University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 13/2019 (298)

Is expansion of overeducation cohort-driven? Evidence from Poland

Jan Baran

Warsaw 2019

WORKING PAPERS 13/2019 (298)

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Is expansion of overeducation cohort-driven? Evidence from Poland

Jan Baran

University of Warsaw, Faculty of Economic Sciences
jbaran@wne.uw.edu.pl

Abstract: The study offers insight into dynamics of overeducation in Poland. The share of overeducated workers among tertiary educated workers grew substantially by about 8 p.p. between 2006 and 2016. In the paper, changing overeducation risk is disentangled into age, period and cohort effects. A strong upward trend in cohort effects is identified for individuals born after 1970, but not for older generations. It suggests that overeducation is a phenomenon which affects more profoundly individuals who entered the labour market after the collapse of the communism. Moreover, the study confirms that overeducation decreases with age, which has been already a well-documented finding in the literature.

Keywords: overeducation, education mismatch, tertiary education, age–period–cohort

JEL codes: I21, J21, J24

Acknowledgments: This study was supported by the Polish National Science Centre [grant number DEC-2016/23/N/HS4/03626].

1. Introduction

Although the education mismatch is well discussed phenomenon in social science literature, it has still attracted relatively little attention in Poland. The scarcity of research referring to Poland is especially surprising in the context of tertiary education boom which took place since the early 1990s and reshaped educational choices of the youth as well as significantly influenced educational structure of the labour supply. Since the beginning of the 1990s the tertiary enrolment rate has risen from around 13% to almost 50% in 2016. The result of the change of educational choices was a large influx of well-educated individuals to the labour market. However as the structure of labour demand upgraded at slower pace, more and more graduates failed to find a job matched to their education.

The consequence of these changes was a dramatic increase in incidence of overeducation in Polish labour market (Kiersztyn, 2013; Baran, 2018). Increasing incidence of overeducation leads to a question about nature of this change. This paper aims at investigating whether age, period and cohort effects can explain the change. To my knowledge, it is the first attempt to analyse simultaneously age, period and cohort effects in the context of changes of overeducation incidence over time. The paper places itself within a broad scope of literature on persistency of overeducation.

As motivation of this paper, identifying age, period and cohort effects has important implications for public policy. If age effect is found a dominant driver of overeducation, it would suggest that overeducation is a transitory state with individuals moving to the better matched occupations with age. Hence, the overeducation would fade over individual’s life cycle. In this case no specific public policy actions would be required. If the period effect is a dominant driver of the phenomenon, it would be an argument in favour of counter-cyclical macroeconomic policy to address education mismatch in the labour market. Finally, if cohort effects are found as an explanation behind the rise of overeducation, it would suggest that overeducation is driven by deep-rooted cohort characteristics and is likely to be persistent over the whole life cycle. In this case the policy should focus on shaping cohorts’ characteristics, especially prior their entrance to the labour market.

Analysing age, period and cohort effects faces collinearity problem as age can be obtained as a difference between year of observation and year of birth (cohort). There is no perfect solution to overcome collinearity problem, but many strategies to do so have been proposed in the literature so far. In this paper I use three different methods, which give similar results, confirming validity of the results.

The paper is organised in the following way. In the section 2, I briefly review literature on overeducation relevant to the topic of this article. Section 3 presents methodology and data used for the empirical analysis. Section 4 presents descriptive statistics on incidence of overeducation in Poland. Section 5 presents results of the econometric analysis. Final section summarises, discusses findings and draws conclusions for public policy.

2. Literature review

The overeducation has been intensively analysed phenomenon for the last three decades. Dynamic aspects of overeducation have been addressed primarily from individual perspective. Addressing dynamics of overeducation from the macroeconomic perspective is, with a few exceptions, still relatively rarely done in empirical research. It is surprising in the context of education shift taking place in developed economies which elevated tertiary education enrolment rates to record high levels. According to my best knowledge, simultaneous analysis of age, period and cohort effects behind overeducation incidence has not been conducted yet.

A vast branch of literature on overeducation centres around an issue whether overeducation is a transitory or permanent state for workers. Career mobility theory by Sicherman and Galor (1990) is frequently mentioned as a theoretical foundation for these considerations. According to Sicherman and Galor, young individuals might be voluntarily willing to take up a job below their competence level at the beginning of their job career. By doing so, they acquire job experience which then results in higher upward promotion probabilities. The career mobility theory implies that overeducation is negatively associated with tenure, and hence with worker's age. In the light of Sicherman-Galor model, overeducation is rather a temporary phenomenon which disappears with getting more job experience. Opposite implications for persistency of overeducation are offered by Thurow's job competition theory (Thurow, 1975). According to Thurow, jobs are ranked and workers form a queue to get high-ranked jobs. Better educated workers get positions higher in the ranking whilst workers with lower education levels are crowded down in the ranking. The model implies that overeducation is rather a permanent phenomenon.

Persistency of overeducation has been empirically tested in numerous papers. Whilst some empirical research paper seem to support career mobility hypothesis (Sicherman, 1991; Robst, 1995a; Frei and Sousa-Poza, 2012), there are even larger number of articles with opposite findings, suggesting that overeducation is a persistent state (Dolton and Vignoles, 2000; Rubb, 2003a; Büchel and Mertens, 2004; Mavromaras and McGuinness, 2012; Baert, Cockx and Verhaest 2013; Kiersztyn 2013; Clark, Joubert and Maurel, 2017; Wen and Maani, 2019).

Studies investigating change of overeducation incidence at macro level are less common. Groot and Maassen van den Brink (2000) conduct meta-analysis which finds no support for significant change in overeducation risk over 1980s and 1990s. In a report prepared for the European Commission, Pouliakas (2013) finds that overeducation risk did not change significantly in EU member states in 2001-2009. Also McGuinness, Bergin and Whelan (2018) claim that there is little or no change in intensity of overeducation in Europe. However, these results heavily depend on the applied methodology of overeducation identification. The result of McGuinness et al. (2018) comes from the fact that authors allow the education requirements to increase in response to the influx of educated graduates, and hence the results seem to be biased due to endogeneity problem. Pouliakas's study suffers from the same problem. Keeping education requirements constant would result in reporting of increasing risk of overeducation.

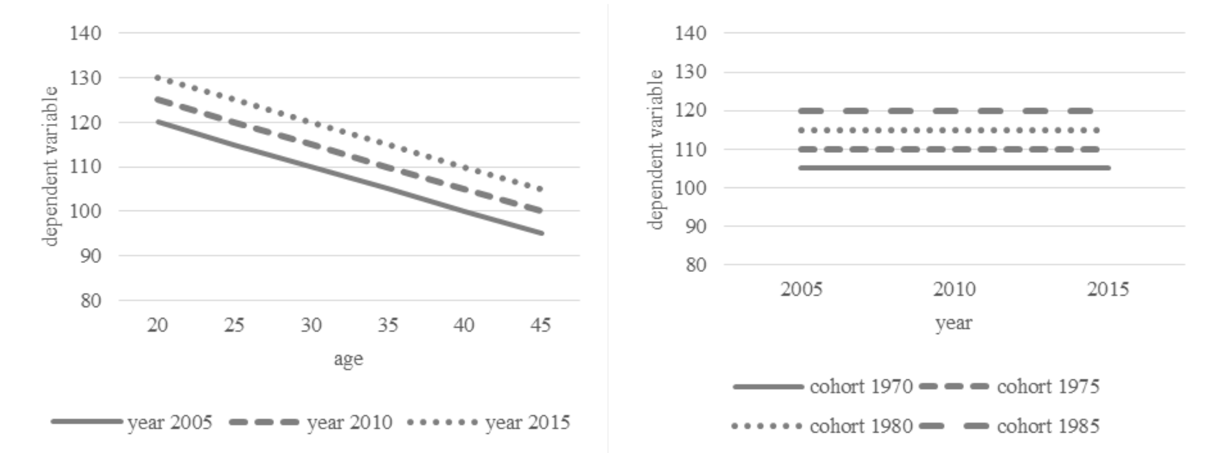
Evidence from different countries suggests overeducation becoming more widespread phenomenon. Green and Zhu (2010) report an increase in overeducation incidence in the United Kingdom between 1992-2006. Korpi and Tåhlin (2009) show that average number of years of overeducation (excess education) steadily increased in Sweden between 1974 and 2000. In Poland, Kiersztyn (2013) reports rising incidence of overeducation for the period 1988-2008. She finds also that the rise of overeducation incidence is associated with upward shift in overeducation risk between cohorts. Baran (2018) finds that overeducation incidence in Poland almost doubled in 2006-2014.

In the context of age, period and cohort effects, abovementioned studies provide only a fragmented evidence. The negative relationship between age and overeducation risk is well documented in these studies. There is also some evidence that overeducation risk is rising over time in some economies. However, none of these studies examines all three effects in one analysis. Because of omitted variable bias, findings of studies with one or two effects analysed might be misinterpreted.

3. Methodology and data

Modelling the age-period-cohort effects faces the problem of collinearity. When three effects are measured with continuous variables, each effect can be obtained as a linear combination of the remaining two. Solving the age-period-cohort effects used to be perceived as quadrature of the circle. However, it is manageable and there is a growing body of literature proposing methods to overcome age-period-cohort collinearity problem and applying them into empirical research. Simultaneous modelling of the three effects has been especially in the area of interest for epidemiology. It also attracted substantial attention in sociology. In economics it is still scarce, but age-period-cohort modelling was applied for instance into analysis of saving rates by Deaton and Paxson (1994).

The ambiguity resulting from collinearity of age, period and cohort is illustrated with Figure 1. It presents hypothetical observations of some dependent variable for individuals in different age in three different moments of time. The left panel suggests that the dependent variable negatively depends on age as well as there is also upward shift for consecutive periods. Hence it might seem that the dependent variable is driven by combination of negative age effect and positive period effect. The right panel of Figure 1 utilises the same data, but depicts them in different manner. Now observations are grouped according to cohort and presented over time. This way we can immediately recognise that dependent variable appears to be driven by positive cohort effects and within cohorts it is stable over time. Both interpretations, i.e. combination of negative age effect and positive period effects or positive cohort change with no period effects are equally reasonable in light of data. It is researcher's role to choose the most plausible interpretation using her/his understanding of a phenomenon which is reflected by the dependent variable.



Source: Own elaboration.

Figure 1. Example illustrating different representations of series of hypothetical data

The simplest way to overcome the problem of collinearity of age-period-cohort effects is to delete one of the three effects. It means that a researcher imposes a zero constraint on one effect. Thus, the model reduces to two-effect model. In the field of labour economics the cohort effect is usually neglected and researchers' attention focuses on effects associated with age and period (which is usually interpreted as impact of business cycle). However, this approach requires that the removed effect has no impact on the outcome variable. Otherwise, the two effects absorb the removed effect and estimated coefficients for them are biased and then it would lead to faulty reasoning and invalid conclusions. To illustrate this, we might think about a phenomenon which is exclusively driven by a positive cohort effect (as depicted in the right panel of Figure 1), but a researcher decides to delete a cohort effect to provide identifiability of the model. Then she or he misleadingly obtains results suggesting that effect is driven by age and period effects (as in the left panel of Figure 1).

When age, period and cohort effects are introduced as dummies, some researchers deal with collinearity problem by imposing constraints on certain parameters. A single constraint is required to obtain model identifiability. For instance, coefficients for two adjacent age groups might be set equal. At the first glance, this approach is less restrictive than deleting the whole effect. However, results obtained in this way are usually very sensitive to the choice of constraints. Utilising different granulation of age, period and cohort effects, which is also frequently used to obtain identification of a model, also constrains model parameters. For instance, cohorts might be defined as five-year time spans, whilst age and period as one-year time spans. It means that estimated effects for one-year generations would be set equal within a five-year cohorts.

The choice of restrictions should depend on a researcher's understanding of the nature of analysed phenomenon to maximise chances that restrictions closely reflect true relationship. Mechanical choosing of constraints is criticised as it might lead to wrong results. Data cannot find themselves which constraint is more appropriate. Harding (2009) points out that all models with one constraint will produce the same goodness of fit. Thus it is advised that constraints should be made on theoretical basis and declared explicitly (Harding, 2009; Bell and Jones, 2013).

An appealing way to overcome the collinearity problem, which I follow in this paper, is to replace either a period or a cohort effect with their proxies. Heckman and Robb (1985) point out that age, period, and cohort effects capture the impact of underlying mechanisms, rather than impact of numbers written in a questionnaire under questions over 'age', 'date of birth', 'date'. According to this reasoning, an age effect merely reflects maturation and physiological changes taking place in our bodies with time, fluctuating macroeconomic conditions are reflected in a period effect, and a cohort effect reflects an impact of the same generation-wide experiences such as economic depressions or wars. Hence, a researcher could replace age, period or cohort dummies with variables directly representing underlying mechanisms. Introduction of proxies of underlying mechanisms instead of age, period or cohort effects means de facto imposition of constraints as they cannot adjust freely for each age, period, cohort of observation, but must follow changes in the proxies for underlying mechanism.

As in economic research a period effect is often associated with an impact of a business cycle, it seems natural to replace period dummies with variables directly describing macroeconomic situation such as GDP growth rate or unemployment rate. The similar could be applied to replace cohort dummies. Potentially good proxies for a cohort effect might be cohorts' tertiary education enrolment rates or unemployment rate at the moment of cohort's education-to-labour market transition.[1] My choice to replace period effects boils down to data availability. Business cycle variables, which I use to proxy period effects, are easily available. Potential proxies for cohort effects are unfortunately not available for Poland prior to 1990.

In recent years, a number of more advanced statistical methods have been proposed to address age-period-cohort identification problem. Out of them I use two as robustness checks: a method based on restricted spline functions and intrinsic estimator.

Carstensen (2007) suggested to model age, period and cohort effects as continuous variables with using spline functions. Spline function is a function which consists of several subpieces defined as low degree polynomials connected in certain points named knots. To illustrate it simply, a natural spline is the shape that would be taken by a flexible rod forced to pass through a number of knots (Sasieni 2012). Number of knots can be chosen arbitrary, with more knots giving greater flexibility. Still, the identifiability of parameters requires to impose some restrictions. By choosing business cycle variables as proxies for period effects in the previous approach, I implicitly assumed that period effects should exhibit cyclical fluctuations. Thus imposing zero slope constraint on period effects seems to be rational choice of constraint for modelling spline functions. To model spline function I use apfit package prepared for Stata by Rutherford, Lambert and Thompson (2010). The procedure works with aggregate data in the form of Lexis diagram which have to be rearranged to provide proper averages for yearly age-

[1] The rationale behind the latter one is that graduating during recession has negative long-lasting impact on individuals' labour market outcomes (Oreopoulos, von Wachter and Heisz, 2012).

cohort-period cells. The detailed discussion of the procedure is provided in Rutherford et al. (2010).

The intrinsic estimator was proposed by Yang, Fu and Land (2004). Intrinsic estimator is seen as a variant principle component regression estimator (Yang et al., 2008). This method of coping with age-cohort-period collinearity problem does not require to declare constraints explicitly by the researcher, but it introduces them implicitly itself. Constraints depend on number of time, cohort and period effects in the data. Sums of parameters for each of age, period and cohort effects are set to zero. According to Harding (2009) intrinsic estimator results can be interpreted as conventional regression estimates. Discussion of properties of intrinsic estimator can be found in Yang et al. (2008) and Yang and Land (2013). There is however controversy about using intrinsic estimator. Luo (2013) criticises this method due to implicitness of constraints used for model identification. What is more, the constraints change for different number of age, period and cohort categories, implying that estimates might differ for different representation of data. Thus intrinsic estimator results should be treated with caution.

To model effects behind changes in overeducation incidence, one should start with definition of overeducation. For the purpose of this research, I define overeducation as a situation that a tertiary-educated individual works in a job which does not require tertiary education. Such binary treatment of education levels, although precludes overeducation among secondary educated individuals who work in jobs requiring only primary education, it allows to focus solely on education mismatches in the context of university boom. To identify mismatches I use data from two surveys: Polish Labour Force Survey (LFS) for 2006-2016 and Balance of Human Capital Survey (Bilans Kapitału Ludzkiego, BKL) for 2010-2014.

Balance of Human Capital Survey provides information on required education level for different occupations based on employers' declarations. In the survey, employers were asked about vacancies in their firms. An ISCO code is assigned to each declared vacancy. Two further questions about vacancies are essential for this study. The first question reveals whether candidate's education level is important for employer: "Please think about a perfect candidate for this [vacant] position. Is education level important?" If an employer answers "yes", second question is asked about expected education level: "What education should that person have?". Answers from different waves of the survey are pooled. Altogether there are 14 586 observations in the sample. Due to large variation in number of observations for different occupations analysis is conducted on 2-digit level of ISCO classification[2]. Based on frequency of employers' answers, I split occupations into two categories: *university jobs* and *non-university jobs*. Jobs with the share of employers' answers stating that tertiary education is required exceeding the threshold value are labelled as university jobs. I apply two thresholds: 50% and 70%. Out of 43 occupations, defined at 2-digit codes of ISCO classification, 13 occupations are assigned as university jobs when the 50-percent threshold is applied, and 9 occupations when the 70-percent threshold is applied. The rest of occupations are considered to be non-university jobs[3]. Education requirements for occupations are kept fixed for the whole period of analysis. In the next step, I cross information on required education in occupations with LFS data which provide information on workers' education. As the interest of this paper is to explain changes in overeducation incidence among graduates over time, I focus solely on tertiary educated individuals.

In the main part of econometric analysis I run binary dependent variable regressions (logistic regressions) with age and cohort dummies and business cycle variables proxying period effects. This analysis is conducted on microdata from LFS survey. Business cycle variables include regional unemployment rate, sourced from Eurostat, and real GDP growth rate, sourced from AMECO database. Moreover, the regressions include set of control variables from LFS survey: gender, migrant status, field of education, size of place of living (degree of urbanisation) and region (voivodship). Regressions are run for three model specifications, differing in number of control variables, and two variants of thresholds used to identify university jobs. Estimations for the broadest set of control variables are run for slightly shorter period (2006-2015), because of changes in classification of fields of studies in 2016 which disables comparability of this variable with previous years.

For the robustness check, I use also constrained cubic spline function, and intrinsic estimator. The robustness checks utilise aggregated data (obtained from LFS data by summing observations into "cells" of yearly combinations of age-period-cohort) and age, period and cohort effects are the only variables included.

The analysis covers the period 2006-2016. Data from Balance of Human Capital on education requirements are available only for period 2010-2014. Hence, there is apparent time discrepancy between BKL data and LFS data. Because there is upward shift in education

---

[2] Even when occupations are aggregated to 2-digit ISCO codes, number of observations per one occupation varies from less than ten to over one thousand.

[3] Eight occupations with number of observations less than 20, were arbitrary assigned. According to LFS data, these occupations correspond to about 3% of workers with tertiary education in 2016.

requirements over time, referred as *inflation of credentials*, one might claim that applying requirements calculated on 2010-2014 data to LFS data for previous periods results in underestimation of true extent of overeducation. I am aware of this problem, however addressing it with shortening the time span of analysis, to make BKL and LFS samples match, would be too costly. Furthermore, time period of 2006-2016 corresponds roughly to length of one business cycle, which is needed to assume zero trend in period effects.

4. Descriptive statistics

Table 1 presents shares of tertiary educated workers according to 50-percent and 70-percent thresholds. The share of tertiary educated individuals who are found overeducated rose significantly in period 2006-2016. For 50-percent threshold this share increased from 24.7% in 2006 to 31.6% in 2016. 70-percent threshold, which puts more occupations into non-university category, produces increase from 30.0% in 2006 to 38.0% in 2016. At the same time number of tertiary educated workers expanded significantly, from 3.3 million to 5.4 million. As a result, number of overeducated individuals doubled (from 0.8 million to 1.7 million for 50-percent threshold, and from 1.0 million to 2.1 million for 70-percent threshold). These statistics prove that expansion of overeducation is a massive phenomenon taking place in Polish economy, hence it is important to explain factors standing behind this change.

Table 1. The incidence of overeducation among tertiary educated workers in 2006-2016

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
number of tertiary educated workers, 10^6	3.3	3.5	3.7	4.0	4.4	4.6	4.6	4.8	5.2	5.3	5.4
share of overeducated individuals											
50-percent threshold	24.7%	25.9%	27.8%	28.0%	29.3%	28.6%	28.0%	29.2%	30.2%	30.7%	31.6%
70-percent threshold	30.0%	31.7%	33.8%	34.2%	35.9%	35.1%	34.3%	35.5%	36.6%	37.0%	38.0%

Source: Own calculations using Polish LFS data for 2006-2016.

Statistical and graphical analysis can provide meaningful impression about age-period-cohort effects (Yang and Land, 2013). Looking at plots describing the evolution of the analysed phenomenon, although is not a sophisticated method of analysis, might be very informative in suggesting the nature of change of analysed phenomenon. Thus, it is recommended that econometric analysis of age-cohort-period effects is preceded by taking careful look at data. In the next paragraphs I follow this advice. Table 2 depicts detailed overeducation rates for each one-year age-period combination. Figure 2 presents shares of overeducated individuals among tertiary educated workers for five-year cohorts in period 2006-2016. Figure 3 presents respective shares for five-year cohorts at different age.

Table 2. Age-specific shares of overeducated individuals among tertiary educated workers for different periods of time

50-percent threshold

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
2006	63%	46%	42%	36%	34%	32%	32%	29%	26%	26%	23%	23%	24%	23%	19%	16%	22%	20%	23%	18%	15%	17%	15%	19%	15%	16%	15%	18%	18%	15%	12%	17%	18%	21%	19%	13%	18%	14%
2007	56%	46%	42%	41%	34%	31%	28%	34%	28%	24%	26%	25%	26%	26%	24%	21%	18%	18%	18%	21%	21%	17%	26%	16%	14%	12%	15%	16%	18%	19%	17%	16%	21%	18%	22%	20%	18%	20%
2008	53%	46%	44%	41%	39%	31%	34%	26%	31%	29%	24%	29%	29%	28%	24%	21%	23%	23%	17%	23%	19%	18%	20%	19%	22%	16%	14%	17%	18%	20%	23%	17%	23%	21%	23%	20%	25%	26%
2009	51%	44%	42%	38%	38%	37%	32%	32%	30%	29%	28%	22%	25%	30%	25%	22%	22%	21%	24%	15%	23%	17%	20%	21%	21%	21%	16%	19%	23%	24%	19%	21%	17%	19%	25%	21%	25%	23%
2010	53%	44%	41%	39%	36%	37%	36%	34%	34%	34%	32%	29%	27%	32%	30%	26%	25%	24%	24%	20%	22%	23%	19%	22%	16%	20%	17%	17%	17%	19%	21%	18%	19%	20%	22%	22%	18%	18%
2011	58%	49%	41%	44%	40%	36%	37%	32%	35%	32%	30%	27%	27%	24%	26%	24%	23%	23%	19%	18%	22%	21%	20%	17%	16%	16%	16%	18%	19%	17%	17%	20%	16%	18%	17%	22%	22%	19%
2012	59%	48%	42%	41%	43%	40%	36%	35%	32%	35%	34%	30%	24%	26%	22%	23%	22%	21%	21%	18%	16%	20%	19%	13%	15%	13%	13%	15%	18%	21%	19%	18%	17%	14%	17%	18%	24%	23%
2013	65%	51%	47%	43%	41%	41%	38%	37%	34%	32%	36%	34%	30%	27%	26%	27%	25%	23%	20%	23%	22%	17%	21%	19%	15%	17%	15%	13%	12%	20%	20%	17%	19%	14%	17%	18%	15%	23%
2014	63%	52%	48%	42%	40%	41%	41%	38%	37%	37%	35%	34%	29%	28%	31%	27%	28%	26%	21%	23%	22%	23%	21%	24%	17%	19%	17%	16%	13%	16%	21%	16%	14%	19%	18%	19%	20%	19%
2015	58%	53%	46%	48%	40%	40%	40%	37%	38%	36%	38%	35%	31%	28%	28%	30%	30%	28%	26%	23%	24%	22%	19%	22%	19%	17%	23%	19%	15%	17%	17%	19%	17%	15%	22%	22%	19%	17%
2016	57%	48%	44%	47%	44%	40%	42%	38%	39%	37%	38%	38%	34%	33%	30%	30%	29%	27%	32%	25%	25%	24%	22%	16%	22%	18%	19%	19%	19%	17%	18%	20%	20%	16%	16%	26%	23%	18%

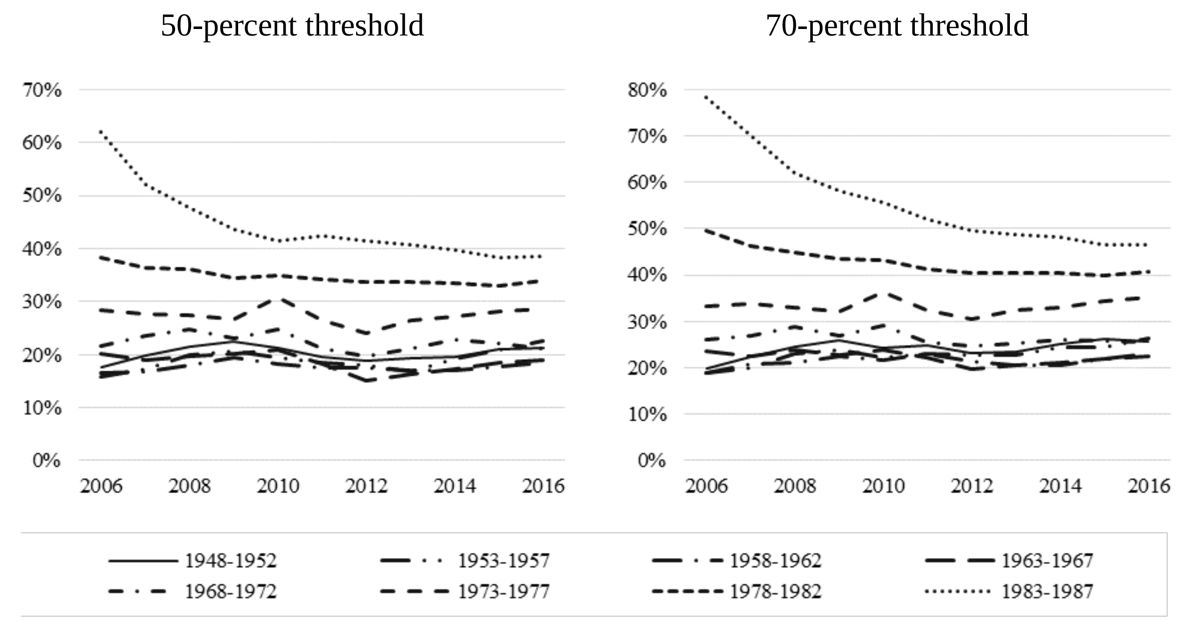
70-percent threshold

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
2006	77%	61%	55%	46%	43%	40%	37%	36%	31%	31%	27%	31%	27%	25%	24%	20%	24%	24%	28%	20%	16%	21%	17%	22%	18%	18%	17%	22%	22%	17%	16%	21%	18%	21%	21%	19%	20%	17%
2007	74%	62%	56%	49%	43%	40%	35%	42%	35%	30%	30%	29%	29%	28%	27%	27%	23%	21%	20%	25%	24%	21%	30%	20%	20%	16%	16%	18%	22%	22%	21%	18%	24%	20%	24%	20%	24%	21%
2008	73%	58%	55%	51%	48%	40%	42%	34%	38%	33%	27%	34%	34%	32%	28%	27%	26%	29%	19%	28%	23%	21%	23%	23%	24%	19%	17%	21%	20%	23%	26%	23%	25%	27%	26%	23%	26%	27%
2009	74%	62%	53%	49%	46%	47%	41%	39%	36%	35%	34%	27%	29%	35%	27%	27%	26%	24%	27%	17%	27%	20%	21%	22%	24%	24%	19%	22%	28%	24%	23%	24%	20%	23%	31%	25%	27%	24%
2010	71%	65%	55%	51%	48%	46%	45%	42%	40%	39%	39%	36%	32%	36%	35%	31%	31%	28%	27%	24%	26%	26%	22%	26%	19%	25%	20%	22%	19%	24%	23%	21%	21%	22%	25%	25%	20%	23%
2011	68%	58%	54%	54%	49%	43%	45%	40%	41%	39%	36%	35%	33%	31%	31%	29%	28%	26%	23%	21%	28%	25%	23%	20%	20%	20%	22%	23%	25%	22%	21%	23%	19%	24%	24%	29%	28%	24%
2012	69%	62%	51%	49%	51%	48%	43%	42%	40%	41%	38%	36%	31%	31%	29%	28%	29%	27%	27%	23%	20%	26%	23%	16%	20%	16%	19%	19%	22%	23%	22%	22%	21%	18%	23%	22%	30%	27%
2013	77%	64%	59%	53%	49%	48%	45%	43%	41%	41%	41%	40%	37%	33%	31%	33%	30%	29%	25%	27%	25%	22%	25%	24%	19%	21%	19%	18%	16%	22%	24%	21%	24%	19%	23%	25%	20%	28%
2014	74%	62%	59%	53%	50%	50%	49%	46%	44%	44%	41%	40%	36%	36%	36%	32%	34%	31%	28%	25%	27%	26%	24%	28%	21%	22%	21%	19%	17%	20%	26%	19%	18%	23%	22%	25%	27%	23%
2015	68%	63%	57%	55%	49%	48%	49%	44%	46%	44%	46%	41%	38%	34%	34%	36%	36%	34%	32%	29%	26%	27%	24%	24%	23%	22%	28%	22%	17%	20%	20%	23%	22%	20%	26%	25%	21%	21%
2016	65%	60%	52%	55%	51%	47%	51%	46%	47%	46%	45%	44%	40%	40%	36%	37%	37%	34%	37%	31%	29%	29%	27%	21%	25%	22%	22%	24%	23%	20%	20%	22%	23%	23%	24%	30%	26%	23%

Source: Own calculation using Polish LFS data for 2006-2016.

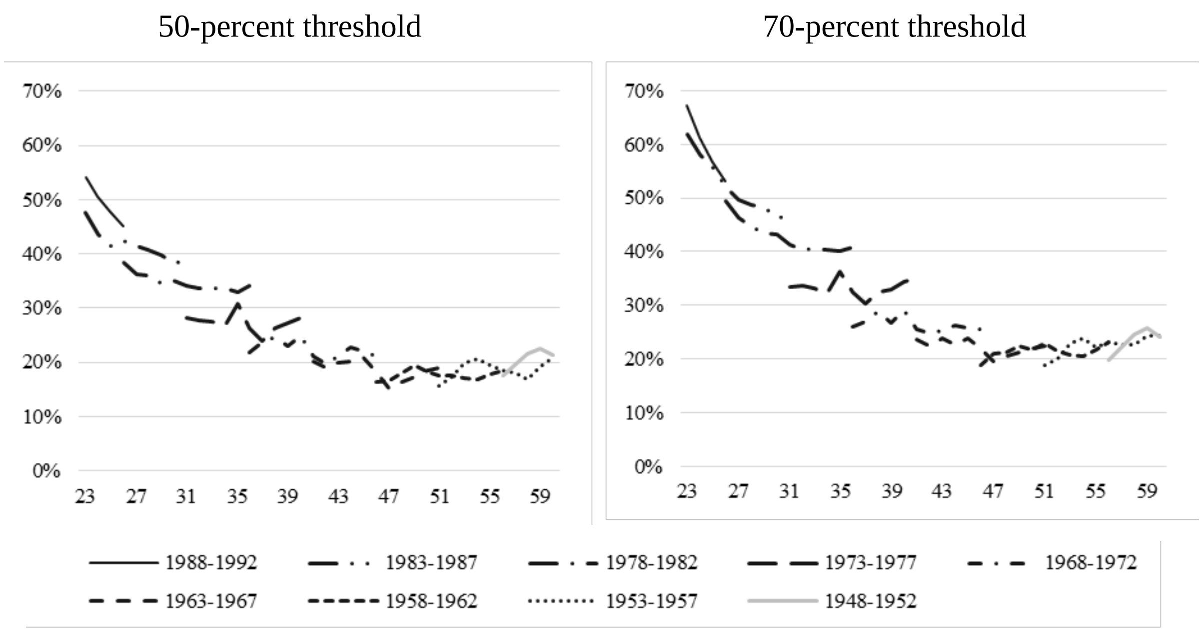
Table 2 illustrates that the youngest workers are at the greatest risk of overeducation. The overeducation risk decreases with age. Workers who are more than 45-year-old face the lowest risk of overeducation. For easier comparison of values in the table, cells are coloured accordingly to the share of overeducated workers for each age-period combination. Coloured cells make easier to notice diagonal patterns of overeducation risks, which are present through the whole age distribution. Overeducation rates seem roughly constant at diagonals, especially in case of workers in their 30s and 40s, whilst there is apparent shift between diagonals. As each diagonal represents a single cohort, it suggests that there might be a cohort effect taking place behind overeducation increase.

Figure 2 also shows that there is visible shift in overeducation risk among tertiary educated workers. It depicts change in overeducation rates over time for five-year cohorts. Cohorts born before 1972 have more or less the same risk of overeducation, which is quite stable over time. Lines for the cohorts born after 1972 are shifted upwards. Workers born 1973-1977 face higher risk than workers born in 1968-1972 for each period of time. The same applies to subsequent cohorts. Workers born in 1978-1982 are more at risk of overeducation than workers born in 1973-1977, whilst workers born in 1983-1987 face larger risk of overeducation than those born in 1978-1982. Furthermore, Figure 3 shows that fragments of age profiles of overeducation incidence which can be obtained from available data are shifted upwards for the youngest cohorts. Such shift is not present for the oldest cohorts. Thus the graphical analysis builds plausible hypothesis of increasing cohort effects behind overeducation risks in working population. The following part 5 addresses it with econometric methods.



Source: Own calculation using Polish LFS data for 2006-2016.

Figure 2. The share of overeducated individuals among tertiary educated workers for five-year cohorts, 2006-2016



Note: In this figure age corresponds to age of individuals of the middle year of a cohort.
Source: Own calculation using Polish LFS data for 2006-2016.

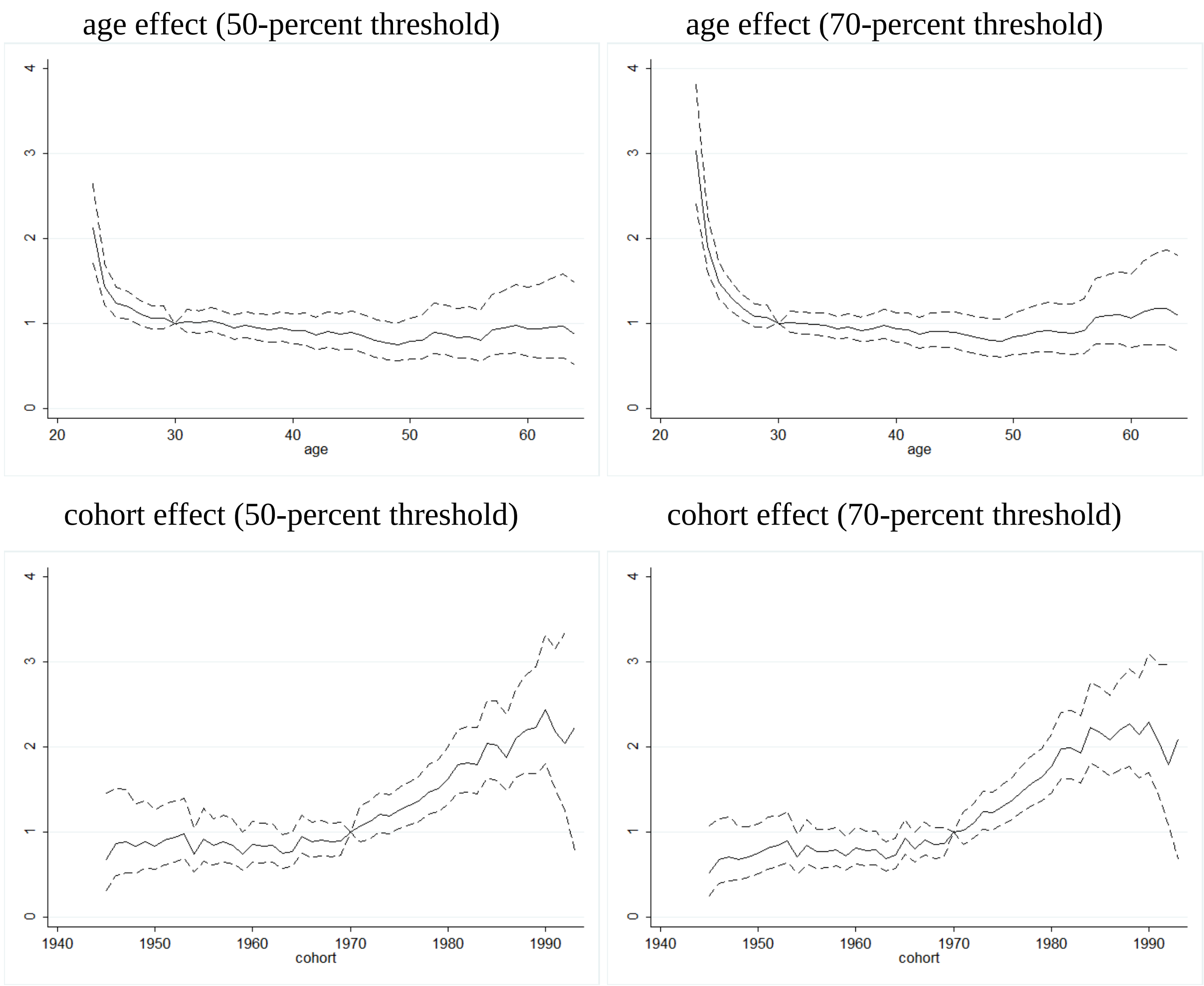
Figure 3. The share of overeducated individuals among tertiary educated workers for five-year cohorts and different age

5. Econometric analysis

In this part of the paper I present results of econometric analysis. Firstly, I present results of logistic regression using business cycle variables which proxy for period effect. It is complimented with two robustness checks: results of analysis using restricted cubic splines and results of analysis using the intrinsic estimator. All three approaches give qualitatively similar results, which justifies setting them together in the paper.

In this section I present results of logistic regression explaining probability of being overeducated among workers with tertiary education. Age and cohort effects are introduced with the set of dummy variables. Due to collinearity the period effect could not be introduced with dummies. Instead, it is proxied with continuous business cycle variables. I use regional unemployment rates and real GDP growth rate. The model is estimated in three specifications. The most parsimonious specification includes age dummies, cohort dummies and business cycle variables. The second specification includes a set of control variables which are: gender, being an immigrant, degree of urbanisation and region. The third specification includes also field of study as an additional control variable.[4] Models are estimated for two definitions of overeducation (using 50-percent and 70-percent threshold definitions of university jobs). Figure 4 depicts odds ratios for age and cohort dummies respectively for the second specification of the model for two definitions of overeducation. Detailed results of regressions are presented in Table A.1 in the Appendix.

[4] Due to data availability for this variable, the analysis for the third specification is shortened by one year, and covers the period 2006-2015.



Note: The age of 30 and year of birth 1970 are chosen as reference levels for age and cohort effects respectively. A solid line represents estimates, dashed lines represent 95-percent confidence intervals.
Source: Own calculations based on Polish LFS data for 2006-2016.

Figure 4. The estimated effects for age and cohort dummies from logistic regression explaining probability of being overeducated among tertiary educated workers, odds ratios

Estimated coefficients for age dummies show decreasing size of the effect over age. The youngest workers have the highest risk to be overeducated. With age, the chance to work in a mismatched job decreases. In the 20s the decrease is relatively fast. After 30, the risk of overeducation decreases more slowly. The lowest risk of overeducation is found for workers aged 49. Then, among the oldest workers the age coefficients start to grow slightly (especially for 70-percent threshold definition of university jobs), but due to widening of confidence intervals it cannot be rejected that actual overeducation risk is the same as for workers in their 30s.

Logistic regression results prove what descriptive analysis has suggested in the previous part that there is apparent shift in cohort effects for individuals born after around 1970. Estimated coefficients for cohorts 1972-1992 are positive and highly statistically significant. Since 1970 is a reference cohort, it means that university graduates born after 1970 face higher risk of overeducation than graduates born in 1970. What is more, for this group of university graduates estimated coefficients are rising for consecutive cohorts. It means that the younger cohort, the higher risk of overeducation. Tertiary educated individuals born in late 1980s face twice the odds of being overeducated compared to individuals born in 1970. This pattern is reflected in all six estimations presented in Table 3. There is less consistent picture for individuals born before 1970. The most parsimonious model specification produces falling odds of overeducation for older cohorts but this effect largely fades out after adding control variables, especially for 50-percent threshold definition of university jobs. For the 70-percent threshold definition of university jobs, more older cohorts experience significantly lower overeducation risk than the reference cohort compared to 50-percent definition. Though, for many older cohorts it cannot be rejected that overeducation risk is the same as for the reference cohort due to widening of confidence intervals.

Business cycle variables explain little variation in overeducation incidence. Out of two variables proxying period effect, the regional unemployment rate is only statistically significant in basic model specification. The estimated coefficient for unemployment rate is positive, thus periods of high unemployment rates are associated with higher risk of overeducation among tertiary educated workers.[5]

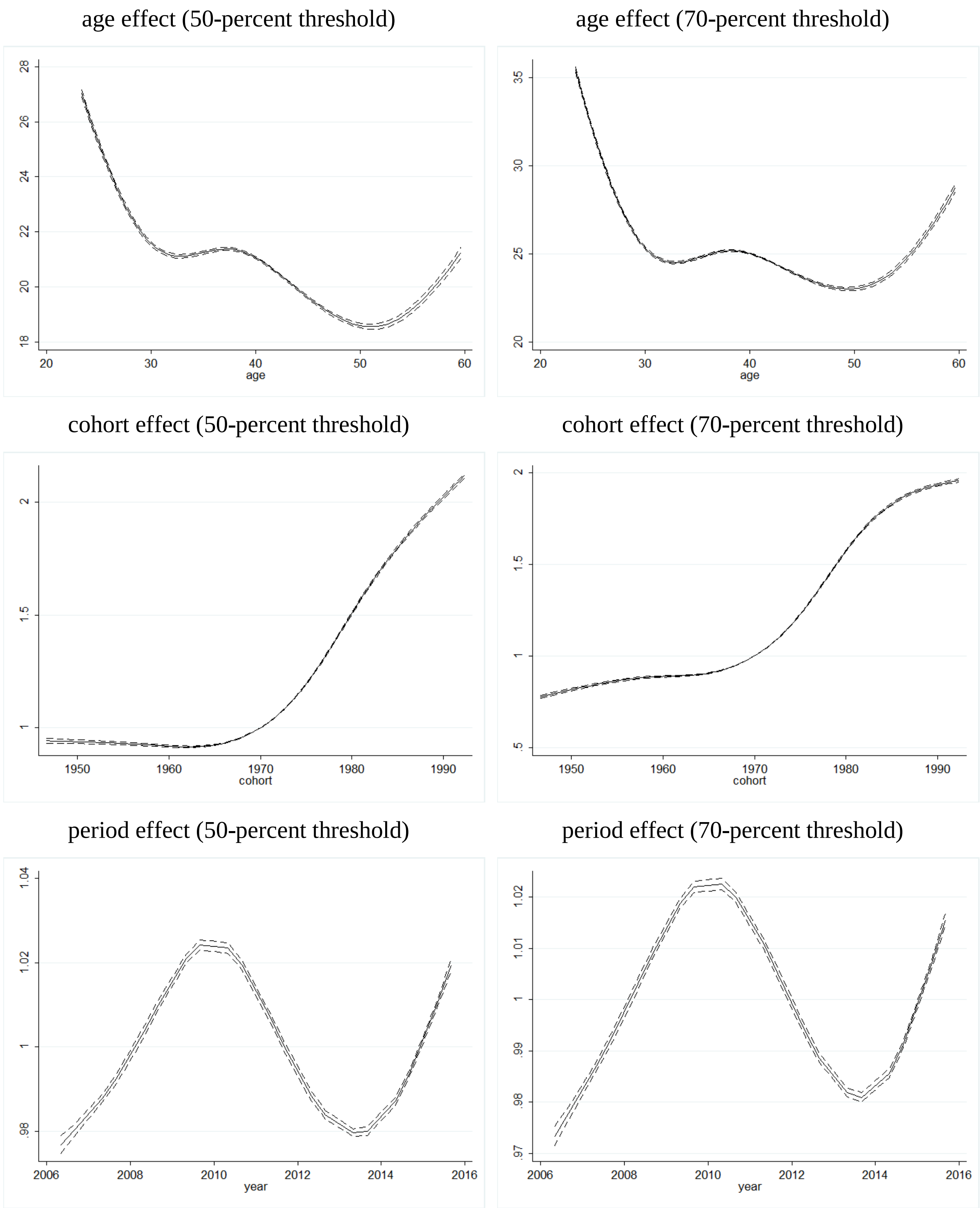
The only statistically insignificant control variable is immigration background. Being a female decreases the risk of overeducation. There is also statistically significant differentiation in overeducation risks depending on regional characteristics: tertiary educated workers in smaller towns and rural areas as well as in less developed voivodships (such as Podlaskie, Lubelskie, Świętokrzyskie and Zachodniopomorskie) face larger risk of overeducation compared to tertiary educated workers in big cities and more developed voivodships. The important factor differentiating risk of overeducation is field of tertiary education, with

[5] This finding seems intuitive. When the unemployment rate is high, the labour market is tight and the unemployed cannot afford long search for jobs properly matching their education as well as mismatched workers are less likely to quit one job and start looking for another, better matched. Thus mismatches are likely to be protracted in periods of high unemployment.

graduates from health and computer sciences facing lowest odds of being overeducated. Adding this control variable substantially improves model's goodness of fit.

5.1. Robustness check: restricted spline function

Figure 5 and Table 4 in the Appendix present results of estimation of age, period and cohort effects using restricted cubic splines. Because the *apcfit* procedure used to model restricted cubic splines works on aggregated data, the data have been collapsed accordingly. For model identification, I constrain period effects to be zero on average with zero slope. Zero slope restriction for period effects can be justified with cyclicality of macroeconomic conditions as in economic research period effects are usually associated with the impact of business cycle. Moreover, it is hard to find similar argument in favour of the zero slope restriction in case of cohort or age effects. Thus zero slope constraint on period effects seems more reasonable. Estimated age effects can be interpreted as overeducation rates for the reference cohort (1970 has been chosen), and cohort effects are rates relative to the reference cohort.



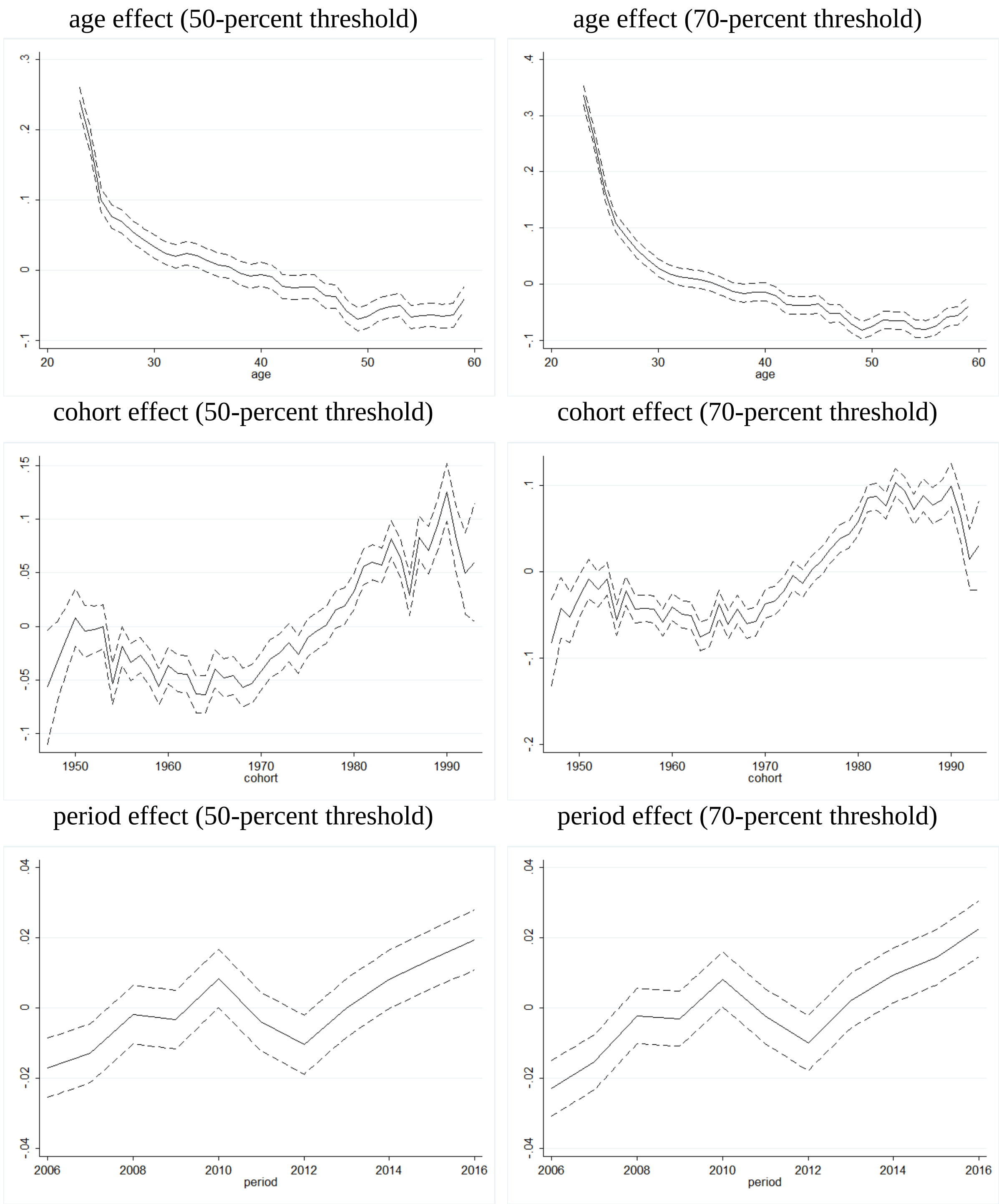
Note: A solid line represents estimates, dashed lines represent 95-percent confidence intervals.
Source: Own calculations using Polish LFS data for 2006-2016.

Figure 5. Age, cohort and period effects estimated with restricted cubic splines

The estimation results show a U-shaped pattern of age effect. The youngest individuals experience highest risk of overeducation. Then the overeducation risk is decreasing with age until around age of 50. Workers in this age experience the lowest risk of overeducation. For workers older than 50-year-old the overeducation risk again grows slightly, but it does not reach levels observed among young individuals. The estimates of cohort effects produce a clear break around year of birth of 1970. For the 50-percent threshold definition of overeducation, all cohorts born before 1970 experience overeducation risk very close to the reference cohort, whilst workers born after 1970 face increasing risk of overeducation. Those born in 1980 have 50% higher overeducation risk compared to the reference cohort, and those born in 1990 have 100% higher overeducation risk. The estimates of cohort effects for the 70-percent threshold definition of overeducation present similar cohort pattern. The notable difference between results for the two definitions of overeducation is that in the latter the oldest cohorts face the lowest risk of overeducation. The estimates of period effects, because of zero-slope restriction, give no trend, although some fluctuations of overeducation risk are observed due to period effects. However, these fluctuations are not large in size compared to cohort effects.

5.2.Robustness check: intrinsic estimator results

Figure 6 and Table 5 in the Appendix report results of intrinsic estimator proposed by Yang et al. (2004). The results are similar to those obtained from previous methods. The age effect steadily decreases over the lifespan. The same as in the previous methods, youngest workers experience the highest risk of overeducation. Though, contrary to previous approaches, intrinsic estimator shows that oldest workers have the lowest risk of overeducation. Estimations of cohort effects show also apparent break after 1970. There is increasing risk of overeducation affecting workers born after 1970. Oldest cohorts also experience slightly increased risk of overeducation compared to workers born in the 1960s. The period effect fluctuates over time in the similar manner as in cubic splines estimates.



Note: A solid line represents estimates, dashed lines represent 95-percent confidence intervals.
Source: Own calculations based on Polish LFS data for 2006-2016.

Figure 6. The effects associated with age, cohort and period estimated using intrinsic estimator procedure

6. Discussion of results and conclusions

The study contributes to better understanding of dynamics of overeducation in Poland, which is an example of country which experienced intensive tertiary education boom. Tertiary educated workers in Poland face increasing risk of being overeducated, i.e. they work in a job which requires lower levels of formal education. The share of overeducated workers among tertiary educated workers grew substantially by about 8 p.p. between 2006 and 2016. This paper investigates age, period and cohort effects behind the rise of overeducation risk. Applying econometric methods to disentangle age, period and cohort effects is a novelty in the overeducation literature. The simultaneous investigation of age, period and cohort effects requires to cope with the collinearity problem. To do so, the study uses three approaches. In the first approach the period effect is replaced with variables controlling for business cycle. The second approach uses restricted cubic spline function, whilst the third one uses intrinsic estimator. All three approaches give consistent results.

The most important finding of this paper is the large upward cohort shift in overeducation risk for individuals born after 1970. What is more, the cohort effect on overeducation risk is increasing, i.e. the younger cohort, the higher risk of overeducation. This finding is robust to the choice of the method used. Moreover, it is coherent with a finding by Kiersztyn (2013), who identifies traces of positive cohort effect in Poland using data for 1988-2008. Existence of strong cohort effects suggests that overeducation is a persistent phenomenon, which is in line with the large body of literature. Persistent overeducation implies that educational mismatch might be a trap in which workers fall and stay for long period of time. The study also confirms that the risk of overeducation decreases with age, what is well-documented in the subject literature. There is little evidence that period effect drives changes in overeducation in Poland.

An interesting finding is that apparent change in trends in cohort effects behind the overeducation risk affects workers born after 1970. Although I am not sure about nature of this change, some possible explanations emerge. It seems that the overeducation increase is associated in some way with the economic transition in Poland. People who had been born in 1970 were the first cohort who entered tertiary education after the regime change and one of the first cohorts who stepped into their first jobs under market economy conditions. Moreover, the economic transformation dramatically reshaped educational choices of young Poles and triggered the tertiary education boom. In 1990 the gross enrolment rate in tertiary education was 13%, compared to 47% in 2016 (Central Statistical Office, 2009; 2017). As a result, the tertiary education boom gradually elevated number of university educated population. It suggests that expansion of tertiary education was not fully accommodated by the demand side.

Rising overeducation risk across cohorts might be also associated with deteriorating quality of tertiary education. Robst (1995b) finds that lower quality of tertiary education increases risk of overeducation. Expansion of tertiary education in Poland was achieved to large extent by gaining popularity of non-stationary studies, which accounted for 34% of students in 2016 (but even for about 50% ten years earlier). Non-stationary studies are believed to offer lower quality of tertiary education compared to regular study programmes (Herbst and Rok, 2010). What is more, tertiary education boom encouraged to establish numerous private universities, which in Poland usually offer lower quality of education compared to public universities. Thus, it suggests that younger cohorts of university graduates obtained on average lower quality of tertiary education compared to older cohorts. Furthermore, unfavourable composition of fields of tertiary education might have also played a role behind cohort shift in overeducation risk. Some fields of studies, for instance humanities, are associated with greater risk of overeducation (Quintini, 2011)[6].

Existence of cohort-persistent overeducation has important implications at individual and macroeconomic levels. It affects individuals' wage paths over life cycle as overeducated workers are penalised in terms of wage compared to properly matched workers (for instance: Duncan and Hofman, 1981; Verdugo and Verdugo, 1989; Groot and Maassen van den Brink, 2000; Rubb, 2003b). Persistent overeducation likely leads to gradual deterioration of human capital (de Grip et al., 2008). Furthermore, it is found that overeducation is associated with dissatisfaction (Battu, Belfield and Sloane, 1999; Green and Henseke, 2016). Tsang and Levin (1985) argue that low job satisfaction can impede firms' productivity.[7] For the economy as a whole overeducation means unused potential of human capital. Large degree of overeducation also means the waste of financial resources spent on university education.

Public policy addressing rising incidence of overeducation among tertiary educated workers might consist of demand and supply side measures. On the demand side, government might stimulate demand for tertiary educated workers. As overeducation is more prevalent in small towns and in rural areas, measures supporting spatial mobility of workers could decrease

[6] Although in logistic regressions I control for fields of studies, still some variation in fields of studies between cohort might be unaddressed due to grouping of fields into rough categories.

[7] This is not in line with findings by Kampelmann and Rycx (2012) or Mahy, Rycx and Vermeylen (2015) according to which overeducation rather improves productivity.

overeducation. The supply side measures should address the quality of tertiary education and field composition. Improved quality of university programmes and promoting education in fields which enjoy lower overeducation risk should decrease overeducation risk for future cohorts of graduates. Curbing the university enrolment rates might be also considered. As cohort effects are identified as an important factor behind overeducation expansion, special attention should be paid those abovementioned policies which shape cohorts' characteristics before their entrance to the labour market.

To conclude, the study sheds light on a relatively neglected aspect of overeducation which is existence of cohort effects. Some studies already suggested that cohort effects might explain differences in overeducation risk across individuals. However, there has been no study which would investigate cohort effects simultaneously with age and period effects. Existence of strong cohort effects behind educational mismatch in the labour market should direct researchers' and policymakers' efforts to understand better long-lasting determinants of overeducation. Although this study has been conducted for Poland, it is very likely that cohort shift in overeducation risks occurred in other post-transition economies which also experienced intensive tertiary education boom.

References

Baert, S., Cockx, B., & Verhaest, D. (2013). Overeducation at the Start of the Career: Stepping Stone or Trap? *Labour Economics*, 25, 123–40.

Baran, J. (2018). A side effect of a university boom: rising incidence of overeducation among tertiary educated workers in Poland. *Economics and Business Review EBR*, 18(2), 41-63.

Battu, H., Belfield, C.R. & Sloane, P.J. (1999). Overeducation Among Graduates: a cohort view, *Education Economics*, 7(1), 21–38.

Bell, A., & Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social Science & Medicine*, 93, 163–165.

Büchel, F., & Mertens, A. (2004). Overeducation, undereducation, and the theory of career mobility. *Applied Economics*, 35(8), 803–816.

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, 26, 3018–3045.

Central Statistical Office (2009). *Szkoły wyższe i ich finanse w 2008 r.* [Higher education institutions and their finances in 2008].

Central Statistical Office (2017). *Szkoły wyższe i ich finanse w 2016 r.* [Higher education institutions and their finances in 2016].

Clark, B., Joubert, C., & Maurel, A. (2017). The career prospects of overeducated Americans. *IZA Journal of Labor Economics*, 6:3.

de Grip, A., Bosma, H., Willems, D., & van Boxtel, M. (2008). Job-worker mismatch and cognitive decline, Oxford Economic Papers, 60(2), 237–253.

Deaton, A., & Paxson, C. (1994). Saving, Growth, and Aging in Taiwan, p. 331-362, in: *Studies in the Economics of Aging*, National Bureau of Economic Research.

Dolton, P., & Vignoles, A., (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review*, 19(2), 179-198.

Duncan, G., & Hoffman, S. (1981). The Incidence and Wage Effects of Overeducation. *Economics of Education Review*, 1, 57–68.

Frei, C., & Sousa-Poza, A. (2012). Overeducation: permanent or transitory? *Applied Economics*, 44(14), 1837-1847.

Green, F., & Henseke, G. (2016). Should governments of OECD countries worry about graduate underemployment?. *Oxford Review of Economic Policy*, 32(4), 514-537.

Green, F., & Zhu, Y. (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. Oxford Economic Papers, 62(4), 740-763.

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

Harding, D.J. (2009). Recent advances in age-period-cohort analysis. A commentary on Dregan and Armstrong, and on Reither, Hauser and Yang. *Social Science & Medicine*, 69(10), 1449-1451.

Heckman J., Robb R. (1985) Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations. In: Mason W.M., Fienberg S.E. (eds) Cohort Analysis in Social Research. Springer, New York, NY.

Herbst, M., & Rok, J. (2010). *The Equity of the Higher Education System: Case-Study Poland*, Document prepared for the World Bank.

Kampelmann, S., & Rycx, F. (2012). The impact of educational mismatch on firm productivity: Evidence from linked panel data. *Economics of Education Review*, 31, 918-931.

Kiersztyn, A. (2013). Stuck in a mismatch? The persistence of overeducation during twenty years of the post-communist transition in Poland. *Economics of Education Review*, 32, 78-91.

Korpi, T., & Tåhlin, M. (2009). Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000. *Labour Economics*, 16(2), 183-193.

Luo, L. (2013). Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem. *Demography*, 50, 1945-1967.

Mahy, B., Rycx, F., & Vermeylen, G. (2015). Educational Mismatch and Firm Productivity: Do Skills, Technology and Uncertainty Matter? *De Economist*, 163(2), 233–262.

Mavromaras, K., &. McGuinness, S. (2012). Overskilling dynamics and education pathways. *Economics of Education Review*, 31(5), 619-628.

McGuinness, S., Bergin, A., & Whelan, A. (2018). Overeducation in Europe: Trends, Convergence and Drivers. *Oxford Economic Papers*, 70(4), 994–1015.

Oreopoulos, P., von Wachter, T., & Heisz, A. (2012). The Short- and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics*, 4(1): 1-29.

Pouliakas, K. (2013). The Skill Mismatch Challenge in Europe. In Employment and Social Developments in Europe 2012. Luxembourg: Publications Office of the European Union.

Quintini, G. (2011). *Over-qualified or under-skilled: a review of the existing literature* (OECD Social, Employment and Migration Working Papers, No. 121).

Robst, J. (1995a). Career mobility, job match, and overeducation, Eastern Economic Journal, 21, 539–50.

Robst, J. (1995b). College Quality and Overeducation. *Economics of Education Review*, 14(3), 221-228.

Rubb, S. (2003a). Overeducation: a short or long run phenomenon for individuals? *Economics of Education Review*, 22(4), 389–394.

Rubb, S. (2003b). Overeducation in the Labor Market: a Comment and Re-Analysis of a Meta-Analysis, *Economics of Education Review*, 22, 621-629.

Rutherford, M.J., Lambert, P.C., & Thompson, J.R. (2010). Age-period-cohort modeling. *The Stata Journal*, *10*(4): 606-627.

Sasieni, P.D. (2012). Age-period-cohort models in Stata. *Stata Journal*, *12*(1), 45-60.

Sicherman, N. (1991). "Overeducation" in the Labor Market. *Journal of Labor Economics*, 9(2), 101-122.

Sicherman, N., & Galor, O. (1990). A Theory of Career Mobility. *Journal of Political Economy*, 98(1), 169-192.

Thurow, L. C. (1975). Generating Inequality. New York: Basic Books.

Tsang, M. C., & Levin, H. M. (1985). The economics of overeducation. *Economics of Education Review*, *4*(2), 93-104.

Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, 629-643.

Wen, L., & Maani, S.A. (2019). Job Mismatches and Career Mobility. *Applied Economics*, 51(10), 1010-1024.

Yang, Y., & Land, K.C. (2013). *Age-Period-Cohort Analysis : New Models, Methods, and Empirical Applications.* Boca Raton, FL: Chapman and Hall/CRC.

Yang, Y., Fu, W.J., & Land, K.C. (2004). A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological Methodology*, 34, 75-110.

Yang, Y., Schulhofer-Wohl, S., Fu, W. J., & Land, K. C. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it. *American Journal of Sociology*, 113(6), 1697-1736.

Appendix

Table 3. Results of logistic regression explaining chances of being overeducated, odds ratios

	50-percent threshold			70-percent threshold		
	(1)	(2)	(3)	(4)	(5)	(6)
age (ref.: 30)						
23	1.904 ***	2.128 ***	2.044 ***	2.770 ***	3.025 ***	3.047 ***
24	1.327 ***	1.441 ***	1.379 ***	1.781 ***	1.906 ***	1.863 ***
25	1.153 **	1.236 ***	1.237 ***	1.387 ***	1.474 ***	1.504 ***
26	1.129 *	1.198 ***	1.197 **	1.239 ***	1.303 ***	1.307 ***
27	1.065	1.116	1.130 *	1.124 *	1.171 **	1.194 **
28	1.029	1.059	1.075	1.056	1.083	1.101
29	1.046	1.060	1.059	1.062	1.073	1.070
31	1.026	1.021	1.031	1.018	1.012	1.017
32	1.036	1.007	1.023	1.025	0.998	1.004
33	1.084	1.032	1.039	1.034	0.987	0.988
34	1.060	0.997	0.985	1.038	0.979	0.973
35	1.016	0.947	0.939	1.004	0.941	0.935
36	1.067	0.974	0.965	1.045	0.959	0.948
37	1.056	0.945	0.928	1.023	0.920	0.907
38	1.058	0.928	0.913	1.066	0.941	0.927
39	1.104	0.943	0.932	1.137	0.980	0.971
40	1.088	0.917	0.896	1.107	0.941	0.921
41	1.099	0.915	0.839	1.100	0.925	0.861
42	1.051	0.865	0.836	1.047	0.869	0.839
43	1.122	0.905	0.899	1.106	0.901	0.900
44	1.108	0.878	0.867	1.130	0.904	0.894
45	1.147	0.891	0.881	1.141	0.898	0.885
46	1.124	0.850	0.874	1.119	0.859	0.871
47	1.056	0.796	0.770 *	1.086	0.827	0.812
48	1.027	0.768 *	0.752 *	1.065	0.806	0.798
49	1.019	0.749 *	0.727 *	1.068	0.795	0.783
50	1.089	0.787	0.774	1.151	0.841	0.829
51	1.124	0.800	0.774	1.204	0.867	0.854
52	1.274	0.896	0.890	1.267	0.902	0.907
53	1.273	0.877	0.853	1.306 *	0.913	0.912
54	1.221	0.836	0.807	1.283	0.890	0.887
55	1.245	0.839	0.808	1.293	0.882	0.875
56	1.229	0.803	0.783	1.373 *	0.915	0.901
57	1.423 *	0.921	0.890	1.624 ***	1.074	1.045
58	1.496 **	0.942	0.841	1.692 ***	1.093	1.012
59	1.559 **	0.976	0.883	1.718 ***	1.102	1.032
60	1.519 **	0.935	0.868	1.677 ***	1.062	1.010
61	1.553 **	0.934	0.814	1.820 ***	1.132	1.063
62	1.623 **	0.957	0.854	1.910 ***	1.171	1.102
63	1.684 **	0.970	0.857	1.966 ***	1.179	1.097
64	1.533	0.872	0.795	1.850 **	1.093	1.029
cohort (ref.: 1970)						
1941	0.316	0.572	0.604	0.204	0.384	0.399
1942	0.449	0.757	0.850	0.388	0.662	0.725
1943	0.323 **	0.517	0.630	0.265 ***	0.430 *	0.504
1944	0.293 **	0.457	0.524	0.255 ***	0.406 *	0.447 *
1945	0.457 **	0.677	0.757	0.345 ***	0.519 *	0.557
1946	0.594 *	0.861	0.966	0.460 ***	0.674	0.730
1947	0.599 *	0.882	1.056	0.478 ***	0.711	0.829
1948	0.579 **	0.827	0.929	0.468 ***	0.678 *	0.748
1949	0.641 **	0.888	0.975	0.503 ***	0.706 *	0.763
1950	0.613 **	0.837	0.904	0.534 ***	0.745	0.797
1951	0.660 **	0.902	0.980	0.589 ***	0.815	0.873
1952	0.694 **	0.940	1.001	0.613 ***	0.840	0.883
1953	0.759	0.978	0.994	0.685 **	0.892	0.902
1954	0.561 ***	0.742 *	0.736	0.532 ***	0.705 **	0.699 **
1955	0.710 **	0.915	0.941	0.653 ***	0.840	0.867
1956	0.665 **	0.838	0.874	0.608 ***	0.768 *	0.793
1957	0.723 **	0.882	0.895	0.636 ***	0.773 *	0.787
1958	0.686 **	0.846	0.833	0.647 ***	0.792	0.774
1959	0.631 ***	0.740 **	0.766	0.614 ***	0.720 **	0.734 **
1960	0.743 **	0.850	0.848	0.705 ***	0.808	0.814
1961	0.737 **	0.836	0.848	0.696 ***	0.784 *	0.807
1962	0.755 **	0.839	0.904	0.712 ***	0.787 *	0.868

1963	0.686	***	0.745	**	0.811		0.636	***	0.687	***	0.765	*
1964	0.712	***	0.766	**	0.820		0.678	***	0.728	***	0.795	*
1965	0.869		0.943		1.001		0.856		0.921		1.000	
1966	0.847		0.884		0.958		0.771	**	0.806	*	0.876	
1967	0.882		0.907		0.944		0.875		0.900		0.964	
1968	0.861		0.881		0.907		0.833	*	0.850		0.878	
1969	0.903		0.896		0.925		0.868		0.862		0.911	
1971	1.105		1.073		1.078		1.050		1.023		1.030	
1972	1.161		1.122		1.086		1.134		1.106		1.087	
1973	1.271	**	1.209	*	1.158		1.288	***	1.236	**	1.191	*
1974	1.237	**	1.185	*	1.073		1.271	***	1.224	**	1.121	
1975	1.370	***	1.259	**	1.124		1.402	***	1.301	***	1.170	
1976	1.442	***	1.307	***	1.198	*	1.495	***	1.370	***	1.274	**
1977	1.531	***	1.359	***	1.224	*	1.647	***	1.483	***	1.347	***
1978	1.680	***	1.470	***	1.322	**	1.779	***	1.576	***	1.442	***
1979	1.754	***	1.514	***	1.357	***	1.872	***	1.640	***	1.487	***
1980	1.917	***	1.629	***	1.477	***	2.046	***	1.766	***	1.621	***
1981	2.151	***	1.791	***	1.610	***	2.328	***	1.974	***	1.806	***
1982	2.221	***	1.809	***	1.636	***	2.394	***	1.986	***	1.818	***
1983	2.208	***	1.791	***	1.672	***	2.337	***	1.931	***	1.829	***
1984	2.540	***	2.038	***	1.898	***	2.735	***	2.232	***	2.083	***
1985	2.568	***	2.020	***	1.893	***	2.694	***	2.168	***	2.034	***
1986	2.373	***	1.878	***	1.756	***	2.581	***	2.079	***	1.976	***
1987	2.737	***	2.100	***	2.030	***	2.795	***	2.192	***	2.098	***
1988	2.858	***	2.192	***	2.111	***	2.901	***	2.268	***	2.197	***
1989	3.034	***	2.223	***	2.052	***	2.875	***	2.147	***	1.992	***
1990	3.337	***	2.440	***	2.492	***	3.065	***	2.286	***	2.358	***
1991	3.019	***	2.177	***	2.646	***	2.815	***	2.059	***	2.380	***
1992	3.028	***	2.042	***	1.996		2.570	***	1.791	**	1.744	
1993	3.287	**	2.223				3.019	*	2.092			
real GDP growth rate	1.010		1.007		1.007		1.010		1.007		1.008	
regional unemployment rate	1.031	***	0.991		0.993		1.030	***	0.992		0.993	
female			0.739	***	0.733	***			0.892	***	0.900	***
immigrant			0.979		1.021				1.000		1.028	
urbanisation. ref. 100 k inhabitants or more												
town, 20k-100k			1.391	***	1.405	***			1.395	***	1.423	***
town, less than 20k			1.388	***	1.406	***			1.354	***	1.390	***
rural area			1.732	***	1.740	***			1.657	***	1.679	***
region (ref.: Dolnośląskie)												
Kujawsko-Pomorskie			1.257	***	1.292	***			1.314	***	1.370	***
Lubelskie			1.387	***	1.445	***			1.394	***	1.456	***
Lubuskie			1.230	***	1.298	***			1.208	**	1.292	***
Łódzkie			0.939		0.963				1.007		1.041	
Małopolskie			0.838	***	0.841	***			0.875	**	0.886	**
Mazowieckie			0.971		0.915				1.016		0.962	
Opolskie			1.097		1.134				1.190	**	1.234	**
Podkarpackie			1.237	***	1.232	***			1.266	***	1.261	***
Podlaskie			1.616	***	1.611	***			1.627	***	1.624	***
Pomorskie			1.137	**	1.130	*			1.169	***	1.168	**
Śląskie			0.987		1.000				1.043		1.059	
Świętokrzyskie			1.420	***	1.439	***			1.600	***	1.642	***
Warmińsko-Mazurskie			1.332	***	1.344	***			1.319	***	1.346	***
Wielkopolskie			0.979		0.955				0.998		0.980	
Zachodniopomorskie			1.404	***	1.342	***			1.411	***	1.361	***
field of tertiary education (ref.: teaching)												
humanities and arts					0.930						1.018	
social sciences, business, law					2.370	***					2.656	***
science					1.229	***					1.366	***
computer sciences					0.826	***					0.926	
engineering, manufacturing					1.269	***					1.328	***
agriculture					2.614	***					2.841	***
health and welfare					0.404	***					0.559	***
services					3.225	***					4.071	***
Constant	0.191	***	0.285	***	0.194	***	0.248	***	0.318	***	0.192	***
Period	2006-2016		2006-2016		2006-2015		2006-2016		2006-2016		2006-2015	
Observations	345 967		345 967		306 595		345 967		345 967		306 595	
pseudo R^2	0.0359		0.0519		0.0912		0.0482		0.0601		0.1027	

*Notes: *** denotes statistical significance at 1-percent level, ** at 5-percent level, and * at 10-percent level.*
Source: Own calculations.

Table 4. Restricted cubic spline results

	50-percent				70-percent			
	coef.	s.e.	95%	c.i.	coef.	s.e.	95%	c.i.
_spA1_intct	-1.607	0.002	-1.610	-1.603	-1.376	0.002	-1.379	-1.373
_spA2	-0.134	0.002	-0.138	-0.130	-0.067	0.002	-0.071	-0.063
_spA3	-0.031	0.001	-0.033	-0.029	-0.080	0.001	-0.082	-0.078
_spA4	-0.011	0.001	-0.012	-0.009	0.003	0.001	0.002	0.004
_spA5	-0.045	0.001	-0.046	-0.043	-0.048	0.001	-0.050	-0.047
_spA6	-0.005	0.001	-0.006	-0.004	-0.006	0.001	-0.007	-0.005
_spP1	0.017	0.000	0.017	0.018	0.016	0.000	0.016	0.017
_spP2	-0.018	0.000	-0.019	-0.017	-0.018	0.000	-0.019	-0.018
_spP3	0.000	0.000	0.000	0.001	0.000	0.000	-0.001	0.001
_spP4	0.006	0.000	0.005	0.007	0.008	0.000	0.007	0.009
_spC1_ldrft	0.020	0.000	0.020	0.021	0.026	0.000	0.026	0.027
_spC2	-0.114	0.001	-0.116	-0.112	-0.078	0.001	-0.080	-0.076
_spC3	0.004	0.001	0.003	0.005	0.018	0.001	0.017	0.019
_spC4	0.019	0.000	0.018	0.020	0.027	0.000	0.026	0.028
_spC5	-0.039	0.001	-0.041	-0.037	-0.031	0.001	-0.032	-0.029
obs.	740				740			

Source: Own calculations.

Table 5. Intrinsic estimator results

	50-percent	70-percent		50-percent	70-percent
	age			cohort	
23	0.336 ***	0.242 ***	1947	-0.082 ***	-0.056 **
24	0.260 ***	0.185 ***	1948	-0.042 **	-0.034 *
25	0.166 ***	0.100 ***	1949	-0.053 ***	-0.013
26	0.108 ***	0.076 ***	1950	-0.028 **	0.008
27	0.085 ***	0.069 ***	1951	-0.008	-0.005
28	0.062 ***	0.054 ***	1952	-0.020 *	-0.003
29	0.045 ***	0.043 ***	1953	-0.008	0.000
30	0.029 ***	0.033 ***	1954	-0.056 ***	-0.054 ***
31	0.019 **	0.024 ***	1955	-0.022 **	-0.019 **
32	0.013	0.020 **	1956	-0.043 ***	-0.033 ***
33	0.011	0.024 ***	1957	-0.042 ***	-0.027 ***
34	0.008	0.021 **	1958	-0.043 ***	-0.038 ***
35	0.003	0.014	1959	-0.059 ***	-0.056 ***
36	-0.005	0.008	1960	-0.040 ***	-0.037 ***
37	-0.013	0.005	1961	-0.049 ***	-0.044 ***
38	-0.016 **	-0.004	1962	-0.051 ***	-0.045 ***
39	-0.014 *	-0.009	1963	-0.075 ***	-0.063 ***
40	-0.014 *	-0.006	1964	-0.070 ***	-0.064 ***
41	-0.021 **	-0.010	1965	-0.037 ***	-0.040 ***
42	-0.037 ***	-0.023 ***	1966	-0.061 ***	-0.048 ***
43	-0.038 ***	-0.025 ***	1967	-0.043 ***	-0.046 ***
44	-0.038 ***	-0.024 ***	1968	-0.060 ***	-0.057 ***
45	-0.036 ***	-0.024 ***	1969	-0.057 ***	-0.054 ***
46	-0.053 ***	-0.037 ***	1970	-0.037 ***	-0.042 ***
47	-0.052 ***	-0.038 ***	1971	-0.033 ***	-0.030 ***
48	-0.071 ***	-0.058 ***	1972	-0.023 ***	-0.025 ***
49	-0.082 ***	-0.070 ***	1973	-0.004	-0.015 *
50	-0.076 ***	-0.066 ***	1974	-0.013	-0.026 ***
51	-0.063 ***	-0.056 ***	1975	0.002	-0.010
52	-0.064 ***	-0.052 ***	1976	0.012	-0.005
53	-0.065 ***	-0.050 ***	1977	0.026 ***	0.001
54	-0.079 ***	-0.067 ***	1978	0.039 ***	0.016 *
55	-0.080 ***	-0.065 ***	1979	0.044 ***	0.019 **
56	-0.074 ***	-0.063 ***	1980	0.058 ***	0.033 ***
57	-0.059 ***	-0.066 ***	1981	0.086 ***	0.055 ***
58	-0.056 ***	-0.064 ***	1982	0.087 ***	0.060 ***
59	-0.039 ***	-0.041 ***	1983	0.077 ***	0.057 ***

period			1984	0.104 ***	0.081 ***
2006	-0.023 ***	-0.017 ***	1985	0.094 ***	0.064 ***
2007	-0.016 ***	-0.013 ***	1986	0.073 ***	0.029 ***
2008	-0.002	-0.002	1987	0.089 ***	0.083 ***
2009	-0.003	-0.003	1988	0.077 ***	0.071 ***
2010	0.008 **	0.008 **	1989	0.084 ***	0.094 ***
2011	-0.002	-0.004	1990	0.100 ***	0.125 ***
2012	-0.010 **	-0.011 **	1991	0.065 ***	0.081 ***
2013	0.002	0.000	1992	0.014	0.049 **
2014	0.009 **	0.008 *	1993	0.031	0.060 **
2015	0.014 ***	0.014 ***	constant	0.329 ***	0.276 ***
2016	0.023 ***	0.019 ***	obs	407	407

*Notes: *** denotes statistical significance at 1-percent level, ** at 5-percent level, and * at 10-percent level.*
Source: Own calculations.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl