Educational mismatches and earnings in Poland: are the graduates more penalized for being overeducated?

Leszek Wincenciak

Faculty of Economic Sciences, University of Warsaw

Jobs Conference: Technology, skills and inequalities
Institute for Structural Research
Warsaw, 27-28 October, 2015

Presentation outline

Motivation

- Mismatch = sub-optimal allocation of resources in the economy
- The question of rationality of spending public money on education
- From the point of view of employers – higher costs of recruitment, require some form of testing and screening to disclose the actual productivity of employees, require large resources to be invested in training and acquisition of skills to bridge the gaps
- Interest of policy makers – an impulse for the reforms of the education system, efficiency of the process of matching human resources through more effective guidance in the selection of educational paths and reduce the waste of common resources devoted to education

Literature findings

- The theoretical grounds for overeducation were laid out by the signalling theory of education (Spence, 1973)
- Alternative view – Thurow's theory of job-competition (1975)
- Duncan and Hoffman (1981) – new empirical approach (ORU) – returns to years of overeducation are half the returns to required years of schooling
- Groot and van den Brink (2000) – positive impact of labour force growth on the incidence of overeducation and a negative effect of unemployment on returns to education
- Allen and van der Velden (2001) – negative wage returns to skill mismatches (underutilisation of skills in current jobs), but wage penalty related to education mismatch is weak. Job satisfaction is also strongly negatively affected by skill mismatches but not by education mismatches

Literature findings

- Bauer (2002) – negative wage premium for the overeducated individuals and a positive one for undereducated ones
- Salas-Velasco (2006) – negative wage returns to underutilization of skills and positive ones to skill deficits
- Budria and Moro-Egido (2008) – wage dispersion within education groups to some extent can be explained by education mismatches. Incorrect qualification and strong mismatches are associated with a wage penalties that range from 13% to 27%
- Aracil and van der Velden (2008) – positive impact on wages of being overeducated as methodological competences pay off even when they are not required

Measurement of the mismatch

Three main approaches to the problem of measurement:

- workers' self-assessments;
- information based on job description;
- information based on realized worker-job matches.

Data for the analysis comes from the Labour Force Survey for Poland for 2013q2 and consists of 86 126 individuals aged 15 or more of which 37 098 are employed. It allows to use the third approach as we have information on the job characteristics (including 3 digit occupations) as well as on education completed.

Data and methodology

Basing on Kiker et al. (1997) – we define a mode of education level based on the education distribution among employees within each 3 digit occupation. Education is measured on a 6 level scale:

- tertiary MA (EDU = 1);
- tertiary BA (EDU = 2);
- secondary vocational (EDU = 3);
- secondary general (EDU = 4);
- vocational primary (EDU = 5);
- primary (EDU = 6).

Overeducation and undereducation dummies

To each 3-digit occupation a number of a dominant education level (EDU^{dom}) was assigned as the representation of required level of education. Then for every individual the following two dummy variables were created:

$$OVEREDUCATION_j = \begin{cases} 1 & \text{if} \quad EDU^{dom} > EDU_j \\ 0 & \text{if} \quad EDU^{dom} \leqslant EDU_j \end{cases}$$

$$UNDEREDUCATION_j = \begin{cases} 1 & \text{if} \quad EDU^{dom} < EDU_j \\ 0 & \text{if} \quad EDU^{dom} \geqslant EDU_j \end{cases}$$

ORU model: years of overeducation and undereducation

Second approach to the problem of measuring overeducation and undereducation relies on classical Duncan and Hoffman paper (1981) and decomposes the completed Y_j years of schooling for j-th individual into three components: Y_j^R — required years of schooling, $Y_j^O = \max(0, Y_j - Y_j^R)$ — years of overeducation, $Y_j^U = \max(0, Y_j^R - Y_j)$ — years of undereducation. The following identity holds (Leuven and Oosterbeek, 2011):

$$Y_j = \underbrace{\max(0, Y_j - Y_j^R)}_{O} + \underbrace{Y_j^R}_{R} - \underbrace{\max(0, Y_j^R - Y_j)}_{U}.$$

Data description and statistics

Subpopulation	Overeducated (%)	Undereducated (%)
Men	28.25	22.95
Women	21.95	28.68
Age up to 25	35.35	38.39
Age 26-35	33.19	20.51
Age 36-45	23.14	23.22
Age 46-55	19.24	24.72
Age 55+	19.43	33.31
Tertiary MA	18.86	0.00
Tertiary BA	51.85	43.42
Secondary vocational	47.10	16.94
Secondary general	41.99	59.51
Primary vocational	0.11	17.47
Primary	0.00	99.95
Total	24.82	24.93

Data description and statistics: variables used in the wage model

Variable name	Values	Share (%)
Sex	Man*	54.43
	Woman	45.57
Education	Tertiary, MA degree	23.95
	Tertiary, BA degree	10.42
	Secondary vocational	24.15
	Secondary general*	9.47
	Primary vocational	26.39
	Primary	5.62
Age	Up to 25	8.89
	26-35	29.16
	36-45	27.00
	46-55	22.22
	55+	12.73
Children under 5 in the household	0*	83.14
	1	14.40
	2	2.36
	3 and more	0.10

Data description and statistics: variables used in the wage model

Variable name	Values	Share (%)
Disability	Severe	0.21
	Moderate	1.45
	Light	1.38
	None*	96.97
Marital status	Single*	24.9
	Married	68.06
	Widowed	1.91
	Divorced, separated	5.13
Class of settlement unit	cities 100 th. and more	30.65
	cities 50-100 th.	9.92
	cities 20-50 th.	12.39
	cities 10-20 th.	6.91
	cities up to 10 th.	5.71
	Rural areas*	34.42

Data description and statistics: variables used in the wage model

Variable name	Values	Share (%)
Sector	Public*	32.72
	Private	67.28
Firm size	Up to 10*	17.85
	11-19	18.44
	20-49	16.86
	50-250	25.35
	251 and more	21.50
Current job tenure	Mean	112.34
	Standard dev.	113.07

Wage model with education dummies: specification

The model is stated in the following form:

$$\ln w_j = x_j\alpha + \beta_1 O_j + \beta_2 U_j + u_{1j},$$
$$y_0 = z_j\delta + u_{2j},$$
$$u_1 \sim N(0, \sigma),$$
$$u_2 \sim N(0, 1),$$
$$corr(u_1, u_2) = \rho.$$

x_j includes: sex, age (years), age squared, tenure in current job (months), education (six levels defined as above), firm size (5 levels), sector (public, private), class of settlement unit (6 levels), NACE code (2-digit level), voivodeship (16 geographical units). Additionally, two dummy variables for overeducation and undereducation are included.
z_j includes: sex, age, age squared, education, marital status, number of children aged 5 or less and a disability.

Wage model with education dummies: regression results

Variables	(1) All	(2) Age < 30	(3) Men	(4) Women
Wage equation				
Sex = woman	−0.1803**	−0.1111**		
	[0.000]	[0.000]		
EDU = Tertiary MA	0.4206**	0.1604**	0.3429**	0.5390**
	[0.000]	[0.000]	[0.000]	[0.000]
EDU = Tertiary BA	0.1652**	0.0294	0.1412**	0.2182**
	[0.000]	[0.229]	[0.000]	[0.000]
EDU = Secondary vocational	0.0326**	−0.0196	0.023	0.0507**
	[0.003]	[0.377]	[0.164]	[0.001]
EDU = Primary vocational	−0.1709**	−0.0904**	−0.1737**	−0.1881**
	[0.000]	[0.001]	[0.000]	[0.000]
EDU = Primary	−0.2970**	−0.1308**	−0.3071**	−0.3225**
	[0.000]	[0.000]	[0.000]	[0.000]
Overeducation	**−0.1427****	**−0.0816****	**−0.1348****	**−0.1607****
	[0.000]	[0.000]	[0.000]	[0.000]
Undereducation	**0.0667****	**0.0161**	**0.0662****	**0.0886****
	[0.000]	[0.445]	[0.000]	[0.000]
Observations	52 813	6 196	22 852	29 961
Uncensored observations	18 249	3 266	9 680	8 569
Wald chi^2	10712.20	940.99	5330.23	6118.29
p-value	[0.000]	[0.000]	[0.000]	[0.000]

Conclusions I

- Substantial and significant returns to formal education. Employees holding MA diplomas earn on average 42.1% more than their colleagues with secondary general education
- Returns to MA degree are also substantially higher among women than men
- There are no significant wage returns to tertiary education with BA diploma (with respect to the reference category) for youngest employees
- Overeducated workers exhibit significantly negative wage premium of 14.3% on average
- The incidence of overeducation is much more common among the young generations on the labour market in Poland, but quite surprisingly, the wage penalty is much lower for this group, reaching only 8.2%

Conclusions I, cont.

- Overeducation seems to penalize women more than men (16.1% and 13.5% respectively)
- Lower wage penalty for graduates suggests that their overeducation may not necessarily be a reflection of their consistent lower ability (which is implied by signalling theory of education). This effect may be due to the fact of more frequent job changes and gaining professional experience before moving to jobs to which they are better matched
- Being undereducated is associated with significant positive wage premium of 6.7% on average
- No significant premium for undereducation for the graduates
- Women exhibit slightly higher wage premium for undereducation than men, that is 8.9% compared to 6.6%

Duncan and Hoffman (1981) specification (ORU model)

Estimation of Duncan and Hoffman (1981) wage equation requires including three variables: Y_j^R — required years of schooling, Y_j^O — years of overeducation, Y_j^U — years of undereducation in place of two dummy variables for overeducation and undereducation status. The wage equation then becomes:

$$\ln w_j = x_j\alpha + \gamma_r Y_j^R + \gamma_o Y_j^O \gamma_u Y_j^U + u_j.$$

As Leuven and Oosterbeek (2011) note, a convenient feature of this specification is that it allows to test standard Mincer specification as a special case of the above, which is done by not statistically rejecting the hypothesis of $\gamma_r = \gamma_o = -\gamma_u$.

Duncan and Hoffman (1981) specification (ORU model): regression results

Variables	(1) All	(2) Age < 30	(3) Men	(4) Women
Wage equation				
Sex = Woman	−0.1743**	−0.1122**		
	[0.000]	[0.000]		
γ_r	0.1104**	0.0532**	0.1008**	0.1313**
	[0.000]	[0.000]	[0.000]	[0.000]
γ_o	0.0458**	0.0042	0.0419**	0.0588**
	[0.000]	[0.507]	[0.000]	[0.000]
γ_u	−0.0557**	−0.0212**	−0.0412**	−0.0785**
	[0.000]	[0.001]	[0.000]	[0.000]
Observations	52 778	6 193	22 818	29 960
Uncensored observations	18 214	3 263	9 646	8 568
Wald chi^2	10563.32	950.95	5321.59	5909.41
p-value	[0.000]	[0.000]	[0.000]	[0.000]
Tests				
$\gamma_r = \gamma_o = -\gamma_u$	994.07	88.68	440.56	608.18
	[0.000]	[0.000]	[0.000]	[0.000]
$\gamma_o = -\gamma_u$	6.59	4.02	0.01	13.70
	[0.0102]	[0.0449]	[0.9043]	[0.0002]

Conclusions II

- Return to each year of required schooling is 11% for entire population, which seems a little higher than what is observed for most of the developed economies
- Women exhibit higher returns than men, although they still do earn less on average. The return to each year of overeducation is roughly 4.6%, which is approximately half the return to required years of education (as is commonly found in the literature)
- The return to each year of required schooling is much smaller for graduates (population under 30)
- There are no wage gains for years of overeducation for the graduates. Undereducation brings a yearly return of about 5.6% — slightly less in absolute terms for the youngest and for men

Conclusions II, cont.

- The hypothesis of Mincer model being a special case of DH81 specification is strongly rejected in all specifications
- Symmetry of returns to years of overeducation and undereducation is not rejected only for men
- Thurow's hypothesis of the years of required schooling being only significant predictors of the wage is also rejected by the data, as in all specifications both return to each year of underschooling and to each year of overschooling are significantly different from zero — excluding the case of return to overschooling for graduates

Summary

- Estimation of wage models resulted in finding significant positive wage returns to undereducation status and also significant wage penalties associated with being overeducated
- The problem of wage penalty for youngest graduates does not seem to be severe, which suggests to reflect that overeducation for them is not necessarily a sign of lower ability, but rather a sign for gathering experience before moving to better matched jobs
- Panel data analysis utilizing information of job transitions would be desired to confirm this hypothesis which is a direction for future research

Summary

Thank you!