DEPARTMENT OF ECONOMICS

# Occupational choice of young graduates: Do job tasks matter?

**Stijn Rocher**

**UNIVERSITY OF ANTWERP**
**Faculty of Applied Economics**

EARNED EXCELLENCE
THE BEST BUSINESS SCHOOLS IN THE WORLD

Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
http://www.ua.ac.be/tew

# FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ECONOMICS

## Occupational choice of young graduates: Do job tasks matter?

**Stijn Rocher**

RESEARCH PAPER 2010-027
NOVEMBER 2010

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be



**D/2010/1169/027**

# Occupational choice of young graduates: Do job tasks matter?*

Stijn Rocher†

University of Antwerp‡

November 27, 2010

**Abstract**

This paper examines the extent to which graduates of higher education direct their own occupational choices. I begin by developing an empirical indicator to identify the relation between occupations based on their task content. To this end, I combine individual education and employment data of UK graduates with ratings on 42 task content areas from the UK Skill Survey. Based on these data, I show that UK graduates who majored in similar fields choose occupations with similar task packages. This is followed by a discussion of the wage implications of entering an atypical occupation relative to the modal graduate from the same field. As such, the indicator can be interpreted within a mismatch context. I find that task mismatch increases the probability of over-qualification, which is subsequently associated with lower wages.

*JEL Classification:* I21, J24, J31

*Key words:* Occupational choice, task content, mismatch

## 1 Introduction

The demand for skills in Western economies is greater than ever. The increase in demand is supported by remarkable developments in educational attainment levels. To illustrate, the tertiary attainment rate amongst the entire population in OECD countries is at 27.5%, with 38.7% of the students currently completing higher education (OECD, 2009). Nonetheless, scholars have questioned whether education continues to provide the right skills for succeeding

*This paper has benefited from the comments of Jan Bouckaert, Thomas Peeters and the participants of the T.A.S.K.S. workshops, particularly Semih Akcomak and Anna Salomons. I am grateful for the financial support received from research grant IAP P6/09 "Higher Education and Research: Organization, Market Interaction and Overall Impact of the Knowledge-Based Era". The research was based on the UK Skill Survey (2006) produced by F. Green, D. Gallie, A. Felstead and Y. Zhou. It was collected by BMRB Social Research and supplied by the UK Data Archive. The data are under Crown copyright and funded by Economic and Social Research Council, Department for Education and Skills, Department of Trade and Industry, Learning and Skills Council, Sector Skills Development Agency, Scottish Enterprise, Learning and Skills Observatory for Wales East Midlands Development Agency, Highlands and Islands Enterprise and the Northern Ireland. Department for Employment and Learning. The author is solely responsible for the content and any views expressed therein.

†e-mail: Stijn.Rocher@ua.ac.be

‡Prinsstraat 13, 2000 Antwerp, Belgium

in today's job market. The expanding body of literature on mismatch addresses this question. This strand of research considers workers employed in occupations below their educational qualification as 'over-educated'. Empirical analysis reveals a substantial incidence of over-education, amounting to 26% of the working population (Groot & Maassen van den Brink, 2000). Furthermore, it appears to be a stylized fact that over-educated people suffer significant wage penalties relative to the earnings of properly matched workers with the same level of education. Nonetheless, over-educated people earn more than their well-matched co-workers do (Sloane, 2003).

According to assignment theory,[1] the under-utilization of acquired human capital may explain the lower wages of mismatched workers relative to their peers. Education provides the skills and knowledge necessary to succeed in the labour market. Returns to educational investments, however, depend on the job in which a worker is eventually employed. While acquired skills are productive and highly valued in some jobs, they may be redundant in others. Within this framework, this paper links skills to job tasks (e.g. like problem solving, numeracy or literacy) and investigates the problem of assigning graduates of higher education to jobs. Gibbons and Waldman (2004) also recognize the possibility that some portion of the human capital of workers could be task specific. They argue that skills acquired on the job are partially specific to the job tasks being performed. However, the accumulation of task-specific human capital occurs even before entry onto the labour market. In this paper, I analyse the extent to which task-specific skills play a role in the allocation of jobs amongst young graduates. During their education, students also acquire skills and knowledge needed to perform job tasks once they are employed. The learning of task-specific skills may differ across the disciplines in higher education. For example, the educational programmes of mathematics majors devote more attention to numeracy training, while literacy is emphasized in the field of humanities. The abundant literature on the determinants of occupational choice notwithstanding, the extent to which young graduates direct their search on the labour market is largely unknown. This is therefore the first issue that this paper seeks to address. I find that graduates do indeed consider the task content of jobs during their job searches.

Within a context of mismatch, I consider the wage effects of occupational selection that result in disparity between the task-specific skills acquired and the task requirements of the job. The quantity of schooling is obviously only one way of considering the match between graduates and their jobs. Recent research has investigated other dimensions of mismatch, including field of study and competence mismatch. Although the various measures appear to be only weakly related, mismatch based on these indicators is also associated with lower wages (Allen & De Weert, 2007; García-Aracil & Van der Velden, 2008; Robst, 2007). Because these new concepts of mismatch are hard to identify, however, they are often studied according to self-evaluation data. In other words, workers are asked to determine the extent

[1] See Sattinger, 1993 for an excellent overview.

to which their educational attainment is related to their jobs, using a scale from complete mismatch to a perfect match. Interpretation of results based on self-reported mismatch is not straightforward. An endogeneity issue could arise if workers rationalize dissatisfaction about their wage or work as skill mismatch. Other underlying reasons (e.g. ability or preference of non-pecuniary benefits over wages) could also be responsible for any relationships between mismatch and wage. Alternatively, labour market experts classify the match between workers and their jobs according to detailed educational and occupational data, although the validity of such a classification is not always clear (Nordin et al., 2010; Sloane, 2009). In this article, I consider these measurement issues and develop a new task-based mismatch indicator. Regression analysis shows that mismatch captured by my indicator is associated with a wage penalty. This indicates that young graduates accumulated task-specific human capital prior to entering the labour market.

The paper is organized as follows. In the following section, I develop a theoretical framework that produces worker-job mismatches on the labour market. Two propositions are derived from this framework with regard to (i) directed job search and (ii) wage effects of mismatch. I then discuss the data used, as well as my empirical methodology in Section 3. Proceeding from the task-approach perspective, I characterize occupations according to their task content. Similar to Gathmann and Schönberg (2010), I construct a distance measure based on tasks to indicate the similarity of occupations. I use this indicator to study the relationship between field of study and the task content of jobs. The obtained results are presented in the fourth section, followed by a final section containing conclusions derived from these results.

## 2 A Simple Model of the Labour Market

### 2.1 Basic assumptions

The simple model of the labour market developed in this section largely follows Moscarini (2001), although I make two simplifications in order to focus on the linkage between skills and tasks. First, the standard matching function that I use in this search model does not create search externalities between types. Second, the model does not include on-the-job searches. In this theoretical framework of the labour market, the economy is populated by many firms and workers. Each firm can hire one worker in order to produce final output. In the tradition of the matching literature, firms therefore post vacancies to attract a (non-random) pool of unemployed workers. Firms are heterogeneous. Each firm can choose the type of vacancy to open. The firm selects one of two available production technologies called occupations ($j = o, n$). In both occupations, workers must perform the same two tasks (e.g. arithmetic, reading or writing). The tasks are thus generic. The worker's occupation determines the importance of a specific task in the total task load, as characterized by an a priori known task-load vector $\mathbf{t}_j = \{t_{j1}, t_{j2}\}$ where $t_{j2} = 1 - t_{j1}$ and $t_{j1} \in [0,1]$. This task-load vector is occupation specific, indicated by index $j$. After the firm has chosen the occupation $j$ in which to create a vacancy, unemployed workers apply for these open positions.

Workers are also heterogeneous. They possess different skill bundles that determine their ability to complete both tasks. Skills are thus task specific. The distribution of skill bundles across workers is exogeneously given by education. Skill endowments depend on the worker's educational specialization $(i = 1, ..., p)$. More specifically, a worker with specialization $i$ is characterized by a two-dimensional skill type $\mathbf{s}_i = \{s_{i1}, s_{i2}\}$ drawn from a bivariate uniform probability distribution function $\psi$. Two specializations $h$ and $i$ are related if $\mathbf{s}_i \approx \mathbf{s}_h$ (i.e., if workers acquire similar skill bundles). Given that matching is profitable, each worker applies for the vacancy, revealing the skill type $\mathbf{s}_i$ to the firm. As in Lazear (2003), after a one-to-one match has been made between a firm with a $j$-type vacancy and an unemployed worker with skill bundle $\mathbf{s}_i$, the flow output $y_{ij}$ is given by:

$$y_{ij} = \mathbf{t}_j \bullet \mathbf{s}_i^T = t_{j1}s_{i1} + t_{j2}s_{i2} = t_{j1}\left(s_{i1} - s_{i2}\right) + s_{i2} \tag{1}$$

The matching function $m\left(u_i, v_{ij}\right)$ describes the continuous matching process between unemployed workers and vacancies. This function is linear and homogeneous, and it satisfies the Inada conditions. Furthermore, the number of realized matches depends on the number of unemployed $i-$type workers $(u_i)$ and the number of $j-$type vacancies $(v_{ij})$ available to these workers. Although the matching technology is the same for all agents in the labour market, the unemployment rate and the number of vacancies created differ between types. This influences the matching rate. A vacancy created in occupation $j$ is matched to workers $\mathbf{s}_i$ at rate $q\left(\theta_{ij}\right)$, with $\theta_{ij} = \frac{v_{ij}}{u_i}$ denoting the job-market tightness ratio. Because the matching function is characterized by constant returns to scale, the function $q\left(\theta_{ij}\right)$ is represented as follows:

$$q\left(\theta_{ij}\right) = m\left(\frac{u_i}{v_{ij}}, 1\right) \text{ where } \theta_{ij} = \frac{v_{ij}}{u_i} \tag{2}$$

Analogously, the rate at which one unemployed worker $\mathbf{s}_i$ meets open vacancies in occupation $j$ can be indicated by $\theta_{ij} q\left(\theta_{ij}\right)$:

$$\theta_{ij} q\left(\theta_{ij}\right) = m\left(1, \frac{v_{ij}}{u_i}\right) \tag{3}$$

A match will hold only if the value of the match is greater than the value of separation. I represent the value of employment for a worker $\mathbf{s}_i$ in occupation $j$ by $E_{ij}$ and the value of being unemployed by $U_i$. I further denote the value for a firm with an occupied job of type $j$ by a worker $\mathbf{s}_i$ as $J_{ij}$ and the value of a vacancy for this firm as $V_{ij}$. Note that vacancies are skill specific, as firms observe skill differences among workers and therefore offer skill-specific contracts. In contrast, unemployed workers are not tied to any particular occupation.

They can search for vacancies in both job types at the same time. The matching condition is formalized as follows:

$$E_{ij} + J_{ij} \geqslant U_i + V_{ij} \tag{4}$$

If this is not the case, no matching will occur. This means that the worker remains unemployed and keeps searching for other vacancies, while the firm holds its vacancy open for other jobseekers. In case of a match between a worker and a job satisfying the matching condition stated in Equation (4), wages are set according to Nash bargaining between the worker and the firm in order to divide the rents of this match. This yields the following equation:

$$(1-\beta)\left(E_{ij} - U_i\right) = \beta\left(J_{ij} - V_{ij}\right) \tag{5}$$

where $\beta \in [0, 1]$ represents the exogenous bargaining power of the worker.

## 2.2 Search and Matching: the Bellman Equations.

In this section, I provide the dynamic arbitrage equations for the search and matching model. In order to do this, I define several additional parameters. The discount rate (denoted by $r$) is assumed the same for all economic agents. The exogenous separation rate of workers and their jobs is represented by $\sigma$. All firms face cost $c$ of opening a vacancy. There is no unemployment benefit. The per-period value of being unemployed for worker $\mathbf{s}_i$ ($rU_i$) can now be formulated according to the following Bellman equation:

$$rU_i = \theta_{io} q\left(\theta_{io}\right) \max\left\{\left(E_{io} - U_i\right), 0\right\} + \theta_{in} q\left(\theta_{in}\right) \max\left\{\left(E_{in} - U_i\right), 0\right\} \tag{6}$$

Consistent with Moscarini (2001), a central assumption of the model is that workers may apply for jobs in both occupations. Given the continuous character of the matching function, the probability that an unemployed worker will find a job in both occupations at the same time is zero. Furthermore, it is not worthwhile for workers to be employed in an occupation in which $E_{ij} < U_i$. In this case, the matching condition specified in (4) will no longer hold, and the worker will cease to search for jobs in this occupation. Likewise, the value of being employed for worker $\mathbf{s}_i$ in occupation $j$ per time unit ($rE_{ij}$), satisfies the following condition:

$$rE_{ij} = w_{ij} + \sigma\left(U_i - E_{ij}\right) \tag{7}$$

where $w_{ij}$ is the wage paid to this worker. It is also possible to express the arbitrage equations for the value of an open vacancy and a filled position. The per-period value of a vacancy created by a firm in occupation $j$ for skill type $i$ ($rV_{ij}$) is defined as follows:

$$rV_{ij} = -c + q\left(\theta_{ij}\right)\left(J_{ij} - V_{ij}\right) \tag{8}$$

Finally, the value of having a filled job in occupation $j$ for the firm for one time unit ($rJ_{ij}$) is explicitly stated as follows:

$$rJ_{ij} = y_{ij} - w_{ij} + \sigma \left(V_{ij} - J_{ij}\right) \tag{9}$$

Firms are free to open or close vacancies. The free entry and exit of firms on all job markets implies that the value of vacancies must equal zero. If this value were to be positive, firms would enter the job market and open vacancies. If the value of vacancies were to be negative, firms would leave the job market and close vacancies. In both cases, the process would endure until the value of a vacancy reached zero. In equilibrium, therefore, the following must apply:

$$V_{ij} = 0 \tag{10}$$

### 2.3 Search behaviour of unemployed workers in equilibrium

Given our assumptions on the matching function, the model of the labour market specified by equations (1) to (10) entails the solution for the worker-specific tightness ratios ($\theta^*_{io}$ and $\theta^*_{in}$) arising on both job markets $j = o, n$ in equilibrium. After specifying $m\left(u_i, v_{ij}\right)$, a unique solution for $\theta^*_{io}$ and $\theta^*_{in}$ can be found by solving the following system of equations:

$$\begin{aligned}
0 &= \theta_{io}\beta c + \frac{(r+\sigma)\, c}{q\left(\theta_{io}\right)} + \theta_{in}\beta c - (1-\beta)\, y_{io} && (11)\\
0 &= \theta_{in}\beta c + \frac{(r+\sigma)\, c}{q\left(\theta_{in}\right)} + \theta_{io}\beta c - (1-\beta)\, y_{in} && (12)\\
\theta^*_{ij} &= \max\left(\theta_{ij}, 0\right) && (13)
\end{aligned}$$

The wage gained by worker $\mathbf{s}_i$ on a $j$-type job can subsequently be calculated as follows:

$$w_{ij} = \beta y_{ij} + \beta c \left(\theta^*_{io} + \theta^*_{in}\right) \tag{14}$$

The matching condition guarantees that $\theta^*_{io}$ and $\theta^*_{in}$ are both non-negative. Certain worker-job matches might not be profitable, however, thus resulting in $\theta^*_{io}$ or $\theta^*_{in}$ being equal to zero. The nature of the equilibrium depends on the worker's skill type, as the search behaviour of workers differs across skill types. I explicitly state the conditions that workers' skill profiles must satisfy in order to determine their behaviour. I, therefore define two critical values $\lambda_{ij}$ ($> 0$) with $j = o, n$ specific to the skill type of the worker. These cut-off values indicate the critical productivity difference $\left(y_{ij} - y_{ik}\right)$ of worker $\mathbf{s}_i$ at both job types with $j, k = o, n$ and $k \neq j$. At the critical productivity difference $\lambda_{ij}$, workers are indifferent between selective and random searching. At this critical value, worker $\mathbf{s}_i$ is indifferent between accepting a $k$-type job or remaining unemployed:

$$E_{ik} = U_i \tag{15}$$

Using this condition to simplify the system of equations (1) to (10), I find that workers indifferent between selective and random searching have the skill profile $\mathbf{s}_i$, thereby satisfying the following condition:

$$\lambda_{ij} = \frac{(r+\sigma)\, c}{(1-\beta)\, q\left(\theta^*_{ij}\right)} \tag{16}$$

$$\text{with } \theta^*_{ij} = \frac{1-\beta}{\beta c} y_{ik} \tag{17}$$

Given a specification for the matching function $m(\cdot)$ and hence $q(\cdot)$, $\lambda_{ij}$ can be defined uniquely. These boundary values allow the definition of three categories of job seekers, as listed in the following lemma:

**Lemma 1** *In equilibrium, unemployed workers with skill profile* $\mathbf{s}_i = \{s_{i1}, s_{i2}\}$ *such that*

- $y_{ij} - y_{ik} < \lambda_{ij}$ *for* $j, k = o, n$ *and* $j \neq k$ : *are random job seekers and search for vacancies in both job types;*
- $y_{io} - y_{in} > \lambda_{io}$*: are selective job seekers and search selectively for o-type vacant jobs;*
- $y_{in} - y_{io} > \lambda_{in}$*: are selective job seekers and search selectively for n-type vacant jobs;*
- $y_{ij} - y_{ik} = \lambda_{ij}$ *for* $j, k = o, n$ *and* $j \neq k$ : *are job seekers indifferent between random and selective search.*

This lemma indicates that three skill categories can be distinguished. For the first category of workers it is beneficial to search for jobs in both occupations. Search by this group of workers is therefore considered random. The consideration of these random job seekers is particularly relevant for the analysis of mismatch. Despite a comparative advantage, such search frictions make them willing to accept any job. As such, mismatches may arise, thus resulting in suboptimal productivity of worker-job matches. The other two categories of workers do not take jobs in both sectors, but specialize in one of two occupations. I refer to this search pattern as selective searching.

This lemma also provides various insights on how skills are linked to tasks. Consistent with the framework developed by Roy (1951), workers prefer jobs in occupations in which they can exploit their comparative advantage. In addition to this framework, however, the model I develop assumes a time-consuming job-search process. Because workers value time, some will accept sub-optimal positions. To illustrate, in the frictionless Roy environment, the number of worker-job contacts per unit of time is infinite. Thus, $\lambda_{ij}$ equals zero when search frictions are absent and all workers self-select into jobs based on comparative advantage. Search frictions limit the number of worker-job meetings in each period. Increases in the constringency of search frictions are associated with higher values of critical productivity difference $\lambda_{ij}$. As a result, more job seekers will randomly search for vacant jobs, thereby reducing the duration of search. As mentioned previously, random job seekers accept any job offer. This may result in inefficient worker-job matches. I define an occupational match (or mismatch) as the job type that maximizes (or minimizes) the flow output of a worker-job match. In other words, $y_{ij} \geqslant y_{ik}$ with $j$ representing the occupational match and $k$ the occupational mismatch. I further define similar (or dissimilar) occupations as occupations for which there is a large (or small) difference between the task loads: $|t_{j1} - t_{k1}|$ is large (or small). This allows the derivation of the first proposition:

**Proposition 1 (Directed search)** *The more (or less) dissimilar the task load between the occupational match and another occupation is, the more (or less) job seekers will direct their searching towards occupational matching jobs.*

Plotting the marginal job seekers specified by equation (16) on a two-dimensional graph helps to clarify this proposition. In Figure 1, marginal workers are depicted in the case when the task loads of two occupations are very similar (solid lines) and when they are very distinct (dashed lines). As indicated by the lemma, random job seekers are situated between each pair of lines. Because the pair of dashed lines lies between the pair of solid lines, the number of random searchers can never be larger when task loads are very different, as compared to the situations in which task loads are relatively similar. As shown in Figure 1, while almost all workers randomly search for jobs when occupations are very similar, they choose to apply selectively for jobs when occupations are very dissimilar. Consider two non-overlapping clusters A and B of related specializations. When task loads of occupations are dissimilar, workers with a specialization located in cluster A self-select into jobs in which Skill 1 is highly rewarded. On the other hand, workers from cluster B specializations will enter only those jobs in which Skill 2 is important.

[Figure 1 about here]

The first proposition clearly provides a more nuanced notion of directed search than does the Roy framework, in which all workers selectively search for jobs.[2] In this proposition, workers face a trade-off between turning down all but the most productive jobs or saving valuable time and accepting the first job offer. The more similar the occupational match and another occupation are to each other, the smaller will be the productivity wedge $|y_{ij} - y_{ik}|$

[2]Following the Roy model, the $45^\circ$ line in Figure 1 assigns workers into occupations.

between both job types. As soon as $|y_{ij} - y_{ik}|$ falls below the boundary value $\lambda_{ij}$, job seekers accept job offers from both types, thus reducing the duration of search. In contrast to the Roy framework, therefore, this model allows for inefficient worker-job matches (i.e. mismatch). The following proposition concerns such mismatch:

**Proposition 2 (Mismatch and wages)** *Mismatched workers earns less than what they could receive when they are properly matched. Furthermore, this wage gap is larger when their current occupations are less similar to the occupational match, but at the same time such matches are also less likely.*

Provided $y_{ij} \geqslant y_{ik}$, equation (14) can also be used to show that $w_{ij} \geqslant w_{ik}$ and that the wage gap $(w_{ij} - w_{ik})$ is increasing in the productivity gap $(y_{ij} - y_{ik})$. Given that $(y_{ij} - y_{ik})$ rises above the critical value $\lambda_{ij}$, however, job seekers will change their search behaviour from random to selective searching. Workers are therefore unlikely to accept worker-job mismatches in the case of large gaps in productivity (and thus in wages).

Having formulated these two propositions, the rest of this paper is devoted to testing them empirically by investigating the occupational choice of young graduates. Higher education prepares students for entry into the labour market by providing field-specific skills necessary to enter a profession related to the chosen major (Heijke, Meng and Ris, 2003; Robst, 2007). In line with Gibbons and Waldman (2004), I argue that at least some of these field-specific skills (e.g. arithmetic, reading or writing skills) are essentially task-specific. Depending upon their choice of major, graduate students acquire different task-specific skill bundles. Young graduates sort themselves across occupations by evaluating their skill bundles. The next section presents an analysis of this sorting process, testing it against the propositions. As suggested by the first proposition, I expect that graduates from similar disciplines will direct their job search towards vacancies with similar task loads. The more an occupational task load differs from the occupational match, the less likely graduates will be to accept job offers within this occupation. Furthermore, I expect that graduates entering jobs that are not typical for the modal graduate in the same field (i.e. graduates in a mismatch situation), will experience a wage penalty.

## 3 Data and Methodology

In order to conduct the analysis, two data sets were merged. The first dataset is the UK Skill Survey that was conducted in 2006. This data set aims to investigate the employed workforce in the United Kingdom.[3] It provides a resource for analyzing skill and job requirements in the

[3] For an in-depth analysis see Felstead et al. (2007) and Green et al. (2006).

British economy and consists of a representative sample of the employed population between the ages of 20 and 60. The total sample included 7,762 working individuals across the UK. The foundation for the questionnaire is similar to that of the Dictionary of Occupational Titles (DOT) in the US and the Quality and Careers Surveys in Germany. It contains detailed questions concerning the types of tasks that are important at the current job of the interviewee. These data are used to construct the task-load vectors of occupations. In addition to the UK Skill Survey, I make use of the Reflex dataset. This dataset contains information on 70,000 graduates from 16 countries, including 1,500 British graduates, concerning their qualifications and employment in 2005, five years after graduation. The survey is appropriate to my purposes, as it contains in-depth information on the chosen discipline and the relation between the jobs and education of a large and representative sample of UK graduates.

The UK Skill Survey contains 42 items that describe the task content of jobs. Individuals use a five-point scale to report the extent to which they consider a particular task essential to their job. Although there are several drawbacks to the use of data based on self-reported responses, they are nevertheless informative. I use explanatory factor analysis to reduce the number of inter-correlated variables and to identify twelve underlying factors. These factors cover the following task fields: Computer, Literacy, Managing, Numeracy, Nurturing, Physical, Problem Solving, Reviewing, Routine, Self-planning, Selling and Teamwork.[4] Table 1 provides an overview of the task-item categorization and reports the task units along which each factor loaded strongly. For further analysis, I select the first component of a principal component analysis on each subgroup of task items as listed in Table 1. This allows each component to be interpreted easily. The scores on the twelve components serve as measures of the task loads executed by workers on their jobs. The Cronbach's alpha score is reported for every task component. In almost all cases, these scores easily satisfy the acceptable level of 0.7. This supports the appropriateness of my methodology. (for further details, see Nunnally & Bernstein (1994)).

[Table 1 about here]

The next step in the analysis involves to define the task content of occupations. The international standard classification of occupations (ISCO) groups jobs in occupations and assigns codes.[5] For my analysis, I retain all three-digit ISCO codes with more than 10 observations.[6] By aggregating the component scores based on workers' task-use responses up to the occupational level, I construct the occupational task load for all twelve task fields. Based on task evaluations from 7,651 observations, I am able to identify the task content

[4] These factors are largely similar as the one indentified by Green (2009) using the same dataset.

[5] The International Standard Classification of Occupations (ISCO) is organised by the ILO and is occasionally updated. Our analysis is based on the 1988 version. For more information, see www.ilo.org.

[6] As a robustness check, I also group observations by four-digit code and raise the minimum number of observations to 25. Although this procedure alters the number of occupations for which I can identify the task content, the ultimate results do not change much.

of 80 occupations. Table 2 presents selected examples that illustrate differences in the task measures amongst occupations that are commonly selected by graduates of higher education. Although the tasks are generic in the sense that they are valued to a certain extent in every job, the extent to which each task is performed varies across jobs. Task intensity is therefore job specific. Given that task indices are obtained from principal component analysis, they have a mean of zero and a standard deviation of one. High (or low) task use is associated with positive (or negative) task-component scores. The occupational groups presented are architects, engineers and related professionals (ISCO 214), health professionals (ISCO 222), finance and sales associates (ISCO 341) and secondary education teaching professionals (ISCO 232).

[Table 2 about here]

The examples presented in Table 2 clearly show the level of variation in task content across occupations. According to the task data, the main tasks of architects and engineers comprise non-routine numeracy and problem-solving tasks often supported by computers. Nurturing tasks are relatively exceptional. Finance and sales associates also spend a considerable proportion of their time on numeracy using computers. For this occupational group, however, selling is much more important than problem-solving, and they perform relatively few physical tasks. Furthermore, the task measures indicate that health professionals engage in considerable teamwork and perform primarily nurturing tasks. In addition, they spend considerable time on reviewing and routine tasks. The task description of secondary education teachers is also quite plausible. They manage, nurture and work in groups often. Planning and organizing their own time are also reported as relatively important features of these jobs.

After specifying the task-load vectors, it is necessary to quantify the similarity of occupations in order to test the propositions formulated above. To this end, I calculate the Mahalanobis distance (MD) between the 12-dimensional task vectors for all pairs of occupations. The MD is calculated according to the following formula:

$$MD_x^y = \sqrt{(\mathbf{x}-\mathbf{y})^T \, \Sigma^{-1} \, (\mathbf{x}-\mathbf{y})} \tag{18}$$

In expression (18), $\mathbf{x}$ and $\mathbf{y}$ indicate the task vectors of two random occupation $x$ and $y$ taken from a task-vector distribution with variance-covariance matrix $\Sigma$. The MD is actually the weighted Euclidean distance, where the weights are determined by $\Sigma^{-1}$, they are therefore superior in data analysis. The computation of the inverted covariance matrix may cause problems in the case of multicollinearity. The MD is also relatively sensitive to measurement error. I therefore choose to use the 12-dimensional task vector instead of all 42 task units. I obtain the MD between all 80 occupations described by the task data. The distances vary from zero to approximately one. The value equals zero for occupations that involve an identical task package and approaches one if the task content is far from similar. The average distance is 0.531, with a standard deviation of 0.135. The next step involves merging the computed distances between occupations with the Reflex dataset. This dataset contains in-depth information for 1,500 British graduates of higher education. From all these observations, I was able to match 1,306 individuals with 59 unique occupations for which task content and distance measures are available. This sample is used for further analysis.

Table 3 helps to provide a better grasp of the distance measure. This table presents details on the selection of occupations used in Table 2 and their distance to closely linked and distant occupational titles. The generic tasks performed by 'architects and engineers' (ISCO 214) relate to the tasks of 'physics and engineering science technicians' (ISCO 311). This is indicated by a relatively low MD of 0.206. The distance measured between 'architects and engineers' and 'library, mail and related clerks' (ISCO 414) is relatively high, with a value of 0.643. The MD also ranges from 0.200 to 0.700 for the other examples of occupational titles. This indicates a similar pattern of close links with some occupations and very loose connections with others.

[Table 3 about here]

# 4 Results

## 4.1 Occupational choice and the task content of occupations

In the following discussion, I draw upon the dataset constructed in the previous section to demonstrate that the task content of jobs influences the occupational choices of higher-education graduates. In accordance with the theoretical model, I expect that students who graduated from the same major will have acquired identical task-specific skills. Furthermore, given the time-consuming nature of the search process, job seekers face a trade-off between waiting for their favoured, most productive job or accepting an early job offer and reducing the duration of search. This choice crucially depends on that graduate's potential productivity on the job that is offered relative to that graduate's occupational match. As stated by the first proposition on directed searching, the greater similarity in the task content of both is expected to reduce this productivity gap and increase the likelihood that worker will accept the job that is offered. I therefore expect that graduates from the same field tend to enter both their occupational match and other closely related occupations. On the contrary, one could assume that productivity does not depend on the task-specific skills of workers. In this case, the job allocation of graduates would not be influenced by differences in task requirements across occupations. As a naive benchmark, the allocation of workers to jobs might be completely random and determined only by the relative size of an occupation. To test this first proposition, the MD is used to quantify the similarity between a graduate's actual occupational choice and that graduate's occupational match. Under the hypothesis that task-specific skills matter for occupational selection, I expect to find that the size of these distances are generally smaller than are those observed under the naive benchmark of random assignment. I therefore compare two distributions of distances: (i) the distribution under observed occupational selection and (ii) the hypothetical distribution under random occupational assignment.

I begin with concise explanation of how these distances are constructed. The first step is to identify a proper match between a graduate from a certain discipline and an occupational field. I use the mode to assign such an occupational match to each field of study.[7] An occupation that attracts the largest share of graduates from a given discipline is regarded as matching. Based on this statistical method of defining matches and mismatches (see also Hartog, 2000), graduates in the field of 'engineering, building and architecture' are assigned to the occupation 'architects and engineers' (ISCO 214) and 'health' graduates are assigned to the occupation 'health professionals' (ISCO 222). A complete list can be found in Table A1 in the appendix. While the identification of proper matches is not an exact science, the results seem plausible. Second, I calculate the distances between the occupational match and (i) the observed occupational choice; (ii) the randomly assigned occupation for this sample of higher-education graduates. The last step consists of plotting both distributions of distances. The kernel density estimates are shown in Figure 2 for graduates who majored in 'engineering, building and architecture'. A Kolmogorov-Smirnov test affirms that both distributions are significantly different at the 1% level. As shown in Figure 2, actual distances tend to be smaller than distances under random assignment. This is statistically confirmed by the Kolmogorov-Smirnov test. These graduates are thus indeed inclined to accept jobs with more similar task content than is predicted by the naive model. Table A2 in the appendix presents comparisons of observed and random distributions for all fields of study, along with the results of equality of distribution tests. The role of job tasks in occupational choice clearly differs among majors. In some cases (e.g. 'arts', 'personal services' and 'social and behavioral science'), the naive model even performs remarkably well. It is also interesting to note the level of variation in the task specificity of jobs. While 'economics, business and administration' or 'engineering, building and architecture' have related job alternatives, 'arts', 'health' or 'law' enter isolated occupations making occupational mobility costly.

[Figure 2 about here]

The framework also predicts that graduates from related specializations ($\mathbf{s}_i \approx \mathbf{s}_h$), will self-select into occupations with a similar task content. The major 'mathematics, statistics and computing' is related to the major 'engineering, building and architecture', but it is quite distant from the field of 'health'. The distances between the occupational match of 'engineering, building and architecture' graduates (i.e. 'architects and engineers') and the

[7] As a robustness check, we manually assigned a matching occupation. The impact on the results was minimal.

observed occupational selection of the three groups of graduates differ. The distribution of these distances is presented in Figure 3 by field of study. As expected, 'mathematics, statistics and computing' graduates are more likely to choose occupations that are similar to those chosen by graduates in the field of 'engineering, building and architecture' than are graduates in the field of 'health'.

[Figure 3 about here]

In Figure 4, I consider the observed and random distribution of distances for the entire sample. In Table 4, I report the corresponding descriptive statistics. These figures indeed indicate that the observed occupational selections of young graduates are directed towards occupations that are more similar. Although the standard deviation is larger (thus indicating a wider spread), the 25th percentile, mean, median and 90th percentile are all lower for the density distribution of the observed distances. Given that the distribution under observed selection is bimodal, these summary statistics should be interpreted with care. For this reason, a Kolmogorov-Smirnov equality-of-distributions test ($p$-value =0.000) is performed. This test also rejects the null hypothesis that both distributions are equal and supports the conclusion that the distribution of the distance measure under observed selection is more right-skewed than it is under random selection.

[Figure 4 about here]

[Table 4 about here]

### 4.2 Mismatch among UK graduates

The second proposition derived from the theoretical model, relates to mismatch. Since the early 1990s, the high incidence of mismatches between skills supply and labour market demand has raised considerable concern among policy makers. Mismatched graduates are not utilizing all of their acquired skills in their jobs. In addition to being an alleged waste of public money, various indicators measuring the degree of over-education show substantial negative wage effects for this type of mismatch, although, differences in measurement generate a range of estimated magnitudes of the wage penalty (Sloane, 2003). Empirical scholars rely on three methods in order to define over-education. First, a subjective measure is based on the self-assessment of workers, who are asked to evaluate their own qualifications relative to the job requirement. In the second method, labour market experts identify the appropriate educational degree necessary to perform a certain occupation. The third method consists of statistical measurement. Workers who have had more (typically one standard deviation) years of schooling than the mean or median worker in a certain occupation are classified as over-educated. In addition to a variety of methods for measuring over-education, researchers have recently argued about the possibility of other dimensions of mismatch, including competence mismatch or mismatch concerning field of study (Allen and De Weert, 2007; García-Aracil and Van der Velden, 2008; Robst, 2007). While these dimensions of mismatch are also penalized in terms of reduced wages, results in this respect should be treated with a degree of caution as they make use of subjective indicators. In a meta-analysis, Groot and Maassen van den Brink (2000) find that measures based on self-reports result in higher incidence of mismatch than do measures that are more objective.

In the remainder of this analysis, I apply the new mismatch indicator based on distances. This statistical measure takes into account potential mismatches between task-specific skills and job-task requirements. This variable is constructed by taking the distance between the occupation chosen by each graduate in the dataset and the respective occupational matches.[8] In other words, graduates are considered mismatched if the tasks that they perform in their jobs differ substantially from those of the modal graduate who majored in the same field. In this section, I discuss how my distance indicator relates to the other mismatch measures, as well as the extent to which I can replicate the finding that the skill match between workers and their jobs affect earnings. This would provide confirmation that graduates possess task-specific human capital.

[Table 5 about here]

Given that my statistical measure of mismatch is new, I report its relation to other measures of mismatch in Table 5. The Reflex dataset contains variables that may be interpreted as subjective measures of mismatch concerning field of study (content match), years of schooling (educational match) and competences (skill match). For the first measure, content match, I use workers' responses concerning what they consider the most appropriate field of study required to do their job in relation to their own fields. The second measure (educational mismatch) makes use of workers' perception of the most appropriate educational level required to do their current jobs in relation to their current level. The last subjective measure of mismatch (skill mismatch) is based on workers' responses with regard to the degree of acquired skills and knowledge used in their current jobs. The correlations between my distance indicator and the three subjective measures from the Reflex dataset are calculated using the coefficient of determination $R^2$ of a linear regression of my indicator on each subjective measure. The correlation with these subjective measures of mismatch is modest, ranging from 0.21 to 0.37. This is not surprising, as previous research has demonstrated weak correlations between different dimensions of mismatch.[9] More importantly, my measure behaves well, because the measure is positively correlated to the other mismatch indicators. Another important finding is that my measure is the most strongly related to content mismatch. Graduates reporting mismatch

[8] The distance indicator has already been used to present the distribution of distances under the heading of observed occupational selections in Figure 3.

[9] For example Allen and de Weert (2007) show that educational mismatches in no way imply mismatches between available and required knowledge and skills. Moreover, graduates indicating an educational match do report skill mismatches.

between their fields of study and their jobs are likely to be performing jobs that require tasks that differ considerably from those demanded in the jobs of workers with identical majors.

Table 6 provides information on personal characteristics that determine the degree of mismatch according to my measure. I estimate four specifications. In the first specification I control only for gender, age, marital status and ethnic background.[10] According to these findings, women are mismatched significantly (at the 0.1% level) more than men are. In addition older graduates are more likely to hold jobs that are unrelated to their fields of study. The rate of mismatch amongst workers with non-European ethnic backgrounds is not significantly different from those of the European reference group, and marital status does not seem to matter. In the subsequent specification, I also control for current educational degree, experience (in months), past periods of unemployment, public sector employment and whether the worker had participated in an internship. The likelihood that graduates will accept unrelated jobs relative to those of their peers is higher for graduates who have experienced a period of unemployment. Internship participation makes graduates less likely to choose jobs that do not relate to their skills. Because vocationally oriented study programmes often provide internships, this could explain why internship participation enhances 'good' matches among graduates. As suggested by the results presented in Table A2, the last two specifications reveal wide variations in the degree of mismatch across fields of study. With the exception of the major 'social and behavioural sciences', none of the other fields is associated with significantly worse matches (on average), relative to the omitted category 'humanities'. Graduates from the field of 'engineering, building and architecture' choose the most similar jobs. Furthermore, graduates who majored in 'agriculture, forestry and fishery', 'arts', 'economics, business and administration', 'education', 'health', 'law', 'mathematics, statistics and computing', 'science' and 'social services' choose occupations in which the task content is more related to their fields of study than is that of the occupations chosen by graduates from the field of 'humanities'. As expected, control variables (e.g. 'internship participation' and 'relative score') reduce the estimates, although the coefficients remain remarkably robust. These results suggests that the degree of vocational orientation in the various fields of study is not the only relevant factor influencing task specialization.

[Table 6 about here]

The final table (Table 7) reports results from a wage regression in which the dependent variable is the log of gross hourly wage. I estimate three specifications. The first specification contains only my distance indicator; specification (2) includes additional standard controls, and the last specification contains various subjective measures of mismatch. These measures relate to the degree of content match, educational match and skill match. In the first specification, my indicator has the expected sign. Following this indicator, young graduates who

[10] The estimates of these coefficients are not reported in Table 6.

perform very different tasks in their jobs as compared to most graduates from their fields are classified as mismatched. In line with the literature on skill mismatch, I find that the mismatched graduates have significantly lower returns to schooling than do properly matched individuals with the same major. A mismatch of one standard deviation higher results in a wage penalty of 8.1% (significant at the 0.1 per cent level).

The second specification includes standard controls in addition to my distance indicator. Graduates in the field of 'economics, business and administration', 'health', 'mathematics, statistics and computing' and 'science' earn a significant wage premium compared to graduates in the field of 'humanities'. On the other hand,'arts' graduates earn significantly lower wages than do graduates in the reference category. Women earn 7.5% less than men do. The earnings of older workers, workers living with a partner or workers with a master's degree are higher on average. Ethnic background, public sector employment and experience had no significant effect on wages. Other variables of interest include a post-graduation period of unemployment, which depresses earnings by 13.1 per cent. Furthermore, a self-reported relative score compared to other students was used as a proxy for ability, and its effects are significant. Graduates indicating that their grades were above average earn 25.6 per cent more than graduates reporting below average grades do. This result is significant at the one per cent level. The final variable included in this specification is a dummy variable indicating the status of the higher education institute. Graduating from a prestigious university tends to increase earnings significantly by almost 20 per cent.

The last specification adds the other mismatch measure to the model, thereby rendering the results of my distance indicator insignificant. Standard or low levels of skill use and lower than tertiary educational requirements are associated with heavy penalties. Other studies have shown similar estimates for over-education and over-skilling amongst research populations in the UK (see e.g. Allen and de Weert, 2007; Sloane, 2003). The subjective measure for content match remains insignificant in this specification as well. Given that my indicator is correlated most strongly with this mismatch measure, I excluded the distance indicator from the regression as a robustness check. This did not alter the finding that content mismatch has no significant impact on earnings. Consistent with Sloane (2003), I therefore conclude that it is not necessarily content mismatch, but rather the experience (or feeling) of being over-educated or over-skilled for the job that results in lower wages. As shown in Table 5, however, my mismatch indicator is weakly but significantly related to both measures of educational mismatch and skill mismatch. In other words, selecting an unrelated job does increase the probability of being over-educated or over-skilled, and it is therefore associated with lower wages, albeit indirectly. In addition, adding these vertical mismatch measures has a negative impact on the proxies for ability (i.e. the graduate's perceived grades relative to other students and the dummy variable for institutional prestige). This indicates that, in addition to content mismatch, low ability is at least one of the factors contributing to over-qualification.

[Table 7 about here]

## 5 Conclusion

The tasks that people must perform in their jobs are key determinants of occupational choice. This paper proposes a matching framework in which workers search for jobs with different task content. Workers prefer to perform job tasks in which they are skilled, although mismatches occur in some cases. Given the time-intensive character of the job-search process, workers may accept sub-optimal job offers in order to reduce the duration of their search. This creates deliberate worker-job mismatches.

This paper offers an empirical analysis applied to young graduates. I merge task data on UK jobs with detailed information on a cohort of young graduates. A novel approach is applied in order to relate occupations, using information on the task content of jobs and summarizing the degree of occupational similarity in a distance indicator. First, the results show that UK graduates who majored in related fields direct their search on the job market towards similar occupations. This finding suggests that the acquisition of task-specific skills differs across the various disciplines. Differences in the accumulation of task-specific human capital drives the sorting process of graduates across jobs. This finding improves our understanding of the occupational choice of young graduates, and it underlines the importance of the task content of jobs in this process. Second, the consequences of task mismatch are investigated. Considerable controversy exists with regard to the extent of mismatch amongst graduates, as well as concerning which dimensions of mismatch are harmful to earnings. At first glance, estimation suggests that task-mismatched graduates suffer a wage penalty. After considering other mismatch measures, however, task mismatch is only indirectly associated with lower wages. It does increases the likelihood of over-education and (to a lesser extent) over-skilling, which in turn depresses wages.

# Bibliography


1. Allen, J. and E. De Weert (2007). "What do Educational Mismatches tell us about Skill Mismatches? A Cross-country Analysis, European Journal of Education, Vol. 42 (1), pp. 59-73.

2. Felstead, A., D. Gallie, F. Green and Y. Zhou (2007). "Skills At Work, 1986 to 2006". University of Oxford, SKOPE.

3. García-Aracil, A. and R. Van der Velden (2008). "Competencies for young European higher education graduates: labor market mismatches and their payoffs", Higher Education, vol. 55 (2), pp. 219-239.

4. Gathmann, C. and U. Schönberg (2010). "How general is Human Capital? A Task-Based Approach", Journal of Labor Economics, vol. 28 (1), pp.1-49.

5. Gibbons, R. and M. Waldman (2004). "Task-specific human capital", American Economic Review, vol. 94(2), pp.203–207.

6. Green, F. et al. (2006). "Skills Survey" [computer file], Colchester, Essex: UK Data Archive [distributor], September 2008. SN: 6004.

7. Green, F. (2009). "Employee Involvement, Technology and Job Tasks", NIESR Discussion Papers, nr. 326.

8. Groot, W. and H. Maassen van den Brink (2000). "Overeducation in the labor market: a meta-analysis", Economics of Education Review, vol. 19 (2), pp. 149-158.

9. Hartog, J. (2000). "Overeducation and Earnings: where are we, where should we go?", Economics of Education Review, vol. 19 (2), pp.131-147.

10. Heijke, H., C. Meng and C. Ris. (2003). "Fitting to the job: The role of generic and vocational competences in adjustment and performance", Labour Economics, vol. 10 (2), 215-229.

11. Lazear, E. (2003). "Firm specific human capital: a skill-weights approach", NBER working paper, nr. 9679.

12. Moscarini, G. (2001). "Excess Worker Reallocation", Review of Economic Studies, vol. 68(3), pp. 593-612.

13. Nordin, M., I. Persson and D. Rooth. (2010). "Education-Occupation Mismatch: Is there an income penalty?", Economics of Education Review, forthcoming.

14. Nunnally, J.C., and I.H. Bernstein. (1994). "Psychometric theory", (3rd ed). New York: McGraw-Hill.

15. OECD (2009). “OECD in figures 2009”, OECD publishing, Paris.

16. Robst, J. (2007). “Education and job match: the relatedness of college major and work”, Economics of Education Review, vol. 26 (4), pp. 397-407.

17. Roy, A. (1951). “Some Thoughts on the Distribution of Earnings”, Oxford Economics Papers, vol. 3 (2), pp. 135-146.

18. Sattinger, M. (1993). “Assignment models of the distribution of earnings”, Journal of Economic Literature, vol. 31 (2), pp.831-880.

19. Sloane, P. (2003). “Much ado about nothing? What does the Overeducation Literature Really Tell Us?” in F. Büchel, A. de Grip and A. Mertens (Eds.), Overeducation in Europe: current issues in theory and policy (pp. 11-48), Edward Elgar Publishing.

20. Sloane, P. and S. McGuinness. (2009). “Labour Market Mismatch Among UK Graduates; An Analysis Using REFLEX Data”, IZA Discussion Papers, nr. 4168.

# Tables & Figures

Table 1: Categorization of Task Items

| | |
|---|---|
| **Computer** ($\alpha = 0.7629$) | **Physical** ($\alpha = 0.8158$) |
| Computer use | physical strength |
| Complexity of computer use | physical stamina |
| Internet use | finger and hand dexterity |
| **Literacy** ($\alpha = 0.8787$) | knowledge of use or operation of tools |
| Reading written information | **Problemsolving** ($\alpha = 0.8636$) |
| Reading short documents | spotting problems or faults |
| Reading long documents | working out cause of problems & faults |
| Write forms, notices or signs | thinking of solutions to problems |
| Write short documents | analyzing complex problems in depth |
| Write long documents | **Reviewing** ($\alpha = 0.7763$) |
| **Managing** ($\alpha = 0.8028$) | noticing when there is a mistake |
| teaching people | checking things to ensure no errors |
| persuading or influencing others | paying close attention to detail |
| making speeches or presentations | **Routine** |
| planning the activities of others | short repetitive tasks |
| **Numeracy** ($\alpha = 0.8124$) | **Self-planning** ($\alpha = 0.8336$) |
| basic arithmetic | planning own activities |
| arithmetic involving fractions | organizing own time |
| advanced mathematics and statistics | thinking ahead |
| **Nurturing** ($\alpha = 0.7505$) | **Selling** ($\alpha = 0.5745$) |
| counseling, advising or caring for clients | Knowledge of products or services |
| dealing with people | Selling a product or service |
| handling feelings of others | Specialist knowledge or understanding |
| managing own feelings | **Teamwork** ($\alpha = 0.8505$) |
| | working with a team |
| | cooperating with colleagues |
| | listening carefully to colleagues |

Table 2: Occupations and their task content

| **Occupation** | **architects and engineers** | **health prof. (except nursing)** | **finance and sales ass.** | **sec. educ. teaching prof.** |
|---|---|---|---|---|
| ISCO - code | 214 | 222 | 341 | 232 |
| | | | | |
| Computer tasks | 0.530 | 0.186 | 0.609 | 0.393 |
| Literacy Tasks | 0.327 | 0.301 | 0.211 | 0.635 |
| Managing Tasks | 0.301 | 0.485 | 0.265 | 1.279 |
| Numeracy Tasks | 0.633 | 0.034 | 0.473 | 0.007 |
| Nurturing Tasks | -0.238 | 0.701 | 0.056 | 0.800 |
| Physical Tasks | -0.374 | -0.092 | -0.811 | -0.358 |
| Problemsolving Tasks | 0.408 | 0.487 | 0.059 | 0.086 |
| Reviewing Tasks | 0.215 | 0.446 | 0.142 | -0.122 |
| Routine Tasks | -0.371 | 0.046 | -0.220 | -0.442 |
| Self-Planning Tasks | 0.277 | 0.153 | 0.410 | 0.709 |
| Selling Tasks | 0.221 | 0.507 | 0.494 | -0.065 |
| Teamwork | 0.194 | 0.564 | 0.023 | 0.534 |
| | | | | |
| share of workers | 1.49 | 1.08 | 3.05 | 1.15 |
| Amount of workers | 114 | 83 | 233 | 88 |

Table 3: Close and distinct occupations

| **architects and engineers (214)** | **distance** |
|---|---|
| *close occupation* | |
| physical and engineering science technicians (311) | 0.206 |
| *distant occupation* | |
| library, mail and related clerks (414) | 0.643 |
| **health professionals (except nursing) (222)** | |
| *close occupation* | |
| health associate professionals (322) | 0.299 |
| *distant occupation* | |
| crop and animal producers (613) | 0.710 |
| **finance and sales associates (341)** | |
| *close occupation* | |
| other specialist managers (123) | 0.225 |
| *distant occupation* | |
| special education teaching prof. (234) | 0.633 |
| **secondary education teachers (232)** | |
| *close occupation* | |
| college, university and HE teaching prof. (231) | 0.233 |
| *distant occupation* | |
| computer associate prof. (312) | 0.692 |

*note:* based on author's calculations; isco '88 coding between parentheses

Table 4: Selected moments of the distance density function under observed and random selection

| | **Observed Selection** | **Random Selection** |
|---|---|---|
| Minimum | 0 | 0 |
| 25th percentile | 0.151 | 0.437 |
| Mean | 0.340 | 0.510 |
| Median | 0.404 | 0.511 |
| 90th percentile | 0.612 | 0.667 |
| Maximum | 0.808 | 1 |
| Standard deviation | 0.223 | 0.129 |
| | | |
| Kolmogorov-Smirnov Test | D = 0.224 | p-value = 0.000 |

*note:* based on author's calculations

Table 5: The distance indicator compared to other mismatch measures
dependent variable: standardized distance indicator

| **Variables** | **(1)** | **(2)** | **(3)** | **(4)** |
|---|---|---|---|---|
| *content match* | | | | |
| exclusively own field req. | ref. | | | ref. |
| own or related field req. | 0.332*** | | | 0.323*** |
| | (0.077) | | | (0.078) |
| very different field req. | 1.020*** | | | 0.947*** |
| | (0.087) | | | (0.090) |
| no particular field req. | 0.925*** | | | 0.785*** |
| | (0.079) | | | (0.092) |
| *educational match* | | | | |
| higher level req. | | ref. | | ref. |
| same level req. | | 0.004 | | -0.111 |
| | | (0.097) | | (0.097) |
| lower tertiary level req. | | -0.352 | | -0.397* |
| | | (0.207) | | (0.202) |
| below tertiary level req. | | 0.642*** | | 0.108 |
| | | (0.104) | | (0.115) |
| *skill match* | | | | |
| intensive skill use | | | ref. | ref. |
| standard skill use | | | 0.325*** | 0.044 |
| | | | (0.062) | (0.065) |
| low skill use | | | 0.897*** | 0.273* |
| | | | (0.088) | (0.108) |
| R-squared | 0.138 | 0.052 | 0.044 | 0.146 |
| correlation | 0.37 | 0.23 | 0.21 | 0.38 |
| N | 1109 | 1109 | 1109 | 1109 |

*note:* stars indicate significance levels: * p<0.05, ** p<0.01, *** p<0.001; robust standard errors are given between parentheses

Table 6: Occupational spread according to different fields

dependent variable: standardized distance indicator

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *educational degree (ref.: bachelor)* | | | | |
| master | | -0.133 | | -0.146 |
| | | (0.085) | | (0.079) |
| phd | | 0.530** | | 0.455* |
| | | (0.194) | | (0.228) |
| *internship* | | -0.467*** | | -0.205** |
| | | (0.066) | | (0.064) |
| *experience (months)* | | -0.004 | | -0.002 |
| | | (0.003) | | (0.002) |
| *unemployed* | | 0.194** | | 0.121* |
| | | (0.063) | | (0.057) |
| *public sector* | | 0.092 | | -0.105 |
| | | (0.059) | | (0.057) |
| *field of study (ref.: humanities)* | | | | |
| agriculture, forestry and fishery | | | -0.444** | -0.293 |
| | | | (0.165) | (0.167) |
| arts | | | -0.362** | -0.363** |
| | | | (0.131) | (0.130) |
| economics, business and administration | | | -0.890*** | -0.866*** |
| | | | (0.108) | (0.112) |
| education | | | -1.144*** | -1.004*** |
| | | | (0.202) | (0.204) |
| engineering, building and architecture | | | -1.192*** | -1.095*** |
| | | | (0.118) | (0.126) |
| health | | | -0.777*** | -0.673*** |
| | | | (0.146) | (0.143) |
| law | | | -0.686*** | -0.709*** |
| | | | (0.196) | (0.196) |
| linguistics | | | -0.161 | -0.173 |
| | | | (0.124) | (0.124) |
| maths, stats and computing | | | -0.587*** | -0.558*** |
| | | | (0.139) | (0.143) |
| personal services | | | -0.324 | -0.220 |
| | | | (0.171) | (0.177) |
| science | | | 0.065 | 0.063 |
| | | | (0.115) | (0.116) |
| social and behavioral science | | | 0.284* | 0.260* |
| | | | (0.114) | (0.114) |
| social services | | | -1.385*** | -1.219*** |
| | | | (0.251) | (0.251) |
| controls | yes | no | no | yes |
| adj. R-squared | 0.019 | 0.063 | 0.232 | 0.253 |
| N | 1109 | 1109 | 1109 | 1109 |

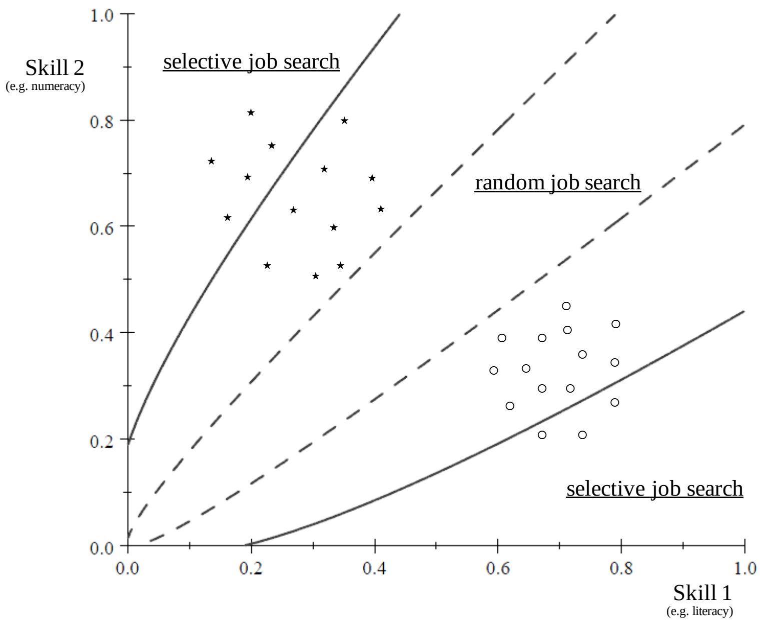
*note:* stars indicate significance levels: * p<0.05, ** p<0.01, *** p<0.001; robust standard errors are given between parentheses; controls are gender, age, marital status and ethnic background.

Table 7: The wage effect of different types of mismatch

dependent variable: log of gross hourly wages

| **Variables** | **(1)** | | **(2)** | | **(3)** | |
|---|---|---|---|---|---|---|
| | **b** | **rob. se** | **b** | **rob. se** | **b** | **rob. se** |
| *stand. distance indicator* | -0.081*** | 0.013 | -0.053*** | 0.015 | -0.027 | 0.015 |
| *field of study* | | | | | | |
| humanities | | | ref. | ref. | ref. | ref. |
| agr., forestry & fishery | | | -0.184 | 0.113 | -0.204 | 0.114 |
| arts | | | -0.146* | 0.072 | -0.105 | 0.067 |
| econ., business & admin. | | | 0.142** | 0.054 | 0.140** | 0.054 |
| education | | | -0.023 | 0.100 | -0.052 | 0.093 |
| eng., building & architecture | | | 0.062 | 0.067 | 0.028 | 0.068 |
| health | | | 0.190*** | 0.055 | 0.142* | 0.057 |
| law | | | 0.063 | 0.081 | 0.044 | 0.078 |
| linguistics | | | 0.010 | 0.066 | -0.029 | 0.065 |
| maths, stats & computing | | | 0.136* | 0.060 | 0.121* | 0.059 |
| personal services | | | -0.021 | 0.071 | -0.046 | 0.075 |
| science | | | 0.115* | 0.056 | 0.088 | 0.057 |
| social & behavioral science | | | 0.039 | 0.051 | 0.020 | 0.050 |
| social services | | | -0.015 | 0.077 | -0.052 | 0.075 |
| *relative score* | | | | | | |
| lower than average grades | | | ref. | ref. | ref. | ref. |
| average grades | | | 0.178 | 0.091 | 0.156 | 0.140 |
| higher than average grades | | | 0.256* | 0.104 | 0.225 | 0.147 |
| *prestigious institute* | | | 0.198*** | 0.041 | 0.162*** | 0.040 |
| *content match* | | | | | | |
| exclusively own field | | | | | ref. | ref. |
| own or a related field | | | | | 0.021 | 0.033 |
| a completely different field | | | | | 0.077 | 0.049 |
| no particular field | | | | | 0.031 | 0.046 |
| *skill match* | | | | | | |
| intensive skill use | | | | | ref. | ref. |
| standard skill use | | | | | -0.102*** | 0.030 |
| low skill use | | | | | -0.146* | 0.068 |
| *educational match* | | | | | | |
| higher level req. | | | | | ref. | ref. |
| same level req. | | | | | 0.035 | 0.053 |
| lower tertiary level req. | | | | | -0.106 | 0.103 |
| below tertiary level req. | | | | | -0.277*** | 0.069 |
| controls | no | | yes | | yes | |
| | | | | | | |
| R-squared | 0.032 | | 0.149 | | 0.224 | |
| N | 1024 | | 1024 | | 1024 | |

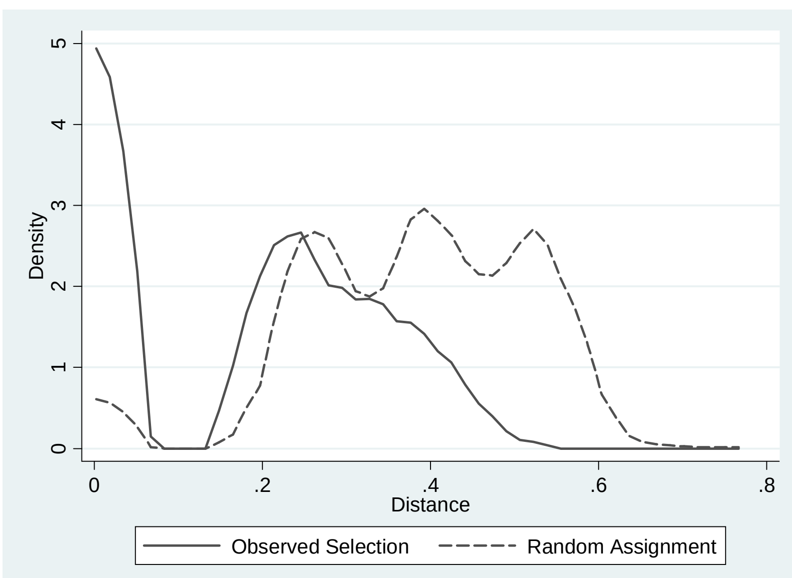
*note:* stars indicate significance levels: * p<0.05, ** p<0.01, *** p<0.001; controls are gender, age, ethnic background, internship, unemployment spell, experience, experience squared, marital status, public sector employment and educational degree figure 1: The Search Behavior of Workers




*notes:* ○ cluster A workers — marginal workers when $|t_{j1} - t_{k1}|$ is small

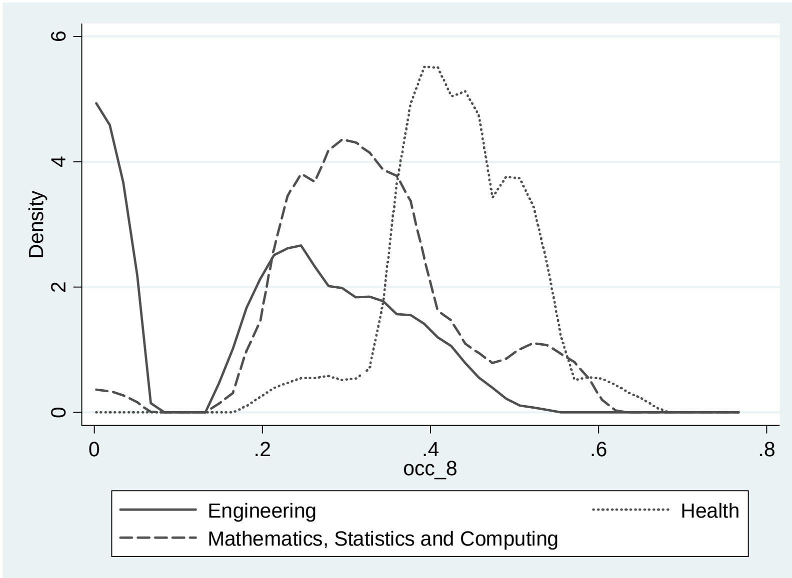
⋆ cluster B workers - - marginal workers when $|t_{j1} - t_{k1}|$ is large

Figure 2: Engineering, building and architecture graduates:
observed selection and random assignment



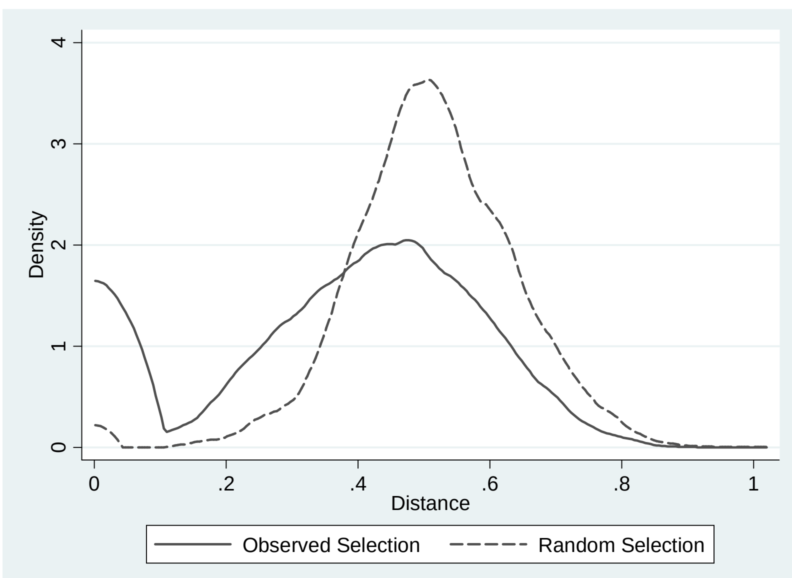

*note:* Epanechnikov kernel function (bwidth = 0.03)

Figure 3: Related and distinct occupational selection




*note:* Epanechnikov kernel function (bwidth = 0.03)

Figure 4: Distance under observed and random selection




*note:* Epanechnikov kernel function
(bwidth observed = 0.04; bwidth random = 0.02)

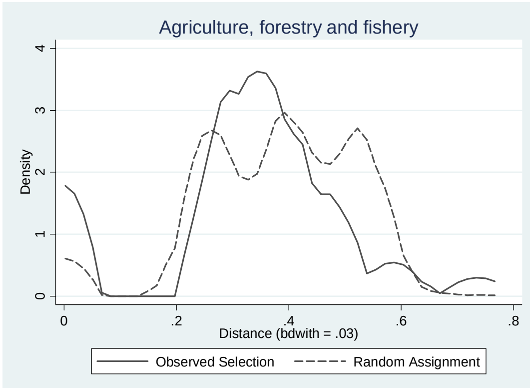
# Appendix A: Tables & Figures

Table A1: Identification of proper match between field of study and occupation

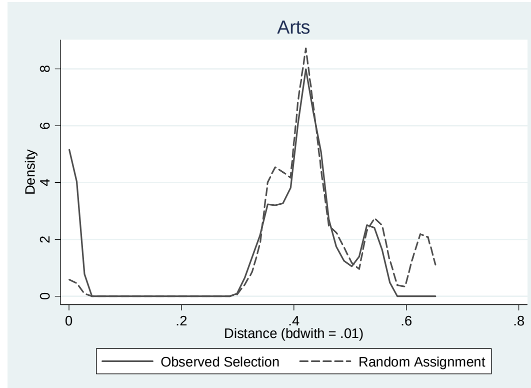
| **Field of study** | **Occupational match (isco)** | **N** |
|---|---|---|
| agriculture, forestry and fishery | Architects, engineers and related professionals (214) | 37 |
| arts | Artistic, entertainment and sports associate prof.(347) | 88 |
| economics, business and administration | Business professionals (241) | 180 |
| education | Primary and pre-primary education teaching prof. (233) | 48 |
| engineering, building and architecture | Architects, engineers and related professionals (214) | 89 |
| health | Health professionals (except nursing) (222) | 87 |
| humanities | Secondary education teaching professionals (232) | 115 |
| law | Legal professionals (242) | 53 |
| linguistics | secondary education teaching professionals (232) | 130 |
| mathematics, statistics and computing | Computing professionals (213) | 90 |
| personal services | Production and operations managers (122) | 34 |
| science | Physicists, chemists and related professionals (211) | 139 |
| social and behavioral science | Social science and related professionals (244) | 198 |
| social services | Health associate professionals (except nursing) (322) | 18 |
| | | |
| total nr. of observations | | 1306 |

*note:* last column indicated by N shows the number of graduates per field of study.

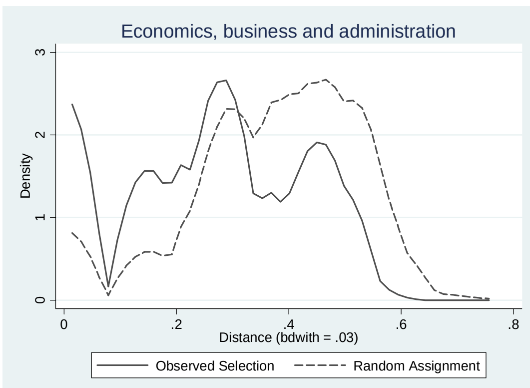
Table A2: Distance density function for different fields of study (1)



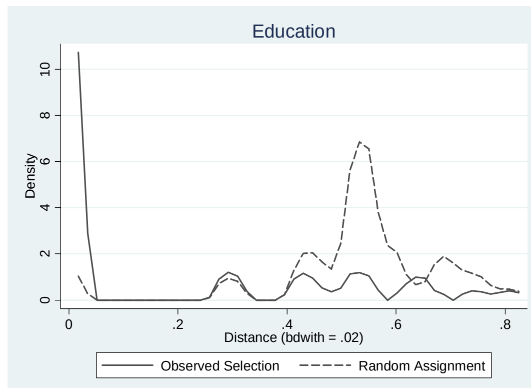

(K-Smirnov: D = 0.244, p-value = 0.018)



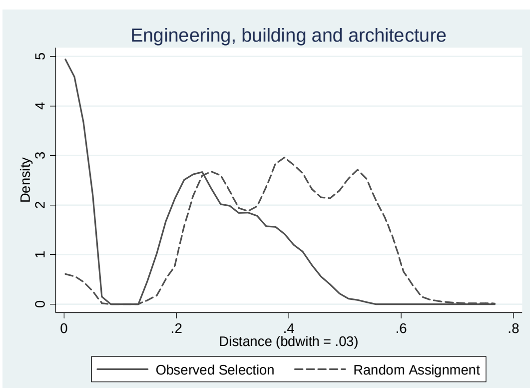

(K-Smirnov: D = 0.225, p-value = 0.000)



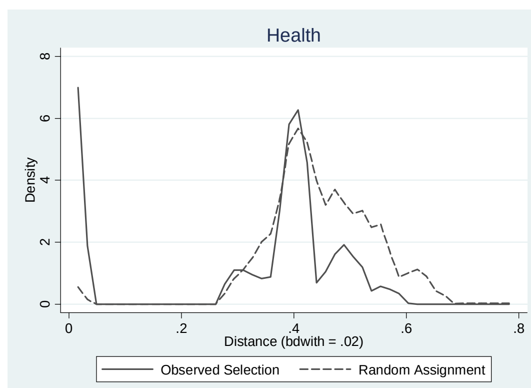

(K-Smirnov: D = 0.414, p-value = 0.000)




(K-Smirnov: D = 0.659, p-value = 0.000)

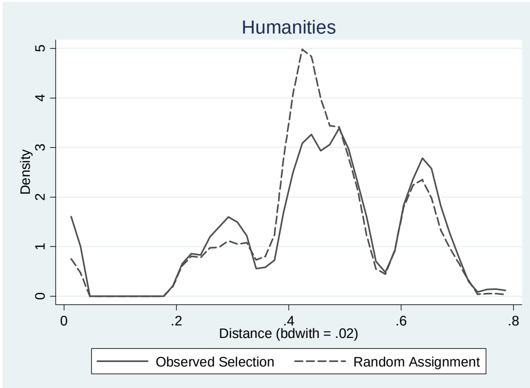



(K-Smirnov: D = 0.561, p-value = 0.000)

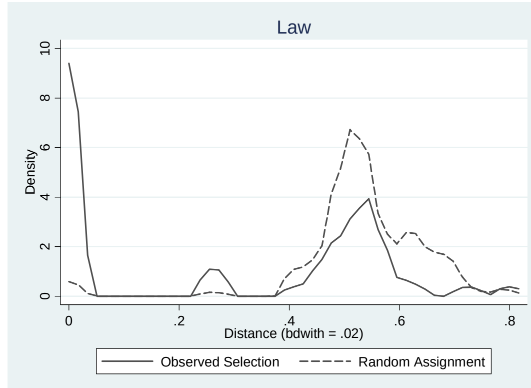



(K-Smirnov: D = 0.537, p-value = 0.000)

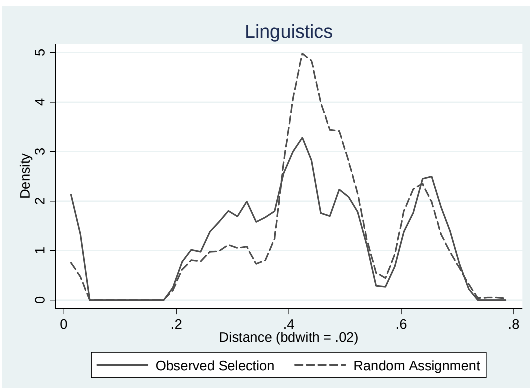
Table A2: Distance density function for different fields of study (2)



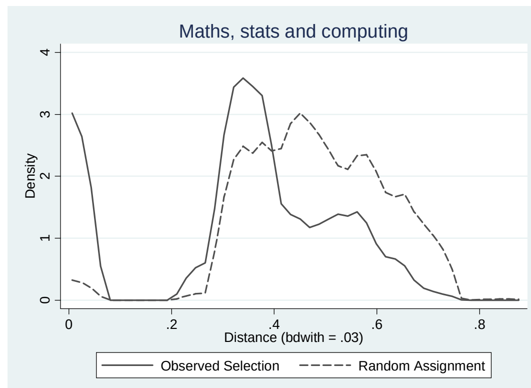

(K-Smirnov: D = 0.107, p-value = 0.149)



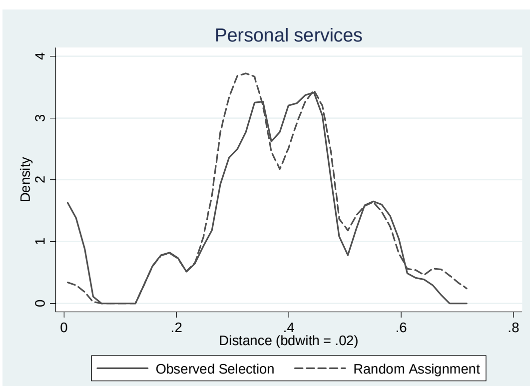

(K-Smirnov: D = 0.512, p-value = 0.000)



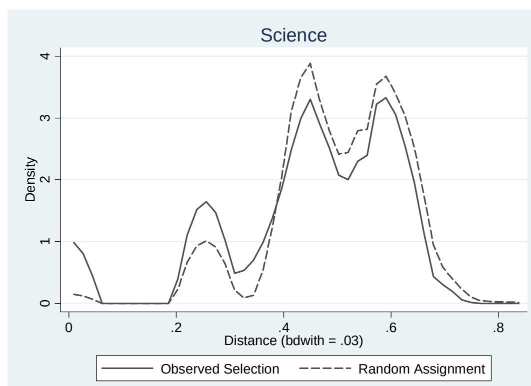

(K-Smirnov: D = 0.225, p-value = 0.000)



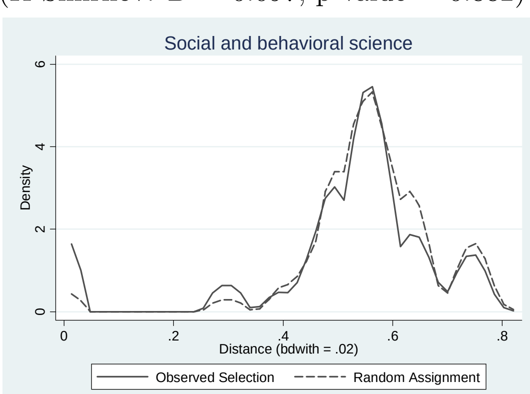

(K-Smirnov: D = 0.446, p-value = 0.000)



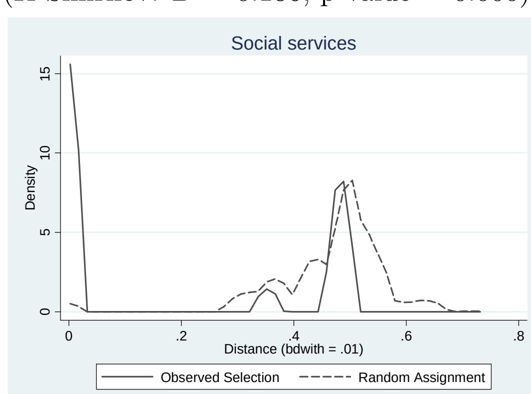

(K-Smirnov: D = 0.097, p-value = 0.882)




(K-Smirnov: D = 0.186, p-value = 0.000)




(K-Smirnov: D = 0.1373, p-value = 0.003)




(K-Smirnov: D = 0.600, p-value = 0.000)