Multi-dimensional heterogeneity and mismatch in a frictional labour market

Joanne Tan*

October 27, 2015

Abstract

This paper examines the sorting of workers and firms when both parties are characterised by multi-dimensional heterogeneity. Firstly, I construct six-dimensional vectors of worker and firm characteristics using NLSY and O*NET data and study the sorting between workers and firms on each dimension. I then propose a tractable, static model of directed search with two-sided multi-dimensional heterogeneity to explain the sorting of job market entrants and their first jobs. The degree of sorting on each dimension is determined by explicit tradeoffs made by the worker in his/her application decision. Shifts in the complementarities between worker skills and job requirements, in the degree of specialization in skills and in the cost of mismatch on production generate key intuitive differences to the application decisions of workers and sorting outcomes. I then estimate the production function structurally. The model could provide a better fit of the observed empirical sorting between worker and job characteristics.

1 Introduction

An extensive literature exists on the incidence of over-educated individuals in the labour market alongside evidence on the lack of key skills and know-how amongst workers[1]. The coexistence of these two phenomena suggests that the human capital of workers comprises of several dimensions and that an excess of human capital on one dimension, such as educational attainment, can exist alongside a deficit in another dimension, such as 'people skills' for instance. Yet, very few papers have examined the mechanisms behind the sorting of workers to jobs in the presence of multidimensional heterogeneity, with the previous literature, until very recently, considering worker skills to be unidimensional.This paper aims to address this gap.

Mismatch can hereby be defined as the excess or deficit of a worker's attribute, be it educational attainment or some work-related skill or knowledge, relative to the level required by his or her job. Evidently, the main difficulty with defining worker-firm mismatch lies with the measurement of worker skills and job requirements. Most of the literature has therefore considered only educational

*I am a first year PhD student at Sciences-Po Paris under the supervision of Jean-marc Robin. This paper is part of my very preliminary work on mismatch between worker skills and job requirements, its evolution over time as well as the investment decisions of workers in their human capital characteristics. I can be contacted at: joanne.tan@sciencespo.fr

[1]For instance, in 'Mind the gaps: The 2015 Deloitte Millennial survey' conducted by Deloitte based on 300 interviews with respondents born after 1982 in 29 countries, gaps were found between millenial's skills and skills valued by businesses, with millenials lacking in 7 skill sets, including creative thinking and communication skills.

attainment, with workers considered mismatched if their educational attainment differed from that required by their job. However, as previously mentioned, education does not sufficiently characterise the human capital of workers and the requirements of their job. Recently, the measurement of the other dimensions of firm requirements has been facilitated by the O*NET database, which provides numerous descriptors of each occupation, thereby allowing for the measurement of the various skill sets required. I use this, along with the National Longitudinal Survey of Youths (NLSY) dataset, to construct a six-dimensional set of worker and firm characteristics - cognitive skills, interpersonal skills, physiological skills, educational level, weeks of relevant work experience and field of knowledge.

Using the constructed dataset, I examine sorting between workers from the 1979 NLSY cohort and jobs. I find that there is significant mismatch on all skill dimensions. Positive sorting between worker and firm characteristics is the strongest on the educational attainment and cognitive skill dimensions, and is weakest on the physiological skill dimension. I also find that there is very little improvement in sorting between workers and their first jobs upon leaving school permanently and their latest recorded jobs in 2012, despite there being, on average, over fifteen years between the two. As this preliminary evidence suggests, mismatch is persistent[2]. The inertia of sorting thus justifies the adoption of a simple static model in this paper. In particular, this paper explains multi-dimensional sorting through a static wage posting directed search model. The choice of the directed search framework over the random search and frictionless search frameworks can justified by the following: 1) The directed search model is arguably a more realistic portrayal of workers' job search process since workers do focus their search on certain subsets of all available jobs, 2) it is tractable, 3) it has yet to be studied in the context of multi-dimensional heterogeneity and 4) since mismatch on the first job persists, how workers sort into their first jobs and how mismatch on each dimension arises initially is important in and of itself.

Very recently, two papers have also examined mismatch in the multi-dimensional setting, also using the NLSY and O*NET databases to construct the characteristics of workers and firms. Lindenlaub (2014) considers the frictionless assignment of workers to firms while Postel-Vinay and Lise (2015) build a model of random search with on-the-job human capital accumulation. However, both models do not sufficiently explain the observed sorting between workers and firms in the data. Lindenlaub's (2014) frictionless assignment abstracts from reality completely, while Postel-Vinay and Lise's (2015) model does not fit the observed sorting in the data very well, particularly for the workers' first jobs. Moreover, since the degree of mismatch in the first job seems to persist over time, it is unlikely that individuals conduct random search for their first jobs. It is plausible that individuals are more selective and strategic when applying for their first jobs, trading off not only between employment probability and wages, but also between sorting in the different dimensions. This aspect is emphasised in this paper.

I therefore construct a static model of directed search with two-sided multi-dimensional heterogeneity. While the model simplifies to that in Shimer (2005) when worker and firm heterogeneity is reduced to a single dimension, the equilibrium assignment is distinctly different with multidimensional heterogeneity. In fact, unless strong restrictions are imposed, there is no objective ranking of workers and firms, and workers have to trade off between sorting in the each of the different dimensions when deciding on their application strategies. I derive the worker's application strategies for a few illustrative specifications for the production function and the distributions of workers and

[2]This finding is in line with those of the existing related literature

firms, and then structually estimate the production function. The model allows for the study of the effect of technological change that is biased towards some skills and not others.

This paper is but a fraction of my ongoing thesis on human capital investment and labour market mismatch, the overarching aim of which is to: i) explain and measure the labour market mismatch, its evolution over time and along the business cycle, ii) examine human capital investment both before entry into the labour market and on-the-job and iii) the implications for efficiency. The static model presented in this paper should therefore be considered as providing a very simplified but clear picture of the how workers sort into jobs when both are characterised by multi-dimensional heterogeneity.

The rest of the paper proceeds as follows. Section 2 consists of the literature review. Section 3 presents some preliminary empirical findings. Section 4 gives an overview of the model, the general solutions to the decentralised case and the Social Planner problem. Section 5 examines a few illustrative examples with closed-form solutions. Section 6 outlines an extension of model to allow for the endogenisation of worker and firm types. Section 7 will consist of simulations and estimation (forthcoming) and the last section concludes.

2 Literature Review

This paper falls squarely at the juncture between the literature on over and under education and the literature on skills mismatch. It can also be linked to the literature on wage polarisation and human capital accumulation.

While much of the first two strands of literature have focused on the empirical measurement of over/under education and over/under skilling, as well as the impact of these on individual wages, their employed methods are debatable on two counts - first, the methods hitherto used to measure over/under education or skilling are controversial and second, the estimation of their impact on wages has been plagued with endogeneity bias. Moreover, a theory of how and why individuals would go to jobs that are not necessarily the best fit for their characteristics has not been thoroughly explored.

On the mismatch in education The existing literature has defined over-education as occuring when a worker's educational attainment exceeds that required by his/her job. This may be determined by subjective individual surveys, an objective measure of job educational requirements such as the Dictionary of Occupational Titles (DOT), or by comparing each individual's educational attainment to the mean or modal educational levels of the population of people in that occupation. The accuracy of each of these measurements of over-education has been debated thoroughly and indeed, as highlighted by Battu et al. (2002) and Hartog (2002), the incidence of over-education often varies slightly according to the measure adopted[3]. Nonetheless, these studies have largely agreed on the impact of over and under education on individual wages, using the standard OLS mincer regression. They find that given the same job, the over-educated earn more than their less-educated but well-matched counterparts. However, given the same education level, the over-educated earn less than their well-matched counterparts and thus face a wage penalty. Evidence for the former is provided

[3]For instance, using subjective individual surveys, a 2013 OECD report puts the average over-education rate amongst OECD countries at 22 percent. For the US, OECD estimates hover at around 20 percent. On the other hand, averaging over the results from several studies on the incidence of over-education, Leuven and Oosterbeek (2011) find the incidence in the US to be 37 percent.

by Hartog (2000) who finds that the wage returns to over-education are positive but smaller (half to two-thirds) than that of required education. On the latter, Battu et Al (2000) find a significant wage penalty for the over-educated relative to well-matched graduates. Conversely, several authors have found that under-educated workers suffer a wage penalty relative to their well-matched colleagues. An obvious pitfall of these estimation exercises is the endogeneity of being over/under educated. Indeed, several papers have suggested that over-educated individuals may be compensating for their lack of work-related skills.

On mismatch in skills Several papers have drawn the distinction between education and work-related skills, on the basis that education does not fully encompass the requirements of a job. However, until very recently, the definition of mismatch between an individual's skills and her job's skill requirements have been based on individual's response to several variants of the same simplistic question - 'Which of the following alternatives best describe your skills in your own work?'. To which the responses follow the likes of: 1)'I have the skills to cope with more demanding duties'(over-skilled), 2)'My duties correspond well with my present skills' (well-matched) and 3)'I need further training to cope well with my duties' (under-skilling)[4]. Based on this, results on the incidence of over/under skilling have been obtained[5]. These papers have shown that the skill mismatch is distinctly different from education mismatch. For instance, according to a 2011 OECD report, 'Right for the Job: Over-qualified or Under-skilled', only 36 percent of over-educated workers are also over-skilled and only about 12 percent of under-educated workers report feeling under-skilled. That workers may be substituting education and skills is suggested by the fact that 30.5 percent of the under-educated consider themselves over-skilled, while 14.2 percent of the over-educated are under-skilled. Previous papers have also conducted the usual Mincer wage regressions to estimate the effect of over and under skilling on individual wages with the same pitfalls as those measuring the effect of over and under education [6]. The major problem with these papers is that precisely what skills are being referred to are completely unspecified. Moreover, these papers have no reliable measure of the degree of skills mismatch between workers and their jobs, but merely whether the worker considers herself under or over skilled.

Recent measurements of skills mismatch In light of the shortcomings in measuring worker skills and job skills requirements, very recent papers have sought to quantify and distinguish the various levels skills required in jobs and possessed by workers, using the Occupational Information Network (O*NET), developed by the US Department of Labor. As its name suggests, the O*NET provides occupation-specific descriptors, provided by incumbent worker surveys and job analysts, that allow for the characterisation of the multi-faceted requirements of each job. One of the first to do so is Lindenlaub(2014), who constructs a measure of cognitive and manual skill requirements of each occupation using principal components analysis (PCA). To measure workers' cognitive and manual skills, she uses the National Longitudinal Survey of Youths (NLSY) to create a crosswalk between individuals' qualifications and occupation codes. She then assumes that an individual who has the qualifications for a given occupation necessarily possesses the level of cognitive and manual

[4]These questions were asked, for instance, in the 2011 OECD report 'Right for the Job: Over-qualified or Under-skilled'?

[5]For instance, the 2011 OECD report found that. for EU 19 countries, Estonia, Norway, Slovenia and Switzerland. For these countries, 33.5 percent of all workers were over-skilled on average.

[6]It seems that the over-educated whose skills are well-utilised face a smaller penalty (relative to the well-matched with the same education level) than the over-educated whose skills are under-utilised. Similarly, Chevalier (2003) finds that the over-educated but not over-skilled graduates face a wage penalty of 5-11 percent (depending on the controls included) relative to well-matched graduates, while the over-educated and over-skilled ('genuinely over-educated') face a wage penalty of 22-26 percent.

skills required for the occupation. Also using the O*NET and NLSY, Postel-Vinay and Lise (2015) use PCA to construct three dimensions of workers' skills and skills requirements for each occupation code - manual, cognitive and interpersonal. Both papers find that sorting between firms and workers is strongest on the cognitive dimension.

Theories of worker-firm mismatch (The unidimensional case) There have been a few theories to explain over-education (and under-education) but these, with their unidimensional setup, are unable to explain the varying degrees and directions of mismatch on the different dimensions of worker and firm characteristics. These include Sicherman and Galor's (1990) model of career mobility, where workers may accept jobs for which they are over-educated in exchange for a higher likelihood of being promoted, Spence's (1973) model of education as a signaling device, Thurow's (1975) theory of job competition, where workers' education levels decide their places in the queue for jobs, models with search frictions, such as that of Dolado et al. (2009) with random search where highly-educated unemployed workers accept low level jobs and subsequently conduct on-the-job search for a better job, as well as that of Shimer (2005) where workers conduct directed search and highly-skilled workers may apply to lower level jobs as they enjoy a higher probability of being employed. What is missing in these models is the explanation for why workers may have an excess of attributes on some dimensions but a lack in other dimensions, which necessitates the explicit modeling on the tradeoff between sorting in different dimensions. This is why the case of mismatch in the presence of multidimensional heterogeneity must be considered on its own terms.

Theories of multidimensional mismatch Lindenlaub (2014) simulates the frictionless assignment of workers and firms and shows that even in the frictionless assignment with positive assortative matching (PAM), there will always be some degree of mismatch since, unless very restrictive assumptions are made on worker and firm type distributions, the level of each worker skill element will not be exactly equal to the level of each corresponding element of job skill requirement. Although her model generates some mismatch, the frictionless assignment is unable to reproduce the mismatch observed in the data. Postel-Vinay and Lise (2015) construct a model of random job search, where workers skills decay or accumulate according to the job they obtain and where workers conduct OTJ search for a better match. Simulations of their model, however, have limited performance in fitting the data on the correlation of worker and job attributes.

Links to the literature on wage polarisation Using data from the Current Population Survey (CPS) Acemoglu and Autor (2011) record the phenomenon of wage polarisation in the US, where wages of low-skilled workers have experienced real declines along with an increase in the college premium over the period between 1973 and 2009. This, they show, has been accompanied by an increase in both high-skill and low-skill occupations and a decline in middle-skill occupations that have been linked to technological progress and offshoring. This has led to increasing numbers of middle-skill workers employed in low-skill jobs. They then propose a frictionless model where workers with low, medium or high skills are hired to perform, low, medium or high skilled tasks and study when happens when an interval of middle skill tasks are taken over by machines. A model allowing for two-sided multi-dimensional heterogeneity can provide a richer explanation of the phenomenon of wage polarisation by allowing for shocks to the productivity of specific elements of worker and firm characteristics, in a more nuanced form of skill-biased technological change.

Multi-dimensional human capital investment Sanders and Taber (2012) provide a brief outline a model of human capital investment where workers decide whether to spend their time working or investing in their any element of their human capital vector, which in turn affects their future wages. A comprehensive model of multidimensional skill mismatch should allow for the endogenisation of worker and firm types, with an examination of the efficiency not only of the sorting between workers and firms, but also of investments in human capital.

Gaps in the literature What the literature still lacks is a model that can sufficiently account for the sorting of workers into firms observed in the data when both sides have multi-dimensional characteristics. Moreover, as will be later shown, there is strong persistence in sorting. The idea that workers search randomly for their jobs, even if the job they match with has lasting impacts on their future job matches, is not too convincing. As such, how labour market entrants choose to match with their first job, including what dimensions they are more willing to be over or under matched in, merits closer examination. Moreover, the endogenisation of worker and firm types is essential to understanding, on the aggregate level, why there may be an surplus or deficit of certain skills in the labour market. This paper aim to address these gaps in the existing literature.

3 Some empirical findings

The question of how to define and measure the vector of worker characteristics and the corresponding requirements of occupations has been addressed by both Postel-Vinay and Lise (2015) and Lindenlaub (2014). Both use O*NET to construct the requirements of each job and use the NLSY to construct workers' skills, albeit differently. As described briefly earlier, Lindenlaub (2014) inferred worker skills from the jobs requirements while Postel-Vinay and Lise (2015) constructed worker skills directly from information provided in the NLSY, such as ASVAB test scores and self-esteem tests, and used the average skill requirements of jobs held by each educational group to impute the cognitive, manual and interpersonal skills of workers. I also use the NLSY and O*NET and construct a six dimensional vector of worker and firm characteristics - cognitive, interpersonal and physiological skills, field of knowledge, educational attainment and weeks of relevant work experience.

The data The NLSY dataset is well-known. It consists of panel surveys of individuals from two cohorts - 1997 and 1979. Individuals are interviewed annually or bi-annually up to the present day. I focus primarily on the 1979 cohort as this panel is much longer. As mentioned previously O*NET contains numerous descriptors of skills and knowledge sets required in each occupation, with information provided by analysts and worker surveys. I use the 15.0 installment of O*NET, which was released in 2010. An older version might have been more appropriate for the 1979 NLSY cohort, since job requirements do change over time, but the older versions of O*NET contain information on far fewer jobs.

The cognitive, interpersonal and physiological dimensions ONET contains 55 descriptors related to cognitive skills, 67 related to physiological skills and 32 related to interpersonal skills required by each job[7]. From the 1979 NLSY cohort, I use the individual's ASVAB test scores[8] and her parents' educational attainment levels to describe cognitive skills, bmi, general health and the presence or absence of physical handicaps to describe physiological skills, and use the respondent's

[7]Cognitive skill descriptors include 'written comprehension', 'critical thinking' and 'deductive reasoning', physiological skill descriptors include 'speed of limb movement', 'gross body coordination' and 'stamina', while interpersonal skill descriptors include 'social perception', 'persuasion' and 'negotiation'

[8]This includes a battery of tests that measures cognitive ability administered by the US army.

scores on Rosenberg esteem scale and Rotter-Locus of control[9] to describe interpersonal skills. I then use PCA separately on each set of descriptors to reduce them into three dimensions of characteristics for workers and firms. I normalise each of these characteristics, such that their measures lie between 0 and 1.

Knowledge Relevance Not only do jobs require certain skills of workers, but they also require them to have the relevant knowledge. O*NET also measures the relevance of 33 fields of knowledge to each job, ranging from 0 to 7, while the NLSY 1979 asks respondents for their college majors for those who ever attended college. I create a crosswalk between the college majors listed in the NLSY and the O*NET fields of knowledge. I then record the relevance of the individual's college major for each job they have occupied after leaving school for good. For now however, I leave out this interesting dimension, as including it would mean discarding all respondents who never went to college.

Educational attainment This was constructed straightforwardly. O*NET gives the minimum level of education required to perform each occupation. From the NLSY, I obtain the maximum educational level attained for each individual at the time when they leave school for good and enter the labour force. The NLSY also provides the number of weeks the individual spends at each job.

Example An example to illustrate what the vectors of workers and firm characteristics looks like concretely, consider the example of respondent 14 of the NLSY 1979 cohort. This is a woman with $X = (x_c, x_e, x_i, x_p) = (0.897, 5, 0.166, 0.0327)$. She worked as an accountant which has requirements $Y = (y_c, y_e, y_i, y_p) = (0.768, 4, 0.626, 0.0431)$. One can observe that with a Masters degree, $x_e = 5$, she is over-qualified for her job, which only requires a Bachelors degree $y_e = 4$. She also has an excess of cognitive skill, but a lack of interpersonal and physiological skills.

Sorting in the first job Table 1 shows the sorting between the characteristics of the 1979 NLSY respondents and their first job obtained after leaving school for good. Sorting is strongest in the educational dimension, with $corr(x_e, y_e) = 0.261$, followed by the cognitive dimension, with $corr(x_c, y_c) = 0.233$, the physiological dimension, $corr(x_p, y_p) = 0.01$, and lastly the interpersonal dimension $corr(x_i, y_i) = -0.147$. Table 1 also suggests some specialisation in worker skills and job skill requirements, with x_p and x_i negatively correlated with other the educational and cognitive dimensions, and y_p negatively correlated with the other Y dimensions, which are in turn positively correlated.

Sorting improvements over time There is very little improvement in sorting over time. This is despite the fact that there was an average of *over 15 years* between the respondents' first and last jobs, and an *average of 12 job changes* during that time. Table 2 shows the correlations between worker and firm characteristics during the most recent wave of interview in 2012. While there are improvements in sorting in the cognitive and education dimensions and minimal improvement in sorting in the physiological dimension, there is also a slight deterioration in sorting in the interpersonal dimensions. However, this finding may be partly explained by the fact that the NLSY data does not record any changes in the respondents cognitive skills and interpersonal skills since their related questions are only asked during the respondents' early youth. Furthermore, by definition, educational attainment does not change once an individual has left school permanently. Only the

[9]The Rosenberg esteem scale measures and individual's self-esteem while the Rotter-Locus of control measures a person's self-motivation and determination through a series of survey questions.

	xe	xc	xi	xp	yc	yp	yi	ye
xe	1							
xc	0.55	1						
xi	-0.28	-0.38	1					
xp	-0.4	-0.08	0.03	1				
yc	0.27	0.23	-0.17	-0.02	1			
yp	-0.2	-0.16	0.09	0.01	-0.52	1		
yi	0.22	0.19	-0.15	-0.02	0.91	-0.48	1	
ye	0.26	0.19	-0.14	-0.01	0.72	-0.56	0.63	1

Table 1: Correlations between worker and firm characteristics are obtained from data on the 1979 NLSY respondents in their first jobs obtained after leaving school for good.

	xe	xc	xi	xp	yc	yp	yi	ye
xe	1							
xc	0.55	1						
xi	-0.28	-0.38	1					
xp	-0.04	-0.08	0.03	1				
yc	0.4	0.39	-0.2	-0.04	1			
yp	-0.31	-0.29	0.14	0.02	-0.50	1		
yi	0.37	0.34	-0.18	-0.02	0.90	-0.50	1	
ye	0.41	0.3	-0.17	-0.02	0.70	-0.50	0.59	1

Table 2: Correlations between worker and firm characteristics are obtained from data on the NLSY 1979 respondents in their most recents jobs.

physiological descriptors are asked consistently through each wave of survey. Knowledge relevance is allowed to change, of course, when an individual changes occupations, but there is no way to know if the individual has acquired a different field of knowledge after leaving college.
That sorting improves minimally over time can also be garnered from Figures 1 and 2. Figure 1 shows the changes in the distribution of mismatch between the respondents first, middle and most recent jobs on all 4 dimensions, while Figure 2 shows the distribution of the changes in mismatch between individuals' first and most recent jobs on all 4 dimensions.

As the empirical findings suggest, the degree of sorting between worker and firm characteristics vary between the different dimensions, and sorting does not seem to improve very much on any dimension, even over a long time. Moreover, for each individual, the degree of mismatch between her skills and her job does not seem to change much between her first and most recent job. In fact, as gleaned from Figure 2, most individuals experience no change in their sorting over time. This lack of mobility in mismatch thus justifies the adoption of a simple static search model in the next section. As such, I adapt from the standard static model of directed search with heterogeneity in the likes of Shimer(2005), introducing multi-dimensional heterogeneity and mismatch costs. The framework of directed search over random search is chosen for 3 reasons: 1) it more realistically describes individuals' job search behaviour, 2) it is highly tractable and 3) it has yet to be explored in the context of multi-dimensional heterogeneity.

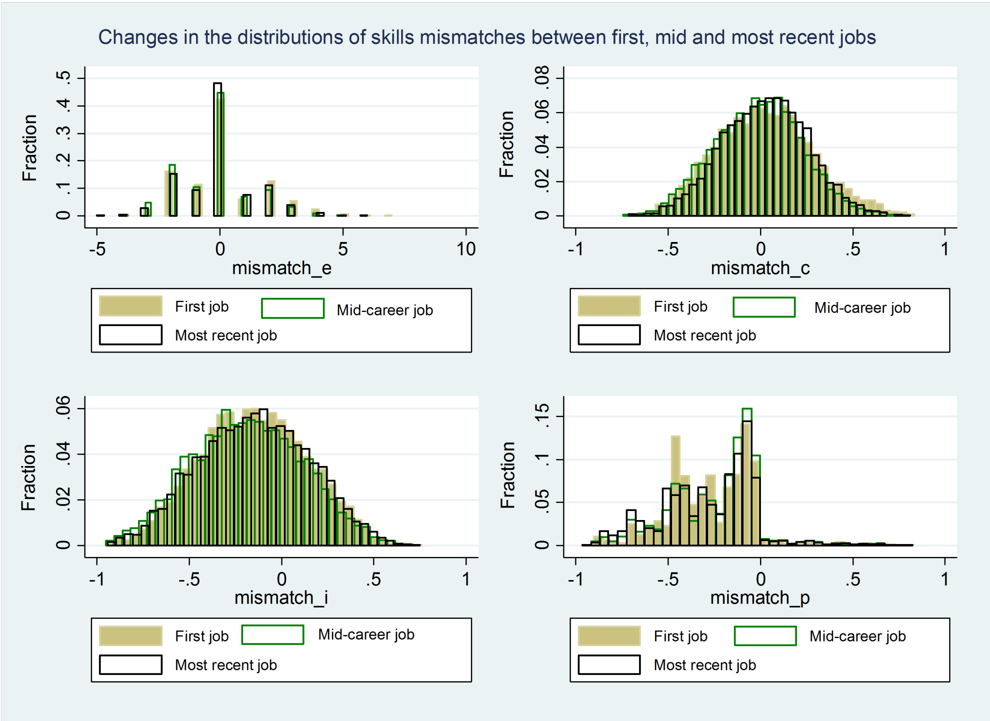


Figure 1: Shift in distribution of skills mismatch between the first job, middle job and the most recent job. 0 means no mismatch, anything to the right (left) implies that that the worker has more (less) of that attribute than what the job requires.

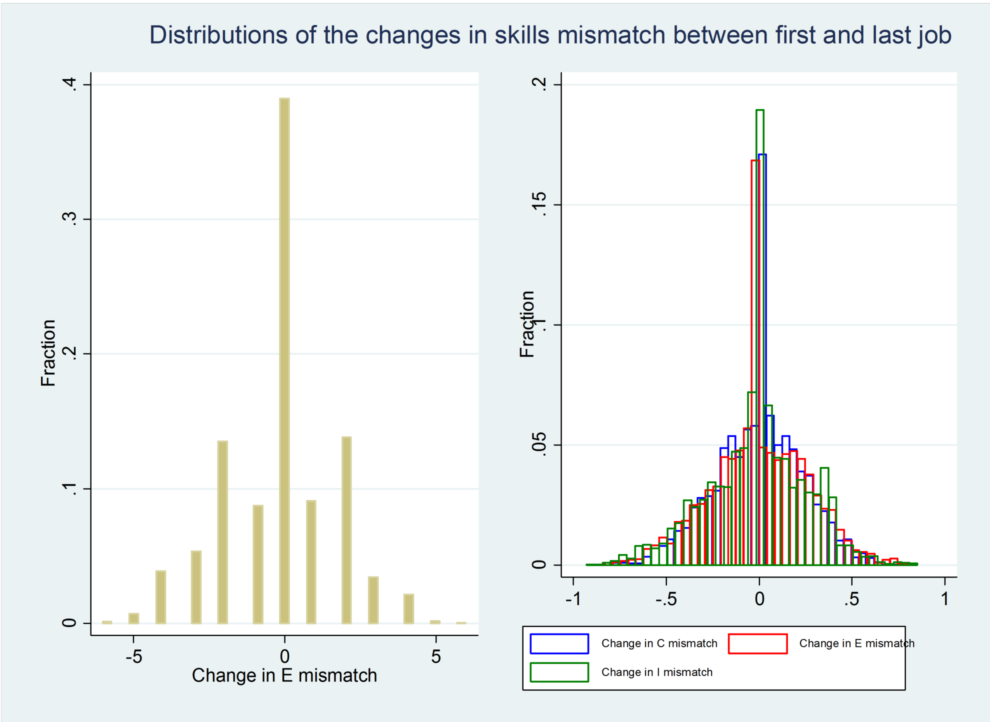


Figure 2: Distributions of the changes in skills mismatch between individuals' first job and the most recent job. 0 means no mismatch, anything to the right (left) implies that that the worker has more (less) of that attribute than what the job requires.

4 The basic model

4.1 The decentralised problem

Setup - workers There is a large number I of risk-neutral first-time workers. At the start of the period, they have just entered the workforce. Workers' characteristics are described by $X = [x_1 \quad x_2 \quad ... \quad x_K]^T$. I define these characteristics broadly as skills, where a skill is loosely defined as any worker attribute that enables the worker to perform his/her job. Let $\vartheta(X)$ be the number of each type X worker. A clarification on how the empirical Xs are considered is timely. With PCA, I had obtained values of x_c/y_c, x_p/y_p and x_i/y_i detailed to several decimal places. In what follows however, I round off these values to the first decimal place. This is done for 2 reasons: 1) to avoid the unhelpful complication of dealing with continuous types and 2) because it is implausible that workers or firms can distinguish or be distinguised beyond a few decimal places. Finally, assume that I is large enough such that workers are unable to coordinate amongst themselves.

Setup - Firms Similarly, jobs are defined by their job requirements $Y = [y_1 \quad y_2 \quad ... \quad y_K]^T$. Let $\gamma(Y)$ be the measure of jobs with characteristics Y and let the number of jobs in total be J, where J is large. Since a single firm can only host one job, I use the terms firm and job interchangeably. For now, I assume that the distributions of jobs and workers are exogenously given. As in Lindenlaub(2014), I define mismatch as follows: a worker and firm are mismatched in the kth dimension as long as $x_k \neq y_k, \forall k$. In addition, I say that there the worker is over-skilled in the kth dimension if $x_k > y_k, \forall k$, and that, conversely, she is under-skilled in the kth dimension if $x_k < y_k$.

The production function A one-to-one match between a worker and a firm results in a the production of a quantity of the final good given by $\Phi(X,Y)$, where I allow $\Phi(X,Y)$ to be potentially affected by the mismatch between workers characteristics and firm requirements, i.e. by $x_k - y_k, \forall k$.

The job search process As in the standard directed search model, a firm hosting job type Y posts a wage $w(X,Y)$ for each type of worker it observes. Let $p(Y|X)$ be the probability that a worker with X characteristics applies to type Y firm. In a given period, the worker can only make a single application to a firm. A given firm then receives a type X worker with probability $q(X,Y) = \frac{P(Y|X)\vartheta(X)}{\gamma(Y)}$. The worker's application probabilities over all firms is such that it maximises the worker's expected utility.

Ranking over workers In order to solve the worker's problem, it is essential to first define the ranking over workers by each firm. There are two notions of ranking that must be distinguished. A ranking is *objective* if it is shared by firms of all types over workers or if conversely, it is shared by workers of all types over firms. On the other hand, a ranking is *subjective* if the ranking over workers by type Y firm is not the same as the ranking over workers by type Y' firm for some $Y' \neq Y$ or if conversely, the ranking over firms by type X is not the same as that by type X' for some $X' \neq X$. While in the unidimensional case with no mismatch cost, as in Shimer(2005), all firms share an objective ranking over workers, the ranking of workers by firms in the multidimensional case is not so clear-cut. For example, how would a firm rank worker $X = \left(\begin{smallmatrix}0.7\\0.6\\0.5\end{smallmatrix}\right)$ against worker $X' = \left(\begin{smallmatrix}0.6\\0.5\\0.7\end{smallmatrix}\right)$? Unless strong restrictions on the distribution of worker and firm types as well as the production function are imposed, a common ranking over workers by firms does not exist. However, one can show that a firm with job Y prefers worker X to X' only if $\Phi(X,Y) > \Phi(X',Y)$ when workers are risk neutral. Following the notation of Shimer (2005), let $Q(X,Y)$ be the expected number of workers preferred to worker X applying to firm Y.

$$Q(X,Y) = \sum_{X'} \mathbb{1}_{\{\Phi(X',Y)>\Phi(X,Y)\}} \frac{p(Y|X')\vartheta(X)}{\gamma(Y)} \tag{1}$$

For simplicity, assume that there is no instance of $\Phi(X,Y) = \Phi(X',Y)$ for $X \neq X$, though this does not affect the result.

The worker's problem The probability that a worker of type X gets the job given he/she applies can be written as $e^{-Q(X,Y)}\frac{(1-e^{-q(X,Y)})}{q(X,Y)}$. This is essentially the probability that no one preferred to X applies and that this individual X gets the job even if other individuals of the same type apply. As such, the expected payoff of a worker to a firm is simply $e^{-Q(X,Y)}\frac{(1-e^{-q(X,Y)})}{q(X,Y)}w(X,Y)$. As is standard in the directed search literature, the worker only applies with positive probability to a job Y if it offers her highest expected utility. Letting the maximimum expected utility of worker X be $U(X)^*$, $p(Y|X) > 0$ only if

$$U(X)^* = e^{-Q(X,Y)}\frac{(1-e^{-q(X,Y)})}{q(X,Y)}w(X,Y) \tag{2}$$

The firm's problem The probability that the most preferred worker a firm of type Y gets is X is $(1-e^{-q(X,Y)})e^{-Q(X,Y)}$. Hence, the expected output of the firm can be written as

$$\sum_{X}(1-e^{-q(X,Y)})e^{-Q(X,Y)}\{\Phi(X,Y) - w(X,Y)\} \tag{3}$$

The firm maximizes this with respect to $w(X,Y)$ subject to the equation 2 for the X types that it wishes to hire. The first order condition gives

$$U(X)^* = e^{-q(X,Y)}e^{-Q(X,Y)}\Phi(X,Y) - \sum_{X'} \mathbb{1}_{\{\Phi(X,Y)>\Phi(X',Y)\}}(1-e^{-q(X',Y)})e^{-Q(X',Y)}\Phi(X',Y) \tag{4}$$

Substituting this into equation 2, the $w(X,Y)$ can be expressed as follows:

$$w(X,Y) = \frac{q(X,Y)}{1-e^{-q(X,Y)}}\{e^{-q(X,Y)}\Phi(X,Y) - \sum_{X'} \mathbb{1}_{\{\Phi(X,Y)>\Phi(X',Y)\}}(1-e^{-q(X',Y)})\frac{e^{-Q(X',Y)}}{e^{-Q(X,Y)}}\Phi(X',Y)\} \tag{5}$$

$\frac{e^{-Q(X',Y)}}{e^{-Q(X,Y)}}$ is essentially the probability that no one better than X' applies given that no one better than X applies, and the expression for $w(X,Y)$ can thus be interpreted as being equivalent to the marginal output of X at job Y. Unless there are strong restrictions placed on the distribution of worker types and the form of the production function, the inability to coordinate amongst themselves compels workers to adopt mixed job application strategies.

4.2 The constrained social planner problem

Having considered the decentralised case, it is pertinent to consider if the assignment is efficient.

Let us therefore assume that the Social Planner cannot distinguish between workers of the same type X and hence can only issue the following order to all workers, " If you are type X, adopt application strategy $p(Y|X)$" for each worker type X over all firms Y. These instructions are given to maximise total output,

$$W = \sum_Y \gamma(Y) \sum_X (1 - e^{-q(X,Y)}) e^{-Q(X,Y)} \Phi(X,Y)$$

$$st \quad \sum_Y p(Y|X) = 1, \quad p(Y|X) \geq 0 \quad (6)$$

As in the decentralised case, the social planner tells the firm to prefer X to X' if $\Phi(X,Y) > \Phi(X',Y)$ $\forall X$. The resource constraint, which says that the expected number of workers of type X queuing at all firms equals to the total expected number of workers of type X, writes as $\sum_Y \gamma(Y) q(X,Y) = \vartheta(X) \leftrightarrow \sum_Y p(Y|X) = 1$. Letting $U(X)^* \vartheta(X)$ be the multiplier on this constraint, the first order condition gives

$$U(X)^* \geq e^{-q(X,Y)} e^{-Q(X,Y)} \Phi(X,Y) - \sum_{X'} \mathbb{1}_{\{\Phi(X,Y) > \Phi(X',Y)\}} (1 - e^{-q(X',Y)}) e^{-Q(X',Y)} \Phi(X',Y)$$

$$P(Y|X) \geq 0 \quad (7)$$

With complementary slackness, which gives the same assignment as in the decentralised case when realised types are known. Yet, while the equilibrium is constrained efficient, the presence of mismatch is certainly not precluded, though extent of mismatch would depend on the distribution of worker and firm types as well as the production function.

5 Illustrative examples with closed-form solutions

With the generic specification above, there is no clear idea of what workers' application strategies (and hence their probability of being mismatched on each skill dimension) look like in equilibrium. Moreover, until this point, how having multi-dimensional heterogeneity instead of just unidimensional heterogeneity changes the equilibrium assignment has not been shown[10].I therefore consider a few illustrative cases that highlight the differences that introducing multi-dimensionality, as well as mismatch costs, bring. For simplicity, I consider just 2 dimensions, but the logic behind could be extended to $K \geq 2$ dimensions. All proofs are in the Appendix.

5.1 Case 1: $\Phi(X,Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2$

In this simple case, there is no mismatch cost. As such, the notion of under/over-skilled does not imply much, since workers with higher x_1, x_2 simply produce more at any given firm Y, $\forall Y$. Insofar as a higher level of a worker's skill enhances the productivity of a job task requiring that skill, the production function can be written as $\Phi(X,Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2$, where $\Phi(X,Y) = 0$ if a the firm hires no one (i.e. $X_0 = [0,0]^T$). Note that if $\beta_2 = 0$, meaning that the second task and skill do not matter for production, or if $x_1 = x_2 \quad \forall X$ and $y_1 = y_2 \quad \forall Y$, the equilibrium application probabilities return to the case in Shimer (2005), where $\Phi(X,Y) = x_1 y_1$. As the equilibrium sorting cannot be determined without specifying the distribution of workers and firms, in particular $corr(x_1, x_2)$ and $corr(y_1, y_2)$, the I derive the equilibrium application probabilities for a few possible specifications.

[10]Indeed if setup was in the model presented above was altered such that workers and firms were only unidimensionally heterogenous, the first order conditions would exactly ressemble that in Shimer (2005).

When $corr(x_1, x_2) = corr(y_1, y_2) = 1$ **:** In this simplest example, there is an objective ranking of workers and firms as $\Phi(X, Y)$ is increasing in both $||X||$ and $||Y||$. The production function is also supermodular. I show that $Q(X, Y)$ is increasing in $||Y||$ when it > 0 and $w(X, Y)$ is increasing in $||Y||$ when $Q(X, Y) > 0$. Workers adopt a threshold application strategy, with worker X applying to some firm with $||Y|| \geq Y_X$, where Y_X is increasing in X. Let X_i denote the worker who has the ith lowest $||X||$. X_i's application probabilities of a worker to a job of type Y can be written as

$$q(X,Y) = \begin{cases} log\frac{\beta_1 y_1(x_{1i}-x_{1i-1})+\beta_2 y_2(x_{2i}-x_{2i-1})}{\beta_1 y_{1X_{i-1}}(x_{1i}-x_{1i-1})+\beta_2 y_{2X_{i-1}}(x_{2i}-x_{2i-1})} \\ \text{if} \quad Y_{X_i} \leq ||Y|| < Y_{X_{i+1}}, \\ log\{\frac{\beta_1 y_1(x_{1i}-x_{1i-1})+\beta_2 y_2(x_{2i}-x_{2i-1})}{\beta_1 y_{1X_{i-1}}(x_{1i}-x_{1i-1})+\beta_2 y_{2X_{i-1}}(x_{2i}-x_{2i-1})} \frac{\beta_1 y_{1X_i}(x_{1i+1}-x_{1i})+\beta_2 y_{2X_i}(x_{2i+1}-x_{2i})}{\beta_1 y_1(x_{1i+1}-x_{1i})+\beta_2 y_2(x_{2i+1}-x_{2i})}\} \\ \text{if} \quad ||Y|| \geq Y_{X_{i+1}} \end{cases} \tag{8}$$

Where $Y_{X_i} = [y_{1X_i}, y_{2X_i}]^T$ is the lowest $||Y||$ firm to which x_i applies, $X_{i-1} = [x_{1i-1}, x_{2i-1}]^T$ is the worker with $||X||$ just one rank below that of X_i and where $X_{i+1} = [x_{1i+1}, x_{2i+1}]^T$ is the worker ranked just above X_i. Note that for the worker with the lowest $||X||$, $||\underline{X}||$, the 'person' below her is just $||\underline{X_1}|| = ||[0,0]^T||$. Likewise, the threshold $||Y||$ for this non-person is simply $||\underline{Y}||$, the firm with the lowest magnitude of requirements. In addition, the probability that Y hires type X is

$$e^{-Q(X,Y)} = \begin{cases} log\frac{\beta_1 y_1(x_{1i}-x_{1i-1})+\beta_2 y_2(x_{2i}-x_{2i-1})}{\beta_1 y_{1X_{i-1}}(x_{1i}-x_{1i-1})+\beta_2 y_{2X_{i-1}}(x_{2i}-x_{2i-1})} \\ \text{if} \quad Y_{X_i} \leq ||Y|| < Y_{X_{i+1}}, \\ log\{\frac{\beta_1 y_1(x_{1i}-x_{1i-1})+\beta_2 y_2(x_{2i}-x_{2i-1})}{\beta_1 y_{1X_{i-1}}(x_{1i}-x_{1i-1})+\beta_2 y_{2X_{i-1}}(x_{2i}-x_{2i-1})} \frac{\beta_1 y_{1X_i}(x_{1i+1}-x_{1i})+\beta_2 y_{2X_i}(x_{2i+1}-x_{2i})}{\beta_1 y_1(x_{1i+1}-x_{1i})+\beta_2 y_2(x_{2i+1}-x_{2i})}\} \\ \cdot \frac{\beta_1 y_1(x_{1i+1}-x_{1i})+\beta_2 y_2(x_{2i+1}-x_{2i})}{\beta_1 y_{1X_i}(x_{1i+1}-x_{1i})+\beta_2 y_{2X_i}(x_{2i+1}-x_{2i})} \\ \text{if} \quad ||Y|| \geq Y_{X_{i+1}} \end{cases} \tag{9}$$

Using the two expressions then, one can obtain the probability that a given person of type X by firm Y, $e^{-Q(X,Y)}(1 - e^{-q(X,Y)})/q(X, Y)$. Also, it can be observed that the application probabilities are dependent on $||Y||$, with $q(X, Y)$ increasing in $||Y||$ for $Y_{X_i} \leq ||Y|| < Y_{X_{i+1}}$. In the case of $||Y|| \geq Y_{X_{i+1}}$, the effect of $||Y||$ is ambiguous and requires more information on the distribution of workers types to become clearer. Likewise, while it is clear that shifts in β_1 and β_2 will affect the application and hiring probabilities, the magnitude and direction of their impacts can only be determined with more information of the distribution of worker types.

Illustration Figures 3a and 4a show the equilibrium application probabilities when each worker has $x_1 = x_2$ and each firm has $y_1 = y_2$. As shown, workers do adopt the aforementioned threshold strategy when applying to firms, with the higher type workers applying only to the higher firm types with positive probability. In equilibrium, higher worker types are most likely to be assigned to higher firm types, although lower worker types still enjoy a positive probability of being assigned to higher firm types. In this specification, an increase is β_1 relative to β_2 does not seem to change the equilibrium application and assignment probabilities much.

When $||X||$ **is equal** $\forall X$ **and** $||Y||$ **is equal** $\forall Y$ **st** $x_1 = ||X|| - x_2$ **&** $y_1 = ||Y|| - y_2$**:** Assume for now that $\beta_1 = \beta_2 = 1$. In this specification, the magnitude of the vector of human capital is the same for all workers while the magnitude of job requirements is the same for all jobs. The difference between workers and firms then lies with the relative sizes of x_1 and x_2 and the relative sizes of y_1 and y_2. Here the notion of supermodularity carries little weight here and there is no objective ranking over workers or firms. In this case one can derive an intuitive overview of how the equilibrium application probabilities would look like. Split the firm into three types, the firms with $y_1 = y_2$, the firms with $y_1 > y_2$ and those with $y_2 > y_1$.

Consider the group of firms with $y_1 > y_2$. Using the FOC from the previous section that gives the expression for $U^*(X)$, one can show that the expected queue length of people better than X at any firm in this group, $Q(X, Y)$, weakly increases with $\frac{y_1}{y_2}$ $\forall X$ when $Q(X, Y) > 0$. By symmetry, $Q(X, Y)$ increases with $\frac{y_2}{y_1}$ $\forall X$ at the group of firms with $y_2 > y_1$. For the firm with $y_1 = y_2$, all workers result in the same amount of output and so is indifferent between all workers. From the worker's problem, a higher $Q(X, Y)$ must be accompanied by a higher posted wage $w(X, Y)$ if the firm is to induce the worker to apply to it with positive probability. However, firms with very high $\frac{y_1}{y_2}$ are not going to offer a high wage to workers with very low $\frac{x_1}{x_2}$ even if they face a low probability of finding the job. As such, those workers will not apply to these firms at all. For a clearer picture, consider the worker with the highest value of $x_2 = \bar{x_2}$, and consequently the lowest value of $x_1 = \underline{x_1}$. Consider also the firm sith the highest value of $y_1 = \bar{y_1}$ and the lowest value of $y_2 = \underline{y_2}$. This firm prefers this worker the least out of all types, and given that it produces so little with this worker, may not be unable to offer a high enough wage without profits becoming negative, to compete with the wages posted by firms with higher y_2.

Using the above reasoning, one can deduce that all workers adopt threshold application strategies, albeit of a more complicated kind than in the case when the first and second dimensions of attributes were positively correlated. Arrange workers on a line according to their x_1, ranging from the worker with the lowest x_1 (and highest x_2) on the left and the workers with the highest x_1 (and lowest x_2) on the right. On a parallel line, arrange firms according to the y_1 in the same logic. The worker with the highest x_1 applies with positive probability only to the right of her threshold $y_{1\bar{x_1}}$. As for worker to her immediate left, she also applies with positive probability only to firms on the right of her threshold, which is to the left of $y_{1\bar{x_1}}$. Similar reasoning applies to workers nearby to the left. At the extreme lefthandside of the spectrum, the worker with the lowest x_1 applies with positive probability only to the left of her threshold $y_{1\underline{x_1}}$ (i.e. she only applies to firms with high enough y_2.). The worker on her immediate right also applies with positive probability only to firms to the left of her threshold, which is to the right of $y_{1\underline{x_1}}$. The same goes for workers nearby on the right. As for the workers in the middle, there is a set of workers around the middle section of the line who apply to all firm types with positive probability as they're offered their highest expected payoff at all firms.

While reasoning behind these threshold application strategies is quite intuitive, a trickier task would be obtaining the closed-form expressions of these strategies. Unlike in the case of perfect correlation between the 2 dimensions, the thresholds of different types overlap, meaning that there is no set of firms that are only applied to by one type of worker. This significantly complicates the calculation of the closed-form expressions. Nonetheless, the results from simulations given in Figure 5a illustrates the threshold application strategies, with workers applying to the left or right of their y_1 threshold depending on whether their x_1 is lower or higher relative to their x_2. One can also observe those with intermediate levels of both skills applying to all y_1 values with positive probability.

Illustration Figures 5a,6a,7a and 8a illustrate the partioning of the labour market in two, where workers with relatively higher x_1 (x_2) applying *and* being assigned almost exclusively to firms with relatively higher y_1 (y_2). When $\beta_1 > \beta_2$, those with relatively lower $\frac{x_1}{x_2}$ seem to focus their applications more towards lower $\frac{y_1}{y_2}$ job types while conversely, those with relative higher $\frac{x_1}{x_2}$ focus their applications more towards higher $\frac{y_1}{y_2}$ job types.

Other worker/firm type distributions : The cases of $corr(x_1, x_2) = corr(y_1, y_2) = 1$ and $corr(x_1, x_2) = corr(y_1, y_2) = -1$ may seem too extreme and unrealistic. In the data for instance, one observes moderate negative correlation between the physiological and cognitive dimensions, and strong but not perfect correlation between the cognitive and education dimensions. As such, Part **a** of Figures 9 to 16 show the equilibrium application probabilities and assignment when dimensions are moderately negatively and positively correlated, and when $\Phi(X, Y) = x_1y_1 + x_2y_2$. Part **b** of Figures 9 to 16 show the same objects but for the case when the first dimension is five times more productive than the second one.

5.2 Cases 2 and 3: cost of underskilling

What about mismatch costs? Consider 2 more specifications of the production function, Case 2 -

$$\Phi(X, Y) = \beta_1x_1y_1 + \beta_2x_2y_2 - \alpha_1^u \mathbb{1}_{\{x_1<y_1\}}(y_1 - x_1) - \alpha_2^u \mathbb{1}_{\{x_2<y_2\}}(y_2 - x_2)$$

and Case 3 -

$$\Phi(X, Y) = \beta_1x_1y_1 + \beta_2x_2y_2 - \alpha_1(y_1 - x_1) - \alpha_2(y_2 - x_2)$$

In both cases, production suffers a extra penalty when workers have less than the required skills for the job. The severity of this penalty depends on the magnitudes of α_1 and α_2. The difference between the two cases lies with the added impact (or not) of having a worker having an excess of skill relative to the job requirement. While in case 2 an over-skilled worker (i.e. with $x_1 > y_1$ and/or $x_2 > y_2$) at Y does not have any additional effect on production apart from a higher $\beta_1x_1y_1 + \beta_2x_2y_2$ than a worker who is not over-skilled, in case 3 she contributes an added $\alpha_k(x_k - y_k)$ to production if she is over-skilled in dimension k. In case 3 then, an over-skilled worker actually contributes an added mismatch 'bonus', while an under-skilled worker gives a mismatch 'penalty' in terms of output [11]. Despite their differences, the production functions in cases 2 and 3 can be treated together.

For simplicity I re-adopt the assumption that $corr(x_1, x_2) = corr(y_1, y_2) = 1$. This assumption allows for an objective ranking over workers and firms. It is worth noting that with these specifications, the production functions are supermodular. For both cases, one can show that $Q(X, Y)$ and $w(X, Y)$ are increasing in Y and that workers adopt threshold application strategies, with worker X applying to Y only if $||Y|| \geq Y_X$, where Y_X is increasing in $||X||$. Let us rank workers according to their $||X||$ in increasing order. The ith ranked worker from the bottom, worker X_i, adopts the following application strategies:

$$\frac{p(Y|X)}{\gamma(Y)} = \begin{cases} log\frac{\Phi(X_i,Y)-\Phi(X_{i-1},Y)}{\Phi(X_i,Y_{X_{i-1}})-\Phi(X_{i-1},Y_{X_{i-1}})} \text{if} \quad Y_{X_i} \leq ||Y|| < Y_{X_{i+1}}, \\ log\{\frac{\Phi(X_i,Y)-\Phi(X_{i-1},Y)}{\Phi(X_i,Y_{X_{i-1}})-\Phi(X_{i-1},Y_{X_{i-1}})} \frac{\Phi(X_{i+1},Y_{X_i})-\Phi(X_i,Y_{X_i})}{\Phi(X_{i+1},Y)-\Phi(X_i,Y)}\} \\ \text{if} \quad ||Y|| \geq Y_{X_{i+1}} \end{cases} \tag{10}$$

[11] Certainly, having a mismatch 'bonus' for the over-skilled obfuscates the notion of mismatch, since firms would then always prefer to hire over-skilled workers rather than workers with just the right amount of skills, which begs the question as to why over-skilled workers are considered **over**-skilled in the first place. Nonetheless, it is not implausible that workers with extra skills on any dimension may contribute a little more in terms of production (i.e. α_1 and α_2 small), which justifies the need for case 3

For the cases of $\Phi(X,Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2 - \alpha_1^u \mathbb{1}_{\{x_1<y_1\}}(y_1 - x_1) - \alpha_2^u \mathbb{1}_{\{x_2<y_2\}}(y_2 - x_2)$ as well as $\Phi(X,Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2 - \alpha_1(y_1 - x_1) - \alpha_2(y_2 - x_2)$. Where $\Phi(X_{i-1}, Y)$ and $\Phi(X_{i+1}, Y)$ are the outputs of the person with $||X||$ just below and just above X_i at Y respectively, and where Y_{X_i} and $Y_{X_{i+1}}$ are the threshold Y job types beyond which X_i and X_{i+1} will not apply respectively. Note that for the worker with the lowest $||X||$, $||\underline{X}||$, the 'person' below her is just $||\underline{X_1}|| = ||[0,0]^T||$. Likewise, the threshold $||Y||$ for this non-person is simply $||\underline{Y}||$, the firm with the lowest magnitude of requirements. The probability that a job of type Y hires a type X worker is therefore

$$\frac{p(Y|X)}{\gamma(Y)}.e^{-Q(X,Y)} = \begin{cases} log\frac{\Phi(X_i,Y)-\Phi(X_{i-1},Y)}{\Phi(X_i,Y_{X_{i-1}})-\Phi(X_{i-1},Y_{X_{i-1}})} \text{if} \quad Y_{X_i} \le ||Y|| < Y_{X_{i+1}}, \\ log\{\frac{\Phi(X_i,Y)-\Phi(X_{i-1},Y)}{\Phi(X_i,Y_{X_{i-1}})-\Phi(X_{i-1},Y_{X_{i-1}})}\frac{\Phi(X_{i+1},Y_{X_i})-\Phi(X_i,Y_{X_i})}{\Phi(X_{i+1},Y)-\Phi(X_i,Y)}\}.\frac{\Phi(X_{i+1},Y_{X_i})-\Phi(X_i,Y_{X_i})}{\Phi(X_{i+1},Y)-\Phi(X_i,Y)} \\ \text{if} \quad ||Y|| \ge Y_{X_{i+1}} \end{cases} \tag{11}$$

These closed-form expressions for the equilibrium application probabilities and probabilities of being hired can be applied to any $\Phi(X,Y)$ that is strictly supermodular when there is an objective ranking of worker and firm types, which in this case, is attained by the assumption that $corr(x_1,x_2) = corr(y_1,y_2) = 1$. The equilibrium application and assignment probabilities in this specification can be viewed in Figures 3c and 4c. As can be seen, including the cost of being underskilled leads to a lowering of application thresholds. However, higher X types still apply with greater probability to higher Y types. Also, lower X types now apply with greater probability to lower Y types than in the first case without underskilling costs. All worker types are also assigned with greater probability to lower firm types.

Illustration Part **c** of Figures 5 to 16 show the equilibrium assignment and application probabilities when $corr(x_1,x_2) = corr(y_1,y_2) = -1$, when $corr(x_1,x_2) = corr(y_1,y_2) = 0.6$ and when $corr(x_1,x_2) = corr(y_1,y_2) = -0.5$.

5.3 Case 4 - $\Phi(X,Y) = c - (x_1 - y_1)^2 - (x_2 - y_2)^2$ (No objective ranking)

In this example, output suffers a mismatch penalty when $x_k > y_k$ or when $x_k < y_k$ for any dimension of skill k. Being over-skilled here therefore creates a loss in output, unlike case 3. It is clear that there is no objective ranking of workers here, even if $corr(x_1,x_2) = corr(y_1,y_2) = 1$, as each firm Y prefers the worker who has the X closest to Y, $\forall Y$. For simplicity, I assume that c is the constant quantity of output produced when there is no mismatch, independent of X and Y. If $c = 0$, for instance, firms only produce non-negative output (0) if $x_1 = y_1$ and $x_2 = y_2$. Assuming that firms will choose to hire if they are indifferent between not hiring and hiring, both of which yield 0 output, a firm Y hires the worker X only if there is perfect match, $X = Y$. Conversely, if $c \to \infty$, firms rank all workers equally. As such $Q(X,Y)$ and $w(X,Y)$ are unchanging with Y and workers apply with probability $\frac{1}{\sum_Y \gamma(Y)} = \frac{1}{J}$ to all jobs.

What happens for a moderate level of c? While deriving the analytical expressions for the equilibrium application and assignment probabilities is a challenge, one can derive a sketch of what these may look like. I use numerical methods to obtain these under four different worker/firm distributions: 1) $corr(x_1,x_2) = corr(y_1,y_2) = 1$, 2) $x_1 = 1 - x_2$, $y_1 = 1 - y_2$, such that $corr(x_1,x_2) = corr(y_1,y_2) = -1$, 3) $corr(x_1,x_2) = corr(y_1,y_2) = 0.6$ and 4) $corr(x_1,x_2) = corr(y_1,y_2) = -0.5$.

Illustration Case **d** in Figures 3 to 16 shows the equilibrium assignment and application probabilities in these 4 worker/firm distribution specifications. For specifications 1,2 and 4, mismatch costs on both dimensions lead to workers applying and being assigned with higher probabilities to firms whose y_1 (y_2) is closest to their x_1 (x_2). This is most stark in specification 4, as seen in Figures 13d, 14d, 15d and 16d.

To conclude this section, the above examples illustrate the differences in the equilibrium in the presence of multi-dimensional heterogeneity compared to unidimensional heterogeneity. The mechanisms behind the determination of the equilibrium assignment are much richer in the multi-dimensional case, with relative skill productivities, the degree of skills specialisation of firms and workers and the distribution of their types playing important roles. The introduction of mismatch costs also affects the pattern of sorting and allows for a more nuanced picture of the interaction between worker and firm types.

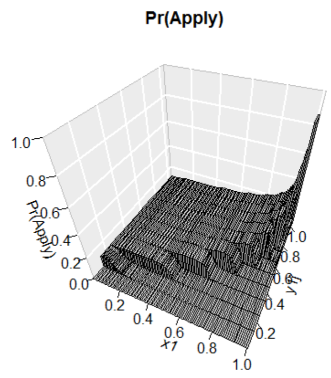
6 Estimation

For the estimation, I use a subset of the observations, namely those who leftschool in the period between January 1983 and July 1990, which was a period of growth, resulting in a sample of 184 respondents. This was done for 2 reasons: 1) The initial sample size of 1979 resulted in matrices that could not be handled easily by the programme and 2) one can expect to find different estimation results when using observations from different parts of the business cycles. Individuals are considered to have 'failed' in their first job search attempt if they experienced an unemployment spell before finding their first job after leaving school, before finding the first job. Among the respondents, 29 percent of them are considered to have 'failed' in their first job search attempt. The others found their first job instantly.

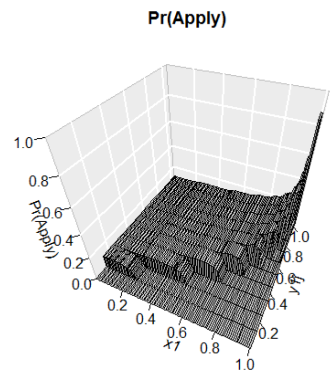
Before proceeding to the estimation, it is useful to skim over the empirical equilibrium assignment Figures 17 and 18 give the number of workers with given values of x_k assigned to firms with given values of y'_k. If the model is a good fit, it should produce assignments that do not deviate much from those that are empirically observed.

In the data, one does not observe the posted wages, the application strategies, the production function and the number of each type Y of job that exists. What one can observe are the accepted wages, existing worker types X, the types Y of the jobs they are assigned to, as well as the types and numbers of workers who experienced an unemployment spell before finding their first job. The estimation of the original Shimer(2005) model has been done in Abowd et al. (2012), but two striking differences exist between their estimation procedure and that discussed here: 1) With matched employer-employee data, Abowd et al. (2012) were unable to observe not only their workers' and firms' types, but they were also unable to observe workers who were unemployed. They therefore had to estimate worker and firm types following the wage decomposition method in Abowd et al. (1999) before estimating the production function. 2) They focus exclusively on the context of unidimensional heterogeneity, which greatly simplifies their estimation of the production function.

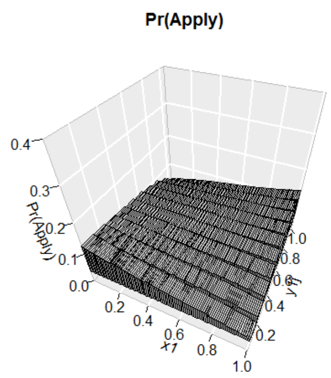
Parameters to estimate : As such, the key parameters to estimate are the production function parameters and the numbers of each type Y job $\gamma(Y)$. With these estimates, one can then obtain information on the posted wages. While all the possible job types are given in the O*NET, it is impossible to identify the $\gamma(Y)$s for the Ys that did not hire any NLSY respondent in the data.



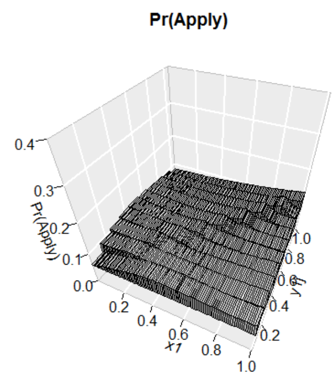
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

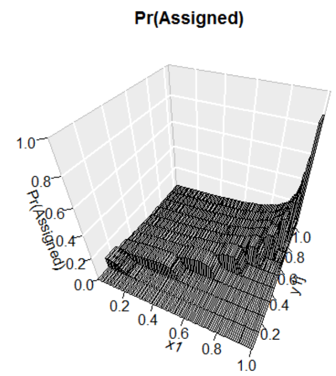


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

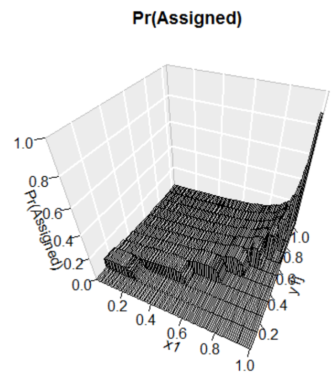


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

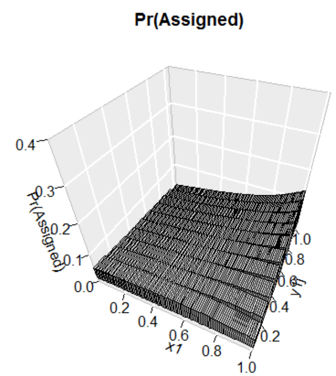
Figure 3: Application probabilities under different production functions when $corr(x1, x2) = corr(y1, y2) = 1$



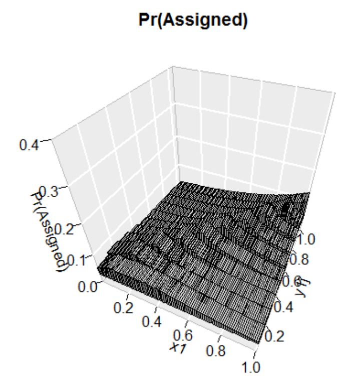
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

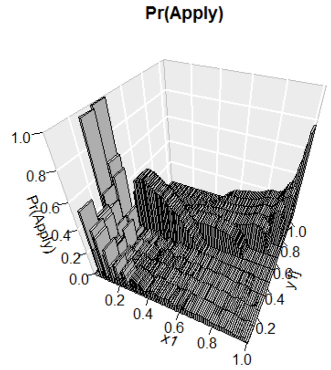


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

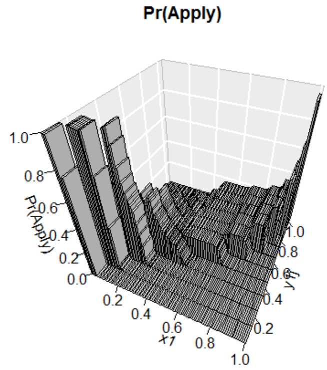


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

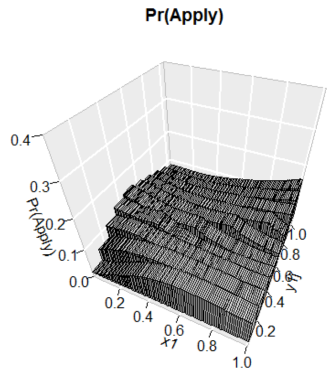
Figure 4: Assignment probabilities under different production functions when $corr(x1, x2) = corr(y1, y2) = 1$



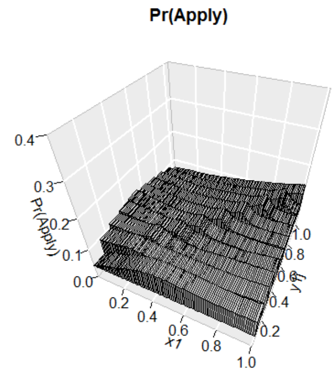
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

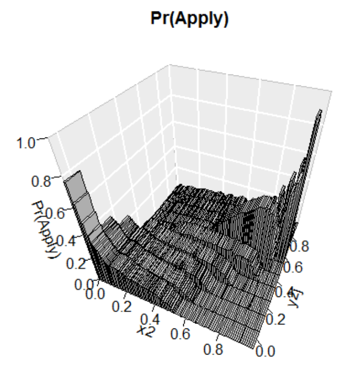


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

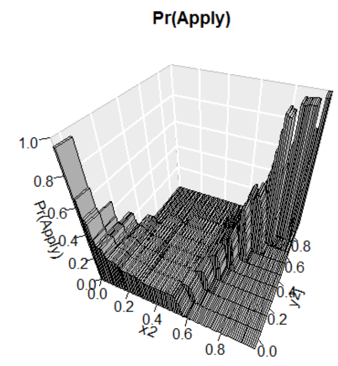


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

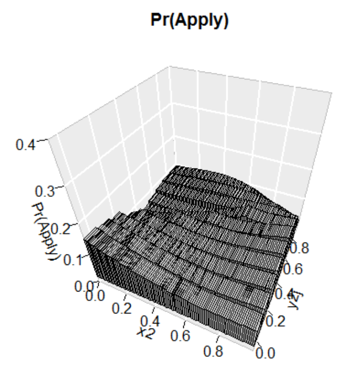
Figure 5: Application probabilities of x_1 to y_1 under different production functions when $x_1 = 1 - x_2, y_1 = 1 - y_2$



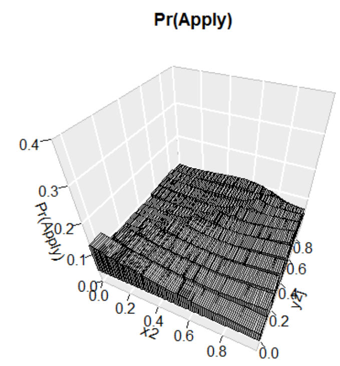
(a) $x_1 y_1 + x_2 y_2$



(b) $5x_1 y_1 + x_2 y_2$

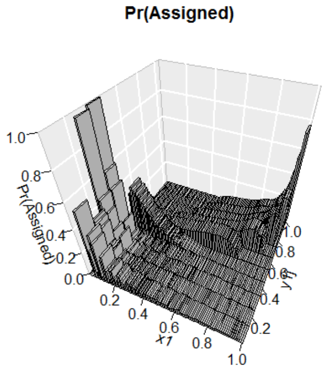


(c) $x_1 y_1 + x_2 y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

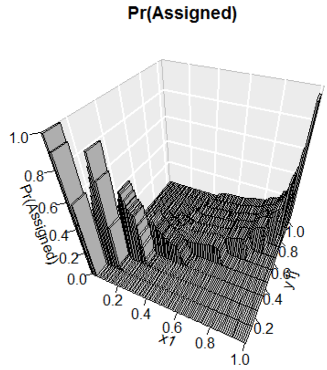


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

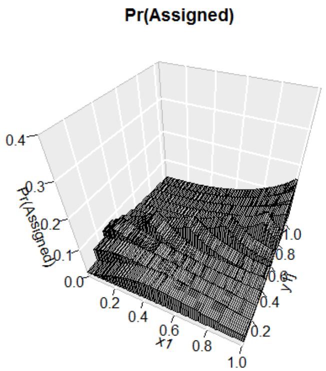
Figure 6: Application probabilities of x_2 to y_2 under different production functions when $x_1 = 1 - x_2, y_1 = 1 - y_2$



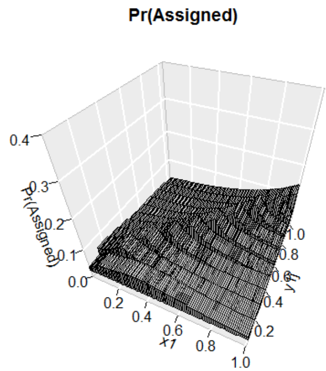
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

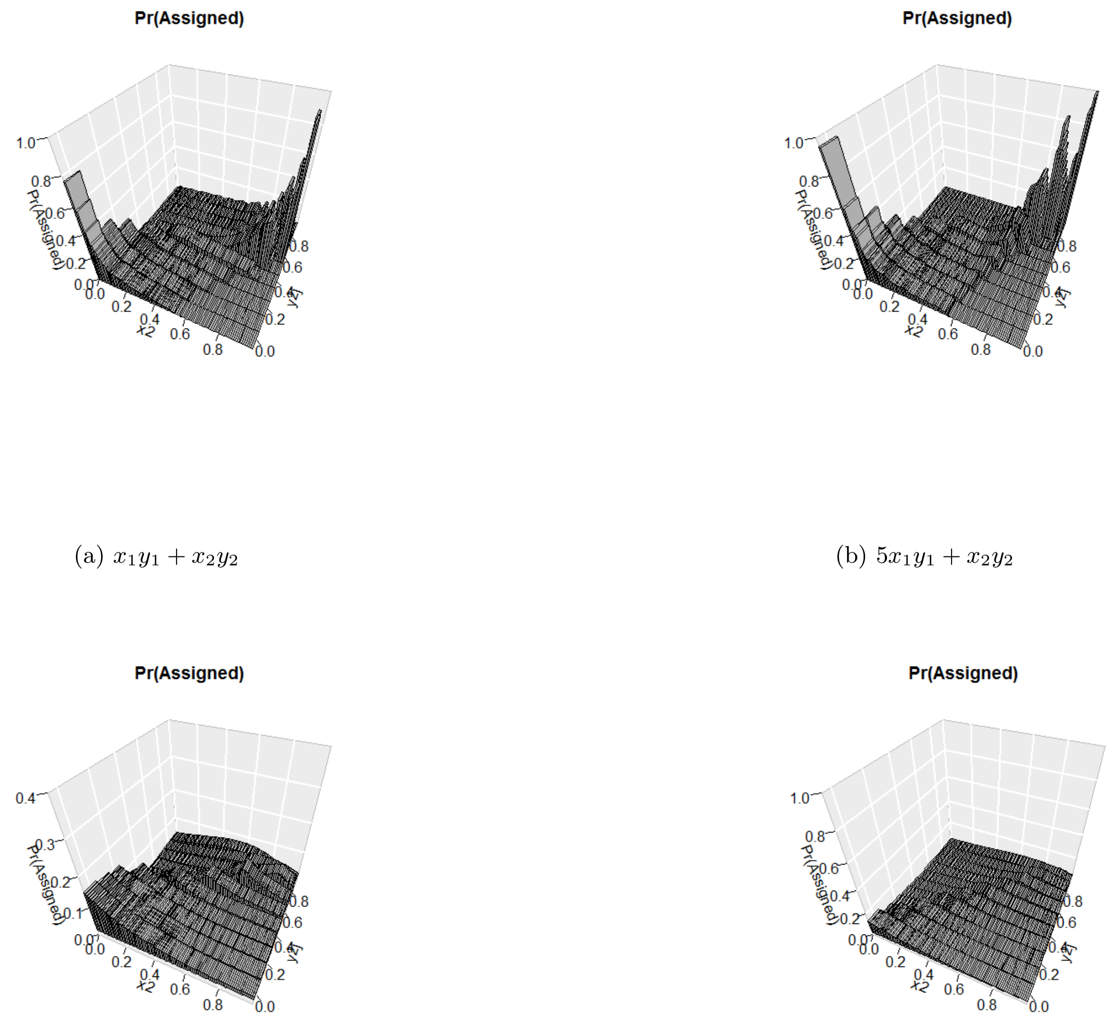


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$



(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

Figure 7: Assignment probabilities of x_1 to y_1 under different production functions when $x_1 = 1 - x_2, y_1 = 1 - y_2$



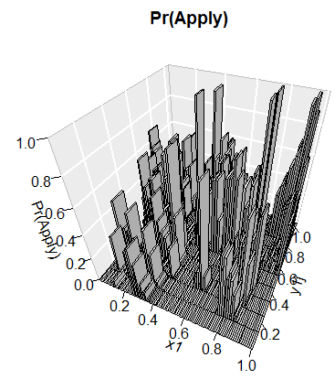
(a) $x_1y_1 + x_2y_2$

(b) $5x_1y_1 + x_2y_2$

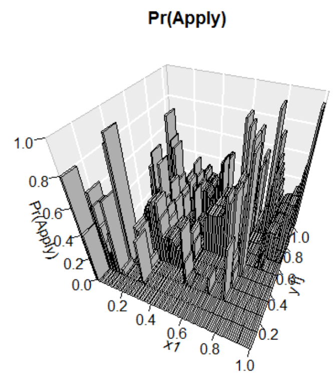
(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

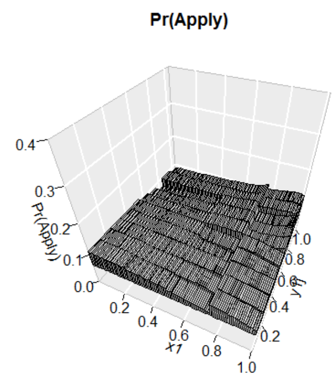
Figure 8: Assignment probabilities of x_2 to y_2 under different production functions when $x_1 = 1 - x_2, y_1 = 1 - y_2$



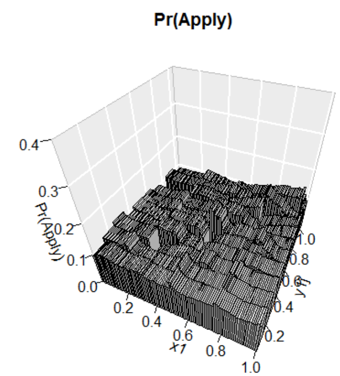
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

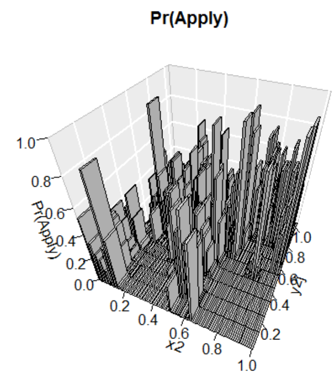


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

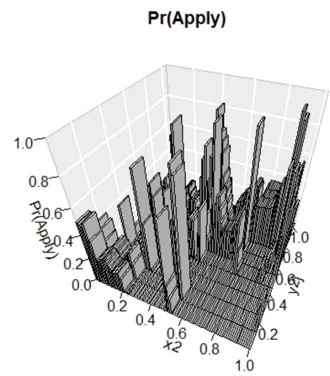


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

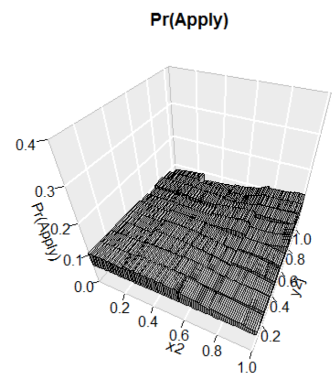
Figure 9: Application probabilities of x_1 to y_1 under different production functions when $corr(x1, x2) = corr(y1, y2) = 0.6$



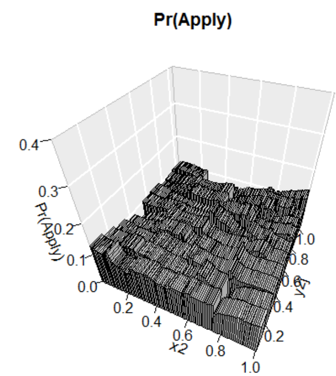
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

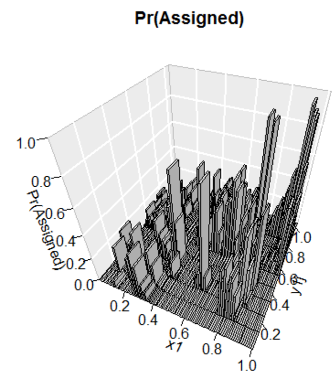


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

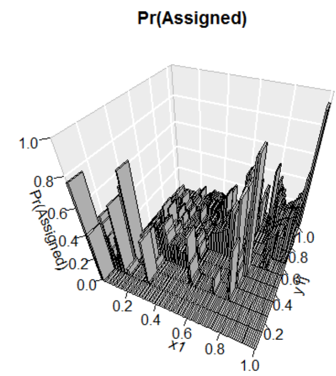


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

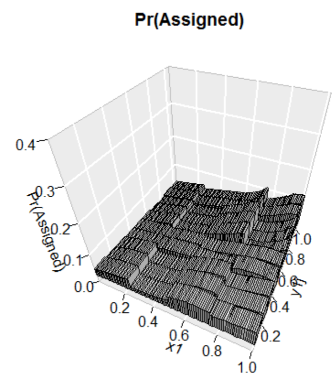
Figure 10: Application probabilities of x_2 to y_2 under different production functions when $corr(x1, x2) = corr(y1, y2) = 0.6$



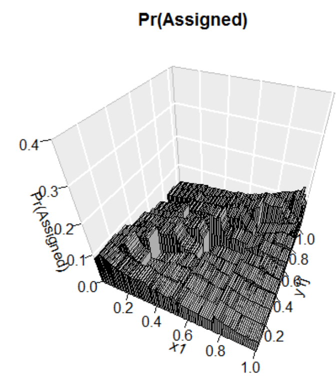
(a) $x_1 y_1 + x_2 y_2$



(b) $5x_1 y_1 + x_2 y_2$

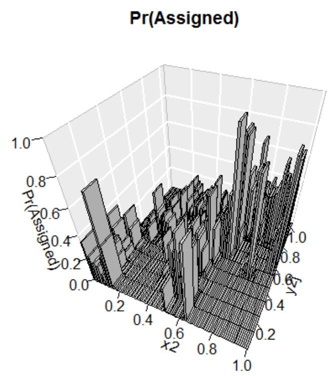


(c) $x_1 y_1 + x_2 y_2 - \mathbb{1}_{x_1 < y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2 < y_2}(y_2 - x_2)^2$

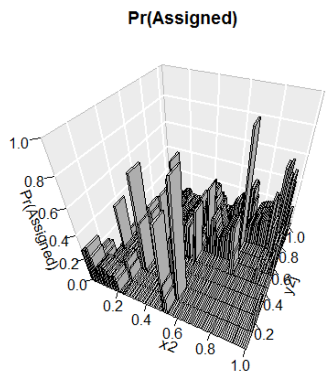


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

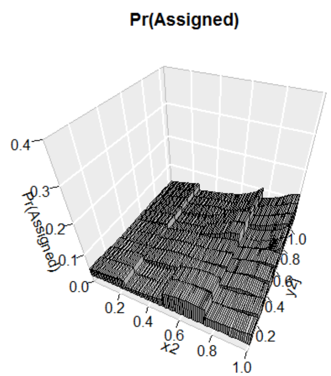
Figure 11: Assignment probabilities of x_1 to y_1 under different production functions when $corr(x1, x2) = corr(y1, y2) = 0.6$



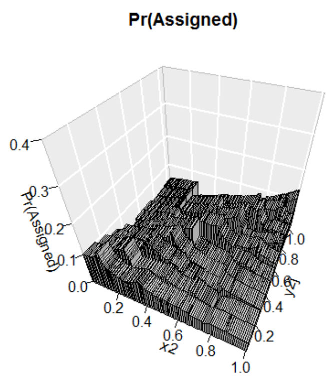
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

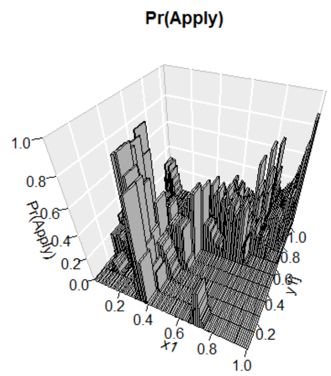


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

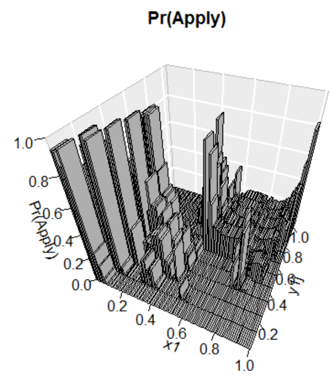


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

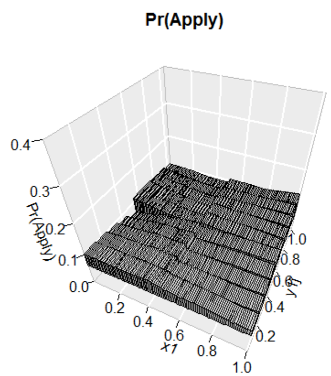
Figure 12: Assignment probabilities of x_2 to y_2 under different production functions when $corr(x1, x2) = corr(y1, y2) = 0.6$



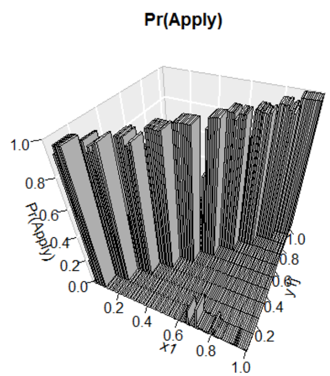
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

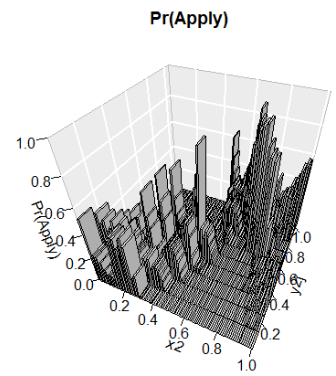


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

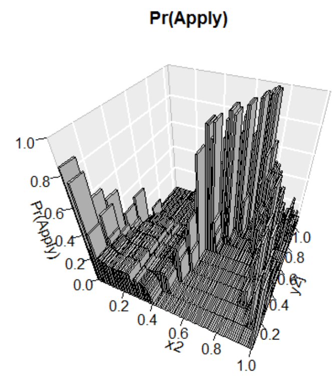


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

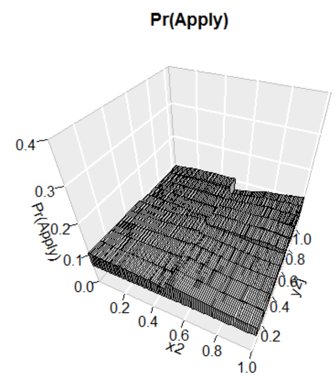
Figure 13: Application probabilities of x_1 to y_1 under different production functions when $corr(x1, x2) = corr(y1, y2) = -0.5$



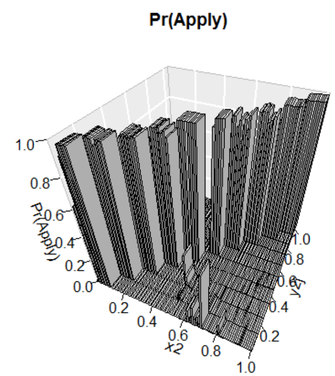
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

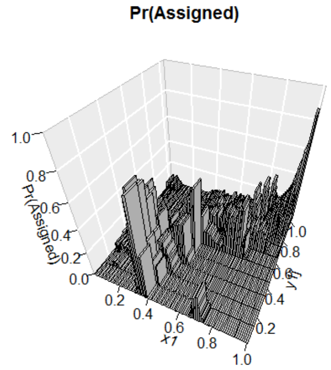


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

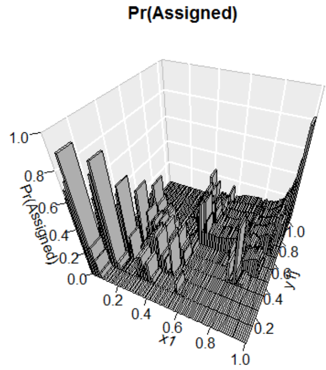


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

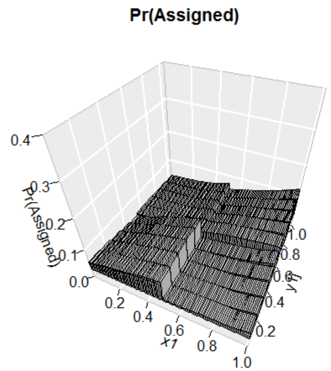
Figure 14: Application probabilities of x_2 to y_2 under different production functions when $corr(x1, x2) = corr(y1, y2) = -0.5$



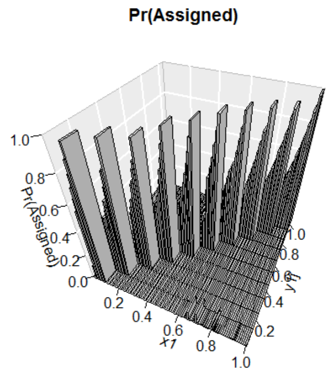
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$

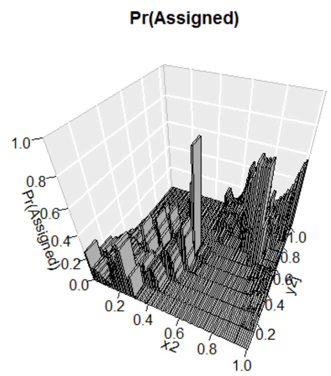


(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$

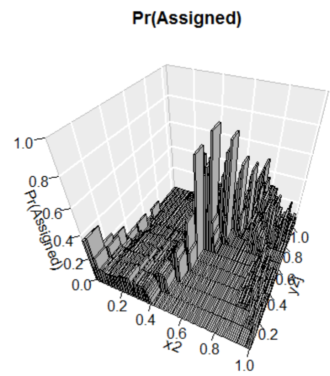


(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

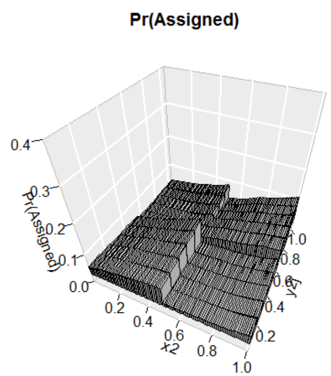
Figure 15: Assignment probabilities of x_1 to y_1 under different production functions when $corr(x1, x2) = corr(y1, y2) = -0.5$



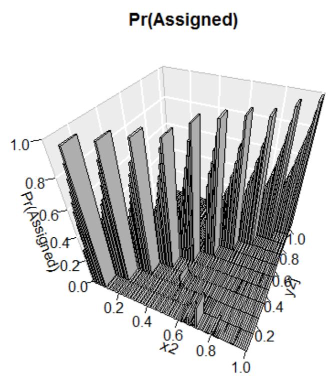
(a) $x_1y_1 + x_2y_2$



(b) $5x_1y_1 + x_2y_2$



(c) $x_1y_1 + x_2y_2 - \mathbb{1}_{x_1<y_1}(y_1 - x_1)^2 - \mathbb{1}_{x_2<y_2}(y_2 - x_2)^2$



(d) $1 - (x_1 - y_1)^2 - (x_2 - y_2)^2$

Figure 16: Assignment probabilities of x_2 to y_2 under different production functions when $corr(x1, x2) = corr(y1, y2) = -0.5$

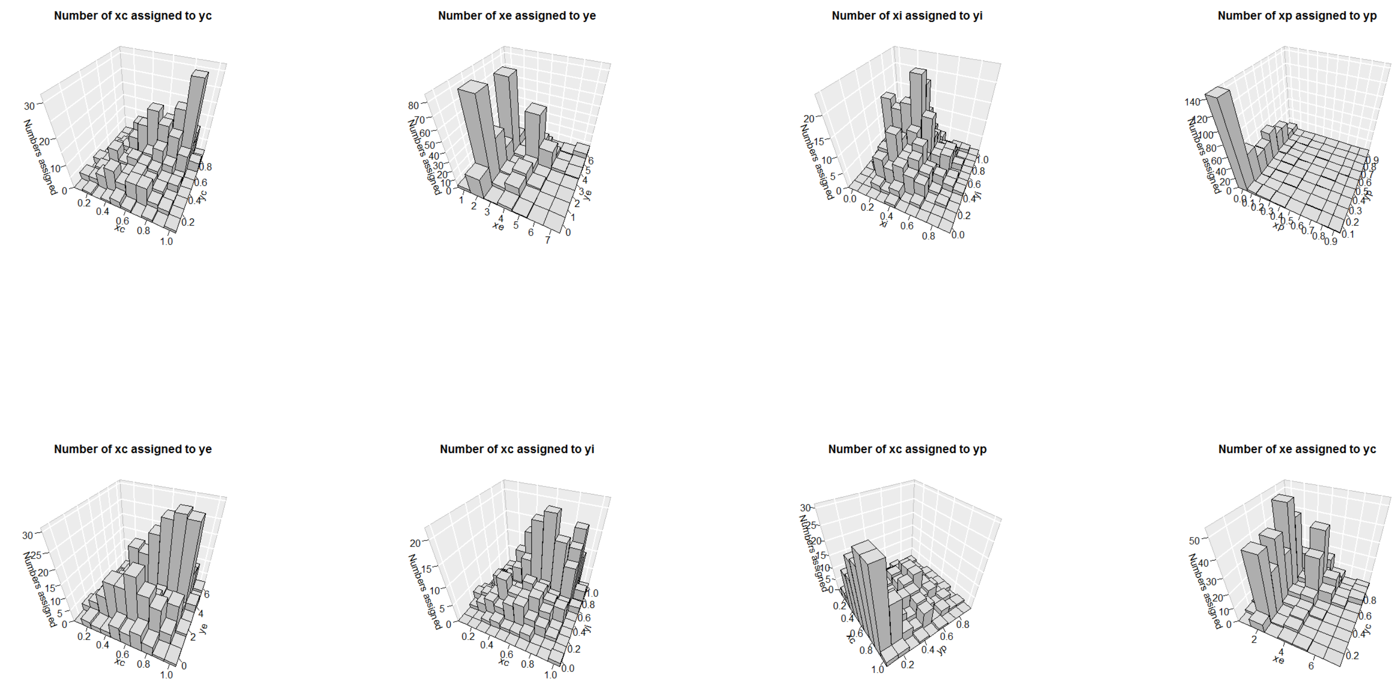


Figure 17: Numbers of workers with skill x_k assigned to jobs with skill requirement y'_k, for $k, k' = c, e, p, i$

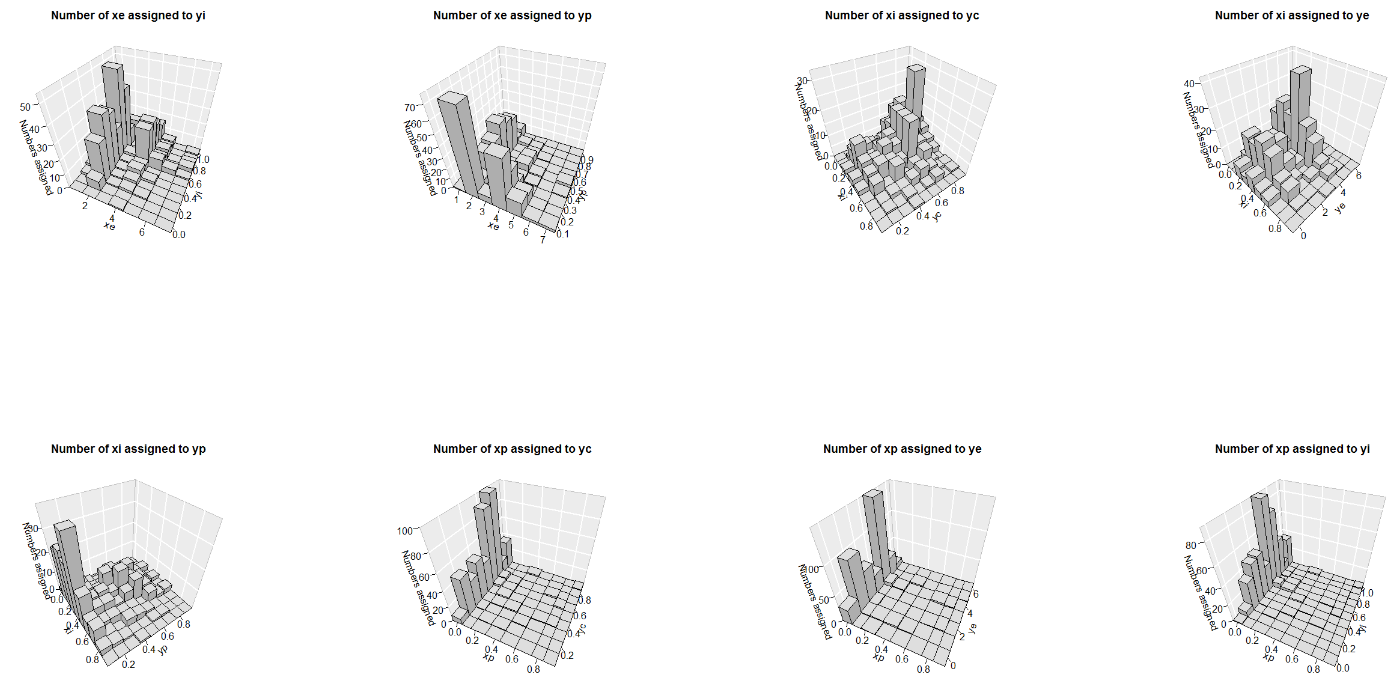


Figure 18: Numbers of workers with skill x_k assigned to jobs with skill requirement y'_k, for $k, k' = c, e, p, i$

Therefore, estimation can only be carried out for the $\gamma(Y)$s of the jobs that were actually observed to have hired someone. The production function is specified for now as:

$$\Phi(X,Y) = \sum_k \beta_k x_k y_k + \sum_k \sum_{l \neq k} \beta_{kl} x_k y_l + \sum_k \alpha_k^o \mathbb{1}_{x_k > y_k}(x_k - y_k) + \sum_k \alpha_k^u \mathbb{1}_{x_k < y_k}(y_k - x_k)$$

Where $\beta_k l$ measures the degree of cross-complementarity between different skill dimensions and where $\alpha_k^o(\alpha_k^u)$ can be interpreted as the effect of an individual having an excess (lack) of skill k relative to what is required[12]

Moments to match : Choosing the following moments, I estimate the model via the Simulated Method of Moments (SMM): 1) $corr(x_k, y_k)$, $\forall k$, 2) $corr(x_k, y_l)$, $\forall k \neq l$, 3) Number of each X type assigned to each Y type, 4) unemployment rate and 5) coefficients on x_k, y_k $\forall k$ from a Mincerian regression of log hourly wages on X and Y and other controls.

Proving indentification This section is pending.

7 Conclusion

To conclude, this paper has presented empirical findings on multi-dimensional mismatch by constructing vectors of human capital for workers and vectors of requirements for each job type. Sorting between workers and firms is stronger on the cognitive and educational dimensions than the other dimensions, although significant mismatch exists. Also, the sorting between workers and firms does not improve much over time and for each worker, the quality of her job match does not change much from her first match. Hence, it becomes important to consider how individuals sort into their first jobs and trade off i) wages and the probability of employment and ii) sorting on each of the different dimensions. I then construct a basic model of directed search in the when workers and firms are characterised by multi-dimensional heterogeneity, adapted from the unidimensional setting in Shimer (2005). I derive the application probabilities of workers and their probabilities of being hired for certain specifications of the production function and restrictions on the distribution of firms and workers. I show how a technological shift that is biased towards some skills but not others, as well as mismatch costs, can influence these probabilities, and hence the equilibrium sorting. I then provided an outline of an extension of the model to allow for the endogenisation of types.

There is still a lot left to be done. A simulation of the full model with endogenous worker and firm types and an estimated production function could allow for a better understanding of i) how well the model fits the data, ii) comparative statics and iii) how the specification of $U(.)$ affects the equilibrium sorting and the fit of the model to the data. Moreover, since human capital investment decisions are based not only on expectations with regards to the first job, but also to the individual's future career, the model has to be dynamic in order for its claims to hold more weight.

[12] The estimation of a more generalised production function is in the works.

8 Bibliography

- Abowd,J., Kramarz, F., Prez-Duarte, S., & Schmutte, I. (2009). A Formal Test of Assortative Matching in the Labor Market.*NBER*

- Acemoglu, D. & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4, 1043-1171.

- Acemoglu, D., & Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, 40(4), 827-849.

- Chevalier, A. (2003). Measuring over-education. *Economica*, 70(279), 509-531.

- Dolado, J. J., Jansen, M., & Jimeno, J. F. (2009). On-the-Job Search in a Matching Model with Heterogeneous Jobs and Workers. *The Economic Journal*, 119(534), 200-228.

- Green, F., & McIntosh, S. (2007). Is there a genuine under-utilization of skills amongst the over-qualified?. *Applied Economics*, 39(4), 427-439.

- Green, F., & Zhu, Y. (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxford Economic Papers*, 62(4), 740-763.

- Hartog, J. (2000). Over-education and earnings: where are we, where should we go?. *Economics of education review*, 19(2), 131-147.

- Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. *Handbook of the Economics of Education*, 4, 283-326.

- Quintini, G. (2011). Right for the Job: Over-qualified or Under-skilled? (No. 120). *OECD Publishing.*

- Lindenalub, I. (2014). Sorting Multidimensional Types: Theory and Application. *Review of Economic Studies (forthcoming)*

- Postel-Vinay, F. & Lise, J. (2015). Multidimensional Skills, Sorting, and Human Capital Accumulation. *Working Paper.*

- Shimer, R. (2005). The assignment of worker to jobs in an economy with coordination frictions. *Journal of Political Economy*, University of Chicago Press, vol. 113(5), pages 996-1025, October.

- Sicherman, N. (1991). " Overeducation" in the Labor Market. *Journal of labor Economics*, 101-122.

- Sloane, P. J., Battu, H., & Seaman, P. T. (1999). Overeducation, undereducation and the British labour market. *Applied Economics*, 31(11), 1437-1453.

9 Appendix

For Section 5

The following proof[13] is valid for obtaining the application strategies when there is an objective ranking of firms and workers and when $\Phi(X,Y)$ is strictly supermodular given this objective ranking. This proof hence applies to the first part of case 1, case 2 and case 3.

Proof that $Q(X,Y)$ and $w(X,Y)$ are increasing in Y $\forall X$: Do proof by contradiction. Let Y' be some job type such that $||Y'|| > ||Y||$. Assume that $Q(X,Y) > Q(X,Y')$. Let X_{-1} be the worker who is ranked just below X by all firms. By transitivity, $Q(X_{-1},Y) > Q(X_{-1},Y')$ also. Since $Q(X_{-1},Y') \geq 0$, one obtains $Q(X_{-1},Y) \geq Q(X_{-1},Y') \geq 0$. This then implies that there is some Y and Y' such that the individual's rationality constraint is not satisfied at the latter but satisfied at the former, implying that

$$U(X)^* = \sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y)}[\Phi(X',Y) - \Phi(X'_{-1},Y)] > \sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y')}[\Phi(X',Y') - \Phi(X'_{-1},Y')]$$

However, by assumption, $e^{-Q(X'_{-1},Y')} > e^{-Q(X'_{-1},Y)}$ $\forall X'$ and by supermodularity, $\Phi(X',Y') - \Phi(X'_{-1},Y') > \Phi(X',Y) - \Phi(X'_{-1},Y)$ $\forall X'$. Hence there is a contradiction and $Q(X',Y)$ cannot be greater than $Q(X',Y')$. Now check that $Q(X,Y) \neq Q(X,Y')$. If this were true, it would imply that

$$U(X)^* = \sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y)}[\Phi(X',Y) - \Phi(X'_{-1},Y)] = \sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y')}[\Phi(X',Y') - \Phi(X'_{-1},Y')]$$

However, by supermodularity and since $e^{-Q(X,Y)} = e^{-Q(X,Y')}$, $\sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y)}[\Phi(X',Y) - \Phi(X'_{-1},Y)] < \sum_{X'=\underline{X}}^{X} e^{-Q(X'_{-1},Y)}[\Phi(X',Y) - \Phi(X'_{-1},Y)]$ and hence there is a contradiction. Therefore, it can only be that $Q(X,Y)$ is increasing in Y. Then, through the worker's individual rationality constraint, since the probability that no one better applies $e^{-Q(X,Y)}$ decreases with Y, $w(X,Y)$ must increase with Y for the constraint to hold.

Threshold strategies: The highest objectively ranked worker $X = \bar{X}$ applies only to the higher ranked firms who can pay her her highest wage and given she applies, she gets these jobs with probability one. Let the lowest firm she applies to be $Y_{\bar{X}}$. For the next best worker, $X = \bar{X}_{-1}$, there is some value for Y such that $e^{-Q(\bar{X}_{-1},Y_{\bar{X}})}w(\bar{X}_{-1},\bar{Y}) = w(\bar{X}_{-1},Y)$. This Y will be the threshold below which $\bar{X}_{-1}$ will apply. Iterating this down the rank of workers, one obtains thresholds for all workers, which are increasing in their rank.

How application strategies are derived I first consider the indifference condition of the lowest ranked worker $\underline{X}$ between her threshold job $Y_{\underline{X}}$ and some $Y > Y_{\underline{X}}$ This gives

$$e^{-Q(\underline{X},Y)} = \frac{e^{-q(\underline{X},Y_{\underline{X}}}}{e^{-q(\underline{X},Y)}} \frac{\underline{x_1}y_{1\underline{X}} + \underline{x_2}y_{2\underline{X}}}{\underline{x_1}y_1 + \underline{x_2}y_2}$$

[13] Note that the first part of the proof, namely that $w(X,Y)$ and $Q(X,Y)$ increase with Y, is an adaption of Shimer (2005).

I then substitute this into the indifference condition of the next lowest ranked worker $\underline{X_{+1}}$ between her threshold job $Y_{\underline{X_{+1}}}$ and some $Y > Y_{\underline{X_{+1}}}$. Substituting iteratively, I get the application probabilities given in case 1 part 1, case 2 and case 3 in Section 5, by adopting the various production function specifications.

For case 1, with $x_2 = 1 - x_1$ and $y_2 = 1 - y_1$ Now I show that $Q(X, Y)$ increases in $\frac{y_1}{y_2}$ for the group of firms with $y_1 > y_2$ and $\forall X$. Before I begin, note that for this group of firms, a worker with a higher x_1 (and lower x_2) is always preferred to one with a lower x_1 and higher x_2, meaning that this group of firms share the same ranking over workers. Now, assume that the opposite is true, i.e. that $Q(X, Y)$ decreases in $\frac{y_1}{y_2}$. Let $Y' = [y'_1, y'_2]$ and $Y = [y_1, y_2]$, where $y'_1 > y_1$ and $y'_2 < y_2$. If the opposite is true, then $Q(X_{-1}, Y') < Q(X_{-1}, Y)$ for the worker X_{-1} who is ranked just below X in terms of $\frac{x_1}{x_2}$. This implies that while firm Y provides X with her maximum expected payoff, Y' does not, which in turn entails that

$$U^*(X) = \sum_{X'=\text{everyone with lower } x_1 \text{ than } X}^{X} e^{-Q(X'_{-1},Y)}(\Phi(X', Y) - \Phi(X'_{-1}, Y)) > \sum_{X'=\text{everyone with lower } x_1 \text{ than } X}^{X} e^{-Q(X'_{-1},Y')}(\Phi(X', Y') - \Phi(X'_{-1}, Y'))$$

Where $Q(X'_{-1}, Y) = Q(X', Y) + q(X, Y)$. Since $e^{-Q(X'_{-1},Y')} > e^{-Q(X'_{-1},Y)}$ by assumption, the above statement can only be true if for some X', $\Phi(X', Y) - \Phi(X'_{-1}, Y) > \Phi(X', Y') - \Phi(X'_{-1}, Y')$. Let us test if this is true. $X = (x_1, x_2) = (x_1, 1 - x_1)$ and $X' = (x'_1, x'_2) = (x'_1, 1 - x'_1)$ be two worker types such that $x'_1 > x_1$, and let $Y = (y_1, y_2) = (y_1, 1 - y_1)$ and $Y' = (y'_1, y'_2) = (y'_1, 1 - y'_1)$ be two firm types such that $y_1 > y_2$.

$$\begin{aligned}\Phi(X', Y') - \Phi(X, Y') < \Phi(X', Y) - \Phi(X, Y) &\leftrightarrow x'_1 y'_1 + (1 - x'_1)(1 - y'_1) + x_1 y_1 + (1 - x_1)(1 - y_1) \\ &< x'_1 y_1 + (1 - x'_1)(1 - y_1) + x_1 y'_1 + (1 - x_1)(1 - y'_1) \\ &\leftrightarrow y'_1(x'_1 - x_1) < y_1(x'_1 - x_1)\end{aligned}$$

This cannot be true since by definition $y'_1 > y_1$. Hence $Q(X, Y)$ must be increasing in $\frac{y_1}{y_2}$ $\forall X$ at the group of firms with $y_1 > y_2$. By symmetry, $Q(X, Y)$ is increasing in $\frac{y_2}{y_1}$ $\forall X$ at the group of firms with $y_2 > y_1$