

## Age Structure of Employment and Wages. An Analysis Across Occupational Groups

Paulina Broniatowska\*, Aleksandra Majchrowska†, Maciej Nasiński‡

Submitted: 21.06.2019, Accepted: 10.06.2020

### Abstract

Recently, in most developed economies, the average age of the workforce has been growing rapidly. Therefore, the questions arise how will it affect the level of wages and the shape of age-productivity and age-wage profiles.

The aim of the paper is to analyse the relationship between changes in the age structure of the employment and wages of individuals in minor occupational groups. Using individual data from the Structure of Earnings Survey in Poland in 2006-2014 we created an unique database of individual wages and the characteristics of employed in occupational groups at 3-digit level of classification. In our analysis we used an extended version of Mincerian wage model where both the characteristics of employees (education, work tenure, age, gender, and type of employment contract) and employers (size and ownership sector) were taken into account.

The results for the whole sample indicate a significant and negative relationship between the proportion of older workers in employment in a given occupational group and individual wages. However, when the analyses were performed separately for each of the 1-digit occupational groups, the results varied significantly. In those groups where knowledge and qualifications of employees are more important than physical strength had to be updated permanently, an increase in the number of the older workers raises the average wages.

**Keywords:** employment structure, workforce aging, wages, occupational groups, Mincer-type wage equation

**JEL Classification:** J24, J31, J14

---

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

†University of Lodz; e-mail: aleksandra.majchrowska@uni.lodz.pl; ORCID: 0000-0002-9759-0830

‡University of Warsaw; e-mail: nasinski.maciej@gmail.com; ORCID: 0000-0001-5236-8040

## 1 Introduction

Recently in most developed economies the average age of the workforce has been growing rapidly. Declining fertility rates along with increasing life expectancy have inevitably led to rising percentage of older people in the total population and to aging of the workforce. The aim of our paper is to answer the question: how changes in the age structure of employment affect the level of individual wages.

Empirical evidence in this area is not straightforward. According to the economic theory, wages should reflect the level of individual' productivity. Although most empirical studies conclude that the workers' productivity peaks around the ages of 30-45 and starts to visibly decline around the age of 50, some find no decrease in productivity levels as the employees get older. The research studies show that potential decrease in the employees' productivity levels depends on several factors including the sector that they work in, their individual tasks as well as human capital, and learning abilities. Although it has been proved beyond doubt that physical performance declines with age, for mental performance there have been mixed results. Despite the vast literature available on the relationship between age and productivity, relatively less attention has been devoted to studying the impact of the changes in the age structure of the workforce on individual wages. This paper aims to fill in this gap.

We add to the literature by modifying the usual approach: we analyse not only the relation between worker's own productivity (measured by both the workers' and employers' characteristics) and wages but also the relation between age structure of the employment in a worker's occupational group and his wage. Our main hypothesis is as follows. The effect of workforce aging on wages of individuals depends on the type of the work performed. In those occupational groups that require constant investment in human capital, the salaries grow with growing share of older employees (*ceteris paribus*). Consequently, in those occupational groups where formal qualifications are not as important as physical strength, productivity decreases with age, implying an overall decrease in individual' wages with growing share of older employees (*ceteris paribus*). Overall effect on average wage level in the economy will then depend on the employment structure.

To prove these hypotheses, we use individual data from the Structure of Earnings survey database for Poland. The time period covers the years 2006-2014. To the traditional augmented Mincerian type wage equation we add a variable which shows the age structure of workers in a given occupational group. We take advantage of a huge number of observations in the database and we are able to analyse the employment structure on the 3-digit level of classification. (According to the International Standard Classification of Occupations ISCO-08, see: [https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms\\_172572.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf)). We use the share of workers ages 55 and more in employment in a given occupational group as our variable of interest. With this approach we are able to answer the question how aging of the

workforce in a given minor occupational groups affects the individual' earnings. Due to our best knowledge it is the first study of a kind.

In favour of choosing Poland as an example in our study speak several facts. Firstly, Poland is a big country covering 35% of total labour force of the new member countries of the European Union (EU). Secondly, we can observe significant changes in the age structure of the workforce and therefore in employment. The path of aging in Poland is one of the highest among European countries: OECD forecasts shows that in 2050 the old-age dependency ratio in Poland at 60.8% will be much higher than OECD average (53.2%). (OECD (2017) Pensions at a glance; [https://www.oecd-ilibrary.org/docserver/pension\\_glance-2017-22-en.pdf?expires=1580297630&id=id&accname=guest&checksum=B7A4C470D48116DA18976A33A7F10EED](https://www.oecd-ilibrary.org/docserver/pension_glance-2017-22-en.pdf?expires=1580297630&id=id&accname=guest&checksum=B7A4C470D48116DA18976A33A7F10EED)). Thirdly, changes in the age structure of workers are not even across occupational groups. In 2014 the share of workers aged 55 and more varied among 3-digit occupational groups between 5 and 40%. Among the occupational groups with the highest median age, there are both elementary occupations as well as managers and sales workers (see Appendix A).

Our results are interesting not only from the labour market participants point of view but also for the policymakers dealing with the problem of population ageing. For the whole sample, the relationship between individual wages and the share of older workers in a given occupational group is significant and negative. Therefore, on average aging of the workforce has a negative effect on individual wages *ceteris paribus*.

The structure of the paper is as follows: Section 1 provides the introduction. Section 2 presents the main findings on the relationship between age and wages provided by the literature. Section 3 describes the data. Section 4 contains a description of the empirical approach used for this paper. And lastly, Section 5 presents empirical results, while Section 6 concludes.

## 2 Aging and wages: theory and empirical evidence

On the individual level, wage is a function of different workers' characteristics. Among them, the literature often mentions productivity, often approximated with age. Therefore, a simple way to assess the relation between age and wage is to compare age-productivity and age-wages profiles. Mincer (1974), who conducted one of the first extensive analyses of age-wage profiles, concluded that older workers can earn more even if they spend less time investing in human capital since they earn the returns to earlier investments.

Since the early 1970s and the introduction of works by Mincer (1974) and Hurd (1971) the 'inverse-U' shape of the age-wage profile is taken for granted. Using this approach, most of the empirical studies prove that in the developed economies the workers' productivity is the highest around the ages of 30-45 and starts to visibly decline around the age of 50 (see e.g. Cataldi et al. 2011; Göbel and Zwick 2009;

Skirbekk 2008). Nevertheless, as workers age, their skills depreciate at different paces. Lis and Magda (2014) point out that although effectiveness of investments in human capital decreases sharply over time, the decline in productivity could be offset by educational investments as employees get older. Lovász and Rigó (2013) prove that in Hungary the age-related gap in productivity levels is higher among skilled workers. There is also evidence in the literature that the rate at which human capital and skills depreciate (and thus productivity declines) is constantly increasing due to the skills-biased technological changes (Börsch-Supan et al. 2005; Bertschek and Meyer 2009).

While empirical research indeed proves that the age-wage profile mirrors to some extent the upward sloping age-productivity profile (see eg. Hellerstein and Neumark 1995 for Israeli data), there are some differences in the shape of both functions. Dostie (2006), using linked employer-employee data for Canada, found that although both wage and productivity profiles are concave, for workers aged 55 and above productivity is diminishing faster than wages. In order to explain this difference, Mitchell et al. (1990) suggest that in some areas, older workers may have abilities that are relatively superior to those of younger workers and their experience could compensate for their declining productivity.

This hypothesis seems to hold as the potential decrease in productivity with age varies among different groups of workers. It depends on the sectors they work in, their tasks, and their human capital. Although it has been established that physical performance declines with age, there are varying results for mental performance (Bazen and Charni 2015; Ng and Feldman 2008; Sturman 2003). Van Ours (2009) analysed the productivity of blue-collar workers (using professional runner as an example) and white-collar workers (where as a measure of productivity, he took the number of publications in economic journals) in the US. He proved that in the case of blue-collar workers, productivity decreases significantly after the age of 40. However, in the case of white-collar workers, he did not confirm a similar dependence, even for people aged 50 and above.

On the other hand, Cardoso et al. (2010), who show age productivity profiles for the Portuguese manufacturing and services sectors, state that that while productivity increases until the age range of 50-54, wages have already peaked around the ages of 40-44. As a result, the average contribution of older employees to firm-level productivity may even exceed their contribution to the wage bill.

The fact that different studies have contradictory results may be justified by difference in occupational structure of the analysed samples. The discrepancies between the results of the age-wage profile analyses may – to some extent – stem from the fact that in different occupational groups, productivity peaks at different ages and it decreases in different pace. In fact, in the aforementioned work, Cardoso et al. (2010) show that productivity of older workers is indeed slightly higher in services than in manufacturing. Van Ours and Stoeldraijer (2011) and Lallemand and Rycx (2009) analysed age-productivity profiles in different group of sectors, which in some

way similar to our approach. Their results also confirm that the impact of aging on wages may vary over different occupational groups. In particular, Van Ours and Stoeldraijer (2011) show separate age productivity profiles for the following sectors in the Netherlands: construction, wholesale trade, retail trade, commercial services, and manufacturing. In their work, they point to clear differences in the age-productivity patterns between these sectors – while for manufacturing value added increases until the age group 50-56, for other sectors the age-productivity patterns are flat. Similar results obtain Aubert and Crépon (2006), who separately consider the impact of age groups on productivity for the French manufacturing, trading, and services sectors. They stress that while in all sectors relative productivity increases until age 35, after this age, older workers' productivity changes depending on the sector. In manufacturing, there is no statistically significant difference in productivity between the age group 35-39 and older workers. In trading, workers aged 40-59 are significantly more productive, whereas in services only workers aged 45-54 are more productive than younger workers.

The hitherto literature suggests that in the occupational groups that require constant investment in terms of human capital, skills and experience increase with age. Therefore, salaries should grow. On the other hand, in the groups where formal qualifications are not as important as physical strength, productivity decreases with age, causing a decrease in wages as well. The problem of declining productivity in manufacturing and mining industry was tackled by Mahlberg et al. (2008), who found that the lower productivity of the aged workers occurs in relation to plant-specific lower productivity because older workers are more often employed in firms with older technologies. After controlling the firms' fixed effects, a positive relationship between productivity and the ratio of older employees is revealed. This implies that in defining relation between the age of the workforce and wages not only age effects, but also cohort effects may occur. Therefore we also try to distinguish between age and cohort effects.

In emerging countries this topic has not been sufficiently studied so far, partly because the problem of population ageing was not so prevalent as in developed economies. Using data for Poland, Zgliczyński (2013) states that among workers aged between 50 and 60, wages vary only slightly and for people older than 60 average wages increase clearly. He explains this fact with the self-selection of employees – in this age group, only specific types of workers are economically active. There are basically two types of older workers: those who are active on the labour market because of the high salary and satisfaction from their work and those who are forced to work due to the financial situation. Majchrowska and Roszkowska (2014), however, argue that wages in Poland increase with age only to a certain point, which confirms the decline in investment in human capital beyond a certain age.

Discussing productivity and age, Veen (2008) distinguishes between occupations that have an increasing productivity with age, occupations that are age neutral, and occupations that have a declining productivity with age. We follow his approach in this paper and analyse the relationship between the age structure of the workforce and wages of individuals in particular occupational groups.

What is new in our paper comparing to the previous research is firstly that we analyse not only the relation between worker's own productivity (measured by both the workers' and employers' characteristics) and his wages but also the relation between age structure of the employment in a worker's occupational group and his wage level. Due to availability of huge database of individual data we are able to analyse the impact of age structure of workers across the minor (3-digit) occupational groups. To our best knowledge this kind of analyses were not performed before. Secondly, the relation between age structure and wages has not been widely discussed in Poland yet. As Poland is currently experiencing fast aging of the population (and hence of the workforce), this topic seems to be of particular importance.

### 3 Data

We use individual data from five rounds of the Structure of Earnings survey conducted by Statistics Poland. Time period covers the years from 2006 to 2014. 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. The database includes 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. These data allow us to account, simultaneously, for both socio-demographic characteristics of employees as well as firm heterogeneity, in explaining earnings.

An added advantage of the SES survey is the high reliability of its data related to wages. Wages are not declared by the respondents but are reported by the accounting departments along with the number of hours worked. Another advantage is the size of the database. On average, between 2006 and 2014, the SES survey covered around 12% of the total number of Polish enterprises whose employees exceeded nine. In a given year, the database contains around 600,000-700,000 of individuals. The disadvantage of the SES database is that it covers only the entities with at least 10 workers.

The aim of the study is to assess the relationship between the age structure of the workforce and wages across minor occupational groups. Due to the changes in classification of occupations in 2010 we had to standardise the structure of occupational groups to achieve comparability of the results over the years (the

actual classification of occupations used by the Central Statistical Office in Poland is structured on the basis of the International Standard Classification of Occupations (ISCO) as recommended by Eurostat to the EU countries. The structure of the former classification was based on ISCO-88 and adapted to suit Polish conditions). Some of the groups were merged and others split because of changes in the international classification.

The actual Classification of Occupations and Specialities in Poland (COS) is structured on the basis of the International Standard Classification of Occupations (ISCO) as recommended by Eurostat to the EU countries. The classification of occupations and specialities is a five-level, hierarchically systemised set of occupations and specialities occurring in the labour market. It organises some occupations (specialities) into gradually more aggregated groups and determines their symbols and names. In 2010 the COS in Poland has been changed. For this reason, we had to standardise the structure of occupational groups over the analysed period to achieve comparability of results over the years. Some groups were merged and others split because of changes in the international classification. A detailed key that was used to rearrange the old classification into the new one, according to the aforementioned change from 2010, is provided in Appendix B. After standardisation, we possess a unique database of wages and characteristics of employed on 3-digit occupational level which is well comparable over time.

The total number of observations in our sample comes from five rounds of the SES survey and ranges around 5 million in which 3.5 million are the original data and 1.5 million come from the merging procedure (see Appendix B). The sample contains workers from both public and private sectors, working full-time and part-time as well as temporary workers and all persons with other forms of employment. The latter decision is justified by the popularity of other forms of employment in Poland. According to Eurostat, in the years 2006-2014, 27% of the total workforce worked temporary (and in other forms of employment). Before starting the estimations we made some adjustments in the data. Firstly, we included only employees between 15 and 65 years of age. The upper bound restriction comes from the official retirement age. (In 2006-2012 the retirement age in Poland was set at 60 years old for women and 65 years for men. In 2013 a phased increase of the retirement age to 67 by 2020 for men and by 2040 for women was introduced. The retirement age was then supposed to increase by three months each year. Consequently, in the years 2013-2014 some of the 64 years old workers had to stay a few months longer at the labour market. For this analysis we decided to set up the upper age bound at the age of 65 for both genders). Secondly, in the database we do not have data on self-employed individuals. Therefore in the analysis we excluded the sixth major occupational group (skilled agricultural, forestry, and fishery workers) as it consists mostly of self-employed workers. All removed observations constitute, however, less than 2% of the entire dataset.

## 4 Empirical strategy

To estimate the relation between age structure of employment in a given occupational group and individual wages we apply the extended Mincer equation. Our dependent variable is the logarithm of real average hourly wage received by an individual. To obtain the level of real wages, we deflate nominal wages by the consumer price index (in constant 2006 prices).

As far as independent variables are concerned, we take the traditional set of variables suggested by the literature for a Mincerian type models. We consider both the characteristics of the employees and employers which are available in our database. For the employees' characteristics we use information about education, work tenure, age, gender, working time, and type of employment contract. Moreover, we consider the characteristics of employers: size of the employer, ownership sector, and the level of wage negotiations prevailing in the firm where the worker is employed. In the next paragraphs follows a discussion of these variables.

Firstly, as far as education is concerned, we divide the workers into three educational levels: low education contains workers with primary education and lower, lower secondary and basic vocational, medium education consists of workers with secondary and post-secondary education, high education consists of workers with bachelor degree and higher. According to the Mincer model we expect a positive correlation between education level and individual' wages.

Secondly, following Mincer, we account for the in-the-job investments in human capital, which we approximate by the work experience. We define tenure as total number of years spent working (the whole job experience, not only in the current place). Following the Mincer (1974) model, we expect a positive correlation between experience and wages up to some age, which we approximate by taking in equation both the number of years of experience and its square.

Thirdly, as the main aim of the paper is to analyse the relation between age structure of employees in a given occupational group and individual' wages, we consider the share of workers aged 55-65 in employment in given 3-digit occupational group. This proportion was calculated for every year and for every of the 110 of the 3-digit occupational group. According to our hypothesis the sign of the relationship between the share of older workers in a given group and individual' wages will depend on the type of job performed. A high proportion of aged workers with low qualifications should lower the average wages in a given occupational group and a high proportion of well-educated older workers should increase the average wages.

In our model, we would like to disentangle age structure effects from the individual effects of tenure, which may be particularly important against the background of seniority wage schemes existing in the public sector in Poland (see also OECD 2014). Therefore, following Mahlberg et al. (2013), we introduce both variables (age structure in a given group and tenure of the worker) into the model.

Fourthly, in the model we introduce a dummy variable indicating gender of an individual. Following the empirical evidence we expect lower wages for women *ceteris paribus*.

Fifth, since we use in the model the average wage level without recalculation into full-time jobs as a dependent variable, we have to control for the working time of an individual. We introduce the variable ‘working\_time’ which shows the working time of individuals in months. (For example, 12 indicates the full-time worker, 6 indicates that the person is a part-time worker with half of the full-time number of hours working). In some 3-digit level occupational groups the proportion of part-time workers reaches 32% of the total employment. We expect the positive relation between number of hours working and the individual’ wage level.

And sixth, we control for the presence of temporary and other ‘atypical’ form of employment (henceforth: temporary). The proportion of temporary workers in Poland is the highest among all EU countries and averages at 27% of the total workforce (Eurostat data). It is also strongly diversified across occupational groups: among managers, in 2014, only 10% were employed on limited duration contracts. Among lower status employees, the proportion of workers working on limited duration contracts reached even 50% of all employed (Eurostat data). We introduce dummy variable ‘temporary’ indicating whether an individual is a permanent or temporary worker. We expect lower individual’ wages in the case of temporary workers.

Moving to the employers’ characteristics, we observe that the average wages in big firms in Poland are much higher than those in small and medium-sized enterprises. The positive relationship between size of the firm and wage level is also well documented in the literature. As a measure of the size of the employer we take the logarithm of number of workers employed in the firm. We also control for the ownership sector and introduced dummy variable indicating whether the firm operates in public or in private sector. 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.

Finally, we take into account the level of wage negotiations prevailing in the firm where the worker is employed. We introduce a dummy variable taking the value 1 for the case where negotiations are provided at the national or sectoral level, and 0 for other possibilities (negotiations at the firm level or no wage agreements). We expect that wages of individuals are higher in the former case. To control for wage differences across different economic sections, we introduce also dummies for 18 NACE sections. To control for wage growth in time we put also dummy variables for years.

We use an extended Mincer equation to estimate the relation between logarithmic wages and the above-described explanatory variables. The analytical form of the

estimated equation is as follows:

$$\begin{aligned}
 \ln(wage_{i,j}) &= \beta_0 + \beta_1 edu_{i,j} + \beta_2 tenure_{i,j} + \beta_3 tenure_{i,j}^2 + \\
 &+ \beta_4 share\_5565_j + \beta_5 gender_{i,j} + \beta_6 working\_time_{i,j} + \\
 &+ \beta_7 temporary_{i,j} + \beta_8 \ln(size_{i,j}) + \beta_9 sector_{i,j} + \\
 &+ \beta_{10} negotiations_{i,j} + \beta_m years + \\
 &+ \beta_n \sum section_{n,i,j} + \varepsilon_{i,j}
 \end{aligned} \tag{1}$$

where:

$wage_{i,j}$  – average real hourly gross wage level of an individual  $i$  working in minor (3-digit) occupational group  $j$  (PLN; deflated by CPI, 2006=1),

$edu_{i,j}$  – educational level of individual  $i$  working in minor occupational group  $j$  ( $low\_edu$  – primary education and lower, lower secondary and basic vocational,  $medium\_edu$  – secondary and post-secondary education,  $high\_edu$  – with bachelor degree and higher level of education),

$tenure_{i,j}$  – total work experience of an individual  $i$  working in minor occupational group  $j$  (in years),

$share\_5565_j$  – share of 55-65 years old workers in employment in a minor occupational group  $j$  (%),

$gender_{i,j}$  – dummy variable indicating sex of an individual  $i$  working in minor occupational group  $j$  (0 for male, 1 for female),

$working\_time_{i,j}$  – length of the working time of an individual  $i$  working in minor occupational group  $j$  (in months, where 12 is a full-time job),

$temporary_{i,j}$  – dummy variable indicating whether an individual  $i$  working in minor occupational group  $j$  is a permanent or temporary worker at time  $t$  (0 for permanent, 1 for temporary and other form of employment),

$size_{i,j}$  – size of the employer of an individual  $i$  working in minor occupational group  $j$  (measured by the number of workers employed),

$sector_{i,j}$  – dummy variable indicating whether an individual  $i$  working in minor occupational group  $j$  works in a public or private sector (0 for public, 1 for private),

$negotiations_{i,j}$  – dummy variable indicating the level of wage negotiations prevailing in the firm where an individual  $i$  working in minor occupational group  $j$  is employed (1 if negotiations are established at the sectoral or national level, 0 for other possibilities),

$section_{n,i,j}$  – dummy variables for  $n$  NACE section in which an individual  $i$  working in minor occupational group  $j$  works (18 dummies),

$years$  – dummy variables for years (5 dummies),

$\varepsilon_{i,j}$  is an error term.

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. In particular, we are able to receive the sign and strength of the relationship between the proportion of workers aged between of 55-65 in a given occupational group and wages of individuals working in this group.

Then, to verify our hypothesis we decided to allow the parameter  $\beta_4$  to vary across major occupational groups (at 1-digit level). We receive then the information how the relation between the share of aged workers and individual wages differ across occupational groups. The parameters of equation (1) are estimated with least squares method with standard errors clustered at the firm level and year. As far as the potential endogeneity problem is concerned, Heckman (2003) underlines that the current empirical debate on the importance of accounting for the endogeneity of schooling is far from settled. Problem with finding good instruments appears in particular in large datasets of individuals like the one the authors use in this paper. Therefore, we decided to use least squares method with clustered standard errors instead of IV method.

## 5 Empirical results

The estimated parameters of equation (1) for the entire sample of workers are shown in Table 1. As expected, we can observe a significant and positive relation between level of education of workers and their wages. Wages of workers with lower level of education are significantly lower than wages of workers with baseline (tertiary) education. There is also a positive relation between workers experience (*tenure*) and wages up to some point of time, the relation between tenure and wages is inversed U-shaped, in line with economic theory.

For the whole sample, the relationship between individual wages and the share of older workers in a given occupational group is significant and negative ( $-0.40$ ), which indicates that on average aging of the workforce has a negative effect on individual wages *ceteris paribus*. It means that the higher is the share of older workers in employment in a given occupational group (at 3-digit level of classification) the lower are individual' wages in that group *ceteris paribus*. To some extent the negative sign may be explained by the fact that majority of older workers in Poland are employed in rather labour-intensive occupations.

The signs of other parameters are in line with economic theory and empirical evidence.

Wages of women are found to be on average approximately 20% lower than wages of men. Those employed in full-time jobs tend to have slightly higher level of wages than those employed part-time. Temporary workers earn on average 20% less than those employed on permanent contracts. The average level of wages increase with the size of employer. Wages in private sector are on average 5% lower than wages in public sector and they are *ceteris paribus* 3% higher in firms where wage negotiations take place at the national or sectoral level.

Table 1: Results of the model (1) for the whole sample of workers

Estimated parameters' value	
Low_edu	-0.769*** (0.003)
Medium_edu	-0.549*** (0.003)
Tenure	0.025*** (0.000)
Tenure <sup>2</sup> * 100	-0.038*** (0.001)
Share_5565	-0.400*** (0.022)
Female	-0.196*** (0.002)
Working_time	0.007*** (0.000)
Temporary	-0.195*** (0.003)
Log(size)	0.054*** (0.003)
Private	-0.070*** (0.007)
Negotiations	0.030*** (0.009)
Constant	2.741*** (0.015)
R-squared	0.496
Section Dummies	yes
Time Dummies	yes
N	5,722,448

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Parameter and standard errors by Tenure<sup>2</sup> was multiplied by 100 due to very low values of the estimated parameters.

In the results discussed above we analysed the average parameters of the estimated wage equation for the whole population of workers. In particular we found on average the negative relation between individual' wages and the share of aged workers in employment in a given occupational group. To check the stability of the results over time we estimated the equation (1) separately

for every period of time analysed. The results (see Table 2) indicate that most of the estimated parameters are stable. In fact, only one parameter which changed significantly, is the parameter of our interest. At the beginning of the analysed period the relation between the share of aged workers and individual wages was positive and significant (0.2) and it turned to be highly negative and significant (in the range between  $-0.8$  and  $-0.6$  in the years 2020-2014).

In the next step we allow the parameter by variable `share_5565` to vary across major (1-digit) occupational groups. The results (see Table 3) indicate significant differences across groups. In most of the groups the parameter by the share of workers aged 55-65 years old is statistically significant at 1% significance level. The exception is the eight group where the parameter is significant at the 10% significance level. In most of the groups the analysed parameter is negative indicating that individual wages are lower in groups with higher share of aged workers.

Table 2: Estimated values of the parameters of equation (1) separately for years 2006-2014

	2006	2008	2010	2012	2014
Low_edu	-0.820*** (0.008)	-0.767*** (0.007)	-0.786*** (0.008)	-0.714*** (0.008)	-0.690*** (0.007)
Medium_edu	-0.569*** (0.006)	-0.552*** (0.006)	-0.574*** (0.006)	-0.514*** (0.006)	-0.505*** (0.005)
Tenure	0.021*** (0.001)	0.024*** (0.001)	0.027*** (0.001)	0.025*** (0.001)	0.025*** (0.001)
Tenure <sup>2</sup> * 100	-0.029*** (0.001)	-0.038*** (0.001)	-0.045*** (0.001)	-0.040*** (0.001)	-0.040*** (0.001)
Share_5565	0.203*** (0.065)	-0.007 (0.057)	-0.779*** (0.053)	-0.618*** (0.050)	-0.568*** (0.035)
Female	-0.178*** (0.004)	-0.200*** (0.004)	-0.197*** (0.004)	-0.205*** (0.004)	-0.197*** (0.004)
Working_time	0.006*** (0.001)	0.010*** (0.001)	0.006*** (0.001)	0.007*** (0.001)	0.005*** (0.001)
Temporary	-0.188*** (0.007)	-0.178*** (0.007)	-0.183*** (0.007)	-0.187*** (0.006)	-0.204*** (0.006)
Log(size)	0.047*** (0.005)	0.068*** (0.004)	0.054*** (0.005)	0.044*** (0.005)	0.050*** (0.005)
Private	-0.029** (0.013)	-0.018 (0.016)	-0.035** (0.017)	-0.059*** (0.019)	-0.064*** (0.021)
Negotiations	0.030 (0.020)	0.015 (0.017)	0.015 (0.014)	0.047** (0.020)	0.031 (0.023)
Constant	2.686*** (0.030)	2.848*** (0.034)	3.061*** (0.030)	3.143*** (0.037)	3.206*** (0.031)
R-squared	0.479	0.479	0.503	0.464	0.469
Section Dummies	yes	yes	yes	yes	yes
N	1,738,543	1,882,005	677,700	710,477	713,723

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Parameter and standard errors by Tenure<sup>2</sup> was multiplied by 100 due to very low values of the estimated parameters.

Table 3: Estimated values of the parameters of equation (1) separately for major occupational groups

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 7	Group 8	Group 9
Low_edu	-0.693*** (0.014)	-0.491*** (0.019)	-0.384*** (0.006)	-0.243*** (0.006)	-0.356*** (0.009)	-0.189*** (0.011)	-0.110*** (0.008)	-0.084*** (0.009)
Medium_edu	-0.445*** (0.006)	-0.407*** (0.004)	-0.193*** (0.003)	-0.151*** (0.003)	-0.248*** (0.007)	-0.123*** (0.011)	-0.058*** (0.007)	-0.052*** (0.010)
Tenure	0.024*** (0.001)	0.034*** (0.000)	0.021*** (0.000)	0.017*** (0.001)	0.010*** (0.001)	0.012*** (0.000)	0.013*** (0.000)	0.004*** (0.000)
Tenure <sup>2</sup> * 100	-0.041*** (0.002)	-0.062*** (0.001)	-0.032*** (0.001)	-0.023*** (0.001)	-0.016*** (0.001)	-0.019*** (0.001)	-0.019*** (0.001)	-0.005*** (0.001)
Share_5565	2.312*** (0.077)	-1.580*** (0.057)	-0.159*** (0.060)	-0.524*** (0.178)	-0.628*** (0.039)	1.354*** (0.062)	-0.155* (0.083)	-0.510*** (0.040)
Female	-0.202*** (0.005)	-0.126*** (0.003)	-0.158*** (0.003)	-0.092*** (0.005)	-0.173*** (0.005)	-0.240*** (0.005)	-0.240*** (0.006)	-0.145*** (0.005)
Working_time	0.003** (0.001)	0.007*** (0.001)	0.010*** (0.001)	0.009*** (0.001)	0.007*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.003*** (0.001)
Temporary	-0.252*** (0.008)	-0.197*** (0.004)	-0.211*** (0.005)	-0.175*** (0.006)	-0.128*** (0.008)	-0.121*** (0.004)	-0.146*** (0.005)	-0.120*** (0.005)
Log(size)	0.089*** (0.005)	0.035*** (0.002)	0.064*** (0.002)	0.044*** (0.004)	0.036*** (0.004)	0.119*** (0.002)	0.115*** (0.003)	0.055*** (0.003)
Private	0.142*** (0.013)	-0.083*** (0.010)	0.003 (0.009)	0.016 (0.010)	-0.170*** (0.012)	-0.003 (0.009)	-0.077*** (0.010)	-0.088*** (0.014)
Negotiations	0.010 (0.019)	0.036*** (0.010)	0.046*** (0.015)	0.073*** (0.014)	0.051 (0.035)	-0.036** (0.016)	-0.007 (0.014)	0.007 (0.009)
Constant	2.168*** (0.039)	2.840*** (0.020)	2.488*** (0.022)	2.281*** (0.031)	2.654*** (0.034)	1.599*** (0.029)	1.856*** (0.028)	2.181*** (0.032)
R-squared	0.325	0.413	0.359	0.340	0.349	0.513	0.525	0.407
Section Dummies	yes							
Time Dummies	yes							
N	380,519	1,625,579	867,598	437,314	769,876	598,419	430,621	612,522

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Sixth major occupational group was excluded from the analyses since most of the workers in those group are self-employed. Parameter and standard errors by Tenure<sup>2</sup> was multiplied by 100 due to very low values of the estimated parameters.

The highest negative value is found in 2<sup>nd</sup> major occupational group (Professionals). Intuitively, we would expect that in this group, together with the increase in workers professional qualifications, wages of individuals will increase among the higher share of more experienced workers. In the authors' opinion, an explanation comes from the fact that this major occupational group is very heterogeneous: on the one hand, there are for example medical doctors or lawyers whose wages grow constantly with age and work experience. On the other hand, there are IT workers, new technology engineers, and others. Due to rapid technological progress over the last years, the qualifications of young or medium-aged workers in these occupations may be higher or more relevant for the market needs than the knowledge of older workers. To check our hypothesis we performed the equation (1) separately for each minor (at 3 digit level) occupational group within the group of Professionals. The results (see Table 4) confirm our intuition. We can find highly negative values of the estimated parameter for the Engineering Professionals (group 214), Electrotechnology Engineers (215), and Software and Applications Developers and Analysts (251). As expected, there are some groups where the analysed parameter is positive: Medical Doctors and Dentists or Teachers, so occupations where knowledge and professional qualifications are increasing in time. The increasing share of the former specialists in the total number of employed in the second major occupational group would explain the negative sign of the estimated parameter. To some extent, this explanation can be confirmed by findings of other authors. Using detailed matched employer-employee data, Lallemand and Rycx (2009) find not only that young workers are significantly more productive than older workers, but also their results show that age structure effects on productivity are much stronger in ICT than in non-ICT firms.

Negative value of the parameter by the share\_5565 variable is observed also for Technicians and associate professionals (3<sup>rd</sup> major occupational group), Clerical support workers (4<sup>th</sup> major occupational group), Service and sales workers (5<sup>th</sup> major largest occupational group), and Elementary occupations (9<sup>th</sup> major occupational group). The two latter (5<sup>th</sup> and 9<sup>th</sup>) are the groups with mostly labour-intensive occupations where no constant improvement of qualifications is not needed and where formal qualifications are not as important as physical strength. For these groups the parameter by the share of workers aged 55-65 years old is relatively high (-0.63 and -0.51 respectively) and confirms the hypothesis that aging of the workforce in these occupations has a negative effect on individual' wages.

There are also two groups where the relation between the proportion of older workers and wages of individuals is statistically significant and positive. There are: Managers (1<sup>st</sup> major occupational group) with the value of the parameter at 2.3 and Craft and related trades workers (7<sup>th</sup> major occupational group) with the value of parameter at 1.35. The latter group is the group of qualified blue-collar workers. It means that even if the employees do not have to poses higher then basic vocational or secondary vocational education, the need to improve the qualifications desired to perform the specific job is required. More experienced, older workers should therefore be paid

more than the younger ones. The group of Managers is a very specific one since the observed level of education among members of this group is very diversified. However, more experienced managers are expected to receive higher wages than less experienced ones. Our findings indicate that in this group a 1 pp higher share of older workers leads to higher by 2.3% wages of managers in a given occupational group.

The parameters by other variables are confirming significant differences between occupational groups. Level of education matters the most for wages of Managers, Professionals and Technicians (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> major occupational groups). On the other hand, among workers in Elementary occupations (9<sup>th</sup> group) the level of education only slightly impacts the level of individual' wages.

Similar conclusions are noted while analysing the relation between the individual' work experience and wages. The lowest values of the parameters are observed in the 5<sup>th</sup> and 9<sup>th</sup> major occupational groups. The highest among Professionals and Managers.

Gender differences are also visible between occupational groups. The highest gender wage gap is observed among qualified blue-collar workers (7<sup>th</sup> and 8<sup>th</sup> major occupational group) and among Managers. The lowest - among clerical workers (4<sup>th</sup> group). The relation between other variables in the model (working time, type of employment contract, size of the employer, ownership sector, and the type of prevailing wage negotiations in the company) are also diversified between major occupational groups (see Table 4).

## 6 Conclusions and discussion

The aim of the paper was to analyse the relationship between the age structure of employees and wages. We used data on individual wages from the Structure of Earnings Survey in Poland to apply to an extended version of Mincerian wage model. The research period covers the years 2006-2014.

We analyse not only the relation between wages and different workers' and employers' characteristics, but also the relation between age structure of the employment in a worker's occupational group and his individual wage level. Our main variable of interest is the proportion of workers aged 55-65 years old in a 3-digit level occupational groups.

The results indicate that the proportion of aged people in the specific 3-digit occupational group has a significant relation with employees' individual earnings. However, the magnitude and direction of this relation depends on the membership to a specific major occupational group.

The results for the whole sample indicate a significant and negative relationship between the wages and the proportion of 55-65 years old workers. However, when the analyses were performed separately for each of the 1-digit occupational groups, the results varied significantly. In those groups, where undoubtedly one type of skills

Table 4: Estimated values of the relation between share of aged workers and individual wages across minor (3-digit level) occupational groups

Number of the group	Name of the group	Estimated value of parameter	Standard errors	R <sup>2</sup>	Number of observations
211	Physical and Earth Science Professionals	-8.548***	(1.033)	0.364	5,278
212	Mathematicians, Actuaries and Statisticians	4.918***	(0.686)	0.546	1,568
213	Life Science Professionals	200.278***	(12.558)	0.437	20,833
214	Engineering Professionals (excluding Electrotechnology)	-17.684***	(1.065)	0.327	108,815
215	Electrotechnology Engineers	-21.076***	(2.663)	0.310	21,074
216	Architects, Planners, Surveyors and Designers	-10.693***	(1.323)	0.272	16,653
221	Medical Doctors	6.048***	(2.093)	0.422	45,543
222	Nursing Professionals	1.618***	(0.326)	0.467	123,151
223	Midwifery Professionals	2.857***	(0.949)	0.469	12,838
224	Traditional and Complementary Medicine Professionals	-2.346	(1.842)	0.611	3,333
225	Veterinarians	3.588**	(1.789)	0.480	1,283
226	Dentists	21.915	(24.088)	0.404	2,640
227	Other Health Professionals	-31.331***	(3.585)	0.388	9,935
231	University and Higher Education Teachers	-7.310	(4.830)	0.303	51,330

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4 (cont.): Estimated values of the relation between share of aged workers and individual wages across minor (3-digit level) occupational groups

Number of the group	Name of the group	Estimated value of parameter	Standard errors	R <sup>2</sup>	Number of observations
232	Vocational Education Teachers	5.334***	(0.598)	0.642	64,400
233	Secondary Education Teachers	36.533***	(2.140)	0.618	124,990
234	Primary School and Early Childhood Teachers	10.411***	(1.285)	0.605	133,477
235	Other Teaching Professionals	1.688	(1.968)	0.346	138,813
241	Finance Professionals	6.465***	(0.658)	0.294	176,841
242	Administration Professionals	5.009***	(0.267)	0.313	199,175
243	Sales, marketing and Public Relations Professionals	-7.539***	(0.579)	0.275	174,939
244	Real Estate Professionals	4.048***	(0.834)	0.276	2,340
251	Software and Applications Developers and Analysts	-36.733***	(5.921)	0.296	49,909
252	Database and Network Professionals	95.212***	(13.276)	0.333	34,657
261	Legal Professionals	-26.793***	(4.902)	0.367	21,732
262	Librarians, Archivists and Curators	6.284***	(0.550)	0.368	12,054
263	Social and Religious Professionals	4.662***	(0.565)	0.360	18,880
264	Authors, Journalists and Linguists	3.348	(2.591)	0.109	7,875

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

(mental or physical) is preferred, empirical model confirms our assumptions that come from economic theory. For example, in the case of Managers where knowledge and qualifications of employees have to be updated permanently, an increase in the number of older workers significantly raises average wages. On the contrary, in those occupational groups where only basic skills are required or there is no need for regular in-the-job investments (Elementary occupations, Service and sales workers), aging of the workforce has a negative effect on the average wages. However, in those groups of blue collar workers where good vocational skills are required (Craft and related trades workers and Plant and machine operators and assemblers) the positive sign of the relationship between share of older workers and wages was obtained.

Slightly different were the empirical results in the 2<sup>nd</sup> major occupational group – Professionals – where the sign is statistically significant but strongly negative. In the authors' opinion it is the result of high heterogeneity of skills inside this large group. Nevertheless, the results obtained in the paper indicate that population and consequently workforce aging have important implications on the level of wages in the economy. The results of the study show that the changing age structure of workers affects their wages. Simultaneously, one has to bear in mind that population aging lead also to changes in the demand for work. The bigger the percentage of oldest people in the population, the greater becomes the need for services in the silver economy. Demographic changes in the labour market may occur relatively slow. But they have consequences for the whole economy are therefore should not be disregarded.

## Acknowledgements

The paper was financed by National Science Center grant 2015/19/B/HS4/03231. We are thankful to the anonymous referee for their valuable comments. We are also very thankful to Paweł Strawiński, for his very valuable help in creating the dataset and many valuable comments and suggestions.

## References

- [1] Aubert P., Crépon B., (2003), Age, Wage and Productivity: Does Productivity Decline at the End of Workers Career?, *Document de Travail de la DESE – Working Papers of the DESE*, 2003-06, Institut National de la Statistique et des Etudes Economiques.
- [2] Bazen S., Charni K., (2015), Do Earnings Really Decline for Older Workers?, *AMSE Working Papers* 1511, Aix-Marseille School of Economics.
- [3] Bertschek I., Meyer J., (2009), Do Older Workers Lower IT-Enabled Productivity? Firm-Level Evidence from Germany, *Jahrbücher für Nationalökonomie und Statistik* 229(2/3), 327–342.

- [4] Börsch-Supan A., Düzgün I., Weiss M., (2005), Altern und Produktivität: Zum Stand der Forschung, *MEA Discussion Paper* 73–2005, Mannheim.
- [5] Börsch-Supan A., (2003), Labor Market Effects of Population Aging, *LABOUR* 17, Special Issue, 5–44.
- [6] Cardoso A., Guimaraes P., Varejao J., (2010), Are Older Workers Worthy of Their Pay? An Empirical Investigation of Age-Productivity and Age-Wage Nexuses, *De Economist* 156 (2), 95–112.
- [7] Cataldi A., Kampelmann S., Rycx F., (2011), Productivity-wage Gaps among Age Groups: Does the ICT Environment Matter? *De Economist* 159(2), 193–221.
- [8] Daveri F., Maliranta M., (2007), Age, Seniority and Labour Costs, *Economic Policy* 49, 118–175.
- [9] Dostie B., (2006), Wages, Productivity and Aging, *IZA Discussion Paper* 2496, Bonn.
- [10] Göbel Ch., Zwick T., (2009), Age and Productivity – Evidence from Linked Employer Employee Data, *ZEW Discussion Paper* 09-020.
- [11] Hellerstein J. K., Neumark D., (1995), Are Earnings Profiles Steeper Than Productivity Profiles? Evidence from Israeli Firm-Level Data, *Journal of Human Resources* 30(1), 89–112.
- [12] Hellerstein J. K., Neumark D., Troske K. R., (1999), Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations, *Journal of Labor Economics* 17(3), 409–446.
- [13] Hurd M., (1971), Changes in Wage Rates Between 1959 and 1967, *Review of Economics and Statistics* 53(2), 189–199.
- [14] Kielczewska A., Lewandowski P., (2017), Population Ageing and the Labour Supply in Poland up to 2050, [in:] *Population Ageing, Labour Market and Public Finance in Poland*, [ed.:] Lewandowski P., Rutkowski J, IBS, European Commission 2017.
- [15] Lallemand T., Rycx F., (2009), Are Young and Old Workers Harmful for Firm Productivity?, *IZA Discussion Paper* 3938, Bonn.
- [16] Lis M., Magda I., (2014), Dynamika płac w cyklu życia a indywidualny stan zdrowia (Life cycle wage dynamics and individual health), *Gospodarka Narodowa* 4(272).
- [17] Lovász A., Rigó M., (2013), Vintage Effects, Aging and Productivity, *Labour Economics*, Elsevier, 22(C), 47–60.

- 
- [18] Majchrowska A., Roszkowska S., (2014), Premia z wykształcenia i doświadczenia zawodowego według płci w Polsce, *Narodowy Bank Polski, Materiały i Studia* 302.
- [19] Mahlberg B., Lindh T., Halvarsson M., (2008), Productivity Consequences of Workforce Ageing: Stagnation or Horndal Effect? [in:] *Population Aging, Human Capital Accumulation and Productivity Growth, Population and Development Review*, [ed.:] Prskawetz A., Bloom D., Lutz W., Supplement to Vol. 34, 238–256.
- [20] Mahlberg B., Freund I., Crespo Cuaresma J., Prskawetz A., (2013), Ageing, Productivity and Wages in Austria, *Labour Economics* 22, 5–15.
- [21] Mincer J., (1974), *Schooling, Experience and Earnings*. National Bureau of Economic Research.
- [22] Medoff K., Abraham K. G., (1980), Experience, Performance, and Earnings, *Quarterly Journal of Economics* 94, 703–736.
- [23] Mitchell O. S., Levine P. B., Pozzebbon S., (1990), Aging, Job Satisfaction, and Job Performance, *CAHRS Working Paper #90-02*.
- [24] Myck M., (2007), Wages and Ageing: Is There Evidence for the Inverse-U-Profile? *IZA Discussion Paper* 2983, Bonn.
- [25] Ng T., Feldman D., (2008), The Relationship of Age to Ten Dimensions of Job Performance, *Journal of Applied Psychology* 93(2), 392–423.
- [26] OECD, (2017), Old-age dependency ratio, in *Pensions at a Glance 2017: OECD and G20 Indicators*, OECD Publishing, Paris.
- [27] Skirbekk V., (2008), Age and Productivity Capacity: Descriptions, Causes and Policy Options, *Ageing Horizons* 8, 4–12.
- [28] Strawiński P., (2015), Krzyżowe porównanie danych o wynagrodzeniach z polskich badań przekrojowych, *Bank i Kredyt* 46(5), 433–462.
- [29] Sturman M. C., (2003), Searching for the Inverted U-shaped Relationship Between Time and Performance: Meta-analyses of the Experience/Performance, Tenure/Performance, and Age/Performance Relationships, *Journal of Management* 29(5), 290–316.
- [30] Van Ours J. C., (2009), Will You Still Need Me – When I’m 64? *CentER Discussion Paper* 09-51, Tilburg University.
- [31] Van Ours J. C., Stoeldraijer L., (2011), Age, Wage and Productivity in Dutch Manufacturing, *De Economist*, Springer, 159(2), 113–137.

- [32] Veen S., (2008), Demographischer Wandel, alternde Belegschaften Produktivitätseffekte, Rainer Hampp Verlag, Munich.
- [33] Zgliczyński W., (2013), Wynagrodzenia w Polsce, *INFOS BAS* 22(159).

## A

Table 5: Share of workers aged 55-65 years old in major (1-digit) occupational level in Poland (on average in 2006-2014)

Name of the group	Number of the major occupational group*	Share of workers aged 55-65 years old (from the highest to the lowest)
Services and Sales Workers	5	0.078
Clerical Support Workers	4	0.089
Professionals	2	0.095
Plant and Machine Operators and Assemblers	8	0.108
Technicians and Associate Professionals	3	0.109
Craft and Related Trades Workers	7	0.110
Chief Executives, Senior Officials and Legislators	1	0.166
Elementary Occupations	9	0.173

\*According to the ISCO-08 Classification.

Table 6: Top five sub-major (2-digit) occupational groups with the lowest and the highest share of workers aged 55-65 years old in Poland (on average in 2006-2014)

Name of the sub-major (2-digit) group	Number of the sub-major group*	Share of workers
Models, Salespersons and Demonstrators	52	0.031
Other Craft and related Trades workers	74	0.057
Tellers, Money Collectors and related Clerks	42	0.063
Handicraft and printing workers	73	0.068
Plant and machine operators, and assemblers	82	0.076
Managers of big and medium sized firms	12	0.163
Managers of small sized firms	13	0.170
Agricultural, forestry and fishery labourer	92	0.184
Elementary Occupations	91	0.211
Chief Executives, Senior Officials and Legislators	11	0.305

\*According to the ISCO-08 Classification.

Table 7: Top five minor (3-digit) occupational groups with the lowest and the highest share of older workers (55-65 years old) in Poland (on average in 2006-2014)

Name of the minor (3-digit) group	Number of the minor group*	Share of older workers
Operators of manufacturing machines	817	0.020
Fashion and other models	521	0.025
Other sales workers	522	0.031
Mining and mineral processing plant operators	711	0.038
Primary school teachers	233	0.041
Protective services workers	515	0.258
Messengers, package deliverers and luggage porters	915	0.287
Legislators and senior officials	111	0.293
Regulatory government associate professionals	345	0.317
Senior officials	112	0.469

\*According to the ISCO-08 Classification.

## B Rearrangement of the old ISCO classification into the new one

To rearrange the old classification into the new one, an associative array that maps keys to values was used. More precisely, an association list was used which is a linked list in which each list element is comprised of a key and value. An association list where unique four digits COS group codes from 2010 are keys and matching groups from 2008 are values was constructed. There are around 500 unique four-digit COS group codes, so searching an association list is not less preferable than searching a binary search tree or hash table. We were able to build the association list because of existence of a transition table provided by CSO. Transition table convey information needed for matching different COS coding in time.

The first association list – transitions:

```
> head(to_2008)
$ '1111'
[1] "1111"
$ '1112'
[1] "1112"
$ '1113'
[1] "1112"
$ '1114'
[1] "1121" "1122" "1123"
$ '1120'
[1] "1211" "1212"
```

```
$ '1311'  
[1] "1221" "1311"
```

For the surveys published after 2008, each observation was replicated in accordance with the COS association list. However, each new observation has some probability of existence because it can be assigned to several groups. Therefore, we built a next association list, which provides probabilities of attendance in each group. Probabilities were calculated using frequencies of workers in each group in the previous survey. It should be obvious that for each observation that was replicated, the probabilities have to add up to one. We also made some assumptions about cases where some occupational groups were lacking in previous surveys.

The second association list – probabilities:

```
> head(kzis_apply_freq(to_2008, summ_bsw_2008))  
$ '1111'  
[1] 1  
$ '1112'  
[1] 1  
$ '1113'  
[1] 1  
$ '1114'  
[1] 0.3333333 0.3333333 0.3333333  
$ '1120'  
[1] 0.8343979 0.1656021  
$ '1311'  
[1] 0.95471934 0.04528066
```

Finally, we get the dataset that has additional rows from a replication process and three supplementary variables – new group, probability, and number of replications. This dataset was built from five SWS surveys published biannually between 2006 and 2014. It contains over 5 million observations and 70 variables.