DISCUSSION PAPER SERIES

IZA

IZA DP No. 2456

Employment Fluctuations and Dynamics of the Aggregate Average Wage in Poland 1996-2003

Michal Myck
Leszek Morawski
Jerzy Mycielski

November 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Employment Fluctuations and Dynamics of the Aggregate Average Wage in Poland 1996-2003

Michal Myck
DIW Berlin and IZA Bonn

Leszek Morawski
Warsaw University

Jerzy Mycielski
Warsaw University

Discussion Paper No. 2456
November 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 2456
November 2006

ABSTRACT

Employment Fluctuations and Dynamics of the Aggregate Average Wage in Poland 1996-2003*

The aggregate average wage is often used as an indicator of economic performance and welfare, and as such often serves as a benchmark for changes in the generosity of public transfers and for wage negotiations. Yet if economies experience a high degree of (nonrandom) fluctuation in employment the composition of the employed population will have a considerable effect on the computed average. In this paper we demonstrate the extent of this problem using data for Poland for the period 1996-2003. During these years employment in Poland fell from 51.2% to 44.2% and most of it occurred between the end of 1998 and the end of 2002. We show that about a quarter of the growth in the average wage during this period could be attributed purely to changes in employment.

JEL Classification: E24, J21, J31

Keywords: wage distribution, aggregation, employment dynamics, transition economies

Corresponding author:

Michal Myck
DIW Berlin
Königin-Luise-Straße 5
14195 Berlin
Germany
E-mail: mmyck@diw.de

* This paper presents some of the results from the research project on "Aggregate wage dynamics in Poland". Financial support from the Polish Scientific Research Committee (Komitet Badań Naukowych) under grant number 2H02C06924 is gratefully acknowledged. Michał Myck would like to thank for financial support through the REVISER project, an RTN project financed by the European Commission (contract no. HPRN-CT-2002-00330). We would like to thank Richard Blundell, Amanda Gosling and Howard Reed from the IFS for their advice in the early stages of the project's development. This paper is a revised version of the DIW Discussion Paper No. 545. The usual disclaimer applies.

1. Introduction

The aggregate average wage is often understood and referred to as an indication of the performance of the economy and its dynamics as a reflection of changes in society's welfare. As such it is given a lot of public attention. As a result it is also used as a reference value for determining values of various fiscal and social policy instruments.[2] It is taken as given that an increase in the average wage is a sign of positive developments in the economy, while it's stagnation reflects a general slowdown in economic development.

It has been recognised for a long time in economic literature that the reported dynamics of aggregate wages may not necessarily play the role it is commonly assigned, and the problem of meaningful aggregation of wages may be more complicated than it is usually perceived (Bills, 1985, Solon et al., 1994, Gossling et al., 2000, Meghir and Whitehouse, 1996).[3] The main issues complicating the interpretation of the aggregate wage as a simple indicator of welfare are:

- the fact that selection into and out of the sample of employees is not random,
- the structure of the employed population changes over time,
- the structure of the wage distribution may change over time.

Each of these factors may on its own affect changes in the observed aggregate wage without any change in individual wages (conditional on characteristics), i.e. without any changes in welfare of employed individuals for whom we observe the wage. In a scenario where there are job losses which are concentrated among people with low human capital, the aggregate average wage will grow without increases in individual wages, and in fact in situation when aggregate welfare is likely to be falling. Similarly if there is an increase in the proportion of young people among the working population the aggregate wage would fall, even if individual wages conditional on characteristics remained unchanged. Along these lines Blundell et al. (2003) demonstrated that when corrected for these factors aggregate wage dynamics behave

[2] For example in Poland the level of income up to which national insurance contributions are paid is 30 times the average gross monthly wage from the previous year. Moreover, the computation of retirement and disability pension entitlements for those who become pensioners relates their earnings and contributions to average monthly gross wages. National insurance contributions paid by the self-employed also depend on average monthly gross wages. See Zdanowicz (2003).

[3] For an excellent survey on aggregation issues in economics see Blundell and Stoker (2005).

significantly differently from the simple average wage calculated for the employed population. While the measured aggregate wage in the UK over the early 1990's rises, the individual wages appear to be essentially flat. Clearly the interpretation of changes in the aggregate wage will be most difficult when the three forces complicating this interpretation undergo important changes. This will therefore apply especially strongly to countries with significant fluctuations in the rate of employment and more broadly to economies which undergo a rapid structural and institutional change.

Table 1. Employment, unemployment and wages in Poland 1996-2003 (winter quarters).

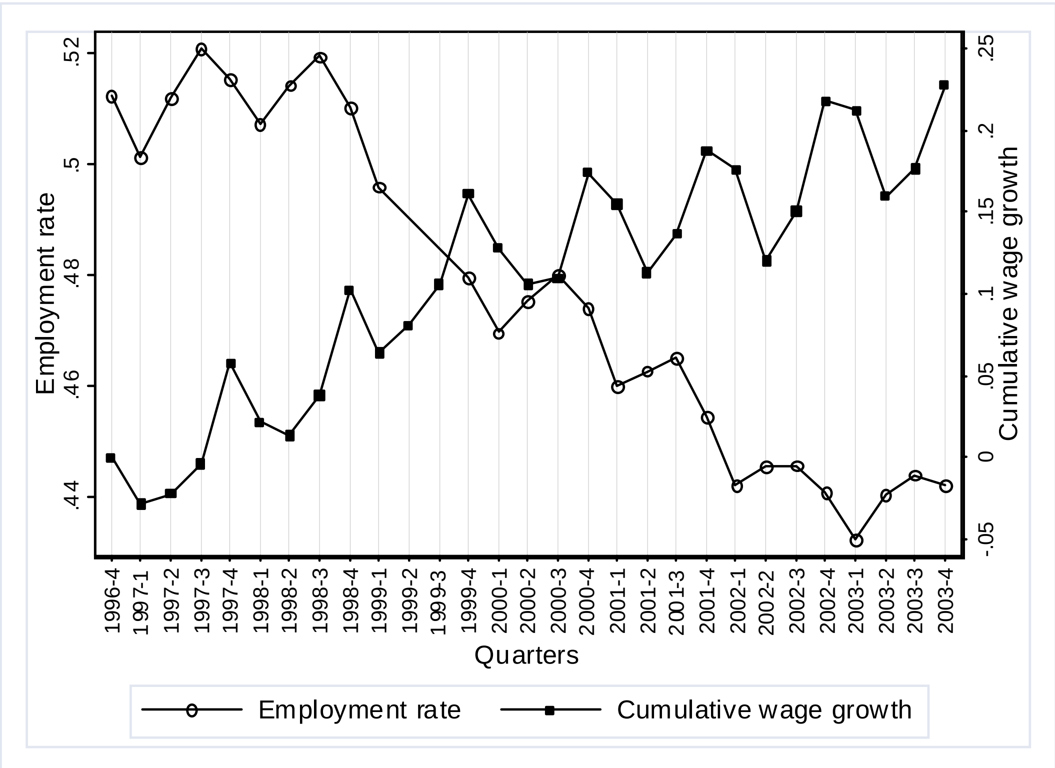
	1996	1997	1998	1999	2000	2001	2002	2003
EMPLOYMENT								
All	0.512	0.515	0.510	0.480	0.474	0.455	0.441	0.442
Men	0.594	0.598	0.589	0.559	0.552	0.525	0.507	0.509
- aged <30	0.496	0.506	0.492	0.466	0.455	0.417	0.394	0.400
- aged 30-39	0.855	0.871	0.873	0.846	0.833	0.808	0.797	0.810
- aged 40-49	0.814	0.813	0.809	0.780	0.774	0.754	0.740	0.735
- aged 50-59	0.595	0.599	0.602	0.568	0.580	0.554	0.534	0.537
- aged 60+	0.202	0.199	0.181	0.161	0.158	0.146	0.135	0.129
Women	0.438	0.440	0.439	0.407	0.403	0.391	0.381	0.382
- aged <30	0.354	0.365	0.373	0.345	0.338	0.322	0.307	0.315
- aged 30-39	0.699	0.699	0.699	0.683	0.671	0.658	0.654	0.651
- aged 40-49	0.734	0.733	0.727	0.683	0.687	0.667	0.670	0.674
- aged 50-59	0.426	0.438	0.452	0.404	0.413	0.420	0.396	0.386
- aged 60+	0.103	0.097	0.092	0.071	0.072	0.064	0.062	0.070
UNEMPLOYMENT								
Registered	0.132	0.103	0.104	0.131	0.151	*0.194*	*0.200*	*0.200*
ILO	0.112	0.099	0.103	0.154	0.160	0.185	0.198	0.193
AGGREGATE WAGE	100	105.8	110.2	116.1	117.4	118.8	121.7	122.7

Notes: employment rate calculated for population aged 15+, ILO unemployment for population aged 15-74, while registered unemployment for population aged 18-60(women)/65(men). Figures for registered unemployment for years 2001-2003 (in italics) are not comparable with earlier ones due to changes in methodology by GUS (the Polish Central Statistical Office).
Source: authors' computations using BAEL data, GUS official statistics.

In this paper we present an illustration of one of the factors driving the aggregate wage dynamics with an application to Polish data. The structure of the employed population in Poland has undergone a significant change since mid-1990s and we postulate that this has had an important effect on the dynamics of the aggregate wage measure. Our analysis uses the framework of a semi-parametric procedure of

DiNardo, Fortain and Lemieux (1996), and introduces employment probability weights to account for changes in employment patterns in the period between the winter quarters of 1996 and 2003. DiNardo, Fortain and Lemieux (1996) (from now on referred to as DFL) present a method which allows counterfactual analysis of changes in the distribution of wages. It facilitates the analysis of the effect of changes in the characteristics of those employed and in institutional features of the labour market assuming all else remained constant. They explicitly consider populations of employed individuals and the issue of selection into and out of employment is not accounted for. As among others Blundell et al. (2003) have demonstrated, the issue of selection is however non-trivial. If movements in and out of employment are non-random with respect to individual wages then fluctuations in employment will be reflected in the distribution of wages and in the dynamics of the measured aggregate average wage.

Figure 1. Employment and growth of real aggregate average wage in Poland: Q4-1996 – Q4-2003



Source: BAEL data sets, GUS official statistics.
Notes: Growth rate of real earnings between Q4-1998 and Q1-1999 assumed the same as a year earlier. Wage growth computed on the basis of CPI-indexed nominal wages.[4]

[4] The definition of the “gross wage” changed in January 1999 in Poland with the introduction of the pension reform. The presented figure is based on separate rates of growth of the aggregate average wage for the period before January 1999 and for the period after. By assuming the rate of growth

The case of Poland is a good example for such an analysis as over a relatively short period of time employment levels dropped significantly and there has been a lot of variation in employment dynamics among various groups of the population. We focus on years 1996 to 2003 with special attention being paid to four years in the middle of this period, i.e. 1998-2002. These four years saw a dramatic fall in the level of employment for men and women and for all age groups. As Table 1 shows between the fourth quarter of 1998 and fourth quarter of 2002 the overall employment rate fell from 51.0% to 44.1%. Men aged 30-39 saw their employment rate fall from 87.3% to 79.7%, while those below 30 from 49.2% to 39.4%. Among women the falls have been smaller, but still very substantial with the overall female employment rate falling from 43.9% to 38.1%. By the end of 2003 employment rate for men aged 40-49 was 73.5%, and for those aged 50-59 53.7%, down from respectively 81.4% and 59.5% in the fourth quarter of 1996.

At the same time the published figures for the aggregate average wage showed an increase of about 23% between end of 1996 and end of 2003 (see Figure 1).[5] Bearing in mind the focus of this paper it is worth noting that a large proportion of the change took place between the end of 1998 and 2002, i.e. at the time of slow growth and substantial falls in employment.

The brief spell of rapid economic growth at the beginning of the analysed period, which lasted for about a year and a half, did not find a reflection in growing employment rates even though there were important reductions in unemployment rates (ILO definition from 11.2% to 9.9% and registered unemployment from 13.2% to 10.3% between the fourth quarters of 1996 and 1997). During the years of rapid economic growth the official aggregate average wage shows significant gains, growing by 5.8% from Q4-1996 to Q4-1997 and reflecting a cumulative growth of

---

between Q4-1998 and Q1-1999 to be the same as that between Q4-1997 and Q1-1998, we can construct a cumulative wage growth over the entire period we examine.

[5] The published information on the aggregate average wage in Poland is based on monthly surveys conducted by the Central Statistical Office. They cover all financial and non-financial enterprises employing at least 50 people and a representative sample of 10% of firms with from 10 up to 50 employees. Wages in very small firms are estimated on the basis of trends from previous years. Average wages are calculated as an arithmetic mean with total wage fund divided by the number of employed people which is in some cases adjusted for number of working hours.

10.2% for the period Q4-1996 to Q4-1998. Detailed quarterly changes in employment and the average wage are presented in Figure 1.

We begin the paper with a presentation of the methodology we apply for the exercise (Section 2) and then a brief description of data sets used in the analysis (Section 3) Because of notorious bad quality of wage data in the Polish labour force survey (BAEL) we rely on the Autumn Earnings Survey (AES) for wage information. This has several advantages, principally that there is no missing information on wages and that wage data is given in terms of gross and not net wages (as in BAEL data). Section 4 presents the constructed ‘base’ wage distribution, while in section 5 we show results of aggregate wage dynamics for the period from the fourth quarter of 1996 to the fourth quarter 2003 under the assumption of constant base wage distribution. Conclusions follow in Section 6.

2. Methodology

Our aim is to illustrate the effect of selection in and out of employment in Poland on the dynamics of the ‘average wage’ in a straightforward but realistic fashion. As we saw in Table 1 and in Figure 1 there was significant variation in employment levels which was observed in Poland at the turn of the century. Reductions in employment levels have most likely been non-random with respect to the distribution of wages and so must have had an effect on the observed wage distribution and the calculated aggregate average wage.

The latter is usually calculated simply as:

$$\varpi_t = \frac{\sum_{i=1}^{I_t} w_{it}}{I_t} \tag{1}$$

where: where w_{it} is the observed wage of individual (*i*) at time (*t*) and I_t is the sample of people employed at time (*t*) (i.e. people for whom we observe a wage) or their sub-sample.

The formula obviously relates only to the wages of those who are employed and is calculated as a simple arithmetic average (usually weighted to account for intensity of employment). This means that although it may be informative of the wage level at a particular point in time, its changes may be difficult to interpret, especially if the population of employees (I_t) changes between (t) and ($t+s$). Because changes in the population of employees, driven by the economic cycle, demand for labour and by individual labour supply decisions, are most likely non-random, at different points in time people from different sections of the wage distribution will leave the sample or join it. The analysis of changes in the average wage, and in fact any analysis which attempts to make welfare assessments on the basis of the observed wage distribution, should therefore take into account the changing composition of the employee population. At different points in the economic cycle people may be fired or hired and/or decide to leave employment or take up a job. This of course presents a difficulty because wages for the non-employed population are not observed. The average wage continues to be calculated in the same way, although the sample I_t changes.

The method presented by DFL allows to recreate the wage distribution under different scenarios by generating counterfactual scenarios along the lines of "how would the density of wages look like at time *(t)* if characteristics of the population remained as at time *(t-s)*". In our case, however, we are not interested in the same sort of counterfactual, since we do not want to analyse how the mean wage would have changed had characteristics of the labour force remained the same and demand affected only the density of wages conditional on individual attributes. This approach implicitly assumes the selection issue away, and we argue that what was observed in Poland (and what often is observed in transition economies) is a significant and semi-permanent change in labour market conditions which leads not only to changes in relative returns to characteristics but also to changes in relative probabilities of having a wage observed, i.e. being employed. So our counterfactual is not "how would the density of wages look like in 2003 if characteristics of the population remained as in the end of 1996", since this counterfactual refers only to the employed population. What we want to show is the effect of changes in the levels of employment on the density of wages and on the aggregate average and thus we need to consider the entire population and not only its employed sub-sample.

Following the notation of DFL we can write the density of wages at a point in time $f_t(w)$ as an integral of the distribution of wages and characteristics at time *t*:

$$f_t(w) = \int_{z \in \Omega_z} dF(w, z \mid t_{w,z} = t) \qquad (2)$$

where as in DFL Ω_z is the domain of definition of the individual attributes.

One could analyse the dynamics of this distribution in time only if either the entire working age population is employed throughout or if entry and exit in and out of employment are independent of characteristics z. The first clearly does not hold and the second has been shown not to hold in several studies (e.g. Blundell et al. (2003), Myck and Reed (2006)). As employment figures for Poland show (Table 1 and Figure 1) it can be safely assumed not to hold in the Polish case as well. Below we show how one could extend the DFL methodology to account for the selection issue.

As in DFL, we can approximate the density of wages by a kernel density estimate $\hat{f}_h$ based on a random sample of employed people $W_1, \ldots, W_n$ of size n, with weights $\theta_1, \ldots, \theta_n$, $\sum_i \theta_i = 1$:

$$\hat{f}_h(w) = \sum_{i=1}^{n} \frac{\theta_i}{h} K\left(\frac{w - W_i}{h}\right) \qquad (3)$$

where h is the bandwidth and K(.) is the kernel function.

If we knew W_i for the entire sample then the kernel density could be estimated for the entire sample even if only some of these wages were to enter the density (since only some individuals are employed). We would then have:

$$\hat{f}_h(w) = \sum_{i=1}^{n} \frac{\gamma_i}{h} K\left(\frac{w - W_i}{h}\right) \qquad (4)$$

where: $\gamma_1,...,\gamma_n$, $\sum_i \gamma_i = 1$ and where:

$\gamma_i = \theta_i$ if $e_i = 1$

$\gamma_i = 0$ if $e_i = 0$

where e_i is a binary indicator of employment and takes value 1 if an individual is employed and value 0 otherwise.

We can use equation (4) to generalise the problem by substituting the binary employment indicator with individual employment probability weights. If we can assign an individual probability of being observed as employed $p_i(e)$ and an individual wage W_i to all individuals in the sample, then we can write the kernel density as:

$$\hat{f}_h(w) = \sum_{i=1}^{n} \frac{\gamma_i(p_i(e), \phi_i)}{h} K\left(\frac{w - W_i}{h}\right) \tag{5}$$

where: $\gamma_1,...,\gamma_n$, $\sum_i \gamma_i = 1$ and γ_i is now written as a function of the probability of being employed $p_i(e)$ and a sample weight ϕ_i.[6] Such reweighing of the distribution, by keeping the underlying wage distribution constant and just changing the probability weights will allow us to measure the extent of changes in the distribution and in its characteristics. More specifically it will facilitate measuring the extent of the effect of changes in employment probabilities on the dynamics of the aggregate average wage. For each period *(s)* equation (5) then becomes:

$$\hat{f}_{h,s}(w) = \sum_{i=1}^{n} \frac{\gamma_i(p_{i,s}(e), \phi_{i,t})}{h} K\left(\frac{w - W_{i,t}}{h}\right) \tag{6}$$

where the underlying wage distribution is constant over time, while employment probabilities change in time according to the labour market dynamics.

[6] Notice that assigning the employment probability weights is essentially the same as the approach taken by DFL to correct for changes in the number of hours worked among the employed population.

3. Data

We use Polish micro-level data from two surveys: the labour force survey – BAEL, and the Autumn Earnings Survey (AES). BAEL is a representative individual level survey (a rolling panel) collected quarterly with a principle focus on labour market status. Each quarter the survey collects information on about 50,000 individuals aged 15 and over. The AES is an annual survey (collected usually in September) which collects data on approximately 700,000 individuals at the company level and focuses on earnings information. The reason for using this joint set up is because wage information in BAEL is generally unsatisfactory.[7] At the same time the AES collects information only on employees and can't be used to analyse changes in participation patterns. Therefore we use the earnings information from the AES data and take advantage of the detailed labour market information from BAEL to study labour market dynamics.

The detailed information on individual wages in the AES is used to generate a wage distribution for the BAEL sample. This distribution is then employed to demonstrate changes in the 'average wage' given the observed trends on the Polish labour market in the period from the last quarter of 1996 to the last quarter of 2003.

As we pointed out in the introduction the analysis is based on the combination of the AES and BAEL data. We use AES collected in September 1996 and on the basis of this dataset generate the expected (gross) wage distribution for the BAEL sample collected in Autumn 1996 (referred to as the 'BAEL base sample' below). Following this exercise we estimate employment probabilities in the BAEL datasets over the period from Q4-1996 to Q4-2003 and use these to generate expected employment probabilities in 'future' BAEL years, which are calculated for the BAEL base sample.[8]

---

[7] As we show in Mycielski et al. (2005) the wage information in BAEL is substantially incomplete. Moreover the AES collects information on gross and not on net wages (as is the case in BAEL). This makes it more comparable with the official average wage statistics.

[8] The BAEL data has seen a significant development and some important changes in the period covered in this paper, however the key variables used in our analysis can be constructed consistently for the entire period. For detailed documentation on changes in the data see Morawski et al. (2005).

Where possible the same sample selection criteria are applied to the AES and BAEL datasets. The most important selection criteria are:

- in both samples we drop people aged less than 18 and over pension age (60 for women, 65 for men),
- we drop the self employed, those who help in family business and full time students (in BAEL only).

In Table 2 we present the basic descriptive statistics for the AES sample and the BAEL base sample (after applying selection criteria). The table also includes descriptive statistics for a sub-sample of the BAEL base sample including only those who are employed in firms employing more than 5 employees. This is the closest we can get to mimic the criteria applied to the creation of the AES sample. The Autumn Earnings Survey collects information only on employees employed in firms with more than 9 employees. We can see that as far as the proportion of higher educated and those with secondary education the selected BAEL sample and the AES sample are very similar. The proportion of men and those with vocational training among the BAEL employees is slightly higher than among the AES employees, and the proportion of those with primary education is somewhat lower. These differences lead to small differences in the earnings distributions generated for AES and for BAEL, but these are not substantial enough to affect our conclusion in any major way.

4. Computing expected wages and employment probabilities

Table 3 presents the results of the wage equation estimation run as truncated regression (on log monthly wage) on AES.[9] Log wages (of 657,613 individuals) are regressed on an age polynomial, education level dummies, region (49 pre-1999 voivodship), a male dummy and interactions of: the male dummy with regional dummies and with education dummies. The choice of these variables is constrained on the one hand by the availability of more information on individuals in the AES dataset, and on the other hand on the need to have the same variables in both the AES

---

[9] The truncation is made at the level of the National Minimum Wage (325PLN) on the left hand side and at the 99^{th} centile of the distribution (2706.80PLN) on the write hand side of the distribution. See figure A1 in the Appendix for the shape of the lower end of the distribution before the truncation.

and the BAEL base sample.[10] For presentation reasons Table 2 does not include the coefficients on the regional dummy variables and the interactions.[11]

Table 2. AES and BAEL sample - descriptive statistics

	AES – 1996[1]	BAEL – Q4-1996[2]	BAEL – Q4-1996[3] (AES selection)
Sample size	667,962	31,273	16,825
Proportion of men	51.29%	47.57%	54.96%
Education:			
- higher	15.72%	10.08%	14.83%
- secondary academic	6.90%	6.81%	6.49%
- secondary vocational	30.46%	27.38%	31.57%
- vocational	29.87%	35.26%	35.05%
- primary or none	17.04%	20.47%	12.05%
Age – men	39.07	40.35	38.03
Age - women	39.34	38.92	38.31

Source: authors' calculation on the basis of AES-1996 and BAEL-Q4-1996.
Notes: 1 - individuals employed in companies with more than 9 employees,
2 - employed and non-employed individuals,
3 - individuals employed in companies with more than 5 employees.

The results are not very surprising. Wages are higher for older and better educated people. The coefficient on the male dummy variable is positive, but men generally get lower returns to age and education.

This estimation is used to generate a distribution of expected wages in AES and in the BAEL base sample (shown on Figure 2).[12] Figure 2b – plotted for the BAEL base sample includes both the employed (in big and in small firms) as well as the non-employed people, while in figure 2c we present the distribution of expected wages in the BAEL sample only for those who are observed as employed. Unsurprisingly the distribution of wages for the employed and the non-employed sample (shown as kernel densities on Figure 2d) are substantially different. The average expected wage for men and women in AES is 862.03 and 708.85 PLN respectively. In the full BAEL base sample these numbers are: 835.90 and 645.75 PLN, while if we just take those who are employed in the fourth quarter of 1996 the expected wages are 843.10 PLN for men and 681.20 PLN for women, while if we limit the sample only to those

---

[10] For example we could not use the work experience information from AES as such information is not available in BAEL.
[11] A significant majority of those coefficients is statistically significant. The full set of results is available from the authors on request.
[12] Expected wages are computed as: $\hat{w}_i = \exp(l\hat{w}_i) * \exp(1/2\hat{\sigma}^2)$. See Blundell et al. (2003).

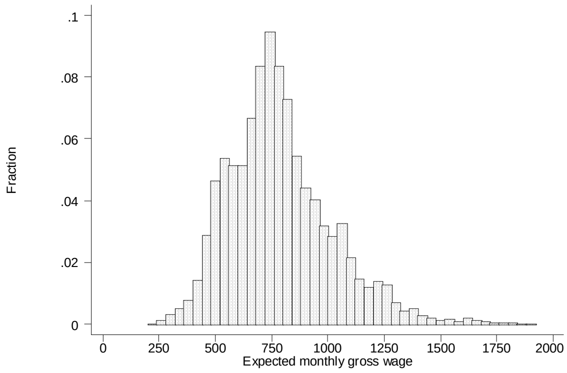
employed in companies employing more than 5 employees the average is very close to that in the AES sample: 868.73 PLN for men and 705.82 PLN for women.

Table 3. Summary results of truncated wage regression in AES, 1996

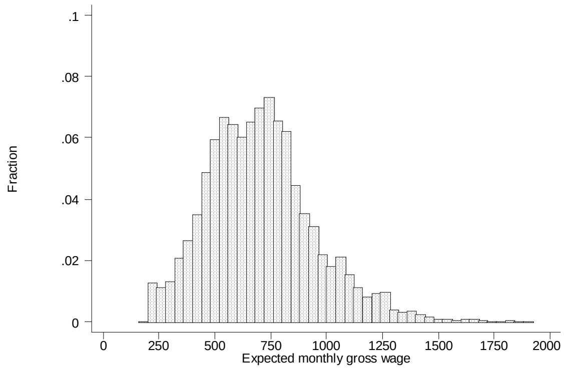
Dependent variable: log monthly gross wage	**Coeff.**	**St. error**	**Significance level**
Age	0.4823	(0.0207)	***
Age^2	-0.0164	(0.0008)	***
Age^3	0.0003	(0.0000)	***
Age^4	0.0000	(0.0000)	***
Education (base cat.: primary)			
- higher	0.6757	(0.0031)	***
- secondary academic	0.4025	(0.0034)	***
- secondary vocational	0.3890	(0.0028)	***
- vocational	0.0648	(0.0033)	***
Male dummy	0.9814	(0.2249)	***
Age*male dummy	-0.0812	(0.0243)	
Age^2*male dummy	0.0040	(0.0010)	***
Age^3*male dummy	-0.0001	(0.0000)	***
Age^4*male dummy	0.0000	(0.0000)	
Education * male dummy			
- higher * male dummy	-0.1801	(0.0041)	***
- secondary academic * male dummy	-0.1957	(0.0059)	***
- secondary vocational * male dummy	-0.1544	(0.0037)	***
- vocational * male dummy	0.0157	(0.0040)	***
Regional dummies		included	
Regional dummies*male dummy		included	
Sigma	0.4095	(0.0005)	***
Number of uncensored observations:		657613	
Number of censored observations:		10349	
Log likelihood		-243465,38	

Source: authors' calculations on the basis of AES 1996.
Notes: Observations truncated at the National Minimum Wage (325zl) and at the top centile of the wage distribution. *** - significant at 1%.

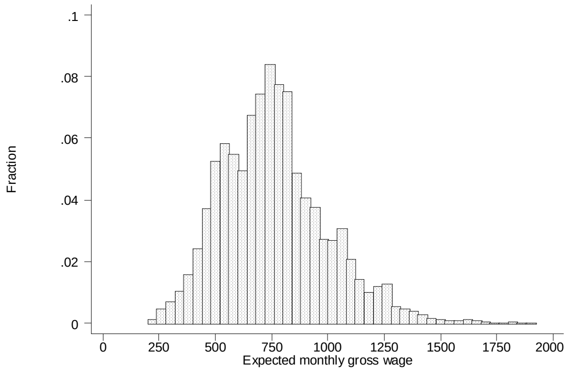
Figure 2. Expected gross monthly wage distributions – AES and BAEL base sample



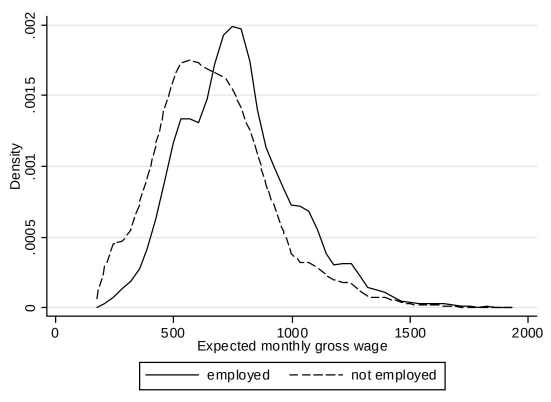
2a - Expected wage distribution in AES



2b - Expected wage distribution in the BAEL base sample



2c - Expected wage distribution in the BAEL base sample – those observed as employed



2d – Kernel densities of expected wage distribution in the BAEL base sample – employed and non-employed

Source: authors' calculations on the basis of BAEL Q4-1996 and AES 1996.

The next step in the methodology is the estimation of employment probability models for BAEL samples over the period Q4-1996 to Q4-2003. Table A2 in the appendix presents the subset of results from the employment probit models run on the winter quarters Q4-1996 to Q4-2003.[13] We thus calculate employment probabilities for people in the BAEL base sample *as if* the conditions they were subjected to were imported from future years. Employment probability changes significantly during the period covered and reflects the trends presented in Figure 1.

5. Employment, wage distribution and aggregate average wage dynamics

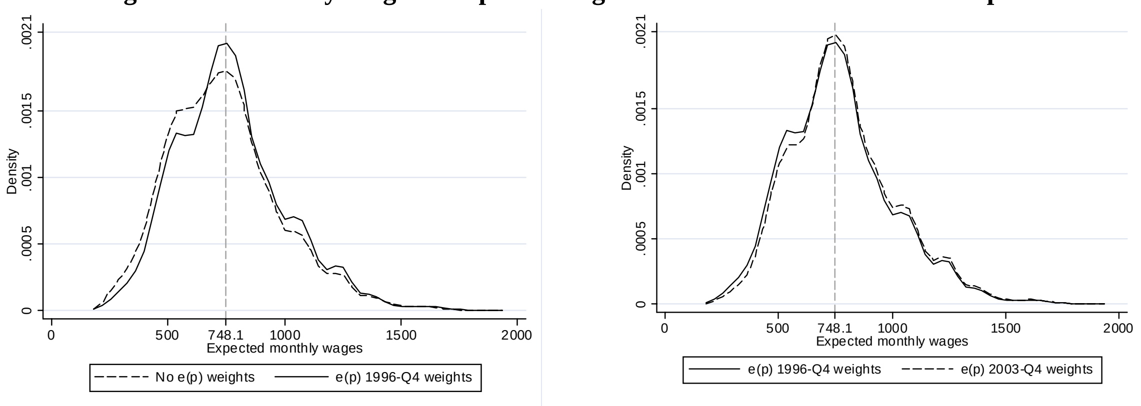
Since we now have a wage measure for every individual in the BAEL base sample, and measures of his/her probability of being employed (at different points in time), we

[13] The sample selection criteria in these cases were identical to those applied to the BAEL base sample. Full set of results for all quarters is available from the authors. Note that BAEL was discontinued in 1999 and there is no data available for the second and third quarter of 1999.

can use equation (6) to derive the counterfactual wage distributions given different employment scenarios and examine aggregate average wage dynamics once our sample is subjected to the dynamics of employment (and all else is held constant).

We start by presenting the difference probability weighting makes to the entire distribution of earnings in the BAEL base sample. This is presented in Figure 4. On panel 4a we show kernel densities of expected monthly earnings for the sample which is unadjusted for employment probability (i.e. everyone in the sample contributes to the density) and the distribution which is adjusted for employment probability using weights from Q4-1996. Notice, that, as we would expect, the latter is almost exactly the same as the density for the employed sub-sample presented in figure 2d. The aggregate average unadjusted wage is 739.27 PLN and the average rises to 767.69 PLN once we adjust for employment probabilities. Probability weighing thus increases the computed average by 3.8%. On panel 4b we again show the expected wage density weighted using the Q4-1996 employment probability weights, and compare it with the density obtained when we use the Q4-2003 probability weights instead. As we expected the density moves to the right suggesting relative employment losses in the bottom end of the wage distribution, and gains around and above the Q4-1996 median (748.10 PLN).

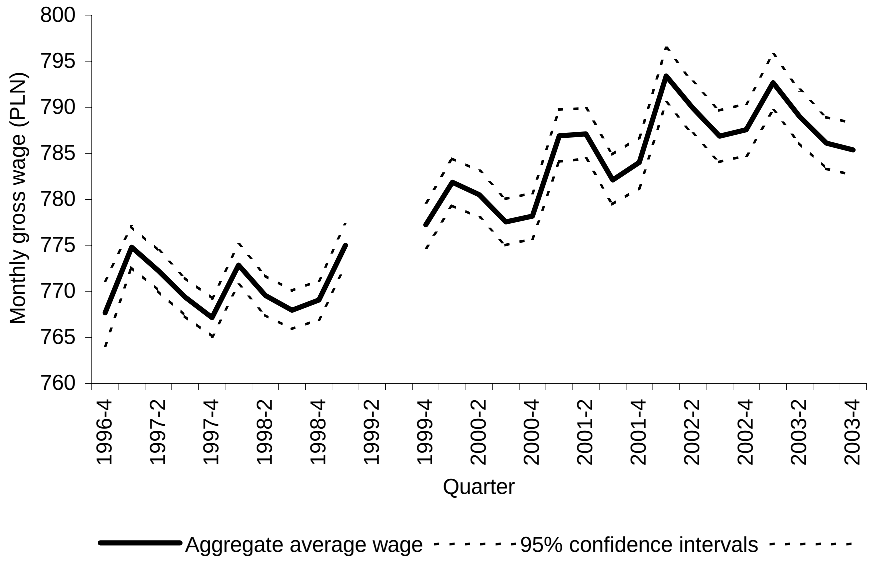
Figure 4. Probability weighted expected wage distributions in BAEL base sample



Source: authors' calculations on the basis of BAEL Q4-1996 and AES 1996.

Density estimates allow us to compute the aggregate average wages for our sample using different employment probability weights computed on the basis of BAEL Q1-1997 to Q4-2003. By changing the probability weights we are able to examine how changes in the probability of employment affect the computed aggregate average wage at the time when the *underlying* wage distribution (i.e. the distribution of expected individual wages) remains unchanged. The only thing that affects the computed average wage is the probability that someone is in the sample of employed individuals. The result of this exercise is plotted on Figure 5. The figure includes also 95% confidence intervals around the computed averages.[14]

Figure 5. Probability of employment and dynamics of aggregate average wage



Source: authors' calculations on the basis of BAEL Q4-1996 and AES 1996.

Changes in employment probability seem to have a significant (and seasonal) effect on the aggregate average wage measure. In a scenario where individual wages remain unchanged and we only change the probability of being employed the computed aggregate average wage rises from 767.69 PLN in the fourth quarter of 1996 to 785.37 PLN in the fourth quarter of 2003, a change of 2.3% (significant at 5%). This is about a tenth of the overall real growth in the aggregate average wage published by GUS, and since our wage equation does not account for unobserved characteristics (as the wage equation is estimated on the AES data, i.e. only on the employed sample)

[14] The confidence intervals were constructed using the non-parametric bootstrap method to account for matching of the data sets and for predicting of employment probability values using data other than the base sample. Given the use of several datasets and levels of analysis we bootstrap at two levels: first we estimate 100 wage equations on 100 bootstrapped samples from the AES. Then we draw 100 bootstrapped samples from each of the BAEL datasets on which we conduct the employment probit analysis. This gives us 10,000 mean wage predictions for each quarter which we use to determine the confidence intervals around the estimated means.

this is most likely the lower bound of the selection effect on the dynamics of the aggregate wage.

The average wage is lowest in the fourth quarter of 1997 (767.14 PLN) and highest in the first quarter 2002 (793.39 PLN) – here the average wage rises by 3.4% in the space of three years just because of changes in the probability of employment. The highest difference in the computed average for two consecutive quarters is between Q4-2001 and Q1-2002. The difference is 1.2% and it is statistically significant at 5%. This difference is driven entirely by changes in individual employment probabilities.

An interesting picture in terms of quarterly changes in wages comes out of the comparison of Figures 1 and 5. Focusing on the changes between the fourth quarter of each year and the first quarter of the following year (which is a comparison we can do for all years in the BAEL series) we can notice, that in all these cases there was a fall in employment levels. This in our analysis is reflected in increases in the computed average wage, suggesting that this selection out of employment is non-random with respect to the wage distribution.[15] The data on the GUS aggregate average wage suggest, however, that the average always *fell* between the fourth and first quarters of consecutive years, which seems inconsistent with our findings. Significant falls in the wages of those employed would be necessary in order for the average to turn negative given the non-random selection out of employment between Q4 and Q1. This can be easily explained by the fact that wages for the fourth quarter include various types of end-of-year bonuses which add an “artificial” seasonal fluctuation in the aggregate wage, which cannot be captured in our analysis. If instead of taking Q4 and Q1 of consecutive years we take Q3 and Q1, then for all five years for which we have the data, aggregate employment falls and the published aggregate average wages rise. Figure 5 suggests that an important part of this increase in the aggregate wage is due to selection out of employment.

Our analysis also suggests that there was little selection driven changes in the aggregate average wage in the second half of the 1990s (i.e. at the time of rapid economic growth of the Polish economy, see Table A1 in the Appendix). In these

[15] In five out of seven years this change is statistically significant. Changes are insignificant between Q4-1994 and Q1-2000 and Q4-2002 and Q1-2003.

years – at least until the end of 1998 employment was stable (see Figure 1) and thus the aggregate average wage should accurately reflect real growth in wages. However, if we look at the following years, the economic slowdown resulted in reduced employment with important consequences for the computed aggregate average wage. Between the fourth quarter of 1998 and the fourth quarter of 2002 the aggregate average wage increased only by about 10.5%, of which, according to our calculations almost a *quarter* was driven by selection out of employment.[16]

6. Conclusion

We have presented an exercise of simulating changes in the aggregate average wage which result purely from changes in the structure of employment with the underlying distribution of individual wages remaining unaffected. This is a simple, but to our knowledge so far unimplemented way of decomposing changes in the aggregate average wage into those which result from actual changes in productivity and those which are sole reflections of changes in the composition of the employed population. The methodology was applied to Polish data on earnings and employment from the Autumn Earnings Survey and the Polish labour force survey (BAEL) respectively.

The analysis shows that changes in employment in Poland have had a significant effect on the observed dynamics of aggregate average wage, especially in the first three years of the twenty first century. Our estimation suggests that this effect was in the range of 2.3% for the overall period from Q4-1996 to Q4-2003 with most of selection-driven growth of the aggregate average wage taking place between Q4-1998 and Q4-2003 (about 2.4%). These estimates suggest that for the overall period about one tenth of the wage growth reported in official statistics is a result of changes in the structure of employment. For the period from Q4-1998 to Q4-2002 almost 25% of real growth of the published aggregate average wage can be assigned to changes in the structure of the employed population. The analysis also points to the important fact that seasonal changes in employment may significantly affect the average wage between two consecutive seasons.

[16] Between Q4-1998 and Q4-2002 employment selection increases the computed aggregate average wage by 2.4% (from 769.06 to 787.56 PLN).

Our results are most likely to be the lower bound estimates of the effect of employment on average wage dynamics. This is because throughout the analysis we have assumed that wages are only determined by observed characteristics. This is clearly a strong assumption. If it does not hold, and unobserved heterogeneity affects both wages and employment (and there is positive correlation between these effects) then the actual effect of changes in employment on the dynamics of aggregate average wage would be even stronger.

The publication of the aggregate average wage is awaited by many analysts of the economy with equally eager anticipation as statistics on unemployment, investments and exports. It is considered to reflect the shape of the economy and is used by many firms as a benchmark for wage negotiations. Apart from this it is also an input used in the computation of several fiscal policy instruments where it is treated as a reflection of changes in economic welfare. Yet, according to our estimates a quarter of the real growth in the aggregate wage between Q4-1998 and Q4-2002 was not related to actual - productivity related - growth in individual wages, but rather was a result of changes in the composition of the employed population and was due to non-random selection out of employment. Though no doubt statistics on the aggregate average wage will continue to be awaited with undiminished eagerness, we suggest caution in their interpretation, especially at the time when labour markets go through periods of rapid increases or falls in the proportion of those employed and changes in their composition.

References:

Bils, M. J. (1985). ‘Real Wages Over the Business Cycle: Evidence from Panel Data’, *Journal of Political Economy*, 93, pp. 666-689.

Blundell, R., Reed, H. and Stoker, T. (2003), ‘Interpreting Aggregate Wage Growth: The Role of Labor Market Participation’, *American Economic Review*, 94(3), pp.1114-1131.

Blundell, R. and T. Stroker (2005), ‘Heterogeneity and Aggregation’, *Journal of Economic Literature*, 28, pp. 347-391.

DiNardo, J., Fortin, N.M., Lemieux, T. (1996), ‘Labor Market Institutions and the Distribution of Wages, 1973-1992: Asemiparametric Approach’, *Econometrica*, 64(5), pp. 1001-1044.

Gosling, A., Machin, S. and Meghir, C. (2000), ‘The Changing Distribution of Male Wages in the UK’, *Review of Economic Studies*, 67(4), pp. 635-66.

Heckman, J. and Sedlacek, G. (1990), ‘Self-Selection and the Distribution of Hourly Wages’, *Journal of Labor Economics,* 8, S329-S363.

Meghir, C. and Whitehouse, E. (1996), ‘The Evolution of Male Wages in the UK: Evidence from Micro-Data’, *Journal of Labour Economics*, 14(1), 1-25.

Morawski,L., Mycielski, J. and Myck, M. (2005), The Polish Labour Force Survey – BAEL 1994 –2004 – database documentation, Warsaw University – Department of Economics, *mimeo*.

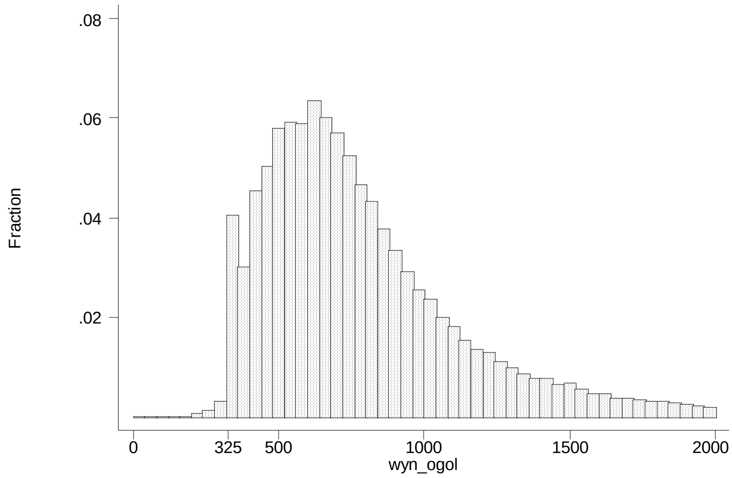
Mycielski, J., Morawski, L. and Myck, M. (2005), Wage data in the Polish LFS – the increasing problem of missing data. Warsaw University – Department of Economics, *mimeo*.

Solon, G., Barsky, R. and Parker, J.A. (1994), ‘Measuring the Cyclicality of Real Wages: How important is Composition Bias’, *Quartery Journal of Economics,* 109, pp.1-25.

Zdanowicz, J. (2003), Wskazniki przecietnego wynagrodzenia liczone przez Glówny Urzad Statystyczny, Warsaw University – Department of Economics, *mimeo*.

Appendix

Figure A1. Lower end of the wage distribution in AES 1996 – before truncation



Notes: 325PLN was the value of the National Minimum Wage in Poland in September 1996.

Table A1. Dynamics of Gross National Product in Poland: 1995-2003. Relative to the same quarter of the previous year.

Year	Q1	Q2	Q3	Q4	Q1-Q4
1996	103.4	105.4	107.2	107.9	106.0
1997	106.9	107.4	106.7	106.3	106.8
1998	106.4	105.2	104.9	103.0	104.8
1999	101.8	103.0	105.0	106.1	104.1
2000	105.8	105.0	103.1	102.4	104.0
2001	102.2	100.9	100.8	100.3	101.0
2002	100.6	100.9	101.8	102.2	101.4
2003	102.3	103.9	104.0	104.7	103.7

Source: GUS.

Table A2. Results from employment probability models - winter quarters 1996-2003.

	Q4 - 1996		**Q4 - 1997**		**Q4 - 1998**		**Q4 - 1999**		**Q4 - 2000**		**Q4 - 2001**		**Q4 - 2002**		**Q4 - 2003**	
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
Age	-0.0781	(0.1721)	-0.1762	(0.1698)	0.0805	(0.1701)	-0.2863	(0.1819)	-0.1463	(0.1849)	0.4836	(0.2091)	-0.1939	(0.2072)	0.1637	(0.2139)
Age^2	-0.0015	(0.0071)	0.0018	(0.0070)	-0.0084	(0.0070)	0.0100	(0.0075)	0.0037	(0.0076)	-0.0224	(0.0085)	0.0068	(0.0084)	-0.0089	(0.0087)
Age^3	0.0002	(0.0001)	0.0001	(0.0001)	0.0003	(0.0001)	-0.0001	(0.0001)	0.0000	(0.0001)	0.0005	(0.0001)	0.0000	(0.0001)	0.0002	(0.0002)
Age^4	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)
Education:																
- higher	1.2957	(0.0488)	1.3699	(0.0490)	1.4227	(0.0476)	1.4804	(0.0517)	1.4528	(0.0499)	1.4218	(0.0511)	1.4279	(0.0493)	1.5436	(0.0497)
- post-secondary	0.9234	(0.0562)	1.0401	(0.0578)	1.1417	(0.0588)	1.1376	(0.0625)	0.9931	(0.0601)	0.9372	(0.0618)	0.9493	(0.0614)	0.8994	(0.0620)
- secondary vocational	0.6908	(0.0344)	0.7182	(0.0349)	0.8091	(0.0353)	0.7912	(0.0392)	0.7477	(0.0390)	0.7429	(0.0407)	0.7153	(0.0408)	0.7843	(0.0426)
- secondary academic	0.5059	(0.0426)	0.5961	(0.0427)	0.6051	(0.0432)	0.6907	(0.0473)	0.6231	(0.0467)	0.5981	(0.0500)	0.6600	(0.0495)	0.6706	(0.0504)
- vocational	0.2819	(0.0340)	0.2983	(0.0346)	0.3920	(0.0348)	0.2991	(0.0394)	0.3049	(0.0389)	0.3228	(0.0401)	0.2804	(0.0403)	0.2987	(0.0421)
Male	-10.370	(1.8830)	-11.580	(1.9109)	-11.160	(1.9257)	-13.179	(2.0606)	-13.021	(2.0725)	-8.2547	(2.3501)	-14.650	(2.3513)	-8.4608	(2.3848)
Town 10k	0.1069	(0.0204)	0.0680	(0.0205)	0.0587	(0.0207)	0.0920	(0.0223)	0.0437	(0.0218)	0.0515	(0.0221)	0.0059	(0.0219)	0.0005	(0.0218)
Town 100k	0.1635	(0.0212)	0.1672	(0.0215)	0.1876	(0.0215)	0.2277	(0.0227)	0.2082	(0.0221)	0.2187	(0.0228)	0.1260	(0.0225)	0.0973	(0.0225)
Age * male	1.0590	(0.2148)	1.1934	(0.2174)	1.1265	(0.2183)	1.4010	(0.2334)	1.3410	(0.2342)	0.7654	(0.2614)	1.5069	(0.2607)	0.8242	(0.2650)
Age^2 * male	-0.0335	(0.0088)	-0.0385	(0.0089)	-0.0350	(0.0089)	-0.0489	(0.0095)	-0.0453	(0.0095)	-0.0207	(0.0105)	-0.0517	(0.0104)	-0.0242	(0.0106)
Age^3 * male	0.0004	(0.0002)	0.0005	(0.0002)	0.0004	(0.0002)	0.0007	(0.0002)	0.0006	(0.0002)	0.0002	(0.0002)	0.0007	(0.0002)	0.0003	(0.0002)
Age^4 * male	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)	0.0000	(0.0000)
Education*male:																
- higher * male	-0.2923	(0.0724)	-0.1726	(0.0767)	-0.2473	(0.0751)	-0.3608	(0.0772)	-0.3667	(0.0750)	-0.1290	(0.0772)	-0.2223	(0.0732)	-0.4184	(0.0725)
- post-secondary * male	-0.0330	(0.1395)	-0.5522	(0.1257)	-0.6210	(0.1243)	-0.4351	(0.1366)	-0.2334	(0.1310)	-0.1324	(0.1255)	-0.1100	(0.1226)	-0.2396	(0.1280)
- secondary vocational * male	-0.1868	(0.0509)	-0.1541	(0.0518)	-0.1525	(0.0524)	-0.0634	(0.0566)	-0.0690	(0.0560)	-0.0630	(0.0583)	-0.0037	(0.0577)	-0.1293	(0.0598)
- secondary academic * male	-0.1165	(0.0841)	-0.3239	(0.0839)	-0.2817	(0.0835)	-0.1271	(0.0859)	-0.0856	(0.0830)	0.0631	(0.0897)	-0.0151	(0.0881)	-0.1112	(0.0862)
- vocational * male	0.0598	(0.0468)	0.0965	(0.0478)	-0.0179	(0.0482)	0.1544	(0.0532)	0.1176	(0.0528)	0.1015	(0.0546)	0.1022	(0.0543)	0.0789	(0.0566)
Invalidity status – 1	-1.7886	(0.0926)	-2.1111	(0.1068)	-2.3031	(0.1274)	-2.3658	(0.1440)	-2.3412	(0.1488)	-2.0266	(0.1281)	-2.0661	(0.1342)	-2.0593	(0.1289)
Invalidity status – 2	-1.7042	(0.0483)	-1.7569	(0.0463)	-1.7604	(0.0463)	-1.5415	(0.0505)	-1.5682	(0.0517)	-1.5000	(0.0558)	-1.3975	(0.0572)	-1.4813	(0.0586)
Invalidity status – 3	-1.3502	(0.0366)	-1.3174	(0.0359)	-1.2506	(0.0365)	-1.1894	(0.0398)	-1.1120	(0.0404)	-1.1700	(0.0420)	-0.9487	(0.0400)	-1.0403	(0.0421)
Married	-0.0189	(0.0215)	0.0339	(0.0214)	0.0442	(0.0212)	0.1097	(0.0225)	0.0699	(0.0218)	0.1182	(0.0224)	0.1852	(0.0220)	0.1757	(0.0222)
Constant	0.5038	(1.4905)	1.5058	(1.4723)	-0.8132	(1.4798)	1.5374	(1.5826)	0.4899	(1.6169)	-5.0449	(1.8549)	0.4678	(1.8417)	-2.4109	(1.9015)
regional dummies	included		included		included		included		included		included		included		included	
Number of observations	31273		31315		30946		26295		27071		25364		25758		26003	
Log likelihhod	-15519.1		-15059.6		-15002.1		-13410.4		-14048.4		-13435.0		-13821.9		-13772.7	
Pseudo R-squared	0.2584		0.2764		0.2697		0.2523		0.2414		0.2324		0.2247		0.2346	

Source: authors' calculations on the basis of BAEL data, winter quarters, 1996-2003.