**Central European Journal of Economic Modelling and Econometrics**

# Wage Effects of Overeducation: Evidence from Poland

Paulina Broniatowska*

Submitted: 30.08.2019, Accepted: 10.09.2020

**Abstract**

The educational expansion in many advanced economies in the past few decades has triggered a debate on overeducation. The aim of the study is to provide an empirical evaluation of the wage effects of overeducation in different occupational groups. We also analyse whether these effects differ between genders. In order to achieve this, we use individual data from the Structure of Wages and Salaries by Occupations database of firms with 10 or more employees in Poland. We use data from the 2006-2014 waves of the survey. We calculate the impact of overeducation on wages using a Mincer-type wage regression model. We show that on average workers are rewarded for being overeducated, but the size of wage effects of overeducation differs among particular occupational groups. We show also that the choice of the method of measurement of overeducation affects the results.

**Keywords:** educational mismatch, overeducation, tertiary education, wage premium

**JEL Classification**: J24, J31, J41, J21, I21

*University of Warsaw; e-mail: paulina.broniatowska@gmail.com; ORCID: 0000-0002-4919-9191

# 1 Introduction

In general, the term overeducation describes the situation when the educational level of the employees exceeds the educational requirements of the type of work they perform. It is considered as a type of vertical mismatch on the labour market. In the to-date literature, overeducation estimates for different countries range from 10% to over 40% of the working population (Groot & van den Brink, 2000) and overeducation is often associated with earnings losses (see eg. Bauer 2002, Boll et al. 2016).
According to OECD (2019) estimates, in 2016 9.2% of workers in Poland had educational qualifications above those required for their current job, which is well below the average for the European Union, that was 14.7%. We claim that this number is different in particular occupational groups and therefore in major and sub-major occupational groups the impact of overeducation on wages varies. This was suggested by Verdugo and Verdugo (1989), who – in one of few studies – find that in the United States earnings of over-, adequately and undereducated workers vary substantially by broad occupational group. The concept and incidence of mismatches in particular occupations in Poland was already subject to some research, but – to the authors best knowledge – wages effect of overeducation have not yet been analysed across occupational groups.
The aim of the study is to provide an empirical evaluation of the wage effects of overeducation on the level of sub-major (two-digit) occupational groups. We also aim to check whether wage effects of overeducation are different for men and women. The classification of occupations and specialities is a five-level hierarchically systemised set of occupations and specialities on the labour market. It organises them into gradually more aggregated groups and determines their symbols and names. Within the range of major occupational groups there are specified sub-major groups which are further divided.
In this study, we use individual data from the Structure of Wages and Salaries by Occupations (SWS) database of firms with 10 or more employees in Poland provided biannually by Statistics Poland. We concentrate on 2-digit occupational groups as the major occupational groups consist of different occupations that are not homogenous. Therefore, the effects of overeducation may vary between particular sub-major occupational groups, which changes the overall effect for the major ones.
We focus on one country – Poland, because of its some interesting features. Poland, as other CEE countries, has experienced a very strong interest in higher education enrollment after transition to a market economy. This caused a higher education boom with a significant rise in the share of higher-educated labour force. At the same time, Poland experienced a growth of labour demand in services, as the country became more familiar with market economy and later entered the European Union. For this reason, services expanded visibly in Poland in the last three decades. To some extent, this could accommodate the increasing share of tertiary educated workers. Ortiz and Kucel (2008) state that the high-job specificity protects graduates of fields like medicine, law or architecture from educational mismatch. Following this, we suppose

that structural mismatches are more common in some occupations. For this reason, we concentrate on the $3^{rd}$ (Technicians and Associate Professionals), $4^{th}$ (Clerks) and $5^{th}$ (Service and Sales Workers) major groups of International Standard Classification of Occupations (ISCO-08). These are the groups where having higher education (bachelor or master) is not necessary to satisfactory perform the job. However, the share of workers that have obtained higher education is in these groups relatively high (in 2014 38.4% in the $3^{rd}$ group, 37.5% in the $4^{th}$ group and 16.4% of the workers in the $5^{th}$ group).
In this study we attempt to verify the following hypotheses:

$H1$ On average, workers in the $3^{rd}$, $4^{th}$ and $5^{th}$ major and their sub-major occupational groups receive a wage premium for being overeducated.

$H2$ Wage effects of overeducation differ between the $3^{rd}$, $4^{th}$ and $5^{th}$ major and between their respective sub-major occupational groups.

$H3$ Overeducation has different wage effects for men and for women in the analysed occupational groups.

$H4$ The choice of the method of measurement of overeducation affects significantly the results.

This study contributes to the literature on overeducation in following ways. Firstly, we control for occupational codes to analyze the data on the level of major and sub-major occupational groups. The usage of occupational classification allows to disentangle the overeducation estimates by the $3^{rd}$, $4^{th}$ and $5^{th}$ major and their sub-major occupational groups, which – to our best knowledge – has not been analysed before for Poland. We follow the existing literature approach to measuring wage effects of overeducation in the European countries, with special reference to the effects on the level of occupational groups. Secondly, we estimate whether wage effects of overeducation in these groups vary between men and women, analysing gender differences in wage premium to overeducation and adding to the broad discussion on gender and wages.
The structure of the paper is as follows. The two next sections provide a literature review with a brief overview of the theories and empirical results of wage effects of overeducation. Section 4 describes the methodology. It provides also a critical assessment to the approaches in measurement of the overeducation. In the same section we describe data and present the model. The results are discussed in section 5 and section 6 concludes.

## 2 Theoretical concept of overeducation

From a broad perspective, the term overeducation can be confusing and to some extent misleading. Is it really possible to get too much education? From a personal perspective, an upgrade in knowledge should bring only positive effects. However, Hartog (2000) noted that the benefits of educational expansion reach their ceiling when expansion outpaces the demand for high-skilled positions. When the demand for highly educated workers is lower than their supply, ones education can be excessive in relation to the job performed.

There are several theoretical constructs that explain overeducation. Traditionally, overeducation has been considered as an exception to the human capital theory. As a mismatch and therefore a market disequilibrium, it was seen as a rather short-term phenomenon. In this context, Becker (1964), notes that human capital is represented not only by the level of education, but also by work experience. According to this theory, overeducation is mostly a consequence of a lack of work experience, which is typical for young people, despite their increasing educational level. Leuven and Oosterbeek (2011) underline that overeducation signals a lack of the work-related component, and should not be seen as a waste of human capital.

The second approach to the theoretical concept of overeducation, the job competition model, refers more to the persistence of overeducation among the adults. Excess schooling is here seen as a consequence of the competition for jobs in presence of rigidity of the demand for educated labour. It causes graduates to accumulate education, which is in some cases more than that requested to get a job, as to have better position in the queue for the job that the other applicants.

Sattinger (1993) reconciles the two previous theories with the assignment theory. Like in the job competition model, his model assumes that there is a limited number of jobs available in the economy, which implies that wage is job specific and independent of the human capital of the individual. At the same time, following the human capital theory, it assumes that with their investment in human capital people are able to compete for the best job offer. Wages are influenced by the human capital level of individuals. In this concept, overeducation arises because wages will neither be entirely related to acquired schooling and other individual attributes, like in the human capital model, nor to the job itself, like in the job competition and job assignment model. Overeducation arises at least partly because the least skilled individuals get the first job offer they get because their reservation wage is low (Caroleo & Pastore, 2013).

Regardless of the theory that explains the roots of workers' overeducation, its persistence may impact the overall competitiveness of an economy in different ways. The existence of overeducation signals that there are some disruptions in the functioning of the labour market. From a macroeconomic perspective, overeducation may reflect an inefficient use of human capital. From a microeconomic perspective, it affects a workers' job satisfaction and can reduce his work motivation.

Sloane (2014) proposes three definitions connected with the concept of overeducation.

He distinguishes between:

– overeducated, if workers' level of education is higher than required,

– overqualificated, if the level of qualification is higher than required,

– overskilled, if the mismatch concerns the level of skills.

In the following work, we concentrate particularly on overeducation in its narrow definition, so we look only at the situation when workers' level of education is higher than required by the type of work he performs. We assume that overeducated workers are often penalized for having acquired too much education by earning less than correctly matched workers with similar educational level. More precisely, we claim that overeducation is associated with wage penalty compared to matched graduates but it gives wage premium when compared to lower educated peers.
The concept of overeducation can also be analysed by looking at gender differences. Gender plays an important role already in choosing the educational path. In general women are characterized by higher participation in higher education. Their choice of study areas is also different than this of men. Rossen et al. (2019) note that gender differences in field-specific overeducation rates could origin in gender-different field-specific enrollment rates and correspondingly different demand/supply ratios on the labor market. This means that, to some extent, field-specific labor market outcomes are not purely causal effects but are partly driven by selection into fields. In this context, gender norms might also impact decisions on family formation and in this way impact educational choices (see eg. Chiappori et al., 2009; Attanasio & Kaufmann, 2017). Moreover, all other things being equal, women, receive lower wages than men (see eg. Goraus & Tyrowicz 2014).

# 3 Review of the empirical literature

The term 'overeducation' was introduced to the economic literature in the 1970s in a book *The Overeducated American*. Its author, Freeman (1976), examined the decreasing wage returns to college graduates in the United States during a period of higher education expansion (1967–1976) and pointed out that in the USA a rising number of university graduates was not matched by labour demand, causing a decrease in tertiary education wage premium.
Since then, many empirical studies have confirmed that when the comparison is made within educational levels, overeducated workers suffer a wage penalty relative to workers who, with same educational level, are properly educated for the job they do. McGuinness (2006) estimated the average wage penalty associated with overeducation, at around 15%, relative to matched workers with similar levels of schooling, on the basis of a dummy variable in a standard wage regression. In general, the literature suggests that wage penalty to overeducation for individuals with the

same educational attainment and in jobs with similar schooling requirements ranges between 13% (Verdugo & Verdugo, 1989) to 20% or more (Chevalier 2000, Robst, 2008).

However, overeducated workers earn a premium over their matched colleagues in the same job. Sloane (2003) suggests that it is ability that can explain why overeducated workers receive a wage premium over their colleagues with the required level of education. To some extent, this assumption is also confirmed by Allen and van der Velden (2001) and Chevalier (2000).

Another approach was presented by Duncan and Hoffman (1981), who introduced an extension to the Mincer wage-regression model. It breaks down the educational level into three components. They used variables capturing the years of education required for a current job, years of overeducation and years of undereducation (the ORU approach). They found that in the United States the individual return to an additional year of surplus education was positive (0.029) and significant for all major demographic groups. The estimated return was also only about half the size of the return to an additional year of required schooling (0.063). Hartog (2000), using the method proposed by Duncan and Hoffman (1981), confirmed positive returns to years of overeducation (although smaller in magnitude to the years of required schooling). General findings in the to-date literature confirmed these results and established a view that the returns to each year of surplus schooling are positive, but they remain smaller than those to required education (Bauer, 2002). They support the view that overeducated workers earn more than their lower educated job colleagues.

The comparability of results of the studies on overeducation is therefore limited by different methodological settings that have been used in the literature so far. The effects of overeducation vary substantially according to the approach applied. Detailed literature reviews on overeducation provide eg. Groot and van den Brink (2000), McGuiness and Wooden (2009) and Leuven and Oosterbeek (2011).

In the European context, most studies have looked at Western European countries. Most of the empirical works for Europe analyse the wage penalty to overeducation. Chevalier (2000) states that overeducated graduates experience a 7% wage penalty compared to matched graduates. Allen and van der Velden (2001) note that skill under-utilisation causes a reduction in earnings by 3.2%. McGuinness and Sloane (2011) found substantial wage penalties for being overeducated in the UK. The model by Duncan and Hoffman was replicated and confirmed for the Dutch labour market by Hartog and Oosterbeek (1988) and for the Spanish labour market by Alba-Ramirez (1993). The wage penalty to overeducation was also confirmed for Portugal by Kiker et al. (1997) and for Germany by Bauer (2002).

The literature suggests also that the returns to education for overeducated workers vary according to their ability and skills (Sloane, 2003). An important role in explaining wage penalties to overeducation may play the skill mismatch (see eg. Chłoń-Domińczak & Żurawski 2017). Following the specification developed by Verdugo and Verdugo (1989), different studies have included dummy variables for

both educational and skill mismatch in the empirical analysis (Allen & van der Velden, 2001; Di Pietro & Urwin, 2006; Green & McIntosh, 2007; Sánchez-Sánchez & McGuiness, 2013; Mavromaras et al. 2013). There is an agreement that overeducation and overskilling have both a negative and statistically significant effect on earnings within the same level of education, with the overeducation effect being higher than the overskilling effect.

As for differences in wage effects of overeducation of men and women, the literature provides different results. Firstly, in the empirical literature, no clear effect of gender on overeducation has been found. For example, Battu et al. (2000) shows that the probability of being overeducated is higher for women. On the other hand, Alba-Ramirez (1993) and Groot (1996) obtain both the results that males face a higher overeducation risk. Some studies find that the effect of gender on overeducation risk is insignificant in a multivariate setup (Boll et al. 2016). Secondly, there are contradictory results on wage effects of overeducation for men and women. Li et al. (2014) suggest that wage effects of overeducation vary significantly by gender, but they are different for given business related major fields of study. While Daly et al. (2000) shows that the wage penalty to overeducation is lower for women than for men, Vahey (2000) reports no lower return to surplus education for overeducated females. As for Polish data, Kucel and Vilalta-Bufi (2012) find that overeducation in Poland is negatively associated with being a woman. Looking at the wages, Wincenciak (2016) finds a wage penalty to overeducation (for a given educational level) that is slightly lower for men (13.5%) than for women (16%). In this context, it is also important to note that in Poland there are gender differences in participation in higher education and in the choice of study areas that may affect wage effects of womens' overeducation. Only few studies analyse wage effects of overeducation in different occupations. Verdugo and Verdugo (1989) find that earnings of over-, adequately and undereducated workers vary substantially by broad occupational group. In their study, overeducation has significant effect on earnings in five of nine occupational categories. Research to-date has shown that overeducation varies by profession and occupation, with managers, leaders, entrepreneurs and workers in positions requiring communication, planning and literacy being more likely to be overeducated (Chevalier & Lindley, 2009; Green & McIntosh, 2007). As for Poland, Baran (2016) found that while in Poland low risk of overeducation is associated with having studied technical and health programmes, such factors as working part-time in small firms in the private sector, and living outside big cities in less developed regions are associated with a higher risk of overeducation. He also notes that age is negatively associated with overeducation. Chłoń-Domińczak and Żurawski (2017) showed that in Poland the sector and occupation difference should be taken into account in analyses of the skills mismatch, as well as policy responses.

# 4 Methodology

## 4.1 Identification of overeducation

The first issue when discussing overeducation is to choose the method of how it is defined and measured. The effects of mismatch on wages depend on how the required schooling is specified and on the definition of overeducation. In the empirical research in this field, there are three main empirical approaches. They vary in the way of defining the required education for a certain job. Verhaest and Omey (2006) show that the correlation between results gained with these different approaches varies from 4% to 72%. Also, each of the methods faces its limitations, either conceptual or practical. In the following paragraphs we provide a compact discussion of the three most common definitions.

The first method, the *job analysis approach*, is based on objective systematic expert evaluations of the level of education needed to perform a certain job. These evaluations are prepared every few years. A worker is considered to be overeducated if his or her educational qualification is above the level required for the performance of their job. While this assessment is objective and professional, the question arises of its adequateness. The educational levels are assorted to the major occupational groups. Therefore, they may remain too general as educational requirements for particular occupations in an aggregated occupational group might be heterogeneous. This approach relies also on arbitrary definitions and does not adapt quickly to the dynamics of occupational and educational change.

The second method, the *workers' self-assessment approach*, remains more subjective as it includes a comparison of the respondents' self-assessment of the educational requirements for their jobs with their actual education. This approach uses information provided by the worker himself regarding the required level of education to get (see eg. Duncan and Hoffman 1981) or to perform satisfactorily (see eg. Hartog and Oosterbeek 1988) the job. Although it reflects better the actual level of education that is needed for performing a certain job, it is subject to some bias coming from possible overestimation of necessary educational qualifications. There is also lack of uniform coding questions (Hartog, 2000).

In the third approach, the *realized-matches method*, the required education is derived from what workers in the respondent's job or occupation usually have attained, e.g. the mean or the mode of that distribution. Any schooling that is above the mode or mean years of schooling for certain occupations is considered to be overeducation. In the literature two options of this approach prevail. Firstly, the specification proposed by Verdugo & Verdugo (1989), where overeducation is defined as more than one standard deviation above the mean years of schooling for workers within the same occupation. However, if an occupational group had a high incidence of overeducated employees, the mean would be affected by this phenomenon and thus underestimating its composition (Dolton and Vignoles, 2000). This method is therefore very sensitive to the labour market situation in the sense that the results may vary with changing

overall trends in education, as Hartog (2000) states. If there is a surplus of qualified workers then workers with a higher qualification that the one required for their job would be hired, so there would be an underestimation of the overeducation effect and *vice versa*. These shortcomings may be to some extent overcome by the approach introduced by Kiker et al. (1997), who use the mode of the years of schooling and classify individuals with an education level higher than the modal within each occupation group as overeducated. This method is less sensitive to the existence of outliers in the educational distribution.
In order to measure overeducation we follow the realised-matches method and construct an objective definition, similar to McGuinness et al. (2017). Firstly, solely for the purpose of obtaining a definition of overeducation, we divide educational levels into five groups, where each group includes possibly similar levels of education:

1) comprehensive primary school & Comprehensive lower secondary school,

2) basic vocational school,

3) secondary vocational school,

4) general secondary school & profiled general secondary school,

5) tertiary education (BA, MA and higher).

For each year, the modal level of education is defined as the most common possessed by workers in each sub-major occupational group. A person is defined as overeducated if his or hers level of schooling lies above the mode. The mode may change over time as more (or less) people with a given level of education join the occupational groups (see Table 4). While by using this approach, we are able to take into our analysis changes in the educational structure of the occupational group, these changes are relatively small as they concern only two sub-major occupational groups. We create a dummy variable, as follows:

$$overeduc_{t,j} = \begin{cases} 1, & \text{if } edu_{i,t} > mode\ edu_{t,j}, \\ 0, & \text{otherwise.} \end{cases}$$

Despite the fact that formal education is an incomplete measure of human capital, as, besides educational mismatches, also workers' skill mismatches play an important part on the labour market (see eg. Chłoń-Domińczak & Żurawski, 2017; Allen & van der Velden 2001), in this study, we use solely the concept of overeducation in the sense of educational mismatches with its definition being a situation when the attained educational level exceeds the level required by the occupation.
Under this definition employees with primary education cannot be overeducated. Table 1 presents the share of overeducated workers by major occupational groups in the whole sample in the analysed period.

Table 1: Share of overeducated workers in major occupational groups in the analysed period

| | 2006 | 2008 | 2010 | 2012 | 2014 |
|---|---|---|---|---|---|
| 1 (Managers) | 0 | 0 | 0 | 0 | 0 |
| 2 (Professionals) | 0 | 0 | 6.69 | 0 | 0 |
| 3 (Technicians & associate professionals) | 30.35 | 7.63 | 8.37 | 9.45 | 10.22 |
| 4 (Clerical suport workers) | 21.11 | 24.71 | 16.21 | 15.36 | 8.64 |
| 5 (Service & sales workers) | 23.25 | 24.80 | 27.34 | 27.33 | 27.06 |
| 6 (Skilled agricultural, forestry & fishery workers) | 21.80 | 22.49 | 26.94 | 27.59 | 23.19 |
| 7 (Craft & related trades workers) | 28.01 | 29.68 | 28.08 | 30.43 | 32.18 |
| 8 (Plant & machine operators & assemblers) | 34.60 | 37.68 | 38.83 | 41.88 | 44.70 |
| 9 (Elementary occupations) | 24.92 | 28.07 | 30.23 | 33.54 | 33.81 |

In the case of the $1^{st}$ and $2^{nd}$ major groups, almost no overeducation exits as these are the groups where the highest educational level is expected to perform the job satisfactory. In the case of $3^{rd}$ and $4^{th}$ groups the share of overeducated workers decreased visibly in the analysed period. In the remaining groups, the share of overeducated workers increased.
As we run the regression also on sub-major of the $3^{rd}$, $4^{th}$ and $5^{th}$ occupational groups, we look at their shares of overeducated workers (Table 2).

Table 2: Share of overeducated workers in sub-major occupational groups in the analysed period

| | 2006 | 2008 | 2010 | 2012 | 2014 |
|---|---|---|---|---|---|
| 31 (Science and engineering associate professionals) | 19.42 | 23.07 | 27.81 | 33.60 | 35.81 |
| 32 (Health associate professionals) | 10.93 | 13.31 | 18.98 | 18.75 | 23.66 |
| 33 (Business and administration associate professionals) | 0 | 0 | 0 | 0 | 0 |
| 34 (Legal, social, cultural and related associate professionals) | 38.89 | 0 | 42.9 | 0 | 0 |
| 41 (General and keyboard clerks) | 22.99 | 26.07 | 30.16 | 34.04 | 36.19 |
| 42 (Customer services clerks) | 15.20 | 20.10 | 27.01 | 33.77 | 32.31 |
| 51 (Personal service workers) | 4.31 | 5.43 | 8.36 | 10.69 | 13.44 |
| 52 (Sales workers) | 6.58 | 7.98 | 12.09 | 17.15 | 17.14 |

Note: Sub-major occupational groups: 35, 43, 44, 53 and 54 were excluded due to insufficient number of observations.

Of particular interest are the $33^{rd}$ and $34^{th}$ sub-major occupational groups (Business and administration associate professionals and Legal, social, cultural and related associate professionals). In the first case, no overeducation was noted while in the latter shares of overeducated people vary from 0 to more than 40%. This results from the fact that the modal educational level in these particular groups is 5 (Tertiary

education; for $34^{th}$ group it is 5 only in years 2008, 2012 and 2014). This explains also why in the Table 1 the share of overeducated workers in the $3^{rd}$ group decreased sharply. It is not a real decrease but a result of the definition used and it results from the change in the modal education of the $34^{th}$ occupational group and not from the actual change in overeducation trend. Tertiary education as a mode in the aforementioned groups means that most workers have obtained higher education, which caused a shift in the education level of the whole group. This problem was mentioned by Dolton and Vignoles (2000), who claimed that if an occupational group had a high incidence of overeducated employees, it would affect its composition.
We perform regressions using also a different definition of overeducation, taken from the job-analysis approach. The alternative version of the variable is constructed as follows. We follow directly the Polish Classification of Occupations and Specializations for labour market Needs (Klasyfikacja Zawodów i Specjalności, KZiS), published in 2014, which is a national adaptation of the International Standard Classification of Occupations, compiled by the International Labour Office, Geneva and assign a level of education to each group, using the five-point scale as described above (see Table 3). As we are analysing only $3^{rd}$, $4^{th}$ and $5^{th}$ major occupational groups, we define the alternative version of overeducation for these groups only.

Table 3: Educational requirements for working in the $3^{rd}$ $4^{th}$ and $5^{th}$ major occupational group

| No. | ISCO Major Group | ISCED 2011 Education Level | Educational level |
|---|---|---|---|
| 3 | Technicians and Associate Professionals | 3, 4 | 4 |
| 4 | Clerks | 3, 4 | 4 |
| 5 | Services and Sales Workers | 3, 4 | 4 |

Therefore, we obtain a dummy variable, where:

$$alt\_over_{i,t} = \begin{cases} 1, & \text{if } edu_{i,t} > 4, \\ 0, & \text{otherwise.} \end{cases}$$

Table 4 shows the share of overeducated workers in sub-major groups of these major occupational groups. For the sub-major occupational groups in which they differ, we will run regressions using both versions of the variable that controls overeducation.

## 4.2 Data description and the model

In the Introduction we briefly presented some facts on the higher education in Poland. We would like to discuss it here in more detail, in order to clarify the background of the study. In the last 30 years Poland experienced an higher education boom. Between 1991 and 2016 the number of students increased by more than twice, despite some decline that was noted in the last years (this decline results from demographic

Table 4: Share of overeducated workers in chosen sub-major occupational groups in the whole period

| Sub-major occupational group | Modal educational level | Share of overeducated workers – basic | Modified educational level | Share of overeducated workers – alternative |
|---|---|---|---|---|
| 31 (Science and engineering associate professionals) | 4 | 27.75% | 4 | 27.75% |
| 32 (Health associate professionals) | 4 | 17.14% | 4 | 17.14% |
| 33 (Business and administration associate professionals) | 5 | 0% | 4 | 57.69% |
| 34 (Legal, social, cultural and related associate professionals) | 4 (2006, 2010) or 5 (2008, 2012, 2014) | 15.25% | 4 | 44.17% |
| 41 (General and keyboard clerks) | 4 | 29.97% | 4 | 29.97% |
| 42 (Customer services clerks) | 4 | 25.83% | 4 | 25.83% |
| 51 (Personal service workers) | 4 | 8.26% | 4 | 8.26% |
| 52 (Sales workers) | 4 | 12% | 4 | 12% |

reasons; see Statistics Poland data). At the same time, the percentage of young population (aged 25-34) that has attained tertiary education rose from about 12% in 1999 to as much as 43.5% in 2016 (OECD data). This caused a significant rise in the share of higher-educated labour force. As a result, Poland experienced a strong increase in the number of tertiary educated workers. For example, Strawiński et al. (2016) showed that the share of workers with tertiary education among technicians increased in the last 15 years from 10 to 40%, and among clerical support workers – from 5 to over 30%. What is important, although tertiary educational attainment rose from the early 90s, the demand for highly educated workers has not increased proportionally.

Another important issue for the analysis of higher education in Poland in the analysed period is its dropping quality. Gaining popularity of private higher schools that offered a possibility to obtain a tertiary education for a fee, contributed to the fast increase of higher educated workforce. However, in the case of some of these institutions, obtaining a diploma did not mean having received proper education on the tertiary level.

Moreover, a fair share of tertiary education graduates has studies fields related to social sciences as new graduates shifted from technical fields towards business fields. This has led to mismatches between educational level and occupations performed by many of the employed. While we do not analyse mismatches in this sense, their increasing number has also added to the problem of overeducation.

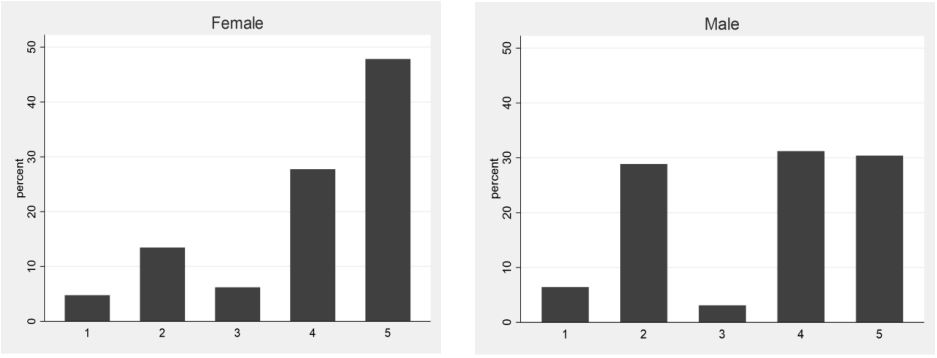
We use data from the Structure of Wages and Salaries by Occupations (SWS) database provided by the Statistics Poland. The survey is carried out with biennial frequency and includes national economic entities from both the public and private sectors whose employees exceed nine persons. We use data from the 2006-2014 waves of the survey. The database covers both full- and part-time employees who worked for the entire month of October in a given year. The database contains information on the individuals' wages and several personal characteristics, such as gender, age, level of education, work tenure, and occupational group. It also includes some employers' characteristics, such as ownership sector, size of the enterprise and its location, as well as the NACE section.

An added advantage of the SWS survey is the high reliability of its data related to wages as compared to other Polish databases of individuals (Polish Labour Force Surveys or Household Budget Surveys) where wages are declared by the responders and are downward biased, especially for higher income workers (see Strawiński, 2015). In the SWS database, wages of individuals are reported by the accounting departments along with the number of hours worked. Another advantage is the size of the database. On an average, between 2006 and 2014, the SWS survey covered around 12% of the total number of Polish enterprises with 10 or more employees. In a given year, the database contains around 600,000-700,000 individuals. Since we have data from five surveys, the total number of observations in the sample ranges around 5 million in which 3.5 million are the original CSO data and 1.5 million come from the merging procedure.

The aim of the study is to assess the wage effects of overeducation across occupations. The Classification of Occupations and Specialities (COS) was changed in 2010 and as a result some groups were merged and others split. Therefore, the structure of occupational groups was standardised over the analysed period to achieve comparability of results over the years. This procedure was conducted as a part of the National Science Center grant 2015/19/B/HS4/03231.

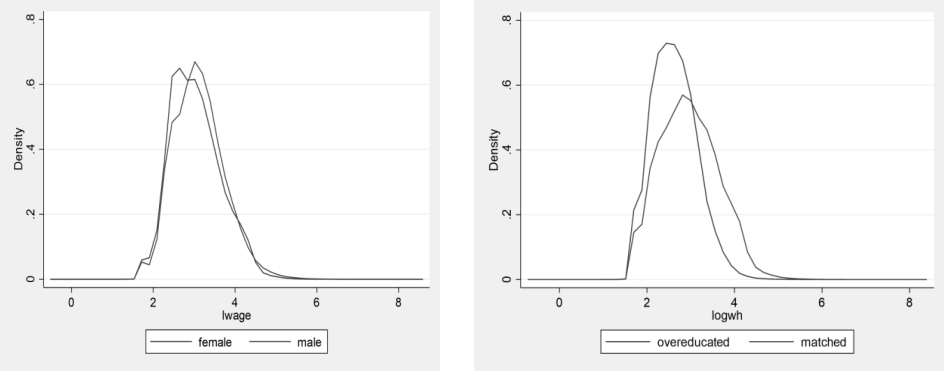
In the whole database, 49.3% records are female and 50.7% are male. However, in the analysed major occupational groups, there are more women than men. Women constitute 60% of the $3^{rd}$, 65% of the $4^{th}$ and 67% of the $5^{th}$ occupational group. With regard to males, 22.2% of the whole sample are overeducated, while only 14.6% of females are overeducated. As for particular major occupational groups, in the $3^{rd}$ and $5^{th}$ occupational groups also more men are overeducated (respectively 16.5% and 9.2% of women are overeducated, compared to 22% and 13.4% of men). This changes in the $4^{th}$ major occupational group, as 33.5% of women compared to 20.7% of men are overeducated.

Figure 1: Shares of men and women in the sample with a given educational level in 2014 (five-point scale, as used to define overeducation)




There are some differences in the educational attainment of men and women in the sample (Figure 1). Almost half of the women in the sample has obtained tertiary education. As for men roughly about one third of them has higher education, one third attender general secondary school or profiles general secondary school, and almost 30% basic vocational school. Education attainment of men is more diversified and it seems to be more directed to vocational education.
Not only educational attainment but also wages vary between genders in the sample. Figure 2 shows the kernel density functions for log hourly net wage distributions for men and women (left) and for matched and overeducated workers with overeducation being defined with the realized matches method (right). In order to calculate the

Figure 2: Kernel density of log hourly wage distribution for men and women



impact of overeducation on wages, we use a Mincer-type wage regression model (Mincer, 1974). Following general practice, we excluded from the analysis major occupational groups 10 (Armed forces occupations) and 6 (Skilled agricultural, forestry and fishery workers), for following reasons. The $10^{th}$ group is usually excluded in wage analyses as the armed forces remunerations do not follow the free market pattern and educational requirements within this particular group vary significantly. As for the $6^{th}$ group, we have obtained relatively little records, so it is not representative. Remunerations and educational attainment of the workers in this group are also not similar to other occupational groups. We estimate the model for the whole sample and for each of remaining sub-major occupation groups separately. Due to data availability we replaced years of schooling in the standard Mincerian earnings equation by education level (see Table 9 in the Appendix).
The functional form of the estimated equation is as follows:

$$\begin{aligned} lwage_{i,j} \quad = \quad & \alpha_0 + \alpha_1\ tenure_{i,t} + \alpha_2\ tenure_{i,t}^2 + \sum_{k=1}^{K} \gamma_k X_{k,i} + \alpha_3\ female_{i,t} \\ & + \alpha_4\ partime_{i,t} + \alpha_5\ public_{i,t} + \alpha_6\ overeducation_{i,t} + \alpha_7\ neg_{i,t} \\ & + \sum_{l=1}^{L} \gamma_l X_{l,i} + \sum_{n=1}^{N} \gamma_n X_{n,i} + \sum_{\nu=1}^{V} X_{\nu,i} + \sum_{y=2006}^{2014} year_{y,i} + \varepsilon_{i,t}. \end{aligned} \tag{1}$$

As the dependent variable we take the log of hourly real base wage (PLN). The explanatory variables on the right-hand side include:

$tenure_{i,t}$ – defined as a total number of years spent working (job experience). Groot (1996) finds that the wage penalty to overeducation is related to tenure. This means that, as time goes by, the employers find out the real productivity of the workers and discriminate those with fewer abilities that the qualifications they possess,

$tenure_{i,t}^2$ – captures diminishing returns to tenure,

$\sum_{k=1}^{K} X_{k,i}$ – a set of dummy variables, indicating the educational level an individual $i$ has obtained (we use eight levels of schooling as presented in Table 9 in the Appendix),

$female_{i,t}$ – a dummy variable for sex (1 for female, 0 for male). In our sample, almost half (49.3%) of the labour force consists of women, however, this variable is strongly diversified between occupational groups. In line with most of the research in this field, we expect lower average wages for women, regardless of the occupational group,

$partime_{i,t}$ – a dummy variable indicating whether an individual $i$ works part-time or full time at time $t$ (1 for part-time, 0 for full time),

$public_{i,t}$ – a dummy variable indicating whether an individual $i$ works in a public or private sector at time $t$ (1 for public, 0 for private). Statistical data indicate that in the cases of some occupational groups, wages to a considerable extent depend on the type of the ownership sector. For example, less educated workers receive higher wages in the public sector as compared to the private sector,

$overeducation_{i,t}$ – dummy variable indicating whether an individual $i$ is overeducated at time $t$ (1 for being overeducated, 0 for not being overeducated; in the basic specification $overeducation_{i,t} = overeduc_{i,t}$; and in the alternative version $overeducation_{i,t} = alt\_over_{i,t}$),

$neg_{i,t}$ – a dummy variable for the level of wage negotiations prevailing in the firm where an individual $i$ is employed (1 if negotiations are established at the sectoral or national level, 0 for other possibilities),

$\sum_{l=1}^{L} X_{l,i}$ – a set of dummy variables, indicating in which major occupational group an individual $i$ is working,

$\sum_{n=1}^{N} X_{n,i}$ – a set of dummy variables for the economic section for the main activity of the firm in which an individual is working at time $t$ (according to the NACE classification),

$\sum_{\nu=1}^{V} X_{\nu,i}$ – a vector of interactions, as we allow for all possible interactions between the variables,

$year$ – dummy variables for years,

$\varepsilon_i$ – an error term.

In the Appendix we provide descriptions of explanatory variables used in the model (Table 8) and key descriptive statistics for the variables used in the model (Table 10). Our estimating strategy is as follows. We start with an estimation of the parameters of the above-presented equation for the entire sample of workers. We derive the average value of elasticity of wages with respect to each of the explanatory variables. Then, we estimate the parameters of equation (1) for the $3^{rd}$, $4^{th}$ and $5^{th}$ major occupational groups and we follow with an estimation of the parameters for their sub-major occupational groups. The parameters of equation (1) are estimated with ordinary least squares (OLS) with clustered (at the firm level) standard errors.

# 5 Empirical results

The estimation of the parameters of equation (1) for the entire sample of workers firms with 10 or more employees in Poland (Table 5) shows a significant and positive relation between overeducation of workers and their wages. The parameter by the variable $overeduc_{i,t}$ shows that for the whole sample, we found an average wage premium of

Table 5: Regression results for major occupational groups

| | (1) | (2) occ3 | (3) occ4 | (4) occ5 | (5) female | (6) male |
|---|---|---|---|---|---|---|
| overeduc | $0.224^{***}$ | $0.120^{***}$ | $0.125^{***}$ | $0.138^{***}$ | $0.189^{***}$ | $0.199^{***}$ |
| tenure | $0.024^{***}$ | $0.026^{***}$ | $0.022^{***}$ | $0.011^{***}$ | $0.026^{***}$ | $0.023^{***}$ |
| tenure2 | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ |
| edu1 | $-0.160^{***}$ | $-0.133^{***}$ | $-0.061^{***}$ | $-0.118^{***}$ | $-0.154^{***}$ | $-0.178^{***}$ |
| edu2 | $-0.123^{***}$ | $-0.163^{**}$ | $-0.047$ | $-0.113^{***}$ | $-0.058$ | $-0.138^{***}$ |
| edu3 | $-0.152^{***}$ | $-0.151^{***}$ | $-0.084^{***}$ | $-0.103^{***}$ | $-0.195^{***}$ | $-0.176^{***}$ |
| edu4 | $-0.099^{***}$ | $-0.044^{***}$ | 0.008 | $-0.028^{***}$ | $-0.095^{***}$ | $-0.170^{***}$ |
| female | $-0.160^{***}$ | $-0.178^{***}$ | $-0.104^{***}$ | $-0.184^{***}$ | | |
| partime | $-0.146^{***}$ | $-0.233^{***}$ | $-0.122^{***}$ | $-0.092^{***}$ | $-0.014^{**}$ | $-0.099^{***}$ |
| public | $-0.128^{***}$ | $-0.055^{***}$ | $-0.076^{***}$ | $-0.193^{***}$ | $-0.125^{***}$ | $-0.120^{***}$ |
| occ1 | $0.908^{***}$ | | | | $0.880^{***}$ | $0.930^{***}$ |
| occ2 | $0.749^{***}$ | | | | $0.798^{***}$ | $0.706^{***}$ |
| occ3 | $0.358^{***}$ | | | | $0.389^{***}$ | $0.345^{***}$ |
| occ4 | $0.212^{***}$ | | | | $0.265^{***}$ | $0.160^{***}$ |
| occ5 | $0.053^{***}$ | | | | $0.100^{***}$ | 0.017 |
| occ7 | $0.163^{***}$ | | | | $0.107^{***}$ | $0.165^{***}$ |
| occ8 | $0.225^{***}$ | | | | $0.307^{***}$ | $0.201^{***}$ |
| neg | $0.063^{***}$ | $0.046^{***}$ | $0.042^{***}$ | $0.065^{***}$ | $0.029^{***}$ | $0.092^{***}$ |
| y2006 | $-0.037^{***}$ | $-0.877^{***}$ | $-0.331$ | $-0.191^{**}$ | $-0.597^{***}$ | $2.000^{***}$ |
| y2008 | $-0.199^{***}$ | $-0.161^{***}$ | $-0.168^{***}$ | $-0.270^{***}$ | $-0.207^{***}$ | $-0.193^{***}$ |
| y2010 | $-0.137^{***}$ | $-0.158^{***}$ | $-0.111^{***}$ | $-0.158^{***}$ | $-0.128^{***}$ | $-0.147^{***}$ |
| y2012 | $-0.064^{***}$ | $-0.060^{***}$ | $-0.02^{***}$ | $-0.079^{***}$ | $-0.060^{***}$ | $-0.067^{***}$ |
| No. obs. | 5242437 | 733049 | 355050 | 314625 | 2465727 | 2776710 |
| R2 | 0.55 | 0.32 | 0.39 | 0.47 | 0.61 | 0.51 |

Note: $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

22.4% for overeducation.

The signs of most of the other parameters are in line with the economic theory and hitherto available empirical evidence. Wages of workers with lower level of education are lower than wages of workers with baseline (tertiary) education. In all specifications, tenure has positive, albeit small impact on wages. The parameter by tenure squared is significant and negative but it remains very small. As we anticipated, for the whole sample wages of women are found to be on average approximately 16% lower than wages of men. Possible explanations have been widely discussed in the literature on the gender wage gap (see eg. Goraus & Tyrowicz 2014; Majchrowska & Strawiński 2018). Working part-time as well as working in the public sector also reflects a negative impact on wages.

In the next step, we performed equation (1) separately for the $3^{rd}$, $4^{th}$ and $5^{th}$ major occupational groups. In all groups there is a wage premium for being overeducated, which partly proves hypothesis $H1$ as for its first part concerning major occupational groups. In all analysed groups wage premium for overeducation is positive, albeit

Table 5 (cont.): Regression results for major occupational groups

| | (7) female occ3 | (8) male occ3 | (9) female occ4 | (10) male occ4 | (11) female occ5 | (12) male occ5 |
|---|---|---|---|---|---|---|
| overeduc | $0.094^{***}$ | $0.144^{***}$ | $0.137^{***}$ | 0.000 | $0.126^{***}$ | $0.160^{***}$ |
| tenure | $0.024^{***}$ | $0.029^{***}$ | $0.021^{***}$ | $0.022^{***}$ | $0.007^{***}$ | $0.019^{***}$ |
| tenure2 | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ |
| edu1 | $-0.219^{***}$ | $-0.148^{***}$ | $-0.082^{***}$ | $-0.057^{***}$ | $-0.026$ | $-0.098^{***}$ |
| edu2 | $-0.084$ | $-0.177^{*}$ | $-0.121^{*}$ | $-0.064$ | $-0.029$ | $-0.103^{***}$ |
| edu3 | $-0.315^{***}$ | $-0.164^{***}$ | $-0.119^{***}$ | $-0.086^{***}$ | $-0.070^{***}$ | $-0.097^{***}$ |
| edu4 | $-0.077^{***}$ | $-0.048^{***}$ | $-0.003$ | $-0.001$ | 0.006 | $-0.022^{**}$ |
| partime | $-0.037^{***}$ | $-0.216^{***}$ | $-0.036^{***}$ | $-0.112^{***}$ | 0.005 | $-0.106^{***}$ |
| public | $-0.052^{***}$ | $-0.053^{***}$ | $-0.084^{***}$ | $-0.048^{***}$ | $-0.135^{***}$ | $-0.269^{***}$ |
| neg | $0.036^{***}$ | $0.063^{***}$ | $0.050^{***}$ | $0.030^{***}$ | $0.071^{***}$ | $0.057^{***}$ |
| y2006 | $-0.568$ | $-1.022^{***}$ | $-1.044^{***}$ | $-0.292^{***}$ | $-0.348^{***}$ | $-0.180^{***}$ |
| y2008 | $-0.147^{***}$ | $-0.183^{***}$ | $-0.170^{***}$ | $-0.164^{***}$ | $-0.255^{***}$ | $-0.270^{***}$ |
| y2010 | $-0.153^{***}$ | $-0.164^{***}$ | $-0.110^{***}$ | $-0.112^{***}$ | $-0.154^{***}$ | $-0.164^{***}$ |
| y2012 | $-0.056^{***}$ | $-0.064^{***}$ | $-0.050^{***}$ | $-0.056^{***}$ | $-0.076^{***}$ | $-0.087^{***}$ |
| No. obs. | 408613 | 324436 | 239730 | 115320 | 183755 | 130870 |
| R2 | 0.29 | 0.29 | 0.43 | 0.33 | 0.47 | 0.48 |

Note: $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

lower than for the whole sample and it varies between groups from 12.0% to 13.8%. Moreover, differences between wages of men and women remained relatively high.

To allow for gender-specified income effects of overeducation, we conducted regressions for women and men separately. In the whole sample overeducated women gain lower wage premium than overeducated men (18.9% compared to 19.9%). Backed by the findings by Boll et al. (2016) this would confirm the hypothesis $H3$. These results hold also for the $3^{rd}$ and $5^{th}$ major occupational groups, where wage premium for overeducation for men is higher than for women. However, in the case of the $4^{th}$ occupational group, there wage premium for men was found to be zero – and insignificant. This group has also the lowest number of observations, what may have caused insignificant results.

The next step is to look at the regression results for particular sub-major occupational groups (Table 6). We assume that the parameter by variable $overeduc_{i,t}$ will vary across sub-major occupational groups. In most cases, the coefficient by variable $overeduc_{i,t}$ remains positive and significant, thus confirming hypothesis $H1$ in its part concerning sub-major occupational groups.

However, wage premium for overeducation varies between particular sub-major occupational groups from 0.06 to 0.17, what can confirm the hypothesis $H2$. Moreover, in the $33^{rd}$ group no significant wage effect of overeducation was found. This results from the fact that following the chosen definition of overeducation, there is no overeducation in these groups, as most of the workers have obtained higher education ($mode = 5$; see Table 4). When we use the alternative measure of

Table 6: Regression results for sub-major occupational groups

| | 31 | 32 | 33 | 34 | 41 | 42 | 51 | 52 | alt_33 | alt_34 |
|---|---|---|---|---|---|---|---|---|---|---|
| overeduc | $0.168^{***}$ | $0.084^{***}$ | 0.000 | $0.155^{***}$ | $0.124^{***}$ | $0.114^{***}$ | $0.062^{***}$ | $0.127^{***}$ | | |
| alt_over | | | | | | | | $0.305^{***}$ | 0.000 | |
| tenure | $0.026^{***}$ | $0.026^{***}$ | $0.028^{***}$ | $0.027^{***}$ | $0.022^{***}$ | $0.021^{***}$ | $0.010^{***}$ | $0.013^{***}$ | $0.025^{***}$ | $0.029^{***}$ |
| tenure2 | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ |
| edu1 | $-0.076^{**}$ | $-0.211^{***}$ | 0.250 | $-0.182^{***}$ | $-0.055^{**}$ | $-0.077^{**}$ | $-0.107^{***}$ | $-0.082^{***}$ | 0.394 | $-0.068^{**}$ |
| edu2 | $-0.278^{*}$ | $-0.467^{***}$ | $0.670^{***}$ | $-0.159^{*}$ | $-0.036$ | $-0.127$ | $-0.034$ | $-0.154^{***}$ | $0.782^{***}$ | 0.009 |
| edu3 | $-0.065^{***}$ | $-0.093^{**}$ | $-0.205^{**}$ | $-0.284^{***}$ | $-0.082^{***}$ | $-0.110^{***}$ | $-0.081^{***}$ | $-0.107^{***}$ | $-0.011$ | $-0.192^{***}$ |
| edu4 | $0.039^{**}$ | $-0.063^{***}$ | $-0.237^{***}$ | $-0.082^{***}$ | 0.008 | $-0.006$ | $-0.032^{***}$ | $-0.028^{**}$ | $-0.043$ | $0.031^{***}$ |
| female | $-0.230^{***}$ | $-0.151^{***}$ | $0.125^{***}$ | $-0.177^{***}$ | $-0.088^{***}$ | $-0.130^{***}$ | $-0.119^{***}$ | $-0.227^{***}$ | 0.066 | $-0.159^{***}$ |
| partime | $-0.254^{***}$ | $-0.114^{***}$ | $-0.063$ | $-0.222^{***}$ | $-0.129^{***}$ | $-0.099^{***}$ | $-0.070^{***}$ | $-0.119^{***}$ | $-0.077^{*}$ | $-0.221^{***}$ |
| public | $-0.022$ | $-0.147^{***}$ | $-0.804^{***}$ | $-0.030^{***}$ | $-0.074^{***}$ | $-0.032^{*}$ | $-0.176^{***}$ | $-0.136^{***}$ | $-0.684^{***}$ | $-0.025^{**}$ |
| neg | $0.061^{***}$ | $0.053^{***}$ | $0.087^{**}$ | $0.034^{***}$ | $0.051^{***}$ | $0.031^{*}$ | $0.063^{***}$ | $0.069^{***}$ | $0.083^{**}$ | $0.036^{***}$ |
| y2006 | $-1.027^{***}$ | $-0.235$ | $-0.481$ | $-0.351^{***}$ | $-0.590^{***}$ | $-0.559^{***}$ | $-0.412^{***}$ | $-0.986^{***}$ | $-0.469^{***}$ | $-0.266^{***}$ |
| y2008 | $-0.208^{***}$ | $-0.165^{***}$ | $-0.298^{***}$ | $-0.145^{***}$ | $-0.179^{***}$ | $-0.144^{***}$ | $-0.270^{***}$ | $-0.244^{***}$ | $-0.268^{***}$ | $-0.125^{***}$ |
| y2010 | $-0.140^{***}$ | $-0.079^{***}$ | $-0.058$ | $-0.199^{***}$ | $-0.119^{***}$ | $-0.091^{***}$ | $-0.166^{***}$ | $-0.147^{***}$ | $-0.080$ | $-0.102^{***}$ |
| y2012 | $-0.059^{***}$ | $-0.013$ | $-0.104$ | $-0.068^{***}$ | $-0.062^{***}$ | $-0.026$ | $-0.079^{***}$ | $-0.077^{***}$ | $-0.114$ | $-0.060^{***}$ |
| No. obs. | 239591 | 82746 | 15027 | 395685 | 267898 | 87152 | 167077 | 147548 | 15027 | 395685 |
| R2 | 0.37 | 0.49 | 0.52 | 0.27 | 0.35 | 0.53 | 0.55 | 0.44 | 0.60 | 0.30 |

Note: $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Table 7: Regression results for males and females in sub-major occupational groups

| | 31female | 31male | 32female | 32male | 33female | 33male | 34female | 34male | alt_33female | alt_33male |
|---|---|---|---|---|---|---|---|---|---|---|
| overeduc | 0.000 | $0.166^{***}$ | $0.054^{**}$ | $0.115^{***}$ | 0.000 | 0.000 | $0.122^{***}$ | $0.183^{***}$ | | |
| alt_over | | | | | | | | | $0.300^{***}$ | $0.307^{***}$ |
| tenure | $0.023^{***}$ | $0.026^{***}$ | $0.025^{***}$ | $0.028^{***}$ | $0.027^{***}$ | $0.026^{***}$ | $0.024^{***}$ | $0.034^{***}$ | $0.023^{***}$ | $0.025^{***}$ |
| tenure2 | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{**}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{**}$ | $-0.001^{***}$ |
| edu1 | $-0.055$ | $-0.080^{**}$ | $-0.261^{***}$ | $-0.230^{***}$ | 0.000 | $0.567^{***}$ | $-0.232^{***}$ | $-0.196^{***}$ | 0.205 | $0.680^{***}$ |
| edu2 | $-0.261^{***}$ | $-0.284^{*}$ | $0.404^{***}$ | $-0.416^{***}$ | $0.895^{***}$ | $0.408^{***}$ | $-0.087$ | $-0.152^{*}$ | $1.067^{***}$ | $0.518^{***}$ |
| edu3 | $-0.123^{***}$ | $-0.069^{***}$ | $-0.209^{***}$ | $-0.118^{***}$ | $-0.157^{*}$ | $-0.244^{***}$ | $-0.351^{***}$ | $-0.295^{***}$ | 0.043 | $-0.049$ |
| edu4 | $-0.002$ | $0.040^{***}$ | $-0.120^{***}$ | $-0.056^{**}$ | $-0.337^{***}$ | $-0.219^{**}$ | $-0.074^{***}$ | $-0.085^{***}$ | $-0.105$ | $-0.029$ |
| partime | $-0.079^{**}$ | $-0.240^{***}$ | 0.008 | $-0.081^{*}$ | $-0.124^{**}$ | $-0.055$ | $-0.055^{***}$ | $-0.211^{***}$ | $-0.118^{**}$ | $-0.077^{*}$ |
| public | $-0.061^{***}$ | $-0.010$ | $-0.131^{***}$ | $-0.210^{***}$ | $-0.623^{***}$ | $-0.861^{***}$ | $-0.027^{**}$ | $-0.025$ | $-0.534^{***}$ | $-0.736^{***}$ |
| neg | $0.038^{**}$ | $0.068^{***}$ | $0.054^{***}$ | $0.046^{***}$ | 0.064 | $0.083^{*}$ | $0.032^{***}$ | $0.048^{***}$ | 0.056 | $0.085^{*}$ |
| y2006 | $-1.030^{***}$ | $-0.037$ | $-0.214^{***}$ | $-0.500^{**}$ | $-0.304$ | $-0.324$ | 0.191 | $-0.310^{***}$ | $-0.407^{*}$ | $-0.449$ |
| y2008 | $-0.194^{***}$ | $-0.210^{***}$ | $-0.120^{***}$ | $-0.263^{***}$ | $-0.482^{***}$ | $-0.173^{*}$ | $-0.148^{***}$ | $-0.145^{***}$ | $-0.393^{***}$ | $-0.183^{*}$ |
| y2010 | $-0.117^{***}$ | $-0.146^{***}$ | $-0.054^{**}$ | $-0.129^{***}$ | $-0.126$ | $-0.026$ | $-0.192^{***}$ | $-0.226^{***}$ | $-0.090$ | $-0.077$ |
| y2012 | $-0.047^{**}$ | $-0.062^{***}$ | 0.003 | $-0.041^{*}$ | $-0.175$ | $-0.042$ | $-0.066^{***}$ | $-0.069^{***}$ | $-0.133$ | $-0.081$ |
| No. obs. | 55930 | 183661 | 47067 | 35679 | 7701 | 7326 | 297915 | 97770 | 7701 | 7326 |
| R2 | 0.33 | 0.34 | 0.36 | 0.56 | 0.62 | 0.49 | 0.30 | 0.21 | 0.67 | 0.58 |

Note: $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Table 7 (cont.): Regression results for males and females in sub-major occupational groups

| | alt_34female | alt_34male | 41female | 41male | 42female | 42male | 51female | 51male | 52female | 52male |
|---|---|---|---|---|---|---|---|---|---|---|
| overeduc | | | 0.154*** | 0.117*** | 0.089*** | 0.123*** | 0.000 | 0.068*** | 0.119*** | 0.164*** |
| alt_over | 0.220*** | 0.236*** | | | | | | | | |
| tenure | 0.026*** | 0.034*** | 0.023*** | 0.021*** | 0.019*** | 0.028*** | 0.010*** | 0.014*** | 0.008*** | 0.029*** |
| tenure2 | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.010** | −0.001*** | −0.001*** | −0.001*** |
| edu1 | −0.137*** | −0.083*** | −0.046*** | −0.055** | −0.107*** | −0.098* | −0.064*** | −0.091*** | 0.001 | −0.078*** |
| edu2 | 0.052 | −0.014 | 0.110 | −0.055 | −0.110** | −0.134 | 0.026 | −0.038 | −0.034 | −0.117*** |
| edu3 | −0.276*** | −0.205*** | −0.103*** | −0.084*** | −0.113*** | −0.110*** | −0.085*** | −0.073*** | −0.070*** | −0.125*** |
| edu4 | 0.008 | 0.021 | 0.005 | −0.003 | −0.006 | 0.012 | −0.008 | −0.020* | 0.010* | −0.015 |
| partime | −0.041*** | −0.204*** | −0.064*** | −0.131*** | 0.005 | −0.115** | −0.033*** | −0.053*** | 0.085** | −0.141*** |
| public | −0.023** | −0.016 | −0.084*** | −0.055*** | −0.042** | −0.041 | −0.180*** | −0.358*** | −0.201*** | −0.129*** |
| neg | 0.036*** | 0.046*** | 0.065*** | 0.033*** | 0.031 | 0.037 | 0.056*** | 0.090*** | 0.081*** | 0.036 |
| y2006 | −0.256 | −0.209*** | −0.264 | −0.564*** | −0.566*** | −0.496*** | −0.716*** | −0.668*** | −0.639*** | −0.578*** |
| y2008 | −0.125*** | −0.128*** | −0.186*** | −0.165*** | −0.145*** | −0.171*** | −0.264*** | −0.286*** | −0.248*** | −0.231*** |
| y2010 | −0.099*** | −0.111*** | −0.121*** | −0.114*** | −0.091*** | −0.100** | −0.153*** | −0.180*** | −0.150*** | −0.132*** |
| y2012 | −0.059*** | −0.061*** | −0.062*** | −0.062*** | −0.027 | −0.035 | −0.068*** | −0.095*** | −0.079*** | −0.071** |
| No. obs. | 297915 | 97770 | 164618 | 103280 | 75112 | 12040 | 71299 | 95778 | 112456 | 35092 |
| R2 | 0.33 | 0.24 | 0.37 | 0.32 | 0.52 | 0.40 | 0.47 | 0.59 | 0.46 | 0.34 |

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

overeducation (as discussed in the section *Identification of overeducation*) and look at the formal requirements and not actual educational level of employees, wage effect occurs. In this case, overeducated workers are rewarded by a wage premium of 30.5% on average. We also applied the alternative measure for the $34^{th}$ sub-major occupational group. Different results obtained for both of these sub-major occupational groups suggest that indeed, the choice of the method of measurement of overeducation can affect the results, as assumed in hypothesis $H4$.
The smallest wage premium for overeducation occurs in the $51^{st}$ sub-major occupational group (Personal service workers). However, working in $51^{st}$ group requires specific qualifications (it consist of such occupations as cooks, hairdressers or building supervisors), so having vocational education may be here of higher importance than higher education.
In the last step, we run the equation (1) on both genders separately, in sub-major occupational groups (Table 7). The parameter by the variable $overeduc_{i,t}$ varies from 0% to 18.3%. There are also significant differences between wage premia between genders in the same sub-major occupational group. Only in one group ($41^{st}$; General and keyboard clerks) the wage premium for overeducation is higher for women (15.4%) than for men (11.7%). In four cases, the effect of overeducation on wages was found to be insignificant: for women in group 31 (Science and engineering associate professionals), for both genders in group 33 (Business and administration associate professionals) and for women in group 51 (Personal service workers). As for group 33, we have come across this issue before, overcoming it by using the alternative measure of overeducation. This approach was used also this time, resulting in alternative wage premia of about 30% for both genders. We assume that no significant effect in group 31 may be connected with the fact that in this group only 25% of the workers are female and at the same time, also most (73%) of the overeducated workers are male. In the remaining sub-major occupational groups, men receive higher wage premium for overeducation than women.

# 6 Conclusions

In this study we attempted to analyse the wage effect of overeducation in firms with 10 or more employees in Poland. We looked at overeducation in general and on the level of chosen sub-major occupational groups. We also checked whether in this sample wage effects of overeducation vary between men and women. The results of the analyses performed in the study show that although in the sample overeducation contributes to higher wages, wage consequences of overeducation indeed vary between major and sub-major occupational groups and between genders as well.
We succeeded in confirming all the initial hypotheses. On average, workers in the whole sample and in particular in the $3^{rd}$, $4^{th}$ and $5^{th}$ major and their sub-major occupational groups are rewarded for being overeducated, but the size of wage effects of overeducation differs between occupational groups (as suggested by Verdugo

& Verdugo, 1989 for the US data). The results are in line with the economic literature, as they show that overeducated workers earn more that their well-matched peers who perform the same job (see eg. Rubb, 2003).
The differences in wage effects of overeducation between men and women occur in all analysed occupational groups and vary among them. These results could be thought as confirmation of Wincenciak's (2016), who finds that wage effects of overeducation are different for both genders in Poland.
We showed also that the choice of the method of measurement of overeducation affects the results. In general, most of the studies that applied more than one approach show that there is a difference in results depending on the method used. We proved it for Poland by introducing two methods of measurement of overeducation. In the case of realized matches method, we showed that this method is sensitive to labour market situation in the sense that if there is a surplus of qualified workers then workers with a higher qualification that the one required for their job would be hired, so there would be an underestimation of the overeducation effect. While the job analysis method provides more reliable results, the question remains open whether a required educational level should be defined at such an aggregated level. This result is also in line with the economic literature, as Mysikova (2016) using Czech data showed that measurements of educational mismatch are highly sensitive to the methodology used. This implies that, when discussing the outcomes of overeducation on wages, one should be very careful when comparing results of different studies.
Nevertheless, these results must be interpreted with caution and a number of limitations should be borne in mind. Firstly, the study has only examined a part of the economy. The database we use is limited only to firms with 10 or more employees. This means, we omit all small firms that constitute an important part of the economy. In 2018, about 25% of all employment in Poland was employment in firms with up to 9 employees. However, with our current dataset we are not able to include also these small enterprises and this issue calls for further research.
Secondly, we have a reference period of eight years. In this period of time the structure of education of workers has also changed thus affecting the results. Furthermore, when discussing the results, the impact of demographic change should be considered. Due to the relatively low birthrate that Poland has been experiencing for almost 30 years, the number of people in the workforce remains relatively stable or is even decreasing. For this reason, it is relatively easier for them to find a job that is well suited to ones' education.

## Acknowledgement

The author is thankful to Paweł Strawiński and two anonymous reviewers for their insightful comments. All the remaining errors are mine. The views expressed in text are those of the author and shall not be related to institutions.

## References

[1] Alba-Ramirez A., (1993), Mismatch in the Spanish labour market: Overeducation?, *Journal of Human Resources* 28, 259–278.

[2] Allen J., van der Velden R., (2001), Educational Mismatches Versus Skill Mismatches: Effects on Wages, Job satisfaction and On-the-Job Search, *Oxford Economic Papers* 53(3), 434–452.

[3] Attanasio O. P., Kaufmann K., (2017), Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations, *Journal of Economic Behavior & Organization* 140(C), 35–55.

[4] Baran J., (2016), A side effect of a university boom: rising incidence of overeducation among tertiary-educated workers in Poland, *University of Warsaw Faculty of Economic Sciences Working Papers* 22/2016 (213).

[5] Battu H., Belfield C. R., Sloane P., (2000), How Well Can We Measure Graduate Over Education and Its Effects?, *National Institute Economic Review* 171(1), 82–93.

[6] Bauer T., (2002), Educational mismatch and wages: a panel analysis, *Economics of Education Review* 21, 221–229.

[7] Becker G. S., (1964), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, University of Chicago Press, Chicago.

[8] Boll Ch., Leppin J., Rossen A., Wolf A., (2016), Overeducation – New Evidence for 25 European Countries, Research Paper no. 173.

[9] Caroleo F. E., Pastore F., (2013), Overeducation at a Glance: Determinants and Wage Effects of the Educational Mismatch, Looking at the AlmaLaurea Data, IZA Discussion Papers 7788, Institute of Labor Economics (IZA).

[10] Chevalier A., (2000), Graduate Over-Education in the UK, *Centre for the Economics of Education Discussion Paper* 07.

[11] Chevalier A., Lindley J., (2009), Over-education and the Skills of UK Graduates, *Journal of the Riyal Statistical Society* 172(2), 307–337.

[12] Chiappori P., Iyigun M., Weiss Y., (2009), Inverstment in Schooling and the Marriage Market, *American Economic Review* 99(5), 1689–1713.

[13] Chłoń-Domińczak A., Żurawski A., (2017), Measuring Skills Mismatches Revisited – Introducing Sectoral Approach, IBS Working Paper 03/2017.

[14] Daly M. C., Buchel F., Duncan G. J., (2000), Premiums and Penalties for Surplus and Deficit Education: Evidence from the United States and Germany, *Economics of Education Review* 19, 169–178.

[15] Di Pietro G., Urwin P., (2006), Education and skills mismatch in the Italian graduate labour market, *Applied Economics* 38(1), 79–93.

[16] Dolton P., Vignoles A., (2000), The incidence and effects of mismatch in the UK graduate labour market, *Economics of Education Review* 19, 179–198.

[17] Duncan G., Hoffman S. D., (1981), The Incidence and Wage Effects of Overeducation, *Economics of Education Review* 1(1), 75–86.

[18] Freeman R. B., (1976), *The overeducated American*, Academic Pass.

[19] Goraus K., Tyrowicz J., (2014), Gender wage gap in Poland – can it be explained by differences in observable characteristics?, *Ekonomia* 36, 125–148.

[20] Green F., McIntosh S., (2007), Is There a Genuine Under-utilisation of Skills Amongst the Overqualified?, *Applied Economics* 39 (4), 427–439.

[21] Groot W., (1996), The incidence of, and returns to overeducation in the UK, *Applied Economics* 28(10), 1345–1350.

[22] Groot W., Maaseen van den Brink H., (2000), Overeducation in the labor market: a meta-analysis, *Economics of Education Review* 19(2), 149–158.

[23] Hartog J., Oosterbeek H., (1988), Education, allocation and earnings in the Netherlands: Overschooling?, *Economics of Education Review* 7, 185–194.

[24] Hartog J., (2000), Overeducation and earnings: where are we, where should we go?, *Economics of Education Review* 19(2), 131–147.

[25] Kiker B. F., Santos M. C., Mendes De Oliveira M., (1997), Overeducation and Undereducation: Evidence for Portugal, *Economics of Education Review* 16(2), 111–125.

[26] Kucel A., Vilalta-Bufi M., (2012), Graduate Labour Mismatch in Poland, *Polish Sociological Review* 3(179), 413–429.

[27] Leuven E., Oosterbeek H., (2011), Overeducation and Mismatch in the Labour Market, IZA Discussion Paper, no. 5523.

[28] Li I., Simonson R., Malvin M., (2014), Over-education and Employment Mismatch: Wage Penalties for College Degrees in Business, *The Journal of Education for Business* 90(3), 119–125.

[29] Majchrowska A., Strawiński P., (2018), Impact of minimum wage increase on gener wage gap: Case of Poland, *Economic Modelling* 70, 174–185.

[30] Mavromaras K., McGuinness S., O'Leary N., Sloane P., Wei Z., (2013), Job Mismatches and Labour Market Outcomes: Panel Evidence on University Graduates, *Economic Record*89, 382–395.

[31] McGuiness S., Wooden M., (2009), Overskilling, Job Insecurity and Career Mobility, *Industrial Relations: A Journal of Economy and Society* 48(2), 265–286.

[32] McGuinness S., Sloane P., (2011), Labour market mismatch among UK graduates: An analysis using REFLEX data, *Economics of Education Review* 30(1), 130–145.

[33] McGuinness S., Bergin A., Whelan A., (2017), Overeducation in Europe: Trends, Convergence and Drivers, IZA Discussion Paper No. 10678.

[34] McGuiness S., (2006), Overeducation in the Labour Market, *Journal of Economic Surveys* 20(3), 387–418.

[35] Mincer J., (1974), *Schooling, Experience and Earnings*, National Bureau of Economic Research, Cambridge.

[36] Mysikova M., (2016), Has Personal Earnings Inequality Become Polarized? The Czech Republic in a Comparative Perspective, *Journal of Income Distribution* 24(3-4), 3-24.

[37] OECD, (2019), OECD Skills Strategy Poland: Assessment and Recommendations, OECD Skills Studies, OECD Publishing, Paris, `https://doi.org/10.1787/b377fbcc-en`.

[38] Ortiz L., Kucel A., (2008), Do Fields of Study Matter for Over-education? The Cases of Spain and Germany, *International Journal of Comparative Sociology* 49, 305–327, DOI: 10.1177/0020715208093079.

[39] Robst J., (2008), Overeducation and College Major: Expanding the Definition of Mismatch Between Schooling and Jobs, *The Manchester School* 76(4), 349–368.

[40] Rossen A., Boll Ch., Wolf A., (2019), Patterns of Overeducation in Europe: The Role of Field of Study, *IZA Journal of Labor Policy* 9(1).

[41] Rubb S., (2003), Overeducation in the Labor Market: A Comment and Re-Analysis of a Meta-analysis, *Economics of Education Review* 22(6), 621–629.

[42] Sánchez-Sánchez N., McGuinness S., (2013), Decomposing the impacts of overeducation and overskilling on earnings and job satisfaction: an analysis using REFLEX data, *Education Economics*, in press, `http://dx.doi.org/10.1080/09645292.2013.846297.`

[43] Sattinger M., (1993), Assignment Models of the Distribution of Earnings, *Journal of Economic Literature*31, 831–880.

[44] Sloane P., (2003), Much Ado about Nothing? What Does the Over-education Literature Really Tell Us?, *International Conference on Over-education in Europe: What Do We Know?*.

[45] Sloane P. J., (2014), Overeducation, Skill Mismatches, and Labor Market Outcomes for College Graduates, *IZA World of Labor* 88, 1–10.

[46] Strawiński P., Broniatowska P., Majchrowska A., (2016), Returns to vocational education evidence from Poland, WNE Working Papers No. 16/2016 (207).

[47] Strawiński P., (2015), Krzyżowe porównanie danych o wynagrodzeniach z polskich badań przekrojowych, *Bank i Kredyt* 46 (5), 433–462.

[48] Vahey S., (2000), The great Canadian training robbery: evidence on the returns to educational mismatch, *Economics of Education Review* 19, 219–227.

[49] Verdugo R., Verdugo N., (1989), The Impact of Surplus Schooling on Earnings: Some Additional Findings, *Economics of Education Review* 22(4), 690–695.

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

[51] Wincenciak L., (2016), Educational mismatches and earnings in Poland: are graduates penalised for being overeducated?, *Ekonomia* 46, 145–167.

# Appendix

Table 8: Definitions of explanatory variables

| Variable name | Values | Share (%) |
|---|---|---|
| Female | Woman* | 50.74% |
| | Man | 49.26% |
| Edu1 | Comprehensive primary school (ISCED 1) & Comprehensive lower secondary school (ISCED 2) | 5.95% |
| Edu2 | Basic vocational school (ISCED 3) | 23.84% |
| Edu3 | Secondary vocational school (ISCED 3) | 4.93% |
| Edu4 | General secondary school (ISCED 3) & Profiled general secondary school (ISCED 3) | 30.33% |
| Edu5* | BA & MA (ISCED 5) | 34.95% |
| Partime | Working part-time* | 7.93% |
| | Working full-time | 92.07% |
| Public | Working in public sector* | 62.37% |
| | Working in private sector | 37.63% |
| Overeduc | Overeducated* | 18.45% |
| | Not overeducated | 81.55% |
| Neg | Wage negotiations are established at the sectoral or national level* | 40.36% |
| | Wage negotiations are not established at the sectoral or national level | 59.64% |

Note: Reference categories for dummy variables are denoted using asterisks.

Table 9: Educational levels used in the model

| | |
|---|---|
| Edu1 | Comprehensive primary school |
| Edu2 | Comprehensive lower secondary school |
| Edu3 | Basic vocational school |
| Edu4 | General secondary school |
| Edu5 | Profiled general secondary school |
| Edu6 | Secondary vocational school |
| Edu7 | BA |
| Edu8 | MA or higher |

Table 10: Descriptive statistics of the variables used in the model (for all periods)

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| lwage | 2.981 | 0.641 | −0.485 | 8.561 |
| tenure | 17.060 | 11.577 | 0.08 | 66 |
| tenure2 | 425.085 | 457.762 | 0.006 | 4356 |
| edu1 | 0.058 | 0.233 | 0 | 1 |
| edu2 | 0.002 | 0.044 | 0 | 1 |
| edu3 | 0.238 | 0.426 | 0 | 1 |
| edu4 | 0.080 | 0.271 | 0 | 1 |
| edu5 | 0.223 | 0.417 | 0 | 1 |
| edu6 | 0.049 | 0.216 | 0 | 1 |
| edu7 | 0.079 | 0.270 | 0 | 1 |
| edu8 | 0.270 | 0.444 | 0 | 1 |
| female | 0.493 | 0.500 | 0 | 1 |
| partime | 0.079 | 0.270 | 0 | 1 |
| public | 0.624 | 0.484 | 0 | 1 |
| overeduc | 0.185 | 0.388 | 0 | 1 |
| sekcjaA | 0.064 | 0.244 | 0 | 1 |
| sekcjaB | 0.042 | 0.199 | 0 | 1 |
| sekcjaC | 0.053 | 0.223 | 0 | 1 |
| sekcjaD | 0.008 | 0.089 | 0 | 1 |
| sekcjaE | 0.051 | 0.220 | 0 | 1 |
| sekcjaF | 0.025 | 0.156 | 0 | 1 |
| sekcjaG | 0.023 | 0.149 | 0 | 1 |
| sekcjaH | 0.022 | 0.147 | 0 | 1 |
| sekcjaI | 0.143 | 0.350 | 0 | 1 |
| sekcjaJ | 0.048 | 0.213 | 0 | 1 |
| sekcjaK | 0.041 | 0.199 | 0 | 1 |
| sekcjaL | 0.052 | 0.223 | 0 | 1 |
| sekcjaM | 0.017 | 0.128 | 0 | 1 |
| sekcjaN | 0.204 | 0.403 | 0 | 1 |
| sekcjaO | 0.015 | 0.122 | 0 | 1 |
| sekcjaP | 0.065 | 0.247 | 0 | 1 |
| sekcjaQ | 0.115 | 0.319 | 0 | 1 |
| sekcjaR | 0.010 | 0.100 | 0 | 1 |
| neg | 0.404 | 0.491 | 0 | 1 |
| y2006 | 0.184 | 0.387 | 0 | 1 |
| y2008 | 0.206 | 0.404 | 0 | 1 |
| y2010 | 0.205 | 0.404 | 0 | 1 |
| y2012 | 0.207 | 0.405 | 0 | 1 |