployed

(c) Transition rates (d) Mean log wages

Figure A.1: Sample description

Figure A.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 observed for fewer years than less educated individuals. This causes the composition of the sample to shift toward less educated individuals as one approaches the end of the observation window. To circumvent this problem, we restrict our estimation sample to the first 15 years (180 months) of the initial sample. This 180-month cutoff is materialized by a thick vertical black line on all panels of Figure A.1.

Figure A.1b shows the non-employment 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 non-employment rate to reach this steady state (roughly ten years). Figure A.1c shows the rates of transition between labor market states. The non-employment exit rate is roughly stable at around 25 percent per month, while the transition rates from job to job and into non-employment decline smoothly over the sample window. Finally, Figure A.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 non-employment rate, they start declining, again a likely consequence of non-random attrition

Table A.2: Examples of skill requirement scores

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

Source: O*NET and authors' calculations

from the sample.

A.2.2 Worker skills and job skill requirements.

Table A.2 lists some examples of the cognitive, manual and interpersonal skill requirement scores we constructed for a few occupations. We denote those scores by $\widetilde{\mathbf{y}} = (\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ and will use them as empirical measures of the model's job attributes $\mathbf{y}$. Examples in Table A.2 include the occupations with the highest cognitive (Physicist), manual (Aircraft Mechanics and Service Technicians), and interpersonal (Preventive Medicine Physicians) skill requirements in the sample, and the occupations with the lowest cognitive (Graders and Sorters, Agricultural Products), manual (Telemarketers), and interpersonal (Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic) skill requirements.

The correlation pattern of workers' *initial* skills and the skill requirements of the *first jobs they are observed in* is described in Table A.3, where workers' initial cognitive, manual, and interpersonal skill indices are denoted by (x_{C0}, x_{M0}, x_{I0}) , while $(\widetilde{y}_C, \widetilde{y}_M, \widetilde{y}_I)$ refer to the empirical measures of job skill requirements in a worker's first job. This correlation pattern reveals several features of the data. First, (x_{C0}, x_{M0}, x_{I0}) are positively correlated in our cross-section of workers. Even though those correlation coefficients are far below one, they suggest that workers with high skills in one dimension tend to have high skills in the other two. Cognitive and manual skills appear slightly more strongly associated with each other than either is with interpersonal skills. Second, $\widetilde{y}_C$ is positively correlated with both $\widetilde{y}_M$ and $\widetilde{y}_I$ in the cross section of workers' first jobs. Even though there is obviously some selection here (as the set of jobs a worker will take up depends on their own skill bundle $\mathbf{x}$), this suggests that jobs requiring high levels of cognitive skills also tend to require high skill levels in one of the manual or interpersonal dimensions. While manual and interpersonal skill requirements are both positively correlated with cognitive skill requirements, they are negatively correlated with each other. Third, $(x_{C0}, \widetilde{y}_C)$, $(x_{M0}, \widetilde{y}_M)$, and $(x_{I0}, \widetilde{y}_I)$ are positively correlated (as expected), and so are $(x_{C0}, \widetilde{y}_I)$, $(x_{M0}, \widetilde{y}_C)$, $(x_{I0}, \widetilde{y}_M)$, $(x_{I0}, \widetilde{y}_C)$. By contrast, $(x_{C0}, \widetilde{y}_M)$ and $(x_{M0}, \widetilde{y}_I)$ are negatively correlated, suggesting that workers select themselves into either manual or non-manual jobs, as fits their skill bundles.

Table A.3: Correlation pattern of initial skills and skill requirements in first job

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

A.2.3 Occupation-specific Wage Deciles

In columns 4–6 of Table 3 we control for a worker's position within their occupation-specific wage distribution. Specifically, we include an indicator variable for which decile of the occupation-specific distribution each worker belongs to. We construct the cutoffs for the deciles using the 1980 Census. From the Census, we restrict the sample to include men, aged 16 to 65 who worked at least 26 weeks in the year. We convert weekly earnings to real 1982 dollars using the CPI to be comparable with the NLSY data. We map the 1980 occupation codes to the 1970 occupation using Table 1 from Census Bureau (1989).

A.3 Construction of Skill Measures

The data sets from which we construct worker skill and job skill requirement scores both consist of a set of P different measures observed for N individuals (workers in the case of the NLSY, and occupations in the case of O*NET). We denote the $N \times P$ matrix of all observations by $\mathbf{M}$. PCA decomposes the matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{FL}$, where $\mathbf{F}$ is the orthonormal $N \times P$ matrix of principal eigenvectors of $\mathbf{M}^\top \mathbf{M}$ and $\mathbf{L}$ is a $P \times P$ matrix of factor loadings. We consider the first 3 principal components only, *i.e.* we consider the decomposition $\mathbf{M} = \mathbf{F}_3\mathbf{L}_3 + \mathbf{U}$, where $\mathbf{F}_3$ is the $N \times 3$ matrix formed by taking the first 3 columns of $\mathbf{F}$ and $\mathbf{L}_3$ is the $3 \times P$ matrix formed by taking the first 3 rows of $\mathbf{L}$.

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

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

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

A.4 Robustness to Alternative Exclusion Restrictions in O*NET

In Table A.4 we present the correlation between our preferred skill requirement measures and skill requirement measures constructed using four alternative exclusion restrictions for each of the cognitive, manual and interpersonal dimensions. Specifically, we consider Mathematical Reasoning, Fluency of Ideas, Written Comprehension, and Oral Comprehension as alternative exclusion restrictions to Mathematics Knowledge to anchor the cognitive skill requirements. We consider Finger Dexterity, Repairing and Maintaining Mechanical Equipment, Arm-Hand Steadiness, and Manual Dexterity as alternatives to Mechanics Knowledge to anchor the manual skill requirements. We consider Selling or Influencing Others, Negotiation, Persuasion, and Speaking as alternative exclusion restrictions to Social Perceptiveness to anchor interpersonal skill requirements. The correlations are all very high, especially for the alternatives we consider for cognitive and interpersonal anchors. For the manual anchor, the correlation is very high for the Finger Dexterity and Repairing and Maintaining Mechanical Equipment alternatives, but somewhat weaker for both the Arm-hand Steadiness and Manual Dexterity alternatives. As a result it is important to keep in mind that our notion of manual skill requirements has more to do with the ability to fix machines than the ability to do very physical labor. This aligns well with our exclusion for the worker skill side in the NLSY79 which measures auto mechanics and machine shop knowledge.

Table A.4: Alternative Exclusion Restrictions in O*NET

Cognitive Skill Requirements	Mathematics Knowledge
Mathematical Reasoning	0.983
	(0.009)
Fluency of Ideas	0.930
	(0.017)
Written Comprehension	0.918
	(0.019)
Oral Comprehension	0.902
	(0.020)

Manual Skill Requirements	Mechanics Knowledge
Finger Dexterity	0.973
	(0.011)
Repairing and Maintaining Mechanical Equipment	0.940
	(0.016)
Arm-Hand Steadiness	0.717
	(0.033)
Manual Dexterity	0.657
	(0.036)

Interpersonal Skill Requirements	Social Perceptiveness
Selling or Influencing Others	0.997
	(0.004)
Negotiation	0.991
	(0.006)
Persuasion	0.979
	(0.010)
Speaking	0.965
	(0.012)

Notes: Correlation with alternative exclusion restrictions. Standard error in parenthesis.

A.5 Identification

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

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

The wage equation (5) can be written as:

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

A first important implication of (16) is that the maximum wage given $(\mathbf{x},\mathbf{y})$ is $f(\mathbf{x},\mathbf{y})$, implying in turn that the maximum wage given $\mathbf{y}$ is $f\left(\mathbf{y},\mathbf{y}\right) = x_T \cdot \left(\alpha_T + \sum_{k=C,M,I}(\alpha_k y_k + \alpha_{kk} x_k \cdot y_k)\right)$. Because $\mathbf{y}$ is observed for all employed workers,[40] and because x_T is also a function of observables (up to the uncorrelated heterogeneity term ε_0), namely the worker's education, initial skill bundle and experience, this proves identification of the parameters α_T, α_C, α_M, α_I, , α_{CC}, α_{MM}, α_{II}, g, ζ_S, ζ_C, ζ_M, ζ_I.

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

First, from the expression of the match surplus (10), one can show that joint observation of $\mathbf{x}$ and the set $\{\mathbf{y} : P(\mathbf{x},\mathbf{y}) = U\}$ allows separate identification of the parameters of $P(\mathbf{x},\mathbf{y})$, *i.e.* the composite parameters $\kappa_k^{u/o} / \left(r + \delta + \mu - g + 2\gamma_k^{u/o}\right)$, $k = C, M, I$.[41] Now, the issue is that we do not directly observe worker skills at all levels of experience: rather, we only observe workers' *initial* skill bundles. However, consider a worker with (observed) initial skill bundle $\mathbf{x}(0)$ starting his working life in unemployment, and who finds a job after an initial unemployment spell of duration $d^{(1)}$. From the human capital accumulation function (6), we know that this worker's skill bundle by the time s/he

[40] What is, in fact, observed, is not directly $\mathbf{y}$ but rather its empirical counterpart $\widetilde{\mathbf{y}}$. With our functional form assumptions on $f(\mathbf{y},\mathbf{y})$ 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.

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

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 composite parameter. This can be done by repeating the latter argument for workers who are initially employed in matches with skill requirements $\mathbf{y}^{(1)}$ for which they are *under*-qualified, *i.e.* such that $x_k(0) < y_k^{(1)}$ for $k = C, M, I$, become unemployed after an initial spell duration of $d^{(1)}$, then find a job again after an unemployment spell of duration $d^{(2)}$. From the human capital accumulation function (6), those workers' skill bundles when they find their second job (at experience $d^{(1)} + d^{(2)}$ is given by $x_k\left(d^{(1)} + d^{(2)}\right) = e^{-\gamma_k^o d^{(2)}}\left[y_k^{(1)} - e^{-\gamma_k^u d^{(1)}}\left(y_k^{(1)} - x_k(0)\right)\right]$. The only unknown parameter in this expression is γ_k^u, which is then again identified from the set $\{\mathbf{y} : P(\mathbf{x},\mathbf{y}) = U\}$.

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

Once the parameters of the human capital accumulation function $\mathbf{g}(\mathbf{x},\mathbf{y})$ are known, we can construct any worker's full path of skill bundles $\mathbf{x}$: consider a worker in his n^{th} spell (which could be a spell of unemployment). Denote the skill requirements in that spell by $\mathbf{y}^{(n)} = \left(y_C^{(n)}, y_M^{(n)}, y_I^{(n)}\right)$ (both equal to 0 if the spell is one of unemployment), the worker's skill bundle at the beginning of that spell by $\mathbf{x}^{(n)} = \left(x_C^{(n)}, x_M^{(n)}, x_I^{(n)}\right)$, and the duration of that spell by $d^{(n)}$. Spell duration $d^{(n)}$ and the vector $\mathbf{y}^{(n)}$ are observed in all spells, while $\mathbf{x}^{(n)}$ is only observed in the initial spell, $n = 1$, where it equals $\mathbf{x}(0)$. Then, using the skill accumulation equation $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x},\mathbf{y})$, we have that $\mathbf{x}^{(n+1)} = \mathbf{X}\left(d^{(n)}; \mathbf{y}^{(n+1)}, \mathbf{x}^{(n)}\right)$, where $\mathbf{X}(\cdot)$ denotes the solution to (13) as explained in the main text. Using backward substitution, we can then construct $\mathbf{x}^{(n)}$ for any spell as a function of the history of durations and skill requirements of past spells and the worker's initial skill level $\mathbf{x}^{(1)} = \mathbf{x}(0)$.

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

Finally, the offer arrival rates λ_0 and λ_1 are identified, conditionally on the rest of the model, from sample U2E and E2E transition probabilities, and the flow value of non-employment, $b(\mathbf{x})$, is identified from the wage of workers exiting non-employment: applying

(16) to workers just exiting non-employment ($\sigma = 0$) yields:

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

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

A.6 Unobserved Heterogeneity ε_0

The distribution of unobserved permanent worker heterogeneity ε_0 can be identified from the comparison of the simulated cross-section distribution of log wages *assuming away unobserved worker heterogeneity (i.e.* setting $\varepsilon_0 = 0$ for all workers) with the empirical distribution of log wages. For example, the variance of ε_0 can be estimated as the difference between the empirical wage variance and the variance of the simulated log-wage distribution without unobserved worker heterogeneity. We apply this strategy and compare the variances of the cross-section distributions of individual-level mean wages. By taking individual-level averages over time, we hope to minimize pollution of our estimate of $\mathrm{Var}\varepsilon_0$ by potential measurement error in the empirical wage data. The variance of $\ln \varepsilon_0$ thus estimated is 0.104 and accounts for 4.5% of the overall variance of individual-level mean wages.

A.7 Computation of Expected Career Output

The expected career output of a worker with current skills $\mathbf{x}$, employed in a job with requirements $\mathbf{y}$ defined in (12) can be equivalently defined in the following recursive way:

$$(r + \delta + \mu) Q(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta V(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} Q(\mathbf{x},\mathbf{y}) \\ + \lambda_1 \int \mathbf{1}\left\{P(\mathbf{x},\mathbf{y}') \geq P(\mathbf{x},\mathbf{y})\right\}\left[Q(\mathbf{x},\mathbf{y}') - Q(\mathbf{x},\mathbf{y})\right] d\Upsilon(\mathbf{y}'), \quad (17)$$

where the dependence of Q on fixed attributes ($\mathbf{x}_0$, education, ε_0) is omitted and where $V(\mathbf{x})$ is the expected career output of an unemployed worker with current skill bundle $\mathbf{x}$. The latter is in turn defined by:

$$(r + \mu) V(\mathbf{x}) = b(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{0}) \cdot \nabla V(\mathbf{x}) + \lambda_0 \int \mathbf{1}\left\{P(\mathbf{x},\mathbf{y}') \geq U(\mathbf{x})\right\}\left[Q(\mathbf{x},\mathbf{y}') - V(\mathbf{x})\right] d\Upsilon(\mathbf{y}'). \quad (18)$$

In those two formulas, $P(\mathbf{x},\mathbf{y})$ and $U(\mathbf{x})$ denote the equilibrium (private) values of, respectively, a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ job, and an unemployed type-$\mathbf{x}$ worker. In both 17 and 18, the last (expectation) term captures the expected continuation value of the worker's career in his future matches. The social values $Q(\mathbf{x},\mathbf{y})$ and $V(\mathbf{x})$ differ from the private values $P(\mathbf{x},\mathbf{y})$ and $U(\mathbf{x})$ precisely because of those expectation terms: under our rent-sharing protocol, existing firm-worker pairs fail to internalize the value generated by the worker in future matches because all of said value is captured by the worker's future employer when he switches jobs. While the economic interpretation of those expectation terms is clear enough, mathematically their impact is to add a non-linear term to the PDEs defining $Q(\mathbf{x},\mathbf{y})$ and $V(\mathbf{x})$, which rules out any closed-form solution. Those equations must therefore be solved numerically, using the following procedure.

We choose a grid $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}\times\{\mathbf{y}_1,\cdots,\mathbf{y}_m\}$ of n worker skill and m job skill requirement vectors. For any point $(\mathbf{x}_i,\mathbf{y}_j)$ on that grid, we approximate $Q(\mathbf{x}_i,\mathbf{y}_j)=\mathbf{\Pi}(\mathbf{x}_i)\cdot\boldsymbol{\phi}_Q(\mathbf{y}_j)$ and $V(\mathbf{x}_i)=\mathbf{\Pi}(\mathbf{x}_i)\cdot\boldsymbol{\phi}_V$, where $\mathbf{\Pi}(\cdot)$ is a basis of complete polynomials of some chosen order, and where $\boldsymbol{\phi}_V$ and $\boldsymbol{\phi}_Q(\cdot)$ are a set of $m+1$ vectors of coefficients that are computed by minimizing the distance between the left- and right-hand sides of (17) and (18) over the grid. Then, for a generic pair $(\mathbf{x},\mathbf{y})$ that is not on the grid, we use the approximations $V(\mathbf{x})=\mathbf{\Pi}(\mathbf{x})\cdot\boldsymbol{\phi}_V$ and $Q(\mathbf{x},\mathbf{y})=\mathbf{\Pi}(\mathbf{x})\cdot\widetilde{\boldsymbol{\phi}}_Q(\mathbf{y})$, where $\widetilde{\boldsymbol{\phi}}_Q(\mathbf{y})$ is a linear interpolation of $\boldsymbol{\phi}_Q(\widetilde{\mathbf{y}}_{j1}(\mathbf{y}))$ and $\boldsymbol{\phi}_Q(\widetilde{\mathbf{y}}_{j2}(\mathbf{y}))$, $\widetilde{\mathbf{y}}_{j1}(\mathbf{y})$ and $\widetilde{\mathbf{y}}_{j2}(\mathbf{y})$ being the nearest two neighbors of $\mathbf{y}$ on the grid.

Solving for $\boldsymbol{\phi}_V$ and $\boldsymbol{\phi}_Q(\cdot)$ involves repeated calculations of three-dimensional integrals like $\int\mathbf{1}\{P(\mathbf{x},\mathbf{y}')\geq P(\mathbf{x},\mathbf{y})\}[Q(\mathbf{x},\mathbf{y}')-Q(\mathbf{x},\mathbf{y})]\,d\Upsilon(\mathbf{y}')$, the computational cost of which increases very quickly as one increases the order of the polynomial basis $\mathbf{\Pi}(\cdot)$ or the size (n,m) of the grid of points upon which the approximation is based. Fortunately, we only need to perform those calculations a limited number of times. In practice, we choose $m=n=15$ and an approximation order of 3. We find that increasing the fineness of the approximation beyond those values makes little difference to the results.

B Web Appendix (not for publication)

B.1 Extension: Worker Bargaining Power

When workers have bargaining power $\beta \in [0,1]$, the dynamic equations (2) and (3) characterizing, respectively, the value of a match and the value of unemployment, must be amended as follows:

$$(r+\delta+\mu)P(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} P(\mathbf{x},\mathbf{y}) \\ + \lambda_1 \beta \mathbf{E} \max\{P(\mathbf{x},\mathbf{y}') - P(\mathbf{x},\mathbf{y}), 0\} \quad (19)$$

and:

$$(r+\mu)U(\mathbf{x}) = b(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{0}) \cdot \nabla U(\mathbf{x}) + \lambda_0 \beta \mathbf{E} \max\{P(\mathbf{x},\mathbf{y}') - U(\mathbf{x}), 0\}. \quad (20)$$

where, in both cases, the last (expectation) term captures the expected surplus share that the worker will extract from future matches thanks to her/his bargaining power.[42] The values defined by those two equations differ from the baseline case (which coincides with $\beta = 0$) precisely because of those expectation terms. While the economic interpretation of those expectation terms is clear enough, mathematically their impact is to add a non-linear term to the PDEs defining $P(\mathbf{x},\mathbf{y})$ and $U(\mathbf{x})$, which rules out any closed-form solution. Those equations must therefore be solved numerically, using the following procedure.

We choose a grid $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \times \{\mathbf{y}_1, \cdots, \mathbf{y}_m\}$ of n worker skill and m job skill requirement vectors. For any point $(\mathbf{x}_i, \mathbf{y}_j)$ on that grid, we approximate $P(\mathbf{x}_i, \mathbf{y}_j) = \mathbf{\Pi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}_P(\mathbf{y}_j)$ and $U(\mathbf{x}_i) = \mathbf{\Pi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}_U$, where $\mathbf{\Pi}(\cdot)$ is a basis of complete polynomials of some chosen order, and where $\boldsymbol{\phi}_U$ and $\boldsymbol{\phi}_P(\cdot)$ are a set of $m+1$ vectors of coefficients that are computed by minimizing the distance between the left- and right-hand sides of (19) and (20) over the grid. Then, for a generic pair $(\mathbf{x},\mathbf{y})$ that is not on the grid, we use the approximations $U(\mathbf{x}) = \mathbf{\Pi}(\mathbf{x}) \cdot \boldsymbol{\phi}_U$ and $P(\mathbf{x},\mathbf{y}) = \mathbf{\Pi}(\mathbf{x}) \cdot \widetilde{\boldsymbol{\phi}}_P(\mathbf{y})$, where $\widetilde{\boldsymbol{\phi}}_P(\mathbf{y})$ is a linear interpolation of $\boldsymbol{\phi}_P(\widetilde{\mathbf{y}}_{j1}(\mathbf{y}))$ and $\boldsymbol{\phi}_P(\widetilde{\mathbf{y}}_{j2}(\mathbf{y}))$, $\widetilde{\mathbf{y}}_{j1}(\mathbf{y})$ and $\widetilde{\mathbf{y}}_{j2}(\mathbf{y})$ being the nearest two neighbors of $\mathbf{y}$ on the grid.

Solving for $\boldsymbol{\phi}_U$ and $\boldsymbol{\phi}_P(\cdot)$ involves repeated calculations of $\mathbf{E}\max\{P(\mathbf{x},\mathbf{y}') - P(\mathbf{x},\mathbf{y}), 0\}$, a three-dimensional integral. Unfortunately, the computational cost of that simulation step quickly becomes prohibitive as one increases the order of the polynomial basis $\mathbf{\Pi}(\cdot)$ or the size (n,m) of the grid of points upon which the approximation is based. This forces us to

[42] With $\beta > 0$, worker-firm pairs thus partly internalize the surplus supplement from the worker's future matches, as the worker now captures a share β of that extra surplus. In the limit $\beta \to 1$, the extra surplus from the worker's future matches is fully internalized by current match partners, and the private match and unemployment values (19) and (20) coincide with the corresponding Planner's values.

limit ourselves to a very sparse grid (in practice, we choose $n = 10$ and $m = 6$) and a low approximation order (in practice: 3), resulting in a very coarse approximation of our two value functions.[43] The results reported in this section should therefore be taken with the appropriate amount of caution. Yet we think it useful to provide an indication of what is likely to change and what isn't, compared to the baseline case $\beta = 0$, when workers are endowed with positive bargaining power. To that end, we re-estimate our model under the assumption that $\beta = 0.5$, using the approximation procedure just described.

Estimated parameters under the assumption $\beta = 0.5$ are reported in Table B.1, alongside our baseline estimates, copied from Table 4 for comparison. Point estimates of the flow surplus parameters (the α's and the κ's) tend to be slightly smaller with $\beta = 0.5$ than in the baseline case $\beta = 0$. Smaller values of the α parameters can be explained by the fact that those parameters are mainly identified off of the levels of wages (see Sub-section 5.3 and Appendix A.5).[44] With positive bargaining power, workers appropriate an extra share of match productivity, which therefore needs to be estimated lower than in the $\beta = 0$ case in order to match the wages observed in the data. Those lower estimated α's have a knock-on effect on the estimated cost of mismatch (the κ's): the cost of mismatch must stay commensurate with the returns on job attributes to rationalize observed mobility patterns.

Having said that, those differences are statistically small: point estimate differences between the two models are generally well within two standard deviations of our baseline estimates (see Table 4). Moreover, the relative values of the various parameters are very close between the $\beta = 0.5$ and $\beta = 0$ cases. As a result, none of the implications discussed above in the context of our baseline $\beta = 0$ case are substantially changed.

Figure B.1 echoes Figure 2 and shows the main aspect of the model's fit in the $\beta = 0.5$ case. A visual comparison of Figures B.1 and 2 suggests that the fit of the $\beta = 0.5$ model is very similar to that of the baseline $\beta = 0$ model, with two main differences. First, the model with bargaining power tends better to capture the wage/experience profile, in particular at low levels of experience (Figure B.1c). As discussed before, this was expected, as positive bargaining power mitigates the tendency of unemployed workers to exit unemployment on very low entry wages by bringing wages closer to match productivity. Second, the model with

[43] Even with those coarse approximation settings, a single evaluation of the $\beta > 0$ version of the model takes about three times as long as a single evaluation of the baseline ($\beta = 0$) model, which involves no approximation. Moreover, estimation of the $\beta > 0$ model takes substantially more iterations to converge than estimation of the baseline model does. Although we cannot prove it, we suspect this is due to the extra noise caused by approximation error in the $\beta > 0$ case.

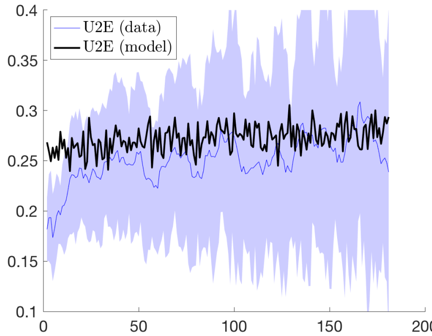
[44] In this version of the model, the wage equation is obtained, as before, by applying the rule $W(\mathbf{x},\mathbf{y},\sigma) = (1-\sigma)U(\mathbf{x}) + \sigma P(\mathbf{x},\mathbf{y})$, with the worker's value function now solving: $(r+\delta+\mu)W(\mathbf{x},\mathbf{y},\sigma) = w(\mathbf{x},\mathbf{y},\sigma) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + \mathbf{g}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} W(\mathbf{x},\mathbf{y},\sigma) + \lambda_1 \mathbf{E} \max\{0, \beta \max\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\} + (1-\beta)\min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\} - W(\mathbf{x},\mathbf{y},\sigma)\}$.

Table B.1: Parameter estimates (with worker bargaining power)

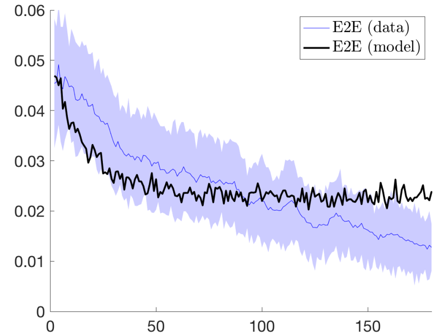
	production function										disutility of work			un. inc.
	α_T	α_C	α_M	α_I	α_{CC}	α_{MM}	α_{II}	κ_C^u	κ_M^u	κ_I^u	κ_C^o	κ_M^o	κ_I^o	b
β= 0	137.5	140.3	64.4	92.4	195.6	10.7	15.4	5,165.3	984.5	337.6	54.1	409.6	171.9	137.5
β= 0.5	122.3	116.2	60.7	91.4	280.7	9.7	14.5	3,901.4	641.3	304.5	53.2	376.1	142.5	122.4

	skill accumulation function							general efficiency			
	γ_C^u	γ_C^o	γ_M^u	γ_M^o	γ_I^u	γ_I^o	g	ζ_S	ζ_C	ζ_M	ζ_I
β= 0	$7.7\text{e}-3$	$2.1\text{e}-3$	$3.4\text{e}-2$	$7.7\text{e}-3$	$1.0\text{e}-3$	$5.8\text{e}-7$	$2.3\text{e}-3$	$2.4\text{e}-2$	0.18	−0.17	0.20
β= 0.5	$7.9\text{e}-3$	$2.1\text{e}-3$	$3.9\text{e}-2$	$8.1\text{e}-3$	$1.1\text{e}-3$	$6.1\text{e}-7$	$2.4\text{e}-3$	$2.4\text{e}-2$	−0.16	−0.22	0.20

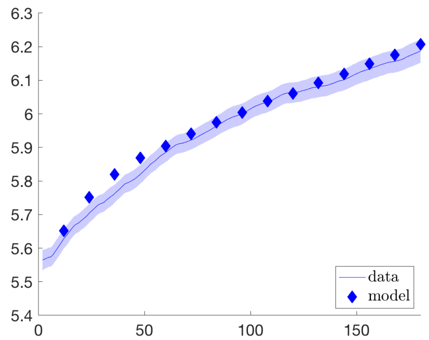
	sampling distribution												trans. rates	
	ξ_C	ξ_M	ξ_I	ρ_{CM}	ρ_{CI}	ρ_{IM}	η_C^1	η_C^2	η_M^1	η_M^2	η_I^1	η_I^2	λ_0	λ_1
β= 0	1.21	0.79	0.88	0.14	0.73	−0.44	1.22	2.86	2.15	2.76	0.93	2.96	0.39	0.16
β= 0.5	1.08	0.76	0.82	0.12	0.72	−0.47	1.23	3.07	2.12	2.90	0.92	3.11	0.41	0.17



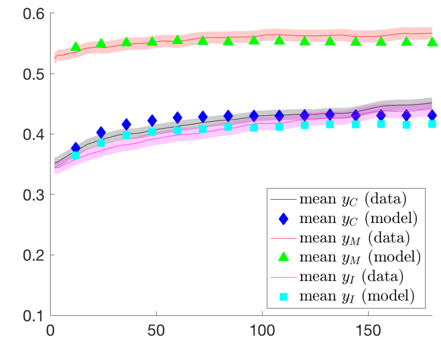
(a) U2E rate



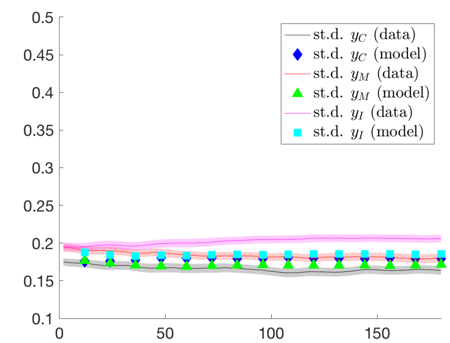
(b) E2E rate



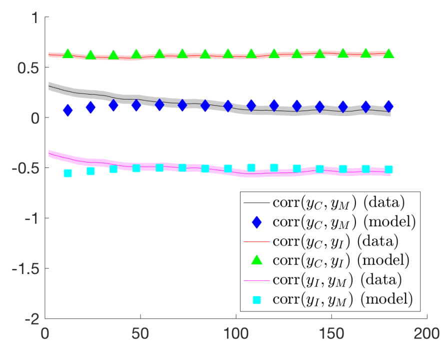
(c) Log wage/experience profile



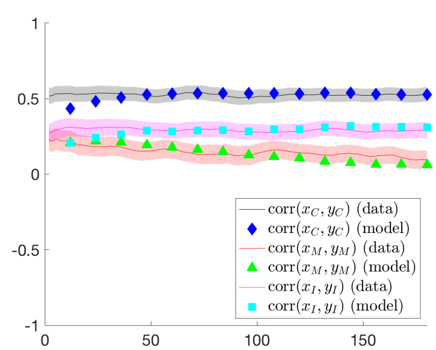
(d) Cross-sectional mean job attributes



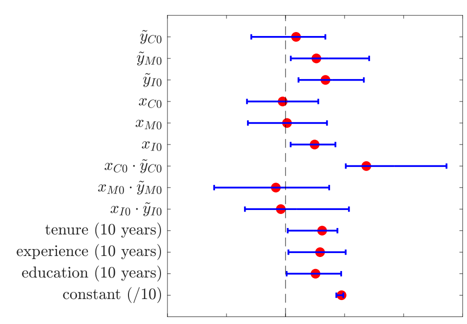
(e) Cross-sectional st.d. of job attributes



(f) Correlation of job attributes



(g) Corr. of job and worker attributes



(h) Descriptive (log) wage regression

Figure B.1: Model fit (with worker bargaining power) bargaining power no longer overstates the estimated returns to tenure to the extent that the baseline model did (Figure B.1h). This is again a consequence of bargaining power shifting weight away from workers' outside options and towards match productivity in the wage bargain, which affects entry wages (for which workers' outside option is low) proportionately more than the wages of longer-tenure workers (whose outside option is on average higher, closer to match productivity).

B.2 Sensitivity of Parameter Estimates to Data Moments

In Table B.2 we present a measure of the local sensitivity of the parameter estimates to the data moments (see Andrews, Gentzkow, and Shapiro, 2017). Specifically, we calculate and report the matrix Λ, defined as follows. Let Θ denote the parameter vector and $\mathbf{m}(\Theta)$ denote the vector of model-based moments we are matching. Next, let $G = \mathbb{E}\left[\partial \mathbf{m}/\partial \Theta^\top\right]$ denote the (expectation of the) Jacobian matrix of the moment function $\mathbf{m}(\Theta)$. Let Ω denote the covariance matrix of the data moments. Define $\tilde{\Lambda} = \left(G^\top G\right)^{-1} G^\top$. The Λ matrix is then the matrix whose ij-th element is given by $\Lambda_{ij} = \sqrt{\Omega_{jj}} \times \tilde{\Lambda}_{ij}$. The ij-th element of Λ can be interpreted as the local approximation to the effect of a one standard deviation change in moment j on parameter θ_i.

For the sake of brevity we omit from the table the five rows for the parameters $[\zeta_S, \zeta_C, \zeta_M, \zeta_I, \alpha_T]$ and the five columns for the moments corresponding to the wage regression coefficients on [years of education, x_{C0}, x_{M0}, x_{I0}, constant] since there is a one-to-one mapping between these parameters and moments.

Table B.2: Sensitivity of parameters to moments

		$U2E$	$E2E_1$	$E2E_2$	$E2E_3$	$E2E_4$	$E2E_5$	$E2E_6$	y_{C1}	y_{C2}	y_{C3}	y_{C4}	y_{C5}	y_{C6}	y_{M1}	y_{M2}	y_{M3}	...
γ_C^u	***	-0.017	0.021	0.016	-0.004	0.003	0.000	0.001	0.008	-0.013	0.001	-0.002	0.010	0.003	-0.039	-0.001	0.011	
γ_M^u	***	0.121	0.206	-0.084	0.200	0.146	0.115	-0.088	-0.145	-0.021	-0.084	-0.055	0.038	0.274	-0.083	-0.061	0.158	
γ_I^u	***	0.235	-0.017	0.033	0.026	-0.012	0.016	0.039	0.066	-0.014	0.029	0.059	-0.042	-0.100	-0.041	0.002	-0.079	
γ_C^o	***	-0.001	0.023	-0.013	0.013	0.008	0.004	-0.009	-0.014	0.005	0.002	-0.006	0.000	0.015	-0.015	-0.019	0.017	
γ_M^o	***	-0.017	0.008	0.018	0.005	0.012	0.002	0.015	-0.005	-0.012	0.010	0.017	0.002	-0.011	-0.026	0.024	0.012	
γ_I^o	***	-0.147	-0.052	-0.020	-0.017	-0.023	-0.003	-0.028	-0.020	-0.022	-0.009	-0.009	0.012	0.038	0.057	0.046	-0.002	
κ_C^u		19.138	-27.787	12.328	-19.626	-13.829	-5.146	10.775	25.275	-5.414	0.325	10.731	-1.540	-26.724	25.866	29.468	-22.689	
κ_M^u		5.454	-8.827	4.687	-2.673	-2.400	-0.139	1.439	4.400	-0.997	-1.417	3.227	-0.434	-5.457	8.745	9.970	-3.805	
κ_I^u		4.227	-0.013	-2.701	3.440	0.615	2.423	-1.090	-0.981	-0.883	-0.795	0.947	-0.762	2.394	2.738	0.955	-0.973	
κ_C^o		0.230	0.024	0.076	-0.126	-0.029	-0.098	0.072	0.104	0.131	0.167	0.111	-0.067	-0.447	-0.032	0.168	0.032	
κ_M^o		2.130	-2.625	1.474	-0.753	-0.867	-0.200	1.266	1.888	-0.716	0.246	1.692	-0.470	-2.710	2.096	3.595	-2.134	
κ_I^o		0.639	-0.748	0.145	-0.435	-0.429	0.013	0.377	0.531	-0.122	0.289	0.692	-0.348	-1.100	0.807	1.028	-0.698	
ξ_C	**	-0.001	0.001	0.056	0.068	-0.026	0.008	0.044	0.376	0.159	0.251	0.299	0.307	0.283	-0.108	-0.029	-0.067	
ξ_M	**	-0.207	-0.320	0.013	-0.118	-0.145	-0.073	-0.057	0.039	-0.026	-0.053	-0.013	0.005	0.007	0.424	0.297	0.065	
ξ_I	**	0.104	0.127	0.206	-0.157	0.009	-0.068	0.080	0.166	0.069	0.115	-0.021	-0.049	-0.254	-0.259	0.016	-0.011	
α_C		0.925	-0.601	-0.031	-0.085	-0.098	0.066	0.335	0.137	0.070	-0.001	0.527	-0.199	-0.511	0.910	0.564	-0.513	
α_M		0.035	-0.527	0.239	-0.390	-0.270	-0.174	0.124	0.295	-0.005	0.114	0.205	-0.041	-0.572	0.347	0.499	-0.171	
α_I		0.577	-0.389	0.037	-0.017	-0.212	0.029	0.107	0.363	-0.152	0.060	0.297	-0.140	-0.479	0.459	0.523	-0.340	
α_{CC}		0.933	-0.822	-0.203	0.004	-0.527	0.110	0.200	0.909	-0.524	-0.309	0.183	-0.054	-0.214	1.128	0.557	-1.039	
α_{MM}		0.049	-0.089	0.016	-0.043	-0.059	0.000	0.049	0.102	-0.033	0.004	0.043	-0.029	-0.110	0.109	0.103	-0.102	
α_{II}		0.182	0.088	-0.130	0.066	-0.017	0.040	-0.034	0.040	0.005	0.001	-0.001	-0.027	-0.035	0.070	-0.037	0.033	
b		0.457	-0.673	0.239	-0.262	-0.307	-0.030	0.330	0.505	-0.169	0.064	0.445	-0.187	-0.709	0.569	0.699	-0.671	
g	***	-0.018	-0.027	0.021	-0.015	-0.012	-0.012	-0.002	0.010	-0.001	-0.003	-0.003	0.003	-0.005	0.002	0.004	-0.012	
η_C^1	**	-0.054	0.051	0.120	-0.081	-0.073	-0.115	-0.013	0.098	-0.004	-0.047	-0.140	0.119	0.087	-0.153	-0.223	0.077	
η_M^1	**	0.366	-0.863	0.310	-0.368	-0.306	-0.324	0.054	0.410	0.091	-0.021	0.090	-0.052	-0.615	0.534	0.752	-0.498	
η_I^1	**	-0.247	-0.103	0.054	-0.321	-0.094	-0.131	0.097	0.208	0.104	0.173	0.015	-0.123	-0.430	0.065	0.205	-0.077	
η_C^2	**	0.287	0.371	0.012	-0.116	0.077	-0.121	-0.096	-0.164	0.306	-0.017	-0.253	0.112	0.209	-0.217	-0.497	0.403	
η_M^2	**	0.333	-0.025	0.467	-0.345	0.140	-0.241	0.306	0.313	0.208	0.235	0.109	0.005	-0.814	-0.026	0.982	0.084	
η_I^2	**	0.291	-0.333	-0.781	0.204	-0.112	0.110	0.002	-0.098	0.049	0.019	0.264	-0.257	-0.206	0.891	0.151	-0.166	
ρ_{CM}	**	-0.076	-0.040	0.019	-0.041	0.003	-0.002	0.017	0.000	-0.004	0.010	0.005	-0.004	-0.008	0.032	0.076	-0.004	
ρ_{CI}	**	0.088	0.023	0.002	0.014	0.003	0.004	0.004	0.015	-0.004	-0.002	0.006	0.003	-0.008	-0.018	-0.018	-0.001	
ρ_{MI}	**	0.052	-0.015	0.013	-0.024	0.000	0.002	0.012	0.023	-0.005	0.008	0.012	-0.006	-0.032	0.016	0.060	-0.016	
λ_0	***	-0.607	0.507	-0.047	0.453	0.341	0.254	0.145	-0.368	-0.036	0.058	0.166	-0.104	0.237	-0.568	-0.364	0.302	
λ_1	***	-0.015	0.009	0.032	-0.003	0.002	-0.011	0.009	0.003	-0.017	-0.007	0.024	0.019	-0.017	-0.042	-0.024	0.019	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	y_{M4}	y_{M5}	y_{M6}	y_{I1}	y_{I2}	y_{I3}	y_{I4}	y_{I5}	y_{I6}	$\mathrm{corr}_1(y_C, y_M)$	$\mathrm{corr}_2(y_C, y_M)$	$\mathrm{corr}_3(y_C, y_M)$	$\mathrm{corr}_4(y_C, y_M)$	...
γ_C^u	***		-0.007	0.023	0.008	0.027	-0.040	-0.010	0.014	0.002	-0.001	0.033	-0.005	-0.011	0.269	
γ_M^u	***		-0.056	-0.018	-0.061	-0.008	0.031	-0.165	-0.063	-0.109	0.250	-1.144	1.585	-0.772	-0.599	
γ_I^u	***		0.090	0.004	0.030	0.102	0.069	0.109	-0.050	-0.059	-0.164	-0.172	-0.664	-0.240	-0.602	
γ_C^o	***		-0.008	0.011	0.007	0.005	0.014	-0.014	0.004	-0.011	0.002	0.044	-0.004	0.067	0.010	
γ_M^o	***		0.037	0.004	-0.029	0.021	-0.036	0.003	-0.014	0.011	0.008	-0.212	-0.094	0.066	-0.070	
γ_I^o	***		-0.032	-0.050	-0.061	-0.085	-0.061	0.004	0.012	0.047	0.097	0.332	-0.076	0.187	0.171	
κ_C^u			4.922	-11.581	-9.329	-0.872	-14.045	20.010	-1.737	14.198	-12.263	-71.440	2.606	-185.169	64.331	
κ_M^u			4.820	-6.664	-6.369	-2.353	-2.127	3.987	-5.177	4.400	1.497	-51.638	33.382	-55.538	-22.411	
κ_I^u			1.029	-3.780	-3.102	-2.496	1.769	1.527	-2.397	-0.337	3.033	-15.434	8.441	-3.368	-46.160	
κ_C^o			0.338	0.096	0.008	0.251	0.121	0.155	-0.105	0.072	-0.446	0.021	-0.711	1.736	-0.107	
κ_M^o			2.523	-1.582	-1.879	0.134	-1.539	2.533	-1.445	1.622	-0.755	-9.238	-4.931	-15.482	-12.182	
κ_I^o			0.445	-0.483	-0.527	0.006	-0.094	1.006	-0.224	0.399	-0.610	-4.386	-4.059	0.127	1.212	
ξ_C	**		0.010	-0.014	0.063	0.064	-0.144	0.046	0.033	0.011	-0.020	0.269	-0.297	-1.208	0.390	
ξ_M	**		0.125	0.090	0.144	-0.213	-0.012	0.071	0.022	0.085	0.103	1.537	-1.494	-0.327	0.335	
ξ_I	**		0.046	0.322	0.247	0.637	0.166	0.301	0.323	0.188	-0.141	0.164	-0.381	-0.457	1.571	
α_C			0.449	-0.543	-0.457	-0.370	0.378	0.539	-0.335	0.236	-0.153	-5.571	-0.302	-3.111	-4.159	
α_M			0.264	-0.176	-0.320	0.036	-0.105	0.285	-0.117	0.256	-0.236	-0.319	-1.188	-0.140	0.854	
α_I			0.340	-0.383	-0.358	0.002	-0.075	0.486	-0.248	0.230	-0.178	-0.991	0.186	-2.727	-1.283	
α_{CC}			-0.301	-1.002	-0.097	-0.487	-0.336	0.669	0.006	0.298	0.068	-1.086	1.168	-12.549	-1.946	
α_{MM}			0.025	-0.067	-0.025	-0.015	-0.049	0.098	-0.010	0.060	-0.051	-0.565	-0.101	-0.464	0.073	
α_{II}			-0.039	-0.064	0.028	0.012	0.127	0.002	0.000	-0.027	-0.072	-0.650	0.754	-0.430	0.014	
b			0.373	-0.394	-0.260	-0.049	-0.242	0.703	-0.184	0.347	-0.338	-2.744	-1.549	-3.647	-0.240	
g	***		-0.002	0.003	0.005	-0.003	-0.010	0.004	0.001	0.006	-0.002	0.199	-0.081	-0.053	0.080	
η_C^1	**		-0.224	0.125	0.273	0.073	-0.076	-0.172	0.142	-0.018	-0.042	1.128	-0.392	-0.012	2.532	
η_M^1	**		0.545	-0.205	-0.255	-0.061	0.020	0.473	-0.339	0.290	-0.288	3.669	-3.843	-0.926	-2.165	
η_I^1	**		-0.029	0.213	0.043	0.221	-0.042	0.108	0.131	0.051	-0.389	0.371	-0.182	0.472	3.642	
η_C^2	**		-0.287	0.319	0.401	0.147	0.411	-0.469	0.080	-0.200	-0.139	-0.601	0.063	3.199	1.625	
η_M^2	**		0.940	0.283	-0.343	0.629	-0.398	0.171	-0.376	0.283	-0.425	-0.982	2.156	1.716	-1.770	
η_I^2	**		0.057	-0.654	-0.601	-0.563	0.800	0.234	-0.153	0.087	0.097	-2.137	0.732	2.714	-0.350	
ρ_{CM}	**		0.006	-0.002	-0.077	-0.009	-0.059	0.003	-0.005	0.026	0.052	-0.378	-0.104	0.112	0.014	
ρ_{CI}	**		0.015	0.004	0.020	0.030	0.022	0.006	-0.007	-0.017	-0.033	-0.068	0.050	-0.165	-0.204	
ρ_{MI}	**		0.027	-0.002	-0.041	0.026	-0.020	0.021	-0.017	0.005	-0.011	-0.155	0.146	0.113	-0.067	
λ_0	***		0.173	0.264	-0.058	0.196	-0.026	-0.268	0.026	-0.181	0.216	-5.066	2.152	1.600	-0.856	
λ_1	***		0.018	0.015	0.014	0.025	-0.028	-0.028	0.015	0.032	-0.021	-0.167	0.029	-0.266	0.203	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	$\mathrm{corr}_5(y_C, y_M)$	$\mathrm{corr}_6(y_C, y_M)$	$\mathrm{corr}_1(y_C, y_I)$	$\mathrm{corr}_2(y_C, y_I)$	$\mathrm{corr}_3(y_C, y_I)$	$\mathrm{corr}_4(y_C, y_I)$	$\mathrm{corr}_5(y_C, y_I)$	$\mathrm{corr}_6(y_C, y_I)$	$\mathrm{corr}_1(y_M, y_I)$	...
γ_C^u	***		-0.122	0.067	0.035	0.071	0.016	-0.174	0.005	0.093	0.054	
γ_M^u	***		-0.373	1.003	1.278	0.065	0.028	-0.124	0.548	-0.325	-0.419	
γ_I^u	***		-0.148	-0.567	-0.463	-0.106	0.039	0.092	0.256	0.244	-0.169	
γ_C^o	***		0.009	0.039	0.020	0.012	-0.055	-0.008	0.004	0.088	-0.004	
γ_M^o	***		-0.119	-0.010	0.064	0.078	0.016	-0.020	-0.032	-0.051	0.061	
γ_I^o	***		0.138	0.330	0.162	0.016	-0.244	-0.031	-0.097	-0.194	0.387	
κ_C^u			-12.538	-66.797	-48.565	4.002	29.870	-41.995	1.328	-55.169	-44.500	
κ_M^u			-22.174	-25.587	9.309	-14.790	15.415	15.545	13.193	-29.270	-12.871	
κ_I^u			16.867	3.233	1.308	-1.150	-10.104	13.003	7.964	-7.932	-17.342	
κ_C^o			0.177	0.557	-0.701	-0.433	-0.373	0.203	0.326	0.778	0.215	
κ_M^o			2.183	-7.926	-6.644	1.621	5.251	3.838	-2.341	-9.208	-4.453	
κ_I^o			1.420	-3.913	-2.963	0.298	-0.354	0.820	-0.093	-0.463	-1.923	
ξ_C	**		0.214	-0.299	-0.131	0.402	0.712	-0.196	-0.174	-0.228	0.150	
ξ_M	**		-0.397	-0.747	-0.383	-0.398	-0.047	0.034	-0.019	0.080	0.795	
ξ_I	**		-0.602	-0.131	-0.530	0.151	0.203	-0.732	-0.167	0.755	0.021	
α_C			2.344	-4.056	-2.122	-1.032	1.887	2.216	-0.450	-1.897	-3.685	
α_M			-2.475	-0.997	-0.560	-0.262	-0.356	-0.037	0.729	0.020	0.826	
α_I			1.547	-0.156	-1.531	0.219	-0.157	0.368	0.840	-0.883	-1.667	
α_{CC}			3.669	-1.560	-0.861	2.555	2.273	-0.092	-0.037	-4.979	-2.727	
α_{MM}			0.380	-0.097	-0.313	0.095	0.088	0.064	-0.122	-0.389	-0.483	
α_{II}			0.396	0.698	0.234	0.147	-0.338	0.075	0.430	0.029	-0.618	
b			1.013	-3.120	-2.427	0.009	1.739	0.572	-0.419	-1.826	-1.590	
g	***		-0.118	-0.161	-0.065	-0.043	0.051	-0.006	-0.026	0.027	0.094	
η_C^1	**		-0.895	-0.703	-0.555	0.770	0.478	-0.801	-0.595	0.109	-0.778	
η_M^1	**		-1.749	0.623	-1.570	-1.449	0.771	0.570	0.658	-0.441	1.435	
η_I^1	**		-0.822	-0.351	-0.174	-0.113	-0.492	-0.817	0.083	1.008	0.737	
η_C^2	**		-1.973	-1.187	-0.784	0.423	-0.726	-0.660	-0.392	1.073	-2.724	
η_M^2	**		-2.213	4.772	1.170	0.148	0.264	-0.150	0.243	-1.454	0.645	
η_I^2	**		1.236	-1.373	0.025	-0.933	-1.417	1.469	2.034	0.545	-1.972	
ρ_{CM}	**		-0.396	-0.005	0.157	0.084	-0.187	-0.081	-0.028	-0.124	-0.023	
ρ_{CI}	**		-0.035	-0.008	-0.166	-0.101	-0.090	-0.186	-0.020	-0.054	-0.152	
ρ_{MI}	**		-0.095	0.145	0.031	0.067	-0.102	-0.122	0.111	-0.026	-0.427	
λ_0	***		0.954	-0.270	0.298	-1.052	1.129	0.931	-0.980	0.086	-1.613	
λ_1	***		0.234	-0.486	0.001	-0.060	0.169	-0.185	-0.127	0.245	0.151	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	$\text{corr}_2(y_M, y_I)$	$\text{corr}_3(y_M, y_I)$	$\text{corr}_4(y_M, y_I)$	$\text{corr}_5(y_M, y_I)$	$\text{corr}_6(y_M, y_I)$	$\text{corr}_1(y_C, x_C)$	$\text{corr}_2(y_C, x_C)$	$\text{corr}_3(y_C, x_C)$	$\text{corr}_4(y_C, x_C)$	...
γ_C^u	***		0.026	-0.001	0.009	-0.116	0.121	-0.276	-0.049	0.001	0.260	
γ_M^u	***		-0.319	-0.240	0.097	-0.539	0.188	-1.305	1.794	2.664	2.638	
γ_I^u	***		0.185	0.061	-0.137	0.300	-0.267	-0.425	-1.146	-0.942	-1.113	
γ_C^o	***		-0.001	-0.097	-0.068	-0.071	0.133	0.002	0.130	0.119	0.136	
γ_M^o	***		0.103	-0.094	-0.045	-0.141	-0.035	-0.164	-0.314	-0.268	-0.049	
γ_I^o	***		-0.164	-0.108	0.166	-0.128	-0.154	-0.063	0.339	0.100	0.470	
κ_C^u			15.933	60.219	149.080	70.769	-148.034	-64.616	-203.664	34.137	-1.100	
κ_M^u			-19.832	10.313	63.632	11.125	-82.512	-31.039	-56.630	48.199	46.060	
κ_I^u			19.458	-2.498	-1.928	2.762	-13.261	-16.094	9.408	19.593	20.369	
κ_C^o			-0.304	-1.337	-0.616	0.867	-0.044	-0.046	-2.213	-1.039	-0.537	
κ_M^o			7.754	4.815	11.901	-0.102	-21.700	-8.933	-29.042	3.186	0.273	
κ_I^o			1.083	0.312	5.318	0.527	-4.938	0.814	-8.780	-3.890	0.391	
ξ_C	**		0.853	0.687	0.035	-0.580	-0.617	-0.770	-0.829	-0.843	-0.633	
ξ_M	**		-1.733	0.977	0.461	-0.069	-0.862	0.236	0.017	0.840	0.066	
ξ_I	**		-0.552	-0.905	0.286	0.465	0.991	1.309	-1.915	-1.201	-0.524	
α_C			1.729	1.664	1.681	1.214	-4.234	0.804	-2.786	2.902	-2.178	
α_M			-1.378	-0.382	2.083	0.712	-2.582	-1.023	-4.590	-1.485	-0.449	
α_I			2.324	-0.540	2.289	1.490	-3.650	-1.968	-4.285	0.369	1.822	
α_{CC}			5.773	4.614	3.703	0.628	-7.980	0.007	-1.912	4.242	0.611	
α_{MM}			0.562	0.366	0.662	0.211	-0.501	0.261	-0.788	-0.026	-0.067	
α_{II}			0.761	-0.594	0.161	0.079	-0.055	0.013	0.531	1.018	1.162	
b			1.812	1.944	3.161	0.457	-4.781	1.217	-5.998	-1.257	-0.755	
g	***		-0.158	0.100	0.079	-0.035	-0.122	0.006	-0.128	-0.008	-0.051	
η_C^1	**		-0.006	1.393	0.975	-0.755	-0.383	-0.588	0.615	0.413	0.212	
η_M^1	**		-5.024	3.476	0.918	1.035	-3.205	-2.468	-4.191	5.796	-0.708	
η_I^1	**		-1.621	-1.142	0.891	0.609	1.127	4.338	-2.749	-1.743	-1.278	
η_C^2	**		-0.691	0.780	0.483	0.123	1.806	1.133	3.395	2.972	2.300	
η_M^2	**		-1.160	-0.784	0.033	0.625	-0.469	-2.687	-6.162	2.350	-0.847	
η_I^2	**		0.494	1.017	-0.565	0.370	-1.039	3.949	-0.245	0.808	-1.880	
ρ_{CM}	**		-0.224	-0.071	0.241	-0.016	0.108	0.244	-0.392	-0.145	-0.045	
ρ_{CI}	**		0.172	0.016	-0.203	0.046	-0.028	-0.403	-0.044	0.103	-0.054	
ρ_{MI}	**		-0.240	-0.208	-0.165	-0.139	-0.332	-0.248	-0.414	-0.059	-0.068	
λ_0	***		2.373	-0.602	-1.417	-1.985	3.592	3.488	-1.385	-3.413	-1.668	
λ_1	***		0.169	0.046	-0.174	-0.106	0.067	-0.297	-0.415	0.033	0.444	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	$\text{corr}_5(y_C, x_C)$	$\text{corr}_6(y_C, x_C)$	$\text{corr}_1(y_M, x_M)$	$\text{corr}_2(y_M, x_M)$	$\text{corr}_3(y_M, x_M)$	$\text{corr}_4(y_M, x_M)$	$\text{corr}_5(y_M, x_M)$	$\text{corr}_6(y_M, x_M)$	$\text{corr}_1(y_I, x_I)$	...
γ_C^u	***		0.014	0.039	-0.200	0.128	-0.048	-0.365	-0.274	0.134	0.107	
γ_M^u	***		-1.025	-1.572	-0.643	7.594	-2.091	2.588	-0.731	1.158	0.695	
γ_I^u	***		0.122	-0.104	-0.392	0.999	1.079	-0.961	0.678	0.106	-0.334	
γ_C^o	***		-0.136	-0.025	-0.137	0.073	0.006	-0.008	0.070	-0.177	0.072	
γ_M^o	***		0.298	-0.029	0.135	0.201	0.224	0.048	0.057	0.052	0.339	
γ_I^o	***		0.257	-0.020	-0.039	-0.490	-0.100	0.158	-0.456	-0.030	0.118	
κ_C^u			98.356	-96.800	296.853	12.780	-73.988	-121.765	-63.652	84.404	-149.724	
κ_M^u			48.435	-35.187	43.499	138.408	-49.088	50.370	5.455	28.385	-47.247	
κ_I^u			10.390	-15.874	-0.052	48.318	13.807	-21.520	1.557	3.153	16.039	
κ_C^o			2.547	2.070	-1.739	-3.980	1.591	-0.516	1.879	-2.444	-1.331	
κ_M^o			23.561	-9.876	21.486	13.048	4.708	-12.011	9.672	-2.122	-4.085	
κ_I^o			8.651	-4.571	10.010	3.727	-2.885	-3.988	3.658	-1.706	-9.986	
ξ_C	**		-0.048	0.458	-2.605	1.335	0.970	-1.123	0.000	0.237	0.082	
ξ_M	**		-0.494	-0.556	-0.253	-1.343	0.073	-0.313	0.155	1.014	-0.645	
ξ_I	**		0.222	1.025	1.389	1.202	-0.833	-0.693	-0.294	-0.509	-1.343	
α_C			5.081	-4.512	13.214	-0.633	1.363	-1.053	3.829	-0.050	-4.607	
α_M			2.264	-0.165	2.481	-1.947	-0.120	0.380	2.427	1.296	-2.661	
α_I			3.728	0.191	2.126	3.191	-0.583	-5.183	3.269	-0.205	0.037	
α_{CC}			-0.240	-1.776	1.444	12.025	-1.153	0.388	-3.233	2.544	-5.926	
α_{MM}			0.750	0.148	0.324	-0.094	0.505	-1.005	-0.290	-0.486	-0.526	
α_{II}			-0.276	0.607	-1.976	1.444	0.563	0.576	-0.020	-1.340	-1.345	
b			6.508	-0.982	6.040	1.941	-0.658	-3.542	1.058	1.008	-4.378	
g	***		-0.065	-0.006	-0.017	-0.085	-0.002	-0.040	0.071	0.079	0.001	
η_C^1	**		-2.779	1.266	-2.814	-1.710	-1.140	-0.382	1.292	-0.940	0.687	
η_M^1	**		1.321	-1.734	-4.767	-1.758	2.664	-1.231	1.949	-0.787	0.087	
η_I^1	**		0.390	-0.108	3.558	0.137	-2.519	0.734	-0.805	-1.592	-0.801	
η_C^2	**		-4.313	1.308	0.942	-3.606	-3.796	2.380	3.586	-1.893	1.377	
η_M^2	**		4.193	-1.345	0.280	-0.173	0.126	2.954	-0.468	-3.038	4.411	
η_I^2	**		-0.036	-3.740	0.101	1.367	-0.808	1.141	3.184	-2.934	2.904	
ρ_{CM}	**		0.113	-0.604	1.386	0.235	-0.665	-0.003	-0.291	0.407	0.767	
ρ_{CI}	**		-0.135	0.052	-0.427	0.242	0.209	-0.218	0.113	0.060	-0.166	
ρ_{MI}	**		0.015	-0.477	0.726	0.454	-0.273	-0.320	-0.131	0.555	0.349	
λ_0	***		2.059	-0.146	-1.337	4.309	0.428	-3.541	2.523	1.810	4.400	
λ_1	***		0.086	-0.001	-0.249	-0.380	0.309	-0.076	0.340	0.395	0.348	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	$\mathrm{corr}_2(y_I, x_I)$	$\mathrm{corr}_3(y_I, x_I)$	$\mathrm{corr}_4(y_I, x_I)$	$\mathrm{corr}_5(y_I, x_I)$	$\mathrm{corr}_6(y_I, x_I)$	$\mathrm{sd}_1(y_C)$	$\mathrm{sd}_2(y_C)$	$\mathrm{sd}_3(y_C)$	$\mathrm{sd}_4(y_C)$	$\mathrm{sd}_5(y_C)$	$\mathrm{sd}_6(y_C)$	$\mathrm{sd}_1(y_M)$	$\mathrm{sd}_2(y_M)$	...
γ_C^u	***		-0.051	0.111	-0.206	0.074	-0.108	0.011	0.010	0.005	-0.001	-0.001	0.000	-0.001	0.001	
γ_M^u	***		1.388	0.629	-1.398	-1.514	1.157	0.125	-0.006	0.049	-0.098	0.056	0.057	-0.069	-0.011	
γ_I^u	***		0.185	1.211	0.256	1.895	0.980	-0.080	-0.034	-0.049	-0.026	-0.049	-0.081	0.022	0.009	
γ_C^o	***		0.057	0.018	0.002	-0.102	-0.049	0.011	0.007	0.004	0.005	0.009	0.009	-0.006	-0.002	
γ_M^o	***		0.043	0.217	-0.131	-0.098	-0.009	-0.009	-0.001	-0.011	-0.011	-0.020	-0.029	0.012	0.008	
γ_I^o	***		0.481	-0.265	0.309	-0.056	-0.082	-0.009	-0.028	-0.021	-0.032	-0.020	-0.017	-0.030	-0.031	
κ_C^u			-139.556	117.349	-55.286	196.038	-13.921	-13.465	-4.817	-6.284	-5.867	-11.025	-10.664	13.386	8.187	
κ_M^u			-6.455	27.399	-1.501	3.281	38.548	-3.998	-3.876	-2.103	-3.702	-3.395	-3.323	4.106	4.000	
κ_I^u			-0.313	-15.235	-5.204	8.704	2.821	-1.549	-1.763	-1.640	-2.728	-1.183	-2.209	-0.849	-1.370	
κ_C^o			-0.851	2.252	-1.257	1.161	-0.545	-0.124	0.044	-0.035	0.101	-0.081	-0.090	0.264	0.286	
κ_M^o			-14.822	19.058	0.518	4.768	1.641	-2.177	-1.045	-1.223	-1.225	-1.936	-2.432	1.898	1.275	
κ_I^o			-2.209	10.034	0.405	4.227	-2.428	-0.827	-0.292	-0.525	-0.323	-0.659	-0.886	0.433	0.213	
ξ_C	**		-0.190	0.292	0.403	0.189	-0.477	-0.058	-0.067	-0.063	-0.095	-0.070	-0.110	-0.046	-0.075	
ξ_M	**		0.125	-0.464	0.602	0.517	-0.470	-0.071	-0.104	-0.047	-0.026	-0.029	-0.011	-0.108	-0.108	
ξ_I	**		-2.547	0.888	0.870	1.253	-0.041	0.063	0.122	0.077	0.136	0.041	0.069	0.180	0.185	
α_C			-0.784	7.971	-3.161	-2.129	4.972	-0.565	-0.177	-0.224	-0.186	-0.395	-0.552	0.346	0.251	
α_M			0.191	2.374	0.266	2.979	-2.260	-0.258	-0.121	-0.158	-0.026	-0.191	-0.171	0.211	0.166	
α_I			-1.603	1.503	-2.695	3.600	-0.822	-0.416	-0.186	-0.293	-0.268	-0.341	-0.462	0.228	0.123	
α_{CC}			-4.785	0.655	5.812	7.499	-4.345	-0.517	-0.554	-0.385	-0.708	-0.321	-0.423	-0.089	-0.427	
α_{MM}			-0.903	-0.001	0.707	0.815	-0.475	-0.074	-0.032	-0.045	-0.035	-0.049	-0.066	0.051	0.008	
α_{II}			0.060	0.148	0.672	0.855	-1.009	-0.009	-0.020	-0.027	-0.038	0.022	0.016	-0.003	-0.009	
b			-3.928	4.977	-0.737	2.123	1.386	-0.602	-0.257	-0.327	-0.282	-0.489	-0.657	0.324	0.161	
g	***		-0.075	-0.107	0.088	0.007	0.060	-0.002	-0.004	0.002	0.005	0.000	0.005	-0.001	0.001	
η_C^1	**		-0.087	-2.248	1.667	0.773	-1.361	0.185	0.132	0.166	0.177	0.203	0.316	-0.058	-0.067	
η_M^1	**		-3.097	3.171	-2.179	2.693	-0.042	-0.348	-0.255	-0.102	-0.023	-0.217	-0.162	0.419	0.442	
η_I^1	**		-2.574	1.440	1.318	2.091	-2.236	-0.016	0.073	-0.021	0.136	-0.014	-0.011	0.173	0.132	
η_C^2	**		-0.072	-3.049	1.291	0.519	-0.676	0.474	0.424	0.450	0.525	0.531	0.770	0.027	0.093	
η_M^2	**		-4.032	4.886	-2.186	0.180	-1.155	-0.036	0.122	0.075	0.119	-0.158	-0.075	0.895	0.894	
η_I^2	**		2.594	3.007	-3.665	2.847	-3.583	-0.388	-0.236	-0.363	-0.188	-0.192	-0.408	-0.165	-0.281	
ρ_{CM}	**		-0.261	-0.146	0.215	0.182	-0.373	0.008	0.006	-0.003	0.001	-0.008	-0.006	0.018	0.002	
ρ_{CI}	**		0.008	0.203	-0.309	0.351	-0.022	-0.001	0.005	0.004	0.000	0.005	0.005	0.006	0.008	
ρ_{MI}	**		-0.297	0.143	-0.230	0.691	-0.269	-0.006	0.002	-0.007	-0.005	-0.012	-0.010	0.028	0.017	
λ_0	***		-0.121	2.378	3.213	-6.672	0.749	0.031	0.045	-0.082	-0.033	-0.058	-0.316	-0.099	-0.130	
λ_1	***		-0.306	0.484	-0.122	-0.188	0.086	0.003	0.008	-0.001	0.006	-0.004	-0.014	0.000	0.006	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued ...

		...	$sd_3(y_M)$	$sd_4(y_M)$	$sd_5(y_M)$	$sd_6(y_M)$	$sd_1(y_I)$	$sd_2(y_I)$	$sd_3(y_I)$	$sd_4(y_I)$	$sd_5(y_I)$	$sd_6(y_I)$	β_{y_C}	β_{y_M}	β_{y_I}	...
γ_C^u	***		0.003	0.002	-0.011	-0.010	-0.014	-0.007	-0.013	-0.018	-0.015	-0.026	-0.097	-0.140	0.064	
γ_M^u	***		-0.100	-0.085	-0.032	-0.040	-0.097	-0.161	-0.106	-0.143	-0.051	-0.110	-0.379	-0.263	0.132	
γ_I^u	***		0.029	0.000	-0.001	-0.014	0.081	0.081	0.085	0.075	0.051	0.054	-0.248	0.695	-1.057	
γ_C^o	***		0.007	-0.002	-0.004	-0.012	0.000	0.004	0.005	0.000	0.002	0.003	0.037	0.078	-0.059	
γ_M^o	***		0.014	0.009	0.005	0.015	-0.010	-0.004	-0.005	-0.005	-0.008	-0.012	0.007	-0.013	0.091	
γ_I^o	***		-0.035	-0.018	-0.007	-0.005	-0.004	-0.014	-0.016	-0.006	0.004	-0.003	-0.081	-0.229	0.116	
κ_C^u			4.614	5.713	7.819	10.926	13.642	10.215	9.158	8.628	6.987	1.393	-202.089	90.022	-148.307	
κ_M^u			-0.443	1.063	4.404	6.715	1.405	-1.009	0.058	0.412	0.806	-0.404	-27.678	-18.971	2.966	
κ_I^u			-1.249	-2.205	-0.366	-1.212	1.840	0.592	1.421	1.188	1.654	1.794	-31.155	15.539	-35.377	
κ_C^o			0.423	0.236	0.180	0.214	0.140	0.279	0.250	0.182	0.091	0.136	-0.601	-0.680	-0.377	
κ_M^o			0.868	0.696	1.410	1.986	0.795	0.712	0.742	0.790	0.312	0.191	-18.517	15.914	-23.835	
κ_I^o			0.564	0.181	0.221	0.099	1.007	0.995	0.984	0.900	0.717	0.685	-7.643	3.784	-8.220	
ξ_C	**		-0.109	-0.055	-0.051	-0.047	-0.021	-0.043	-0.039	-0.029	-0.044	-0.060	-0.510	-0.368	-0.236	
ξ_M	**		-0.155	-0.067	-0.034	-0.068	0.083	0.050	0.034	0.069	0.076	0.065	0.631	0.439	-0.019	
ξ_I	**		0.231	0.173	0.045	0.108	-0.088	-0.008	-0.056	-0.060	-0.108	-0.127	0.122	-0.248	0.095	
α_C			0.161	0.060	0.275	0.261	0.506	0.431	0.530	0.478	0.337	0.452	-8.148	1.614	-7.041	
α_M			0.235	0.189	0.201	0.260	0.283	0.296	0.270	0.260	0.224	0.147	-1.592	1.583	-1.975	
α_I			0.209	0.023	0.137	0.100	0.459	0.416	0.438	0.358	0.287	0.245	-3.484	4.150	-5.695	
α_{CC}			-0.706	-0.418	-0.002	-0.065	0.504	-0.003	0.182	0.249	0.331	0.121	-8.709	5.791	-7.558	
α_{MM}			0.024	0.006	0.025	0.026	0.075	0.055	0.055	0.061	0.042	0.038	0.028	0.478	-0.345	
α_{II}			0.057	-0.025	0.013	-0.041	0.122	0.089	0.129	0.080	0.103	0.089	-0.482	1.057	-1.160	
b			0.123	0.095	0.177	0.207	0.435	0.366	0.380	0.398	0.236	0.230	-5.775	4.291	-6.649	
g	***		-0.011	0.003	0.003	0.007	-0.009	-0.009	-0.013	-0.007	-0.008	-0.008	0.066	-0.038	0.034	
η_C^1	**		-0.123	-0.022	-0.047	-0.040	-0.188	-0.154	-0.217	-0.196	-0.167	-0.163	-0.233	-1.579	1.310	
η_M^1	**		0.248	0.303	0.464	0.554	0.206	0.283	0.212	0.206	0.147	0.107	1.572	3.402	-1.525	
η_I^1	**		0.370	0.222	0.067	0.057	0.296	0.337	0.299	0.292	0.251	0.186	0.341	-0.871	0.853	
η_C^2	**		0.122	0.077	0.017	-0.014	-0.290	-0.148	-0.236	-0.265	-0.193	-0.115	0.870	-1.125	2.317	
η_M^2	**		0.884	0.687	0.673	1.053	-0.231	-0.008	-0.075	-0.131	-0.172	-0.278	-0.411	0.626	-0.324	
η_I^2	**		0.228	-0.161	-0.015	-0.425	1.308	1.136	1.281	1.134	1.161	1.146	0.489	1.843	-0.915	
ρ_{CM}	**		0.021	0.016	0.008	0.022	0.006	0.001	-0.007	0.005	0.015	-0.007	0.015	-0.131	-0.109	
ρ_{CI}	**		0.006	-0.001	0.000	-0.004	0.007	0.008	0.011	0.000	0.000	-0.002	0.000	0.093	-0.037	
ρ_{MI}	**		0.031	0.015	0.011	0.016	0.024	0.021	0.019	0.014	0.022	0.003	0.049	0.144	-0.234	
λ_0	***		0.066	-0.066	-0.194	-0.233	-0.062	-0.064	0.008	0.038	-0.061	0.038	0.214	-0.153	0.266	
λ_1	***		0.005	0.006	-0.006	-0.011	-0.015	0.001	-0.001	-0.010	-0.019	-0.020	0.026	-0.057	0.045	

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows).
Values multiplied by *10, **100, ***1000.

Table B.2: ... continued.

		...	$\beta_{x_C y_C}$	$\beta_{x_M y_M}$	$\beta_{x_I y_I}$	β_{ten}	β_{exp}
γ_C^u	***		-0.141	-0.104	0.034	0.000	0.000
γ_M^u	***		-1.281	-1.119	1.000	0.000	0.000
γ_I^u	***		-0.394	0.313	-0.810	0.000	0.000
γ_C^o	***		0.029	0.090	-0.012	0.000	0.000
γ_M^o	***		0.053	-0.017	0.014	0.000	0.000
γ_I^o	***		-0.101	-0.136	0.059	0.000	0.000
κ_C^u			-191.574	24.635	-286.850	0.016	-0.011
κ_M^u			-36.001	-39.612	-5.983	0.001	0.000
κ_I^u			-43.820	1.720	-35.270	0.000	0.001
κ_C^o			-0.397	-1.641	-0.819	0.000	0.000
κ_M^o			-19.606	2.550	-31.138	0.002	-0.001
κ_I^o			-8.613	0.876	-10.993	0.001	0.000
ξ_C	**		-0.987	-0.371	0.048	0.000	0.000
ξ_M	**		1.194	0.373	-0.234	0.000	0.000
ξ_I	**		0.271	-0.335	0.025	0.000	0.000
α_C			-9.582	-2.246	-9.639	0.000	0.000
α_M			-1.269	1.147	-3.237	0.000	0.000
α_I			-4.147	2.498	-6.379	0.000	0.000
α_{CC}			-10.309	2.008	-10.820	0.000	0.000
α_{MM}			0.093	0.308	-0.341	0.000	0.000
α_{II}			-0.696	0.515	-1.023	0.000	0.000
b			-6.469	1.417	-8.799	0.001	0.000
g	***		0.142	0.046	-0.020	0.000	0.000
η_C^1	**		-0.560	-1.402	1.306	0.000	0.000
η_M^1	**		2.935	3.579	-1.895	0.000	0.000
η_I^1	**		0.622	-0.707	0.626	0.000	0.000
η_C^2	**		1.435	-0.975	1.762	0.000	0.000
η_M^2	**		-1.014	1.684	-0.016	0.001	0.000
η_I^2	**		0.468	2.013	-0.469	0.000	0.000
ρ_{CM}	**		-0.001	0.025	-0.041	0.000	0.000
ρ_{CI}	**		0.005	0.079	-0.044	0.000	0.000
ρ_{MI}	**		0.070	0.220	-0.175	0.000	0.000
λ_0	***		0.434	0.942	0.000	0.000	0.000
λ_1	***		0.012	0.009	0.069	0.000	0.000

Note: The change in each parameter (rows) induced by a one standard deviation change in each moment (rows). Values multiplied by *10, **100, ***1000.

B.3 Comparing the unconditional and the conditional variance decompositions (Table 5)

Conditioning on broad levels of education reduces the share of variance explained by initial skill bundles, as those are correlated with education. The basic reason is that education explains a fair share of the variance in $\mathbf{x}_0$, and a smaller share of the overall variance in $\ln Q$. Therefore, once one conditions on education, the residual variation in $\mathbf{x}_0$ explains a smaller share of the (conditional) variance of $\ln Q$.

This can be expressed formally, taking up the notation from Section 7 in the paper (and dropping indices to de-clutter the notation):

$$\frac{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0)\right]}{\mathbf{Var}\ln Q} > \frac{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\}}{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}\left[\ln Q \mid \text{ed}\right]\right\}}$$

where our results in Section 7 say that the l.h.s. is about 0.65 while the r.h.s. is slightly below 0.3. Now, the l.h.s. can be further decomposed as:

$$\frac{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0)\right]}{\mathbf{Var}\ln Q} = \frac{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\} + \mathbf{Var}_{\text{ed}}\left\{\mathbf{E}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\}}{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}\left[\ln Q \mid \text{ed}\right]\right\} + \mathbf{Var}_{\text{ed}}\left\{\mathbf{E}\left[\ln Q \mid \text{ed}\right]\right\}} \tag{21}$$

Applying our estimates to those decompositions of the numerator and denominator of the fraction above, we find that:

$$\begin{aligned} \mathbf{Var}_{\text{ed}}\left\{\mathbf{E}\left[\ln Q \mid \text{ed}\right]\right\} &\simeq 0.507 \times \mathbf{Var}\ln Q \\ \mathbf{Var}_{\text{ed}}\left\{\mathbf{E}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\} &\simeq 0.779 \times \mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0)\right] \end{aligned} \tag{22}$$

i.e. education explains much more of the variance of $\ln Q$ conditional on $\mathbf{x}_0$ than in the whole sample. Substitution of (22) into (21) implies:

$$\begin{aligned} \frac{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0)\right]}{\mathbf{Var}\ln Q} &\simeq \frac{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\} \times \left(1 + \frac{0.779}{1-0.779}\right)}{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}\left[\ln Q \mid \text{ed}\right]\right\} \times \left(1 + \frac{0.507}{1-0.507}\right)} \\ &\simeq \frac{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}_{\mathbf{x}_0}\left[\mathbf{E}(\ln Q \mid \mathbf{x}_0) \mid \text{ed}\right]\right\}}{\mathbf{E}_{\text{ed}}\left\{\mathbf{Var}\left[\ln Q \mid \text{ed}\right]\right\}} \times 2.23 \end{aligned}$$

which is roughly the ratio found in Table 5.