Education Surplus, Skills Mismatch and Technological Change[1]

Ceren Ozgen [a,2] Marco G. Ercolani [b] Raymond J.G.M. Florax [c,d]

Abstract

This paper considers if the 'routinisation' of many occupations, driven by factors such as outsourcing and the advent of the World Wide Web, has caused overeducation, undereducation and a wage penalty for these employees. The resulting polarisation of the labour market is further impacted by supply-side factors, such as the influx of migrants, which may have adverse effects on wages. We estimate a three-equation trivariate probit system with overeducation, undereducation and probability of receiving a wage penalty as binary dependent variables. Furthermore, overeducation and undereducation also enter as explanatory variables in the wage penalty equation. Our approach differs significantly from the traditional ORU (overeducation, required education and undereducation) estimates where overeducation and undereducation variables only enter in the system as dependent variables to avoid endogeneity bias. To aid convergence in the trivariate probit we apply appropriate exclusion restrictions and we apply constraints on unwarranted combinations of education and skills mismatch. Our analysis is based on exceptionally rich Dutch micro-data that uniquely enables us to match each employee's occupation to its required level of education (as defined by Statistics Netherlands). This matching using official data enables us to overcome the typical self-reporting bias found in much of the overeducation literature. Our results suggest that routinisation has mainly led to overeducation and, improved, undereducation.

Keywords: overeducation, skills mismatch, routinization, wage premium

JEL-classification: J24, J31, J61, J62, R23

---

[1] We have profited substantially from the comments and suggestions of Jacques Poot and Stan Siebert. Ceren Ozgen gratefully acknowledges the Marie-Sklodowska Curie Individual Grant for MAStErS project (H2020-MSCA-IF-2015, No. 705366) from the European Commission.

[2] Corresponding author: c.ozgen@bham.ac.uk; Tel. +44 12141446235.
[a] Department of Economics and IRiS, University of Birmingham, UK; and IZA Bonn, Germany.
[b] Department of Economics, University of Birmingham, UK
[c] Department of Agricultural Economics, Purdue University, USA
[d] Department of Spatial Economics, Vrije Universiteit Amsterdam, and Tinbergen Institute, Netherlands

1 Introduction

Technological improvements and the changing geography of production both complement and threaten the matching of employees' skills with jobs. Given that skills groups are imperfect substitutes in production, relative supply-demand changes shift wage and employment outcomes for these groups. It is well-documented in the literature that the decline of middle segment jobs (mostly routine-task intense) is an implication of technological change and of the changing nature of production (Acemoglu and Autor, 2011). These changes could play a significantly detrimental role by increasing overeducation. This is assuming that options like on-the-job training or enrolling in education, which could potentially ensure upskilling of employees cannot easily outpace the changing demand for skills.

This study offers an innovative perspective on introducing task complexity and demand-shifts into the over-education literature through the estimation of a probit model. We explore the role of technological changes, like routinisation, on the prevalence of skills mismatch and the consequent wage premium the employees receive. We focus mainly on over-educated employees (*i.e.* those in jobs below their educational attainment) but we also provide substantive discussion on under-educated employees (*i.e.* those who work in jobs requiring higher level of education than they have). Our research follows two related lines of inquiry. Firstly, how do the changing demand-side conditions with respect to technology (*e.g.* firms' decision to outsource routine tasks) influence the prevalence and persistence of overeducation? Secondly, and simultaneously, what is the response of wages to evolving technological changes for these over-educated employees? Our approach enables us to extend the standard discussion beyond education and occupation mismatches and provides further knowledge on describing a mechanism of how becoming over- and under-educated is shaped by external technological changes. We use exceptionally rich administrative data obtained from the Statistics Netherlands under strict confidentiality agreements. Our analysis uses linked employer-employee data, which not only permits incorporating technological-change induced determinants of skills mismatch but also allows for firm heterogeneity and the inclusion of wages for these employees.

Many aspects of technologically-induced skills mismatch have previously been overlooked. Among these is that demand-side dynamics are mostly beyond the influence of individual employees and they are likely to exacerbate the skills mismatch. In parallel to computers replacing the tasks of mostly middle-skilled employees, the changing dynamics of international trade through outsourcing and offshoring impacts the structure of employment. In particular jobs with routine tasks are likely to be outsourced through the changing nature of production, both in manufacturing and in information-based tasks.[3] Firms' decisions to outsource routine segment jobs, especially internationally, may lead either to increased unemployment or higher competition for jobs within occupations. It becomes increasingly challenging for employees in vanishing routine-task complexity jobs to acquire positions in higher complexity jobs while facing competition from similarly skilled employees. Finally, the type of skills employees can supply to the market in response to technological changes is unlikely to meet the immediate needs of firms.

Internationally mobile labour, which we discuss later and which we incorporate in our modelling, can be an imperfect substitute to address this immediate need. The hunt for international talent and qualifications requires extra effort to attract the most talented employees and to select the 'correct' international immigrants. Evidently, labour migration policies in developed countries increasingly attract highly skilled immigrants by offering generous subsidies and funding schemes for employees and talented students (e.g. Green card, EU Blue card and others). The migration literature documents that many skilled people however struggle to find jobs within host countries or take jobs well below their educational level (Eurostat, 2011). When employees are overqualified for their jobs there are substantial private and social losses such as deferred labour market assimilation and crowding-out. As a whole, nevertheless, the cost of mismatch is shared between individuals, firms and economy. In our analysis, we devote particular attention to how 'global skills' act to counterbalance the impact of technological changes on the prevalence of overeducation.

Human capital theory predicts higher earnings for greater educational achievement (Becker, 1964). The empirical evidence however contradicts the theory by indicating an earnings penalty for overeducated employees such that excess schooling is associated with lower payoffs than required level of education (Leuven and Oosterbeek, 2011). Moreover, returns to schooling have decreased in many countries while the number of schooling years has increased steadily across developed countries (Katz and Autor, 1999). The literature provides substantial empirical evidence on the individual returns to overeducation (Chiswick and Miller 2009; Hartog 2000; Nordin et al. 2010). Driven by data availability, most of these studies are conducted at the individual level. Typically, a so-called ORU specification is estimated to predict the returns to wages. Many studies apply ORU specification while at the same time controlling for individual characteristics, regional effects or firm effects. Predominantly, the skills mismatch literature approaches overeducation as an incident due to a surplus of skills. The typical ORU setup is based on employee attributes as a marker for identifying the job-education level mismatch. A better understanding of the observed earnings penalty for

---

[3] One particular advantage of outsourcing these jobs is the almost costless nature of their transportation across countries. So routinisation is likely cut across skill levels but is confined to certain sectors. Moreover, Autor *et al*. (2013) show that the impact of computerization and trade exposure have shifted from the production to the services sector for routine tasks and have been asymmetric across gender, age and skills groups in the labour market.

over-educated employees is warranted. At the same time is there is a need to jointly incorporate demand and supply side factors into the analysis.

Hartog and Oosterbeek (1988) pointed out an important derivation of ORU models, based on the job competition theory. Their approach not only opens many new research avenues for skills mismatch research, but could also offer further insights on demand-side effects, where empirical evidence is rather limited. The stylized facts in many countries reveal a number of important dynamics within labour markets. Firstly, among these is that in recent decades there has been a significant increase in the average level of education. Secondly, returns to human capital have been decreasing in many countries. In an elegantly simple exercise on the Netherlands for 1982-1988, Teulings (1995) shows that, even before the advent of the World Wide Web, the skill level of employees rose more rapidly than the complexity of job requirements. This means that employees are increasingly unlikely to find jobs that satisfy their skills. Moreover, the globalization of production decreased demand for blue-collar jobs via international outsourcing, mostly in the manufacturing sector. Given that different skill groups are imperfectly substitutable, relative supply-demand changes are likely to affect wage and employment outcomes.

The literature consistently uncovers job polarisation due to decline of jobs in the middle-skilled (routine-tasks) segment of the labour market. Potentially this could play a detrimental role by increasing over-education in all segments of the labour market. An extra effect on the middle-skilled employees is also observed because they are typically not endowed with the skills to quickly adapt to new technologies and are unable to take up positions in the higher skills segment. Instead they end up competing with their peers for over-subscribed routine-task jobs.

Clearly, these dynamics cannot be controlled by the individual employee while may have serious consequences for her earnings. Interestingly, despite a strong skill upgrading of jobs and rapid occupational change and hollowing out of middle segment routine-jobs, low-skilled labour supply stayed relatively stable (Oesch, 2013). These changes bring into focus other influences such as international migration (which is a source of low-skilled labour supply) or offshoring (which had a relatively small impact on low-skilled labour) that are beyond the simple supply-demand equilibrium within local, domestic labour markets. In the light of the above-mentioned developments, overeducated employees may be heterogeneous with regard to their occupation and with regard to their employer’s location, size and market orientation. Moreover, increasing wage dispersion within groups and changing demand for the complexity of the labour-market tasks are potential forces that cause the asymmetry in returns to overeducation. We analyse these intriguing forces in more detail in our econometric analysis.

The rest of the paper is organised as follows. Section 2 explains the modelling specification and identification. Section 3 describes the matching firms to employee level micro-data and provides a descriptive analysis. Section 4 presents the empirical findings of the multivariate probit model for the joint probability of being over- and undereducated and their respective effect on wages. Section 5 concludes.

2 Trivariate probit model

This analysis is possible thanks to a unique dataset we have constructed by linking individual employee and employer data (LEED) for the Netherlands. These links are possible thanks to being given access to confidential employer and employee identifiers. Using these linked data, we can model the level of job-skill mismatch using variables such as skill heterogeneity and self-selection. Furthermore, our dataset enables us to address some common measurement error problems found in the literature (see Beckhusen *et al.* 2013 for a detailed review). One such problem is that over-education has often been measured by employees' subjective self-assessments. Another one is that it has sometimes been measured by comparing each employee's education to the average for her occupation group, known as the 'realised matches method'. The fact that these measures are problematic is, evident from the low correlations among different measures of mismatch (see Verhaest and Omey, 2006). Yet another source of bias is the failure to control for other endogenous variables in the system.

Our study addresses these problems in several ways. First, to mitigate the measurement bias our study integrates an objective job analysis approach, which does not suffer from self-reporting bias. Instead it is based on a classification of occupations by the required level of education level in the Netherlands, all compatible with the most detailed level of ISCO and ISCED classifications. The (mis)match between the highest degree of education achieved by an employee and the required level of education in the respective occupation identifies each employee as over-, required or under-educated and the extent of any over- or under-education (in terms of years). Second, we deal with the inference problem arising from the endogeneity of ability that typically plagues the estimates of wage equations by using a trivariate probit model, explained in detail below.

2.1 The model

We specify a modified ORU system of equations to estimate the probability of being overeducated (y^{o*}), undereducated (y^{u*}), or of receiving a positive wage premium (y^{w*}). The excluded alternatives are required-education and not receiving a positive wage premium. Estimation of the resulting trivariate probit system is by simulated maximum likelihood (SML):

$$y^{o*} = \alpha_1 + \alpha_2 x'_i + \alpha_3 x'_k + \alpha_4 x'_l + \alpha_5 s'_{jl} + \alpha_6 o_{jl} + \eta_s + \eta_l + \varepsilon^o, \quad (1)$$

$$y^{u*} = \beta_1 + \beta_2 x'_i + \beta_3 x'_k + \beta_4 x'_l + \beta_5 s'_{jl} + \beta_6 u_{jl} + \theta_s + \theta_l + \varepsilon^u, \quad (2)$$

$$y^{w*} = \gamma_1 + \gamma_2 y^{o*} + \gamma_3 y^{u*} + \gamma_4 e_i + \gamma_5 m_i + \gamma_6 a_i + \xi_s + \xi_l + \varepsilon^w, \quad (3)$$

where the subscripts index an individual employee i, working in occupation j, who is employed in firm k, located in labour market area l for the firm's location within forty 'Corop' regions based on daily commuting areas across the Netherlands.

Exogenous regressors in equations (1) and (2) include vectors of personal (x'_i), firm (x'_k), labour market area characteristics (x'_l), and demand and supply shifters (s'_{jl}). The demand shifter is an index measure of routinisation, whereas the supply shifter is the ratio of foreign-born with specific educational attainments and skills. Exogenous regressors in equation (3) include work experience (e_i), employment tenure in the same firm (m_i) and age (a_i). All three equations also include fixed effects for sector $(\eta_s, \theta_s, \xi_s)$ and location $(\eta_l, \theta_l, \xi_l)$. The errors (ε) are potentially correlated.

Over-education (y^{o*}) and under-education (y^{u*}) are included as endogenous explanatory variables in equation (3). Therefore, in equations (1) and (2) we include variables that meet the 'exclusion restriction' of being uncorrelated with the positive wage premium (y^{w*}) but of being significant regressors for y^{o*} and y^{u*}. We construct these exclusion restriction variables using detailed data on years-of-schooling and level-of-schooling in each occupation by labour market area. The exclusion restriction variable in equation (1) is the fraction of people with the same level of schooling and working in the same city that are *not overeducated* (o_{jl}). This is an indicator of the ease with which one could find a job for which one has the required education; in other words the ease with which one can move up the job/education ladder. We therefore expect the estimated coefficient on o_{jl} to be positive. The exclusion restriction variable in equation (2) is the fraction of people with the same level of schooling and working in the same city that are *not undereducated* (u_{jl}). This is indicative of the same moving up the job/education ladder phenomenon. In those cases where there are many people with the same level of schooling who are not undereducated it is then easier for the employee to move up, and hence the probability of one being undereducated is smaller. So, we expect the estimated coefficient on u_{jl} to be negative.

Estimation of this trivariate system of probit equations by SML is computationally intensive because it is nonlinear and because it is a system of simultaneous equations. Part of this complexity can be seen in Table 1, which illustrates that for each individual there are eight potential outcome combinations given the three binary dependent variables (2^3=8). However, we can aid the estimation by limiting the number of outcome combinations to just six. This is because, as Table 1 illustrates, the first two outcomes are logically inconsistent as they assume a person is both overeducated and undereducated. In order to exclude these inconsistent combinations, we built restrictions into the likelihood estimator. We estimate this system of equations using the package `mvprobit` in the Stata statistical software.[4] We also carry out simple consistency checks by estimating a linear probability model approximation of this system using the Stata

[4] mvprobit uses marginal probit regressions as starting values for the SML estimator.

command `reg3`. These linear probability model estimates are typically much less computationally intensive than the probit estimates but are good approximations to the probit estimates if the mean value of the dependent variable(s) is close to 0.5 (see Gunderson (1980) and Hellevik 2009).

Table 1: Potential outcomes and inconsistent outcomes

Outcome	y^o	y^u	y^w	Comment
1	1	1	0	Inconsistent
2	1	1	1	Inconsistent
3	1	0	0	Overeducated, no wage premium
4	1	0	1	Overeducated, positive wage premium
5	0	1	0	Undereducated, no wage premium
6	0	1	1	Undereducated, positive wage premium
7	0	0	0	Required education, no wage premium
8	0	0	1	Required education positive wage premium

2.2 The variables

The binary wage premium variable y^{w*} is operationalized by defining a premium threshold value that is the median wage per occupation j in labour market area l. The threshold value is constructed by using a 1600-cell grid of 2-digit ISCO classification by 41 labour market areas in the Netherlands. In order to ensure there is a sufficient number of observations in each grid-cell, we combine the 2009, 2010 and 2011 waves of Labour Force Surveys. As mentioned above, following the career mobility theory (Sicherman and Galor, 1990), we include experience and tenure as regressors in the wage equations to control for the possibility of employees being at different stages of their career. The cut-off point determining the wage premium is defined on the basis of the median wage by three-digit occupation *j,* across forty labour market areas *l*:

$$y_i^{w*} = \begin{cases} 1 & \text{if} \quad w_i > \text{ median}(w_{\mathrm{j}l}) \\ 0 & \text{if} \quad w_i \leq \text{ median}(w_{\mathrm{j}l}) \end{cases} \tag{4}$$

Therefore, when employee i's wage w_i is below the cut-off point of his occupational and market median, then his binary wage premium y_i^{w*} equals 1, and 0 otherwise. Defining y_i^{w*} in this way also ensures that the mean value of y_i^{w*} is 0.5, which conveniently means the linear probability estimator should be a good approximation for the probit estimates of equation (3).

The variables for employee, firm and municipal characteristics are as follows. For employee characteristics we use: age, gender, ethnicity, employment tenure and labour market experience. These demographic characteristics may affect the probability of an employee being correctly matched. Firm characteristics are: firms' size and each firms' average wages paid by skill level. Economic characteristics for the labour market areas are operationalized using the usual indicators for specialization and diversity, i.e. a location quotient that varies over sectors and an inverse Hirschman-Herfindahl index that varies over labour market areas.

The immigration variable, which is a proxy for supply side factors, also varies across occupations j and labour market areas l in period $t-15$. This accounts for the response of factor mobility to the changing conditions of the local labour markets.

We expect the demand shift variable (i.e. routine-task intensity index) to predict a polarized occupational distribution at the bottom and at the top of the occupational ladder at the expense of the middle. This mechanism potentially impacts the probability of overeducation but the extent of this impact is an empirical matter.

Firstly, we construct a routinisation measure lagged by 15 years through assigning each occupation to a routine-task intensity level. This measure depicts the extent of routinisation in each occupation and is expected to have a positive correlation with the probability of overeducation. In other words, the higher the vulnerability of occupation j to computerization of tasks, the higher the likelihood of employee i experiencing skill mismatches. This variable is related to the content of the tasks (analytic, manual, routine) used in an occupation and how these tasks change over time (towards routine tasks or not) in the same occupation. It refers to the degree of routine tasks in order to perform a certain occupation. The level of routinisation can be independent of the education required for an occupation.

In order to avoid correlation of the routinisation index with contemporaneous production dynamics, and to account for long-term impact of routinisation, we follow Autor et al. (2013) and lag this routinisation variable by 15 years. We calculated the degree of routine-task intensity n, where $n = 1,2,\dots,N$. For this we used the importance scores from the O*NET database, in each occupation j in time $t-15$. This construct is not contaminated by simultaneity as it excludes the transition of employees from different phases of routine-task levels based on their unobserved ability component. The five commonly used categories of routineness and non-routiness in the literature are routine cognitive (RC), routine manual (RM) non-routine manual (NRM) non-routine analytic (NRA), and non-routine interactive (NRI).

The five components of routiness and non-routiness are constructed as sums of the following standardised measures of routiness that have been calculated for each and every ISCO'08 code at the four-digit level. Note that each one of these standardised measures of routines and non-routiness have also been normalised using the same formula in equation (6) before they are included in the equation (5).

To reduce *data dimensionality,* following Autor and Dorn (2013), we combine these five indicators (at the four-digit ISCO level)[5] into a single composite measure of routine task-intensity (RTI) as shown below:[6]

---

[5] The Autor et al. (2003) study uses US SOC, while Dutch occupations are compatible with the ISCO. Therefore we used SOC-ISCO crosswalk provided by the US Labour Bureau to translate the degree of routine-task intensity in SOC into ISCO.

$$\mathrm{RTI}_{j,t-15} = \mathrm{RC}_{j,t-15} + \mathrm{RM}_{j,t-15} - \mathrm{NRM}_{j,t-15} - \mathrm{NRA}_{j,t-15} - \mathrm{NRI}_{j,t-15} \quad (5)$$

where RTI varies across occupations j in period $t - 15$. However, before estimation this measure is normalised using the standard normalisation formula:

$$\mathrm{RTI}_{j,t-15} = (\mathrm{RTI}_{j,t-15} - \mathrm{mean}[\mathrm{RTI}_{j,t-15}]) / \mathrm{s.d.}(\mathrm{RTI}_{j,t-15}) \quad (6)$$

where mean(...) denotes the mean value of the variable and s.d.(...) denotes its standard deviation. This normalisation ensures the estimated coefficient on RTI will be within a 'reasonable' range so we can identify the direction and statistical significance of its effect but not its scale effect.

3 Data

In this research we make use of unique micro-data sets obtained under confidentiality agreement with Statistics Netherlands. Although we utilize and link several micro-data sets, the core of the analysis is based on Dutch Labour Force Surveys (LFSs) because the education and occupation level of employees in the Netherlands can only be obtained through the Dutch LFSs. These Dutch Labour Force Surveys are a rotating panel of around 80,000 observations per wave, per annum spanning 1995-2016. We then link these Dutch LFSs with data from tax and municipal registrations.[7]

Data on the educational attainment and occupation of employees are obtained from the 2011 Dutch Labour Force Survey (LFS), which provides an employee's highest educational degree as well as their most recent occupation(s). We use the lowest possible level of occupational disaggregation when we construct our measures of over- and undereducation. These measures are then constructed by matching the three-digit International Standard Classification of Education (ISCED) codes at the four-digit International Classification of Occupations (ISCO) level.[8] An indicator for the required

---

[6] Using US based task-complexity definition can be argued to be sub-optimal given the different structure of the both economies. We would nevertheless argue that pre-WWW period the task content of the occupations should be similar across countries that are at the similar development level compared to today. Today given the pace of technology adaptation the job contents within the same occupations across countries should indicate a larger dispersion. In order to cross-validate this expectation, to the extent possible, we resort to OECD's findings from the Survey of Adult Skills (PIAAC) where OECD (2014) compares countries by the skills used at the workplaces. It is shown that the degree of Task discretion; ICT skills and Dexterity (which compares closest to Autor et al (2003) classification of routineness of tasks) in the two countries –US and the Netherlands- compares almost identical, with only Dexterity being more frequently encountered than it is the Netherlands.

[7] These datasets include: 1) EBB: Labor force survey (80,000 obs); 2) BAANKENMERKENBUS: Characteristics of jobs (10 million obs); 3) BAANSOMMENTAB: Earnings and structure of earnings (10 million obs); 4) GBA: Municipal registrations (20 million obs).

[8] The ISCED distinguishes 9 1-digit and 15 2-digit categories of education levels in the Dutch LFS. The ISCO-08 distinguishes 10, 43, 130 and 433 occupations at the first through fourth digit level. (*QC: this needs more detail that can be provided in an Appendix.*)

level of education by occupation provided by Statistics Netherlands[9] allows for the operationalization of the latent variables for over- and undereducation into binary indicators. Data on individual wages are obtained by linking the LFS data to a comprehensive tax database of all employees in the Netherlands. These linked data reveal detailed information on individual wages and jobs, including the sector and location of firms.

Dutch Tax records (based on BAAN datasets) include about 10 million observations based on annually collected information on all employed tax payers and their job characteristics, though the self-employed are not included. Each employee is observed repeatedly over time in different jobs. We therefore observe earnings, job spells and information on employers such as detailed sector classification, size and location by postal code. Given that the education, occupation, sector and location coding in these datasets is available at the most disaggregated level of classification, this enables us to meaningfully construct covariates for the estimations and enjoy a certain degree of flexibility for robustness checks.

The municipal registrations, which are based on about 20 million observations covering the entire population in the Netherlands, offer a wide range of information on the ethnic background, household characteristics and residential location of the employees. We retrieve employees' country of birth, age, gender and marital status by linking tax and municipal registrations. The linking process takes Dutch LFSs as the reference dataset because the variables of interest, namely education and occupational background, can only be observed there. Therefore, in order to create a linked employee-employer dataset (LEED), firstly the 2011 LFS is merged to the tax records and then to municipal registers. Finally, combining these micro-data a rich collection of information is obtained both for firms as well as employees on their mobility, performance, location and various other characteristics. In our study, a job an employee holds refers to an occupation, which we can observe in the tax records with a start date and job identification number. An employee is regarded as foreign if he was born outside of the Netherlands.[10]

Our study makes use of LEED data from 2011, meaning that the left-hand side variables are defined for this year. While demand (routine-task intensity index) and supply (net immigration stock per occupations) shifters variables are constructed for 1996 which is more than a decade earlier. The motivation behind this setup is to isolate

---

[9] This classification of occupations by required level of education is kindly provided by Statistics Netherlands. This variable is produced through a joint work of Statistics Netherlands together with Research Centre for Education and Labour Market (ROA Maastricht), and is based on the ISCO classification.

[10] Unfortunately, we only observe time spent in the host country for a limited number of foreign employees. We do however observe whether an employee is a 1st or 2nd generation immigrant. Inclusion of the generation variables improved the estimates and provided a meaningful description of how sojourn improved labour market outcomes.

the factors that may influence an employee to improve his/her skill sets to more recent task requirements due to computerization and changing trade patterns. Therefore, all individual variables are obtained from the 2015 LFS while occupational and regional level variables are lagged by 15 years.

Although we repeat our analysis for the entire sample for the generic model, for further analysis we focus on the 35-64 years-old age group for two reasons. Firstly, this group is clearly of working age population and supposedly with a lower tendency to invest in additional education. This age group is therefore more likely to hold a stable position in the labour market in terms of matches with jobs and educational attainment. In other words, they are less likely to transition between phases of being over-educated, required-educated and under-educated. Secondly, they have a long work experience. Therefore the possibility of facing skills mismatch due to being a young entrant with insufficient knowledge about the jobs and dynamics of labour markets is much lower.

3.1 Descriptives

Based on Dutch LFS data for 1996-2016 there appears to be a persistent secular trend in the increase of overeducated and the decrease of undereducated employees in the labour market. In our sample from 2011 we obtained 73,604 workers from the 2011 wave of the Dutch LFS. On average 23 percent of employees are overeducated and 10 percent of them are undereducated. The average age is 39 and slightly less than half the sample is female. We observe that 8 percent of employees are foreign-born that is slightly lower than the national average. Median annual earnings before tax is 27,521 euros. 29 percent of the employees are working in high-skilled occupations, while more than 31 percent of the employees hold a graduate or higher level diploma. Similarly 34 percent of the employees are working in middle skilled jobs while 43 percent of the labour force obtained a degree in the middle segment.

In the next section we discuss the results of the multivariate probit analysis.

4 Empirical Findings

This section has not yet been fully cleared for release by the Statistics Netherlands; therefore the tables will be shared at a later stage.

The main findings, based on statistical significance, of the trivariate probit model are as follows. The probability of being undereducated decreases with work experience and increases with age. Overeducation behaves in exactly the opposite way, as expected. Similarly, longer job tenure decreases the probability of being overeducated, increases the probability of earning a wage premium above the respective occupation-location grid median wage and increased the probability of being undereducated.

Females are more likely to be mismatched both in terms of being undereducated and overeducated. First generation immigrants are significantly more likely to be overeducated, while second generation immigrants seem, *ceteris paribus,* to be no different to other natives. This suggests that second generation immigrants are assimilated into the labour market and face the same probability of being undereducated as other natives. In the overeducation equation, the sector for which one works significantly alters her/his chances relative to the reference group.

The ratio of foreign-born workers supply derived from the 1600-cell grid of 2-digit ISCO classification by 41 labour market areas is insignificant in the overeducation equation, has a significant negative effect on the probability of being undereducated, and has a significant and positive effect on the probability of earning a wage premium.

We now turn out attention to the influence of the two endogenous variables on the probability of receiving a wage premium. Overeducation has a positive effect on the probability of receiving a wage premium while undereducation has a negative effect.

Finally, RTI as we hypothesized has the effect of decreasing the probability of being undereducated, increasing the probability of being overeducated, and decreases the probability of receiving a wage premium.

5 Conclusion

This section has not yet been cleared for release by the Statistics Netherlands; therefore the magnitude of the predictions will be shared at a later stage.

References

Acemoglu D, Autor D. (2011) Skills, Tasks and Technologies: Implications for Employment and Earnings, *Handbook of Labor Economics,* Vol 4b: 1044-1166.

Autor D, Dorn D, Hanson H.G (2013) Untangling Trade and Technology: Evidence from Local Labor Markets, NBER No. 18938.

Becker G (1964) *Human Capital*. 2nd edition. Columbia University Press, New York.

Beckhusen J, Florax RJGM, Poot J, Waldorf BS. (2013) Attracting Global Talent and Then What? Overeducated Immigrants in the United States. *Journal of Regional Science, 53*(5): 834–854.

Chiswick BR, Miller PW. (2009) Educational mismatch: are high-skilled immigrants really working at high-skilled jobs and the price they pay if they aren't?. In Barry R.

Chiswick (ed.), *High Skilled Immigration in a Global Labour Market,* DC: American Enterprise Institute Press, pp. II1-54.

Eurostat. (2011) *Migrants in Europe: A statistical portrait of the first and second generation*. Eurostat Statistical Books.

Gunderson M. (1980) "Probit and Logit estimates of Labor Force Participation" *Industrial Relations,* Vol.19, No.2.

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

Hartog, J, Oosterbeek H. (1988) Education, Allocation and Earnings in the Netherlands: Overschooling? *Economics of Education Review* 7(2): 185–194.

Hellevik, O. (2009) "Linear versus logistic regression when the dependent variable is a dichotomy", *Quality and Quantity,* 43(1), pp.59-74.

Katz LF, Autor D (1999) Changes In The Wage Structure And Earnings Inequality. *Handbook of Labor Economics,* Edited by O. Ashenfeher and D. Card, Vol 3A:1464-1548.

Leuven E, Oosterbeek H. (2011) Overeducation and Mismatch in the Labor Market, In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (ed). *Handbook of the Economics of Education,* pp. 283–326. Elsevier.

Nordin M, Persson I, Rooth DO. (2010) Education-occupation mismatch: Is there an income penalty? *Economics of Education Review,* 20 (2010):1047-1059.

Quintini, G. (2014) Skills at Work: How Skills and their Use Matter in the Labour Market, *OECD Social, Employment and Migration Working Papers,* No. 158, OECD Publishing. http://dx.doi.org/10.1787/5jz44fdfjm7j-en.

Oesch D. (2013) *Occupational Change in Europe: How Technology and Education Transform the Job Structure*. Oxford University Press.

Sicherman N, Galor O. (1990) A theory of career mobility, *Journal of Political Economy,* 98(1): 169–92.

Verhaest D, Omey E. (2006) The impact of overeducation and its measurement. *Social Indicators Research,* 77(3): 419–448.

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy,* 103(2):280–315.