View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elektronische Publikationen der Wirtschaftsuniversität Wien

Romanian Journal of Regional Science
The Journal of the Romanian Regional Science Association
Vol. 5. No. 1, Summer 2011

DECOMPOSING INCOME DIFFERENTIALS BETWEEN ROMA AND NON-ROMA IN SOUTH EAST EUROPE

Susanne Milcher

Address for correspondence:
Vienna University of Economics and Business
Nordbergstr. 15/4/A, A-1090 Vienna, Austria
Email: susanne.milcher@gmail.com

Biographical note:

Susanne Milcher holds a master degree in economics from Maastricht University, Netherlands, and currently works as a Policy Specialist in Social Inclusion and Poverty Reduction at the United Nations Development Programme in Bratislava. She focuses on poverty analysis, human development and minority integration. This paper is part of her PhD thesis at Vienna University of Economics and Business focusing on vulnerability and labour market discrimination of Roma in South East Europe, in progress at the Vienna University of Economics and Business.

Abstract. The paper decomposes average income differentials between Roma and non-Roma in South East Europe into the component that can be explained by group differences in income-related characteristics (characteristics effect), and the component which is due to differing returns to these characteristics (coefficients or discrimination effect). The decomposition analysis is based on Blinder (1973) and Oaxaca (1973) and uses three weighting matrices, reflecting the different assumptions about income structures that would prevail in the absence of discrimination. Heckman (1979) estimators control for selectivity bias. Using microdata from the 2004 UNDP household survey on Roma minorities, the paper finds that a large share of the average income differential between Roma and non-Roma is explained by human capital differences. Nevertheless, significant labour market discrimination is found in Kosovo for all weight specifications and in Bulgaria and Serbia for two weight specifications.

JEL classification: J71, C11, C50, O52

Keywords: Labour market discrimination, income differential decomposition, Heckman selection model, Roma, Europe

1 Introduction

Roma are one of the main poverty risk groups in South East Europe. Indeed, Milcher (2009) has shown that they are both poorer than non-Roma living in a similar socio-economic environment and are more likely to fall into poverty, or remain poor in the future. The sources of Roma vulnerability to poverty are intertwined and lead to a vicious circle of Roma poverty and labour market disadvantage. Poverty can be partly explained by the education gaps of Roma vis-à-vis the non-Roma that lead to low skill levels and subsequent weak labour market chances. In fact, according to UNDP (2006) Roma in South East Europe are disproportionately employed in low-quality jobs and depend primarily on income from work in the informal sector (34%) and on social transfers (33%). Further, even for Roma that achieve higher education levels, employment opportunities improve more slowly than for non-Roma with similar levels of education. Further, Milcher (2006) empirically established that the incidence of income poverty for the Roma is likely to be higher than for the non-Roma *irrespective of educational achievement.*

The apparently lower returns to education vis-à-vis incomes for Roma indicate that other factors, such as labour market discrimination may also be responsible for existing income differentials between employed Roma and non-Roma. Labour market discrimination in this context is generally referred to when some workers have higher wage incomes than others with the same endowment of productive economic characteristics by virtue of some non-economic personal characteristic (such as race, sex, class, caste, etc.). While low levels of educational achievement are considered as one of the most significant factors explaining labour market disadvantage of Roma in South East Europe, it is less known what portion of the existing income differential between employed Roma and non-Roma is attributable to differences in human capital characteristics, and what portion can be said to be due to other factors, i.e. labour market discrimination.

Given the increased attention and public expenditures pledged to Roma education in South East Europe within the political framework of the Decade of Roma Inclusion[1], this paper shares the

[1] Since 2005 governments of Albania, Bulgaria, Bosnia and Herzegovina, the Czech Republic, Croatia, Hungary, the Former Yugoslav Republic of Macedonia, Montenegro, Romania, Serbia and Slovakia have pledged to close the gap in welfare and living conditions between Roma and the non-Roma in their countries, and to break the vicious circles of poverty and social exclusion (http://www.romadecade.org).

ambition to analyse education- versus labour market discrimination-based explanations of income differentials between Roma and non-Roma in Albania, Bulgaria, Croatia, Kosovo and Serbia. The focus of this study is on income differentials (differences in average wage incomes) rather than wage differentials due to data limitations. This paper is based on statistical decomposition analysis, popularized by Oaxaca (1973) and Blinder (1973) but departs from this previous work in several respects. *First*, most empirical work on labour market discrimination (see, for example, Patrinos and Sakellariou 1992, Kimmel 1997, Maani 2002) assumes that discrimination penalises the minority group by preventing them from earning wage incomes according to the majority income structure. One could also argue, however, that discrimination gives the majority group an undeserved advantage and results in higher incomes for this group. According to Oaxaca and Ransom (1994) such an assumption would explain why majority groups resist antidiscrimination policies. Therefore, the model employed in this paper specifies the different assumptions about the competitive wage structure in the absence of discrimination, according to Blinder (1973), Oaxaca (1973) and Reimers (1983).

Second, wage incomes are observed only for people who are participating in the labour force and this might be a selective group. Consequently, the parameter estimates of the wage income characteristics can be biased and inconsistent. In order to control for potential sample selection, which may result from decisions people make about labour market participation, the paper includes a selection-correction variable in the wage determination equation (see Heckman 1979).

Third, the paper uses the most recent available comparative data source, the 2004 United Nations Development Programme (UNDP) dataset. This survey gives information about the living standard of Roma compared to non-Roma across South East Europe, including Albania, Bulgaria, Croatia, Kosovo, and Serbia. The survey has been designed in close cooperation with Roma experts and the Roma community in order to overcome problems with self-identification and sampling.

The remainder of the paper is organised as follows. The section that follows briefly describes the standard Blinder-Oaxaca approach to decomposition analysis. The model specification reflects three assumptions about the competitive income structure in the absence of discrimination. Section 3 outlines the selectivity bias adjustment that is pertinent to the decomposition analysis in this study. Section 4 proceeds to describe the variables and data, and section 5 presents the empirical results. Finally, Section 6 closes the paper.

2 The standard approach to wage decomposition analysis

The seminal work of Blinder (1973) and Oaxaca (1973) introduces the decomposition of the mean wage differential between demographic groups into a part that is attributable to group differences in human capital endowments (characteristics effect), and a part attributable to group differences in returns to these endowments (coefficients or discrimination effect). The underlying assumption of this methodology is that despite equal ability or human capital endowments, two demographic groups receive unequal earnings. The part of the wage gap that cannot be attributed to the characteristics effect, may therefore be attributed to wage discrimination on the basis of gender, race and/or ethnicity.

According to Becker (1975) and Mincer (1974) the wage rate of population group *j* (*j*=1, 2) reflects the productivity potential based on various human capital characteristics, as expressed in the linear regression relationship in Eq. (1)

$$Y_j = X_j \; \beta_j + \varepsilon_j \qquad j \in \{1,2\} \tag{1}$$

where Y_j represents the *nj*-by-1 vector of wage rates measured in logarithmic terms. X_j is a *nj*-by-*k* matrix containing $k-1$ column vectors representing human capital attributes, such as education, work experience, occupation that determine the wage rate of individuals in group *j*, as well as a column vector of ones related to the intercept term. Note that the individuals in both groups are characterized by the same $k-1$ attributes. β_j is a *k*-by-1 parameter vector reflecting the responsiveness of wages to the various characteristics, and the constant. The error term ε_j reflects the measurement error and is assumed to be independent and identically distributed.

The generalised wage decomposition, suggested by Oaxaca and Ransom (1994), is obtained by estimating separately Eq. (1) for Roma (*j*=1) and non-Roma (*j*=2) using consistent parameter estimates of β_j (*j*=1, 2) from ordinary least-squares estimation. Eq. (2) expresses the average wage differential $\overline{Y}_2 - \overline{Y}_1$ between Roma and non-Roma as the difference in the linear prediction at the group-specific means $\overline{X}_j$ (*j*=1, 2) of the regressors

$$\overline{Y}_2 - \overline{Y}_1 = (\overline{X}_2 - \overline{X}_1)\,\beta^* + [\overline{X}_2\,(\hat{\beta}_2 - \beta^*) + \overline{X}_1\,(\beta^* - \hat{\beta}_1)] \tag{2}$$

where the first $[(\overline{X}_2 - \overline{X}_1)\beta^*]$ and the second $[\overline{X}_2\,(\hat{\beta}_2 - \beta^*) + \overline{X}_1\,(\beta^* - \hat{\beta}_1)]$ components of the average log wage differential $(\overline{Y}_2 - \overline{Y}_1)$ represent the characteristics effect (denoted by *C*) and the coefficients or discrimination effect[2] (denoted by *D*), respectively. β^* is a *k*-by-1 vector of non-discriminatory coefficients and the constant, which reflects the wage structure that would prevail in the absence of discrimination. However, β^* is unknown and needs to be estimated. Oaxaca (1973) suggests that either the Roma ($\beta^* = \hat{\beta}_1$) or non-Roma ($\beta^* = \hat{\beta}_2$) wage structure would prevail in the absence of discrimination. This is considered the "index number problem" (Cotton, 1988; Neumark, 1988). Cotton (1988) argues that neither the group *j*=1 nor the group *j*=2 wage structure would prevail in the absence of discrimination. Instead, the non-discriminatory wage structure lies somewhere in between the population groups' wage structures ($\beta^* = 0.5\hat{\beta}_1 + 0.5\hat{\beta}_2$), as proposed by Reimers (1983).

As Oaxaca and Ransom (1994) propose, Eq. (2) can also be expressed as

$$\overline{Y}_2 - \overline{Y}_1 = (\overline{X}_2 - \overline{X}_1)[(W\hat{\beta}_2 + (I - W)\hat{\beta}_1)] + [(\overline{X}_2(I - W) + \overline{X}_1 W)](\hat{\beta}_2 - \hat{\beta}_1) \tag{3}$$

---

2 If in the absence of discrimination Roma and non-Roma would receive identical returns for the same characteristics, and differences in wages would thus be due only to differences in pay-related characteristics, then this coefficients effect can be interpreted as the part of the log wage differential due to discrimination. This is the essence of the Blinder-Oaxaca approach (Neumark 1988). However, unobserved factors, such as cultural differences, lifestyle, work ethics or prior discrimination in the education system are not accounted for in the wage equation but may exert influence on wages and thereby cause omitted variable bias and may overestimate the discrimination estimate. Therefore, it is suggested to consider this component of the wage gap as an 'upper bound' estimate of labour market discrimination.

where W represents a k-by-k matrix of relative weights given to the coefficients of the non-Roma and I is a k-by-k identity matrix.

The assumption about the proper choice of the weights depends on assumptions about the wage structure that would prevail in the absence of discrimination[3]. In the two cases proposed by Oaxaca (1973), the weighting matrix, W, is equal to the null matrix or equal to the identity matrix, respectively (W=I is also suggested by Blinder 1973). In the case of W=I, it is assumed that discrimination penalizes the minority group by preventing minority workers from receiving wage incomes according to the majority wage income structure. MacIssac and Patrinos (1995) argue that this would constitute a situation, whereby the majority workers (non-Roma) would not have any objections to ending discrimination, since their own wage incomes would not be affected. On the contrary, it could be assumed that discrimination gives the majority workers an undeserved advantage, and that they receive higher wage incomes than what they would get in the absence of discrimination (W=null matrix). In this case, it seems that the minority workers (Roma) would not have any economic reason for desiring discrimination to end, since their wage incomes would be unaffected by the change. The approach, however, hints that the "true" non-discriminatory wage structure lies somewhere in between the two population groups wage structures. Therefore, Reimers (1983) uses W=0.5I. Cotton (1988) proposes using the relative group size as a weight. Alternatively, Neumark (1988) proposes using the coefficients from a pooled model for both groups, which is also proposed by Oaxaca and Ransom (1994).

While decomposition is intuitive, most empirical studies omit drawing inferences regarding statistical significance of the decomposition components. Oaxaca and Ransom (1998) propose an asymptotic approximation to the variance of the decomposition effects estimates, based on a linear Taylor series expansion around the true - but unknown - parameter vector, given by

$$\mathrm{var}\left(\hat{D}\right) = \left(\hat{D}+1\right)^2 \ \bar{X}_2' \left(\Sigma_1 + \Sigma_2\right) \bar{X}_2 \tag{4a}$$

---

3 Examples for β^* are: $\beta^* = \hat{\beta}_1$ corresponds to W=null matrix while $\beta^* = \hat{\beta}_2$ corresponds to W=I (Oaxaca 1973, Blinder 1973) and $\beta^* = 0.5\hat{\beta}_1 + 0.5\hat{\beta}_2$ corresponds to W=0.5I (Reimers 1983).

$$\Sigma_1 = \hat{\sigma}^2_{\varepsilon 1} \left(X_1' X_1\right)^{-1} \tag{4b}$$

$$\Sigma_2 = \hat{\sigma}^2_{\varepsilon 2} \left(X_2' X_2\right)^{-1} \tag{4c}$$

where the noise variance estimates $\hat{\sigma}^2_{\varepsilon 1}$ and $\hat{\sigma}^2_{\varepsilon 2}$ are typically constructed using the least-squares residuals from the group 1 (Roma) and group 2 (non-Roma) regressions, respectively.

The computation of the decomposition components is straightforward, if the process governing the decision of labour market participation is random. In reality, however, this process can depend on a variety of characteristics, such as household size, number of children, health, poverty or marital status, etc. If this is the case, Eq. (1) may be subject to selectivity bias.

3 Selectivity bias

Since wage structures may be affected by decisions people make about labour market participation, selection bias (Heckman, 1979) plays an important role in estimating unbiased parameters of wage rate characteristics. The classical decomposition model, presented above, does not take into account sample selection bias that may occur, if those individuals that do not participate in the labour force are not a random sample. Therefore, Eq. (1) further depends on labour market participation choices of the individuals in group j (j=1, 2) expressed as

$$S_j = Z_j \gamma_j + \upsilon_j \tag{5}$$

where S_j is a nj-by-1 vector associated with labour market participation for individuals in population group j. Note that S_j>0 indicates labour market participation of individuals in group j. The nj-by-k_S matrix Z_j contains $k_S - 1$ column vectors[4] representing socio-economic attributes determining labour market participation characterizing individuals in population group j, as well as a column vector of

[4] Note that the individuals in both groups are characterized by the same k_S attributes. Further k_S is different from k, since labour market selectivity bias is driven by the notion that some of the socio-economic characteristics of the individual determining the probability to participate in the labour market are different from those determining the wage rate.

ones related to the intercept term. γ_j is a k_S-by-1 parameter vector, which reflects the responsiveness of labour market participation to the socio-economic attributes for the two demographic groups, as well as the constant. The *nj*-by-1 disturbance vectors υ_j and ε_j from Eq. (1) follow a bivariate normal distribution $(0, 0, \sigma^2_{\upsilon_j}, \sigma^2_{\varepsilon_j}, \rho)$ where ρ is the correlation between the disturbance vectors.

The probability of labour market participation for individuals in group *j*=1 (Roma) and *j*=2 (non-Roma) is expressed as

$$\begin{aligned} \Pr(S_j > 0) &= \Pr(\upsilon_j > -Z_j\gamma_j) \\ &= \Phi(Z_j\gamma_j) \end{aligned} \tag{6}$$

where Φ is the *standard normal cumulative density function*. Wages are only observed for individuals in group *j* who are participating in the labour market, so that their expected wage rate is determined according to

$$\begin{aligned} E(Y_j \mid S_j > 0) &= X_j\beta_j + E(\varepsilon_j \mid \upsilon_j > -Z_j\gamma_j) \\ &= X_j\beta_j + \lambda_j\theta_j \end{aligned} \tag{7}$$

where $\theta_j = \rho\sigma_{\varepsilon_j}, \lambda_j = \phi(Z_j\gamma_j) / \Phi(Z_j\gamma_j)$, and ϕ is the *standard normal density function*. The term *"λ_j"* refers to the Inverse Mills Ratio (IMR), reflecting the probability of participation.

If participation in the labour market is not random, given the observed characteristics, so that $E(\varepsilon_j \mid \upsilon_j > -Z_j\gamma_j) \neq 0$, the average observed wage, as well as the least-squares estimates of the coefficients of Eq. (1) are subject to selectivity bias. Consistent estimates of Eq. (1) can, however be obtained using the procedure suggested by Heckman (1979), which first estimates Eq. (5), the probability of labour market participation, and then adds the Inverse Mill's Ratio as an additional explanatory variable into Eq. (1). The Mill's ratio, a proxy for the probability of labour force participation, then controls for the expected error in the wage, given that the individual worked in the wage sector so that her wage is observed. Equation (1) for workers of group *j*=1 (Roma) and *j*=2 (non-Roma), participating in the labour market, is then expressed as

$$Y_j \left| (S_j > 0) = X_j \beta_j + \lambda_j \theta_j + \mu_j \right. \tag{8}$$

where the matrix X_j contains $k-1$ column vectors representing human capital attributes, such as education, work experience, occupation that determine the wage rate of individuals in group *j*, as well as a column vector of ones related to the intercept term. β_j is a *k*-by-1 parameter vector reflecting the responsiveness of wages to the various characteristics, and the constant. λ_j is a *nj*-by-1 vector of the inverse Mill's ratio from the sample-inclusion probit for *nj* individuals in demographic group *j*, θ_j is a scalar reflecting the covariance $\rho\sigma_{\varepsilon_j}$ between the disturbance vectors in the participation and wage equations (to be estimated), and μ_j is an $N(0,\Sigma_j)$ error term.

The most straightforward approach to control for selectivity bias in the decomposition is to subtract the selection effects from the overall wage differential and then to apply the standard decomposition to this adjusted differential (Reimer 1983)[5]. The selectivity adjusted wage decomposition differentials of interest are obtained as

$$(\overline{Y}_2 - \overline{Y}_1) - (\overline{\lambda}_2 \hat{\theta}_2 - \overline{\lambda}_1 \hat{\theta}_1) = \underbrace{(\overline{X}_2 - \overline{X}_1)[(W\hat{\beta}_2 + (I-W)\hat{\beta}_1)]}_{C} + \underbrace{[(\overline{X}_2(I-W) + \overline{X}_1 W)](\hat{\beta}_2 - \hat{\beta}_1)}_{D} \tag{9}$$

where $\overline{\lambda}_j \hat{\theta}_j$ represent the estimate of the selectivity bias in the average observed wage for the two demographic groups with $\overline{\lambda}_j$ representing the average of the Inverse Mills Ratio, and $\hat{\theta}_j$ the estimate of $\rho\sigma_{\varepsilon_j}$. $(\overline{X}_2 - \overline{X}_1)[(W\hat{\beta}_2 + (I-W)\hat{\beta}_1)]$ and $[(\overline{X}_2(I-W) + \overline{X}_1 W)](\hat{\beta}_2 - \hat{\beta}_1)$ represent the familiar characteristics and discrimination components, *C* and *D*, respectively.

5 Other approaches have been proposed for example by Yun (2007) who suggests a 'generalized selection bias' approach using the sample average of residuals of wages for estimating the selection effects. Neuman and Oaxaca (2004) propose further decomposition of the selection bias into characteristics and coefficients effect.

4 Data and variables

The paper applies this selectivity bias adjustment to the decomposition analysis of wage differentials between Roma and non-Roma in five South East European countries, including Albania, Bulgaria, Croatia, Kosovo and Serbia. We use three weighting matrices, W=0, W=I and W=0.5I, reflecting different assumptions about wage income structures that would prevail in the absence of discrimination, as proposed by Blinder (1973), Oaxaca (1973) and Reimers (1983). Data used come from the 2004 UNDP survey of Roma minorities. The survey followed the structure of an integrated household survey and thus contains, for each country, individual and household level information.

In the five countries under study, in total 9,889 Roma and 7,438 non-Roma respondents were surveyed[6]. The overall samples are representative of the Roma population living in Roma settlements or areas of compact Roma population (where the share of Roma population at least equals the national share of Roma population in the given country). Since the census is not a reliable source on the absolute size of the Roma population in these countries, the sampling methodology had to rely on various sources, including experts' estimates of Roma population shares. Since Roma who are more dispersed in the general population fall out of this sampling methodology, the samples cover roughly 85 percent of Roma in each country. The non-Roma samples – households living in close spatial proximity (same municipalities or administrative units) to Roma households – were constructed in each country using similar procedures as for the Roma samples. In order to overcome the possible distrust to enumerators, Roma interviewers were trained as interviewers (see UNDP 2006 for more information on the overall survey).

The analysis is limited to working-age individuals with age between 16-65 years who have declared wage from employment as their main source of income. This restriction ensures that the sub-sample, used in this study, only includes employed individuals and excludes self-employed, agricultural workers, those working only for subsistence and in the shadow economy. Missing data

6 The overall samples for the countries subject to this study are: 2,479 Roma and 1,876 non-Roma individuals in Albania, 2,176 Roma and 1,302 non-Roma individuals in Bulgaria, 1,252 Roma and 715 non-Roma individuals in Croatia, 2,223 Roma and 2,275 non-Roma individuals in Kosovo as well as 1,759 Roma and 1,270 non-Roma individuals in Serbia.

on some independent variables did lead to a further reduction in the country-specific sample sizes. As a result, the final sub-samples selected for this study comprise 289 Roma and 570 non-Roma individuals in Albania, 241 Roma and 370 non-Roma individuals in Bulgaria, 77 Roma and 219 non-Roma individuals in Croatia, 123 Roma and 280 non-Roma individuals in Kosovo as well as 111 Roma and 341 non-Roma individuals in Serbia. The differences in sub-sample sizes between Roma and non-Roma populations are due to smaller proportions of Roma with wage income as major source of income in the respective countries.

A major drawback of the UNDP dataset is that no data on actual wages have been collected. Hence, in this study wage income is used as a proxy for actual wages. Sources of income may, however be different for Roma and non-Roma, which could seriously bias estimates of the rates of return to education, according to Blinder (1973). However, the construction of the sub-samples of Roma and non-Roma workers, as described above, justifies the use of income as a proxy for wages.

The natural logarithm of income is regressed on six explanatory variables and an intercept term, specified by the matrix X_j (j=1, 2) in Eq. (8). Table 1 provides the full list of variables used in the analysis. The education variable measures the number of years of schooling. The UNDP survey does not contain information on the actual number of years of work experience. Therefore, age is used as a proxy for potential work experience. The square of this variable is included to capture decreasing marginal returns to experience. Two dummy variables measure the occupational status of the individuals, full time, indicating whether the individual works full time, and high skills, indicating whether the individual is occupied in a skilled (blue or white collar) employment. These variables control for human capital characteristics differencing the Roma and non-Roma population groups. A male dummy is added to control for gender-specific effects[7]. The matrix Z_j (j=1, 2) in Eq. (5) is specified by the same six independent variables, but one more dummy variable is added, capturing the poverty status of the individual. The dummy variable takes the value of one, if the individual is living in a poor household, and zero otherwise. S_j (j=1, 2) in Eq. (5) is taking the value of one, if the individuals in group j are participating in wage employment, and zero otherwise.

[7] The male dummy only accounts for gender effects in the wage regression relationship of Roma and non-Roma. It cannot account for gender wage discrimination among both groups. This requires separately estimating wage equations for Roma women and Roma men (non-Roma women and men, respectively). In the context of this study, small sample sizes do not allow to pursue this further.

Table 2 presents descriptive statistics of the dependent and explanatory variables. The descriptive statistics illustrate that Roma in all five countries have lower average wage incomes than non-Roma. The average income differentials range from 0.26 in Croatia to 0.70 in Albania. Roma are also worse off than non-Roma with respect to productive endowments. The average level of schooling of the two groups differs substantially. Roma in Albania, for example have only half of non-Roma's average years of schooling. Roma are also less represented in high skills and high quality forms of employment than non-Roma.

Table 1 Variables used in the analysis

Variable	Variable definition
Wage income estimation	
Income	natural log of wage income [in Euro] per month
Education	number of years of schooling
Work experience	age of individual in years [potential work experience]
Work experience squared	age (in years) squared in 100
Full time	a dummy variable taking the value of one if the individual works full time, and zero otherwise
High skills	a dummy variable taking the value of one if the individual is engaged in a skilled occupation, and zero otherwise
Male	a dummy variable taking the value of one if male, and zero otherwise
Labour market participation estimation	
Labour market participation	a dummy variable taking the value of one if the individual is participating in wage employment
Education	number of years of schooling
Work experience	age of individual in years [potential work experience]
Work experience squared	age (in years) squared in 100
Full time	a dummy variable taking the value of one if the individual works full time, and zero otherwise
High skills	a dummy variable taking the value of one if the individual is engaged in a skilled occupation, and zero otherwise
Male	a dummy variable taking the value of one if male, and zero otherwise
Poverty	a dummy variable taking the value of one if the individual is living in a household with equivalent household expenditures below 4.3$PPP, and zero otherwise

Table 2 Description of the variables

	Albania		Bulgaria		Croatia		Kosovo		Serbia	
	Roma	Non-Roma	Roma	Non-Roma	Roma	Non-Roma	Roma	Non-Roma	Roma	Non-Roma
Variables (means and standard deviations in brackets)										
Log income	4.47	5.17	4.26	4.75	5.86	6.12	4.75	5.27	4.87	5.19
	(0.69)	(0.63)	(0.55)	(0.46)	(0.62)	(0.62)	(0.82)	(0.82)	(0.74)	(0.73)
Education	6	12	7	12	9	13	7	12	9	13
(no. of school years)	(3.65)	(2.83)	(3.09)	(2.60)	(3.05)	(2.69)	(3.16)	(2.54)	(3.09)	(2.55)
Work experience	36	41	38	40	32	38	35	38	39	41
(age in years)	(10.37)	(10.37)	(11.10)	(10.20)	(9.77)	(11.65)	(11.19)	(11.74)	(10.50)	(10.49)
Work experience	14	18	16	17	11	16	14	16	16	18
squared in 100	(7.99)	(8.22)	(8.69)	(8.27)	(6.45)	(9.36)	(8.59)	(9.50)	(8.12)	(8.47)
Dummy variables (percentage of sample, with each level of variable)										
Full time work										
yes	53	89	71	95	87	93	54	82	68	94
no	47	11	29	5	13	7	46	18	32	6
High skills										
yes	69	89	20	74	44	93	27	68	47	94
no	31	11	80	26	56	7	73	32	53	6
Male										
yes	73	61	66	51	71	53	90	83	82	55
no	27	39	34	49	29	47	10	17	18	45

5 Empirical results

Selectivity bias occurs in Albania for both populations groups and in Kosovo and Serbia for the non-Roma population group[8]. Consequently, a decomposition analysis based on ordinary least squares estimators would yield biased and inconsistent results. In these cases, the Heckman estimators based on Eq. (8) are used for the decomposition analysis, specified in Eq. (9). In the other cases, the decomposition analysis is based on consistent estimates from ordinary least-squares estimation (Eq. (1)).

The results of the wage income regression for Roma and non-Roma are summarized in Table 3. The first four columns present the results of estimating labour market participation (Eq. (5)). The last four columns present the parameter estimates of the wage income regression for the two demographic groups (j=1: Roma, j=2: non-Roma) along with p-level calculations (in brackets) and standard errors.

The variables which are significant in predicting the incomes of Roma workers are country-specific and reflect structural differences between the national economies. Work experience of Roma is associated with positive, yet diminishing returns in Croatia, whereas Roma in Albania, Bulgaria, Kosovo and Serbia are not rewarded for work experience. Education has only positive and significant impacts on Roma income in Albania and Bulgaria but not in Croatia, Kosovo and Serbia. With the exception of Albania, the occupation status (working full time and in a skilled occupation) has a positive and significant impact of incomes of Roma in all countries. Only in Croatia, the full time variable is not significant. In Albania and Croatia, Roma men have higher incomes than Roma women. Only in Albania, labour market participation of Roma exhibits sample selection. The coefficients of the labour market participation estimation in Albania demonstrate that the probability of Roma participating in wage employment does not rise with education or work experience. This probably reflects the low returns of education in the formal labour market and high opportunities in the informal labour market. Roma in Albania are more likely to participate in formal wage employment, if working full time and in a skilled occupation. Poverty is also a significant factor in lowering the odds of labour market participation for Roma in Albania.

8 The likelihood-ratio test for $\rho = 0$ (correlation between the disturbance vectors of Eq. (1) and Eq. (5)) for Roma rejects its null only in Albania. The likelihood-ratio test for $\rho = 0$ for non-Roma rejects its null in Albania, Kosovo and Serbia.

Table 3 Wage income equation corrected for selectivity bias

	Labour market participation estimates				Wage income estimates			
	Roma (*j*=1)		Non-Roma (*j*=2)		Roma (*j*=1)		Non-Roma (*j*=2)	
	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.
(a) Albania (*n*1=289, *n*2=570)								
Constant	-1.021 (0.067)	0.558	-2.584 (0.002)	0.832	3.891 (0.000)	0.501	3.039 (0.000)	0.321
Education	0.010 (0.532)	0.016	0.081 (0.002)	0.026	0.025 (0.024)	0.011	0.027 (0.003)	0.009
Work exp.	0.032 (0.302)	0.031	0.106 (0.009)	0.041	0.030 (0.285)	0.028	0.060 (0.000)	0.015
Work exp.^{2}	-0.041 (0.324)	0.041	-0.144 (0.006)	0.052	-0.034 (0.366)	0.038	-0.068 (0.000)	0.019
Full time	0.557 (0.000)	0.114	0.955 (0.000)	0.167	0.072 (0.440)	0.094	0.286 (0.002)	0.091
High skills	0.723 (0.000)	0.113	0.321 (0.089)	0.189	0.115 (0.147)	0.079	0.157 (0.086)	0.092
Male	0.289 (0.010)	0.112	0.364 (0.011)	0.143	0.201 (0.012)	0.081	0.357 (0.000)	0.049
Poverty	-0.759 (0.000)	0.103	-0.960 (0.000)	0.165				
(b) Bulgaria (*n*1=241, *n*2=370)								
Constant	-	-	-	-	3.355 (0.000)	0.521	3.463 (0.000)	0.379
Education	-	-	-	-	0.024 (0.039)	0.011	0.041 (0.000)	0.009
Work exp.	-	-	-	-	0.021 (0.435)	0.027	0.016 (0.324)	0.016
Work exp.^{2}	-	-	-	-	-0.027 (0.416)	0.033	-0.020 (0.332)	0.021
Full time	-	-	-	-	0.386 (0.000)	0.089	0.144 (0.381)	0.164
High skills	-	-	-	-	0.255 (0.001)	0.074	0.287 (0.000)	0.050
Male	-	-	-	-	0.057 (0.339)	0.060	0.233 (0.000)	0.042

Table 3 *ctd.*

	Labour market participation estimates				Wage income estimates			
	Roma (j=1)		Non-Roma (j=2)		Roma (j=1)		Non-Roma (j=2)	
	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.
(c) Croatia (n1=77, n2=219)								
Constant	-	-	-	-	3.142 (0.000)	0.684	5.068 (0.000)	0.588
Education	-	-	-	-	0.023 (0.450)	0.031	0.072 (0.000)	0.016
Work exp.	-	-	-	-	0.113 (0.002)	0.035	-0.019 (0.496)	0.028
Work exp.^{2}	-	-	-	-	-0.153 (0.004)	0.051	0.028 (0.420)	0.034
Full time	-	-	-	-	0.259 (0.236)	0.217	0.458 (0.010)	0.176
High skills	-	-	-	-	0.396 (0.017)	0.162	-0.068 (0.620)	0.137
Male	-	-	-	-	0.262 (0.032)	0.120	0.085 (0.295)	0.081
(d) Kosovo (n1=123, n2=280)								
Constant	-	-	-1.366 (0.126)	0.892	3.373 (0.000)	0.644	3.945 (0.000)	0.611
Education	-	-	0.131 (0.000)	0.036	0.027 (0.192)	0.021	-0.046 (0.107)	0.029
Work exp.	-	-	0.001 (0.988)	0.046	0.043 (0.209)	0.034	0.095 (0.001)	0.028
Work exp.^{2}	-	-	0.005 (0.929)	0.058	-0.064 (0.143)	0.043	-0.114 (0.001)	0.034
Full time	-	-	0.856 (0.000)	0.178	0.760 (0.000)	0.131	-0.209 (0.141)	0.142
High skills	-	-	-0.108 (0.537)	0.175	0.477 (0.000)	0.127	0.258 (0.008)	0.098
Male	-	-	0.216 (0.335)	0.224	0.018 (0.918)	0.171	0.321 (0.036)	0.153
Poverty	-	-	-0.328 (0.048)	0.166				

Table 3 *ctd.*

	Labour market participation estimates				**Wage income estimates**			
	Roma (*j*=1)		Non-Roma (*j*=2)		Roma (*j*=1)		Non-Roma (*j*=2)	
	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.	Coefficient (*p*-level)	Std. Err.
(e) Serbia (*n*1=111, *n*2=341)								
Constant	-	-	-0.578 (0.513)	0.884	2.891 (0.000)	0.641	4.657 (0.000)	0.654
Education	-	-	0.030 (0.191)	0.023	0.023 (0.213)	0.018	0.045 (0.014)	0.018
Work exp.	-	-	0.005 (0.919)	0.046	0.048 (0.182)	0.036	-0.008 (0.770)	0.028
Work exp.^{2}	-	-	0.000 (0.998)	0.057	-0.040 (0.394)	0.047	0.014 (0.690)	0.034
Full time	-	-	0.216 (0.454)	0.288	0.470 (0.000)	0.119	0.202 (0.378)	0.229
High skills	-	-	0.617 (0.006)	0.225	0.394 (0.002)	0.123	-0.034 (0.815)	0.144
Male	-	-	-0.006 (0.963)	0.131	0.069 (0.691)	0.173	0.186 (0.019)	0.079
Poverty	-	-	-0.966 (0.000)	0.234				

Note: Heckman estimators for Roma and non-Roma in Albania, and for non-Roma in Kosovo and Serbia. OLS estimates for the other cases.

Table 4 shows the observed and the selection-corrected income differentials between Roma and non-Roma in the five countries. Further, the table presents the discrimination effects estimates for the three weighting matrices (*W=0, W=0.5I, W=I)* reflecting the assumption of the non-discriminatory wage income structure. Table 4 illustrates that the selectivity bias widens income differentials in Kosovo and Serbia where non-Roma appear to have a larger negative selection bias than Roma, and narrows the income differential in Albania. How important is labour market discrimination in producing these income differentials, and to what extent do discrimination effects estimates depend on assumptions about the income structure that would prevail in the absence of discrimination? Only in Bulgaria, the three estimates of discrimination, based on *W=0, W=0.5I, W=I,* produce quite similar results. The discrimination effect explains 28 (0.135/0.489), 24 (0.119/0.489) and 21 percent (0.104/0.489) of the log income differential (0.489), respectively. It seems that in Bulgaria changing the weights has offsetting effects with respect to different parameters.

Table 4: Income differentials between Roma and non-Roma and the extent of discrimination

	(1) Observed average income difference	(2) Income differential corrected for selectivity bias	(3) Discrimination effect		
			(*W*=0)	(*W*=0.5*I*)	(*W*=*I*)
Albania	0.701	0.343	0.126	0.075	0.024
Bulgaria	0.489	-	0.135	0.119	0.104
Croatia	0.261	-	0.029	-0.008	-0.044
Kosovo	0.517	0.719	0.199	0.551	0.903
Serbia	0.320	0.547	0.144	0.263	0.381

Note: Discrimination effects estimates are based on Heckman estimators for Roma and non-Roma in Albania, and for non-Roma in Kosovo and Serbia. In the other cases, the discrimination effects estimates are based on consistent estimates from OLS.

In the other four countries, however, the choice of the weighting matrix has a significant impact on the size of the discrimination effects estimates. In Kosovo and Serbia, the discrimination effects are larger when the decomposition is evaluated at the non-Roma income structure, while in Albania and Croatia the discrimination effect is larger when assuming the Roma income structure as the competitive one. In these countries, Roma appear to have larger parameters than non-Roma in several variables where they are at a disadvantage relative to non-Roma. Therefore, the effects of changing weights do not offset each other.

The largest estimate of discrimination is found in Kosovo (0.903) where a larger part of the income differential can be attributed to differences in parameters of the income equation, rather than differences in the characteristics of the groups. In contrast, the lowest estimate of discrimination (negative discrimination) can be found in Croatia (-0.044), where the difference in parameters of the income equation goes in favour of Roma for those characteristics where they are at a disadvantage. This implies that Roma in Croatia apparently would be receiving higher wage incomes compared to non-Roma in Croatia, if they had the same characteristics. The large difference in the decomposition effects between Kosovo and Croatia is not surprising, given that several indicators of poverty and unemployment of Roma in Kosovo point to worse socio-economic conditions than in the other countries of South East Europe, while in Croatia socio-economic conditions of Roma are better than in the rest of South East Europe (UNDP 2006).

Table 5 presents country-specific decompositions of the log income differential into characteristics and discrimination effects for the three sets of weights. Inferences regarding statistical significance of the characteristics and discrimination effects estimates can be drawn based on asymptotic variance approximation[9]. The reported probabilities indicate the existence of significant characteristics effects in the five countries under consideration and for all three weighting matrices, with the exception of *W=I* in Kosovo and Serbia. They also show that the discrimination effect is significant for all three weighting matrices in Kosovo, and for at least two weighting matrices in Bulgaria and Serbia.

The results indicate that there is no consistent pattern of the two effects across the countries. However, some general conclusions can be drawn. *First*, the decomposition effects estimates are very sensitive to the assumption of the non-discriminatory income structure. *Second*, the characteristics effects appear to make up most of the average income differential in South East

9 Standard errors and significance levels are computed based on Oaxaca and Ransom (1994).

Europe. *Third*, there may exist labour market discrimination against Roma in Bulgaria, Kosovo and Serbia.

A more detailed view for each country provides some more interesting findings. In *Albania*, selection bias narrows the overall income differential between Roma and non-Roma to 0.343 log points, out of which about 93 percent is explained through differences in average characteristics between Roma and non-Roma, rather than through differing returns to these characteristics.

Table 5 Country-specific characteristics and discrimination effects estimates

	$\hat{C}$	Std. Err.	*p*-value	$\hat{D}$	Std. Err.	*p*-value
Albania (*n*1=289, *n*2=570)						
W=0	0.216	0.077	0.005	0.126	0.095	0.183
W=0.5*I*	0.267	0.052	0.000	0.075	0.077	0.325
W=*I*	0.319	0.066	0.000	0.024	0.087	0.781
Bulgaria (*n*1=241, *n*2=370)						
W=0	0.354	0.073	0.000	0.135	0.071	0.057
W=0.5*I*	0.369	0.049	0.000	0.119	0.055	0.030
W=*I*	0.385	0.062	0.000	0.104	0.073	0.155
Croatia (*n*1=77, *n*2=219)						
W=0	0.232	0.107	0.030	0.029	0.114	0.800
W=0.5*I*	0.268	0.072	0.000	-0.008	0.094	0.935
W=*I*	0.305	0.088	0.001	-0.044	0.116	0.703
Kosovo (*n*1=123, *n*2=280)						
W=0	0.520	0.104	0.000	0.199	0.113	0.078
W=0.5*I*	0.168	0.093	0.070	0.551	0.116	0.000
W=*I*	-0.184	0.156	0.239	0.903	0.179	0.000
Serbia (*n*1=111, *n*2=341)						
W=0	0.404	0.099	0.000	0.144	0.128	0.258
W=0.5*I*	0.286	0.078	0.000	0.263	0.116	0.023
W=*I*	0.167	0.117	0.152	0.381	0.148	0.010

Note: *n*1=Roma, *n*2=non-Roma. Standard errors and significance levels based on Oaxaca and Ransom (1994)

The result is based on the assumption that the non-Roma income structure would hold in a non-discriminatory world (*W*=*I*). The contribution of the characteristics effect to the income differential reduces to 63 percent, when the decomposition analysis is based on the assumption that non-Roma are overpaid (*W*=0). Regardless of the assumption of the non-discriminatory income structure, the discrimination effect is not significantly different from zero.

In *Bulgaria,* the characteristics and discrimination effects explain 76 and 24 percent of the log income difference (0.489), respectively, if the competitive income structure lies in between the Roma and non-Roma income structure. In this case, the discrimination effect is significantly different from zero at the five percent level, whereas when non-Roma are overpaid (W=0), the discrimination effect is only weakly significant at the 10 percent level, and when Roma are underpaid (W=I), it is not significantly different from zero. While the extent of wage discrimination is roughly of equal size, the significance of the effect's estimate is sensitive to the assumption of the income structure that would hold in the absence of discrimination.

As in Albania, the discrimination effect identified in *Croatia* is not significantly different from zero, regardless of the assumption of the non-discriminatory income structure, but the characteristics effect is. This effect largely contributes to the ethnic income differential. Note that the size of the effect is largest when the decomposition analysis is evaluated at the non-Roma income structure.

The sensitivity of the size and significance of the decomposition estimates to the assumption of the competitive income structure is largest in *Kosovo*. The characteristics and discrimination effects explain 72 and 28 percent of the log income difference (0.719), respectively. This is based on the assumption that the Roma income structure holds in a non-discriminatory world. In this case, the discrimination effect is weakly significant at the 10 percent level. The size and significance of the effects estimates change considerably when assuming W=0.5I. In this case, the characteristics and discrimination effects explain 23 and 77 percent of the log income difference, respectively, but the characteristics effect is only weakly significant and the discrimination effect is significantly different from zero at the one percent level. For W=I, the characteristics effect turns insignificant, and the income differential is explained solely by discrimination. The results clearly indicate that in Kosovo, a country where poverty among the Roma is highest, out of the five countries, labour market discrimination explains a large share of the income differential for all three weights.

As in Kosovo, in *Serbia* the decomposition effects estimates are highly sensitive to the choice of the weighting matrix. Assuming that non-Roma are overpaid (W=0), the characteristics and discrimination effects explain 74 and 26 percent of the log income differential (0.549), respectively.

In this case however, the discrimination effect is not significantly different from zero but the characteristics effect is. Both effects are statistically significant when the non-discriminatory income structure is assumed to lie in between the income structure of the two groups, and they contribute roughly equally to the income differential. In contrast, about 70 percent of the log income differential can be explained by discrimination when it is assumed that Roma are underpaid (*W*=*I*). Note that in this case, the characteristics effect is even insignificant. The empirical results are of course conditional on the methodology and sample used. Given the sample, the findings depend on the weight and are rather country-specific. Thus, only a range of discrimination estimates can be reported in Kosovo, and weak discrimination estimates are found in Serbia and Bulgaria. Country differences can be explained by differences in structures of national economies. In general, the labour market performance (measured in terms of unemployment or incomes) of Roma is worst in Kosovo and best in Croatia which is supported by the findings on income differential decomposition in this paper. Further, the methodology employed takes into account only formal labour market participation. However, the shadow economy is important in all these national economies. Differences in findings between countries may also result from country differences in participation rates in the shadow economy that have indirect effects on labour market discrimination, apart from the direct effects estimated in this paper.

6 Closing remarks

This paper used the selectivity bias adjustment, as suggested by Heckman (1979) and Reimers (1983), to a Blinder-Oaxaca decomposition analysis of income differentials among Roma and non-Roma population groups in five South East European countries, using data from the 2004 UNDP survey. Three weighting matrices, reflecting differing assumptions about the income structure that would prevail in the absence of discrimination, have been chosen for the decomposition of income differentials.

The selection-corrected approach accommodates bias in the parameters of wage characteristics in the cross-sectional semi-log regression relationships, when wage structures are affected by non-random decisions people make about labour market participation. Unbiased parameter estimates of the wage characteristics can be obtained using the Heckman (1979) procedure, which first estimates the probability of labour market participation, and then adds the Inverse Mill's Ratio, a proxy for the probability of labour force participation, as an additional explanatory variable into the wage equation. Based on this, the Blinder-Oaxaca type of decomposition analysis with the familiar characteristics and discrimination effects can be obtained by subtracting the selection effects from the overall differential (Reimers 1983).

The results point to a large sensitivity of the characteristics and discrimination effects estimates size and significance with regard to the choice of the weighting matrix. Only in Bulgaria, changing weights does not affect much the size of the estimate of discrimination. However, a presence of statistically significant discrimination effects can be found in Bulgaria, Kosovo and Serbia. In Bulgaria, the most significant discrimination effect occurs when the competitive income structure lies in between the Roma and non-Roma income structure, while in Serbia and Kosovo discrimination is most significant when it is assumed that Roma are underpaid. In the latter two countries, the characteristics effect is even insignificant, pointing to discrimination as the main explanation for the income differentials among Roma and non-Roma. In the other countries, however, differences in measured characteristics and not labour market discrimination are the overwhelming reason for the shortfall in incomes for Roma.

Inferences regarding statistical significance of the decomposition components have been drawn using asymptotic variance approximation, as proposed by Oaxaca and Ransom (1994). The limitation of this approach, however, is that the approximation requires an assumption of an asymptotic multivariate normal distribution for the parameter vector, as well as the use of the variance-covariance matrix for the parameter estimates. Keith and LeSage (2004) have demonstrated that the presence of heteroscedasticity or outliers, which may occur especially in the case of small samples, can have an adverse impact on the characteristics and discrimination effects estimates when using least squares.

References

Becker, G.S. (1975), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, New York: National Bureau of Economic Research and University of Chicago Press

Blinder, A.S. (1973), “Wage discrimination: reduced form and structural estimates”, *Journal of Human Resources*, Vol. 8, No. 4, pp. 436-455

Cotton, J. (1988), “On the decomposition of wage differentials”, *The Review of Economics and Statistics*, Vol. 70, No. 2, pp. 236–243

Heckman, JJ. (1979), “Sample selection as a specification error”, *Econometrica*, Vol. 47, No. 1, pp. 153-162

Keith, K. and LeSage, J.P. (2004), “Robust decomposition analysis of wage differentials”, *Journal of Economic and Social Measurement*, Vol. 29, No. 4, pp. 487-505

Kimmel, J. (1997), “Rural wages and returns to education: differences between whites, blacks, and American Indians”, *Economics of Education Review*, Vol. 16, No. 1, pp. 81-96

Maani, S.A. (2002), “Education and Maori relative income levels over time: the mediating effect of occupation, industry, hours of work and locality”, New Zealand Treasury, Working Paper 02/17, Wellington, New Zealand

MacIsaac, D. and Patrinos, H. (1995), “Labor market discrimination against indigenous people in Peru”, *Journal of Development Studies*, Vol. 32, No. 2, pp. 218-33

Milcher, S. (2006), “Poverty and the determinants of welfare for Roma and other vulnerable groups in Southeastern Europe”, *Comparative Economic Studies*, Vol. 48, No.1, pp. 20-35

Milcher, S. (2010), “Household vulnerability estimates of Roma in Southeast Europe”, *Cambridge Journal of Economics*, Vol. 34, No. 4, pp. 773-792

Mincer, J. (1974), *Schooling, experience, and earnings*, New York: National Bureau of Economic Research and Columbia University Press

Neuman, S. and Oaxaca, R.L. (2004), “Wage decompositions with selectivity-corrected wage equations: a methodological note”, *Journal of Economic Inequality*, Vol. 2, No. 1, pp. 3–10

Neumark, D. (1988), “Employers’ discriminatory behavior and the estimation of wage discrimination”, *The Journal of Human Resources*, Vol. 23, No. 3, pp. 279-295

Oaxaca, R.L. (1973), “Male-female wage differentials in urban labour markets”, *International Economic Review*, Vol. 14, No. 3, pp. 693-709

Oaxaca, R.L. and Ransom, M. (1994), "On discrimination and the decomposition of wage differentials", *Journal of Econometrics*, Vol. 61, No. 1, pp. 5-21

Oaxaca, R.L. and Ransom, M. (1998), "Calculation of approximate variances for wage decomposition differentials", *Journal of Economic and Social Measurement*, Vol. 24, No. 1, pp. 55–61

Patrinos, H.A. and Sakellariou, C.N. (1992), "North American Indians in the Canadian labour market: a decomposition of wage differentials", *Economics of Education Review*, Vol. 11, No. 3, pp. 257-266

Radchenko, S.I. and Yun, M.S. (2003), "A Bayesian approach to decomposing wage differentials", *Economics Letters*, Vol. 78, No. 3, pp. 431-436

Reimers, C.W. (1983), "Labor market discrimination against Hispanic and Black Men", *The Review of Economics and Statistics*, Vol. 65, No. 4, pp. 570-579

UNDP (2006), *At risk: Roma and the displaced in Southeast Europe*, United Nations Development Programme, Regional Bureau for Europe and the Commonwealth of Independent States, Bratislava

Yun, M.S. (2007), "An extension of the Oaxaca decomposition using generalized residuals", *Journal of Economic and Social Measurement*, Vol. 32, No. 1, pp. 15-22