UNIVERSITY OF JYVÄSKYLÄ
SCHOOL OF BUSINESS AND ECONOMICS

TUOMO SUHONEN
&
JAAKKO PEHKONEN
&
HANNU TERVO

ESTIMATING REGIONAL DIFFERENCES IN RETURNS TO EDUCATION WHEN SCHOOLING AND LOCATION ARE DETERMINED ENDOGENOUSLY

WORKING PAPER
N:O 363/2010

University of Jyväskylä
School of Business and Economics
P.O. Box 35, 40014 University of Jyväskylä
E-mail: tuomo.suhonen@econ.jyu.fi, jaakko.k.pehkonen@jyu.fi,
hannu.tervo@econ.jyu.fi

ISBN: 978-951-39-3943-4
ISSN: 1239-3797
Jyväskylä 2010

ABSTRACT

The empirical evidence on differences in location-specific returns to education is potentially biased if the endogenous nature of either schooling or location is neglected during the estimation. In this paper, we apply the nonparametric approach of Das, Newey and Vella (2003) to correct for this double selectivity bias when estimating the effect of job location on returns to education in Finland. Using register-based data from six yearly cross-sections between 1980 and 2004, marginal returns at different amounts of schooling and average marginal returns are compared between the Helsinki metropolitan area and a group of university regions. While the standard OLS estimates suggest that there have been no differences in the region-type-specific returns, certain temporary differences are found after correcting for the selectivity. Furthermore, we find strong evidence that the marginal return has not been constant over different amounts of schooling attained.

JEL: C14; C31; J24; J31

Keywords: returns to education; regional differences; sample selection models; nonparametric estimation

1 Introduction

The demand and supply of skills acquired by formal education differ across locations, which, according to the basic theory on skill prices (Acemoglu, 2002), can generate regional differences in individual returns to education. In this framework, a number of studies have examined the causal responses of returns to education to changes in a variety of aggregate variables, including school quality (Card and Krueger, 1992), unemployment (Ammermueller et al., 2009), and local housing prices (Black et al., 2009). To address these questions, the studies have first estimated location-specific returns and, for the sake of simplicity, dismissed the potential endogeneity of educational choices (Griliches, 1977; Card, 2001) as well as the possible bias arising from selective migration between locations (Falaris, 1987; Dahl, 2002). In past studies, certain attempts have been made to address either one of these two endogeneity problems when estimating within-country regional variation in the returns (Bennett et al., 1995; Dahl, 2002; Tokila and Tervo, 2010), but none of the studies have treated both of these variables as endogenous. Thus, it is unclear how well the estimates on regional differences, in fact, reflect the actual differences in local skill prices.

In this paper, we study the effect of job location on returns to education while paying serious attention to the econometric problems arising from the unobserved individual heterogeneity. In the spirit of Das, Newey and Vella (2003), we specify an econometric model in which both schooling and location are determined endogenously, and in which the years of schooling are allowed to affect both the wage and the choice of location in a nonlinear manner[1]. We apply the model for estimating differences in marginal and average marginal returns to education between two "job region types" in Finland, the Helsinki metropolitan area and a group of university regions, using six yearly cross-sections between 1980 and 2004. The empirical analysis of this particular case provides an engrossing experiment, as the metropolitan area has been the leading region in both the supply of and demand for highly educated

---

[1] Here, we follow Harmon et al. (1999), Das et al. (2003) and Trostel (2005), who have provided evidence that the return to an additional year of schooling is not constant over different amounts of schooling attained.

workforce in Finland. Thus, on theoretical grounds, the returns could have been either higher or lower in the metropolitan area compared to the other central regions.

For the empirical analysis, we employ a register-based data set from Statistics Finland, including yearly cross-section samples of young men aged 25–35 and classified as private-sector wage earners. For these samples, we estimate the model of wage, schooling and location using the nonparametric selectivity correction approach of Das et al. (2003). The advantage of this approach, compared to the more traditional selectivity correction methods, is that it allows us to simultaneously address the endogeneity of individual choices and the possible nonlinearity of schooling. We use the information on individuals' region of birth as an instrument for identifying the causal effect of both endogenous variables on wages. As we include a large set of controls for the potential indirect linkages between the region of birth and wages, we expect this variable to serve as a legitimate instrument. We also contrast the average marginal return estimates from the Das-Newey-Vella procedure to those estimated by the standard OLS method in which the exogeneity and linearity assumptions are imposed.

The paper proceeds as follows. In Section 2, the econometric model and estimation strategy are discussed in more detail. The data and variables used in the estimation are described in Section 3. In Section 4, we present the marginal and average marginal return estimates from the Das-Newey-Vella procedure and compare these to the estimates from the standard OLS regressions. Section 5 concludes the paper.

2 Methodology

2.1 Econometric model

In this section, we specify a model of wage, schooling, and location, which is then used for the empirical analysis of the region-type-specific returns[2]. The model is based on the idea that individuals first choose the amount of schooling, which then goes on to affect their choice of region type. Finally, the wage is determined on the

---

[2] The model is similar to the one in the empirical example of Das et al. (2003), with the exception that the model of labor market participation is replaced with a model of region type choice.

basis of the endogenous schooling and location choices, as well as other variables such as age and tenure.

To formalize, let the wage of individual i working in region of type j be defined as follows:

(1) $$w_{ij}^* = x'_i \beta_j + g_j(s_i) + \varepsilon_{ij} \qquad (i = 1,\ldots,n; j = 1,2)$$

(2) $$w_{ij} = d_{ij} w_{ij}^*,$$

where w_{ij}^* denotes the logarithmic monthly wage which is observed if region type j was chosen $(d_{ij} = 1)$; s_i denotes the years of schooling, vector x_i includes other observed wage covariates (plus a constant term), and ε_{ij} is the error term. In equation (1), the relationship between the wage and schooling is given by function $g_j(s_i)$ which is of an unknown functional form. We are interested in estimating two parameters, the derivative of this function at a particular amount of schooling s_x and the average derivative, i.e.,

(3) $$\left.\frac{\partial E\left[w_{ij} \mid x_i, s_i, d_{ij} = 1\right]}{\partial s_i}\right|_{s_i = s_x} = g'_j(s_x)$$

(4) $$\frac{1}{n_j} \sum_{i=1}^{n_j} g'_j(s_i),$$

where n_j is the number of observations for region type j. For the sake of simplicity, we assume that the marginal and average marginal return to education are given by the estimates of (3) and (4), although these neglect the marginal costs of education arising, e.g., from foregone earnings. Furthermore, we compare these estimates between region type 1 (metropolitan area) and 2 (university region) to assess the effect of job location on returns to education.

We address the possibility that there may be at least two self-selection mechanisms that cause the standard nonparametric estimate of $g_j(s_i)$ to be biased, implying that the estimates of (3) and (4) are also biased. First, s_i is likely to be endogenous in the wage equation due to some unobserved variables, e.g., the innate ability, which affect both wage and schooling choice (Griliches, 1977; Card, 2001). To highlight this problem, let us assume that the years of schooling are determined by the following reduced-form *schooling equation*:

$$(5) \qquad s_i = z'_i \gamma + \upsilon_i ,$$

where z_i is a vector of observed variables determining the choice (including a constant term) and υ_i is an unobserved component. Formally, the endogeneity of schooling arises from the correlation between υ_i and ε_{ij}.

Second, the choice of region type j may also be non-random and correlated with unobserved wage determinants, which causes selectivity bias in the estimate of $g_j(s_i)$[3]. In modelling the choice of region type, we consider the following *region type equation*, in which the schooling variable is allowed to enter the model in a nonlinear manner:

$$(6) \qquad d_{ij} = y'_i \theta_j + h_j(s_i) + \omega_{ij} ,$$

where y_i is a vector of variables explaining the choice, and $h_j(s_i)$ is an unknown function of the years of schooling. At this point, an important simplifying assumption is made. Namely, we assume that the choice of region type is binary, i.e., over the two alternatives considered in this study. Although a more realistic model would allow a multinomial location choice, the complexity involved in the estimation of such a model would be substantial[4]. Furthermore, as our main interest is in measuring "the

[3] In his study of differences in returns to education between the U.S. states, Dahl (2002) found evidence that the selectivity of the state-specific samples results in an upward bias in the OLS estimates of regional returns.

[4] In the estimation of models with multinomial selection rules, it is common to impose simplifying assumptions due to "the curse of dimensionality". An underlying assumption in our approach is that the ratio of two choice probabilities is independent of other alternatives. This is also assumed when using choice probabilities estimated by the multinomial logit model for the selectivity correction (see Bourguignon et al., 2007). Another underlying assumption is that the selectivity correction is only a function of the "first-best" alternative. This assumption has been previously used in the approaches of Lee (1983) and Dahl (2002).

effect of switching from one region type to another", we find it relevant to concentrate on the binary case. With the binary treatment assumption, equation (6) can be treated as a linear probability model with the propensity score for the choice of region type j defined as

$$(7) \qquad p_{ij} \equiv E\left[d_{ij} = 1 \mid y_i, s_i\right] = y'_i\,\theta_j + h_j(s_i)$$

For the propensity scores, the binary treatment assumption implies that $p_{i2} = 1 - p_{i1}$. Likewise to the case of the wage equation, it is relevant to suspect that the schooling variable is endogenous also in the region type equation (see Sanders, 2004). This possibility is also taken into account in our estimation procedure.

2.2 Estimation strategy

In order to overcome the endogeneity problems discussed above, we apply the nonparametric selectivity correction approach of Das, Newey, and Vella (2003). The method involves estimating the conditional expectation of the wage equation error term, which is assumed to be a function of two arguments, the propensity score of the choice of region type $p_{ij},$ and the schooling residual υ_i, i.e.,

$$(8) \qquad E\left[\varepsilon_{ij} \mid x_i, y_i, z_i, s_i, \upsilon_i, d_{ij} = 1\right] = \lambda_j(\upsilon_i, p_{ij}).$$

According to Das et al. (2003), if the model disturbances are independent of z_{1i} and z_{2i}, and an additive structure is imposed, the conditional expectation of the wage is given by

$$(9) \qquad E\left[w_{ij} \mid x_i, y_i, z_i, s_i, \upsilon_i, d_{ij} = 1\right] = x'_i\,\beta_j + g_j(s_i) + \lambda_j(\upsilon_i, p_{ij}).$$

Thus, an unbiased estimate of $g_j(s_i)$ can be obtained by estimating function $\lambda_j(\upsilon_i, p_{ij})$ nonparametrically. The identification requires an exclusion restriction for the wage equation; that is, we need instruments that affect the selection of region type and schooling but are independent of the wage. The discussion of the instruments used in the present study is left for Section 3.

The fact that we allow an endogenous schooling variable to enter the region type equation is taken into account by specifying that the propensity score also depends on the schooling residual in accordance with an unknown function $\pi_j(\upsilon_i)$. Therefore, equation (7) is rewritten as

$$(10) \qquad p_{ij} \equiv E\lfloor d_{ij} = 1 \mid y_i, z_i, s_i, \upsilon_i \rfloor = y'_i \theta_j + h_j(s_i) + \pi_j(\upsilon_i)$$

To ensure identification, we also impose an exclusion restriction in the region type equation by including certain variables in z_i that are excluded from y_i.

We estimate the model in three steps. The first step involves the estimation of the reduced-form schooling equation, which provides the predicted residual $\hat{\upsilon}_i$. In the second step, the region type equation is estimated while including functions of s_i and $\hat{\upsilon}_i$ in the model. This produces the estimate of the propensity score $\hat{p}_{ij}$. The final step is the estimation of the wage equation while including functions of s_i, $\hat{p}_{ij}$ and $\hat{\upsilon}_i$ in the regression. All three steps are estimated by OLS, and power series approximation is applied for the nonparametric estimation of the unknown functions. That is, we include polynomials of s_i, $\hat{\upsilon}_i$ and $\hat{p}_{ij}$ in the regressions. Thus, in the form of selectivity-corrected regressions, the wage and region type equations can be rewritten as

$$(11) \qquad w_{ij} = x'_i \beta_j + \sum_{k=1}^{K} \phi_{1k} s_i^k + \sum_{l=1}^{L} \phi_{2l} \hat{p}_{ij}^l + \sum_{m=1}^{M} \phi_{3m} \hat{\upsilon}_i^m + \sum_{r=1}^{R} \phi_{4r} \hat{p}_{ij}^r \hat{\upsilon}_i^r + \xi_{ij}$$

$$(12) \qquad d_{ij} = y'_i \theta_j + \sum_{t=1}^{T} \phi_{5t} s_i^t + \sum_{u=1}^{U} \phi_{6u} \hat{\upsilon}_i^u + \eta_{ij}$$

where ξ_{ij} and η_{ij} denote the independent error terms. As our schooling variable only takes five different values (9, 12, 14, 15, and 17), the possibilities of us non-parametrically estimating functions $h_j(s_i)$ and $g_j(s_i)$ are limited. With five data points, the maximum order of the schooling polynomials would be four, but, in practice, we found it appropriate to restrict the order of the polynomials to three[5] at most. Then again, functions $\pi_j(\upsilon_i)$ and $\lambda_j(\upsilon_i, p_{ij})$ can be estimated in a more flexible manner, as there are large continuums of values for $\hat{\upsilon}_i$ and $\hat{p}_{ij}$.

In each case, the number of polynomial terms is determined by cross-validation (CV). For calculating the CV values, we use the method of 10-fold cross-validation, which involves a significantly smaller computational burden compared to the leave-one-out method used in the empirical example of Das et al. (2003)[6]. In the 10-fold cross-validation, the data is randomly split into ten equal-size groups and the model is estimated ten times, each time ignoring one of the groups in the estimation (validation set) and predicting the mean squared error for this group using the rest of the observations (training set). The CV value is, then, the average of these predictions over the ten groups. As the accuracy of the procedure depends on the randomization used in choosing the folds, it is common to repeat the procedure a number of times and use the average of these calculations as the final CV value (Witten and Frank, 2005). Here, we repeat the procedure ten times.

Two final remarks are made about the estimation. First, due to the need to restrict the order of the schooling polynomials, we do not include polynomial terms in addition to those suggested by the CV values, although “undersmoothing” has been recommended in the literature (Newey et al., 1999; Das et al., 2003). Second, since the Das-Newey-Vella procedure involves using predictions $\hat{\upsilon}_i$ and $\hat{p}_{ij}$ in the regressions, the OLS standard errors are, in general, biased. Therefore, we base our statistical inference on bootstrapped standard errors calculated using 500 bootstrap replications.

3 Data and variables

The register-based data set of this study is provided by Statistics Finland. The data includes panel information on a seven percent random sample drawn from the Finnish

[5] The fitted fourth-order polynomials for schooling were found to produce unstable estimates, as well as large confidence intervals. Thus, with five data points, it seems appropriate to limit in using only linear, quadratic and cubic polynomials.

[6] According to the evidence of Breiman and Spector (1992), the k-fold methods can also perform better in the model selection framework than the leave-one-out method.

population in 2001[7]. From this data, we draw cross-section samples for the years 1980, 1985, 1990, 1995, 2000, and 2004. The samples include men aged between 25 and 35 and reported as private-sector wage earners[8]. One of the advantages in our data set is that it allows us to use information on individuals' job location, instead of the location of residence. Therefore, the measured regional wage variation in our data is not biased by the fact that many individuals commute long distances to work. In the data, the definition of job location is based on relatively small regional units, i.e., the 72 sub-regions of Finland. We restrict our analysis to those workers whose job region is classified as either "metropolitan area" or "university region"[9]. These two region types cover the largest centres of Finland; according to the data, 63 percent of the young workforce (aged 25–35) worked in these regions in 2004.

The descriptive statistics in Table 1 provide an overview of the samples used. The sample size varies between 4 186 and 4 990 for the metropolitan area, and between 2 851 and 3 622 for the university regions. Since only employed individuals are included in the analysis, the sample sizes do not increase monotonically over time but rather fluctuate according to the ongoing employment situation[10]. Table 1 shows that both the average monthly wage and average educational level have been higher among the workers in the metropolitan area in comparison to those working in the university regions through the whole observation period. However, the gap in the average years of schooling has clearly reduced towards the end of the period. Another finding from Table 1 is that the average tenure, measured by the duration of the

---

[7] Since the sample is drawn from those alive and living in Finland in 2001, the data is not a fully representative sample of the population in any of the observation years. The decision to use only young individuals (aged 25–35) in the analysis was made partly due to this problem, as the possible selection bias is likely to be less severe among the samples of young individuals.

[8] We concentrate on private-sector workers alone, as their wages are more likely to reflect the local supply and demand conditions than the wages of public-sector workers. Entrepreneurs are dropped out for the reason that their wage income is inaccurately recorded and more likely to reflect tax-minimization than the amount of work effort.

[9] The metropolitan area consists of the regions of Helsinki, Lohja, Riihimäki and Porvoo, while the group of university regions includes Turku, Tampere, Vaasa, Kuopio, Joensuu, and Oulu.

[10] The use of employed individuals only poses another possible source of selectivity bias. As we already address two sources of selection in our estimation procedure, we dismiss this issue in order to avoid the increase of estimation complexity.

Table 1
Samples sizes and variable means.

	1980	1985	1990	1995	2000	2004
Metropolitan area						
Monthly wage (100 €)	7.88	13.05	19.48	19.94	27.92	31.11
	(.04)	(.10)	(.15)	(.13)	(.41)	(.46)
Years of schooling	11.85	12.21	12.57	12.67	12.89	12.99
	(.04)	(.04)	(.03)	(.04)	(.03)	(.04)
Age	30.13	30.02	29.85	30.32	30.12	30.02
	(.05)	(.05)	(.05)	(.05)	(.04)	(.05)
Tenure			45.59	51.10	43.34	47.39
			(.67)	(.69)	(.60)	(.58)
Sample Size	4186	4334	4623	4257	4990	4468
University regions						
Monthly wage (100 €)	7.06	11.28	16.44	18.31	24.33	26.50
	(.05)	(.12)	(.13)	(.14)	(.40)	(.32)
Years of schooling	11.61	11.98	12.42	12.61	12.76	12.94
	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
Age	30.05	30.05	29.99	30.25	30.10	29.92
	(.05)	(.06)	(.06)	(.06)	(.06)	(.05)
Tenure			54.05	55.43	50.56	51.50
			(.94)	(.93)	(.81)	(.69)
Sample Size	3221	3100	2851	2563	3027	3622

Note: Standard deviations in parentheses.

current employment in months, has been lower in the metropolitan area, which most likely reflects higher job turnover among the firms of that area.

We then turn the discussion to the variables used in the estimation of the sample selection model described in Section 2 by equations (5), (11), and (12)[11]. The data contain information on the annual earned income and the months of employment, which are subsequently used to calculate the dependent variable of the wage equation, i.e., the monthly wage. The key explanatory variable, the years of schooling, is constructed using the information on the highest level of education achieved. Naturally, the resulting measure of education includes some inaccuracies, for instance, as it only takes into account the completed educational qualifications, but no preferable measure is available in the data. Alternatively, one could measure the "sheepskin effects" of education using dummy variables for different educational levels (e.g., see Harmon et al., 2003). This would, however, add complexity to the

[11] The detailed definitions of the variables are included in Appendix 1.

estimation of the sample selection model, as there would be a discrete multinomial choice of schooling involved. For this reason, we prefer to use a single "continuous" measure of education in the analysis.

Slightly different sets of control variables are included in different models. The wage equations include controls for age, tenure, marital status, mother tongue, parents' education/socioeconomic status, and job location. Of these, tenure and job location are not included in the selection equations, since they are regarded as variables that are determined ex-post during the working career. The rest of the differences in the controls arise from two exclusion restrictions, on which our identification strategy is based. The first restriction, for allowing the years of schooling to be endogenous with respect to the choice of region type, is imposed by excluding parents' education from the region type equation. In the absence of a well-established theoretical linkage between parents' education and the choice of job location, we expect parents' education to serve as a legitimate instrument. Although we find small positive correlation (ranging from 0.05 to 0.13) between the propensity to select into the metropolitan area and parents being highly educated, we argue that this correlation mainly arises through individuals' own level of education and the region of birth, which correlate highly with both parents' education and job location[12].

The second exclusion restriction, for allowing schooling and region type to be endogenous with respect to wage, is imposed by excluding the region of birth from the wage equation. As the region of birth is likely to be associated, e.g., with individuals' access to higher education and preferences over job locations, this variable is expected to be a relevant instrument in this case[13]. The tests for the instrument relevance (Section 4, Table 3) also support this hypothesis. However, some concerns about the instrumental validity may be presented: although no direct linkage between the region of birth and wage would exist, indirect correlation through

[12] Previously, Sanders (2004) has used individuals' smoking behaviour as an instrument for education when studying the effect of education on the choice to live in metropolitan areas in the U.S. In his study, parents' education variables were included in the choice model, but these were found to have no significant impact on the choice.

[13] Several studies have applied a related strategy by using a measure of school proximity (or availability) as an instrument for educational choices (Kane and Rouse, 1993; Card, 1995; Bedi and Gaston, 1999; Long, 2008). The region of birth has also been used directly as an instrument by Kruhse-Lehtonen (2007) in her study of returns to education by time and cohort in Finland.

regional variation in family background, wage level, or school quality might cause the validity to break down (Card, 1995). In order to minimize the risk of the first two sources of correlation, we include a set of controls for parents' education and socioeconomic status, as well as additional dummy variables for the location of job, in the wage equation. The third potential source of correlation, through school quality, is more problematic, as we are not able to produce controls for school quality from our data. With the Finnish data, however, this problem may be relatively small, as nearly all of the formal education in Finland is arranged in public schools that can be thought of as being relatively homogeneous in quality[14]. Moreover, the international evidence on the linkage between school quality and earnings has not been very strong (see Betts, 1995; Heckman et al., 1996).

4 Return to education estimates

In this section, we report the results from the Das-Newey-Vella procedure (Das et al., 2003). As the analysis produces a large number of estimates, we concentrate on reporting only the parameters of interest; that is, the implied marginal and average marginal returns to education. The regression coefficients from the different models are excluded from the paper.

In the first two steps of the procedure, we estimate the reduced-form schooling equations and region type equations for each year. This yields estimates for the schooling residual and propensity score. Before estimating the region type equations, we use cross-validation (CV) to fix the polynomial orders for the years of schooling and schooling residual in these models. The CV results (Table 2) indicate that allowing the nonlinearity of the years of schooling in the region type equation is relevant, as the inclusion of higher order terms for this variable is supported in each case. In the models of 1985, 1990, and 2004, the selected polynomial is quadratic, and cubic in the rest of the cases. The CV results also suggest that a correction for the endogeneity of the years of schooling is necessary for the models of 1990, 2000, and

[14] According to the evidence of Kirjavainen (2009), there are modest efficiency differences between upper secondary schools in Finland, but these differences are related to students' past performance and their parents' educational background, not to school resources.

Table 2
Model selection for the region type equation by 10-fold cross-validation. The CV-minimizing models.

Year	Number of polynomial terms		CV
	Years of schooling	Schooling residual	
1980	3	0	.1600
1985	2	0	.1456
1990	2	1	.1467
1995	3	0	.1455
2000	3	1	.1437
2004	2	7	.1495

Table 3
Tests of instrument relevance. F-test statistics.

	1980	1985	1990	1995	2000	2004
Schooling equation						
Region of birth	4.55	9.43	14.23	13.70	15.46	10.33
Parents' education	36.64	46.84	55.19	50.36	55.31	65.15
Region type equation						
Region of birth	195.39	242.23	211.34	200.54	231.14	234.89

2004, as the inclusion of the schooling residual terms is supported in these models. During the estimation of the first-step and second-step models, we examine the relevance of our instrumental variables, the region of birth and parents' education dummies, by calculating tests for the joint significance of these in the models. In general, the F-test statistics (Table 3) are highly significant, indicating that the chosen instruments are relevant.

The third step in the Das-Newey-Vella procedure involves the semiparametric estimation of the region-type-specific wage equations for each year. Both unknown functions $g_j(s_i)$ and $\lambda_j(\upsilon_i, p_{ij})$ are estimated by power series approximation, and the correct polynomial orders for the years of schooling, schooling residual and propensity score, as well as the number of interactions between the two latter, are once again chosen on the basis of the cross-validation results. For the sake of comparison, Table 4 lists the CV-minimizing models with linear, quadratic, and cubic specification of schooling for each year and region type. The models with only a linear term for the years of schooling do not achieve the lowest CV value in any of the

Table 4
Model selection for the wage equation by 10-fold cross-validation. The CV-minimizing models with linear, quadratic and cubic schooling.

Year	Region type	S	R	P	I	CV
1980	Metro area	1	6	0	0	.1109
		2	7	0	0	.1107
		3	**1**	**0**	**0**	**.1104**
	Univ. region	1	6	0	0	.0902
		2	0	1	0	.0901
		3	**1**	**0**	**0**	**.0898**
1985	Metro area	1	4	0	0	.1146
		2	4	0	0	.1145
		3	**0**	**0**	**0**	**.1141**
	Univ. region	1	6	0	0	.1066
		2	1	0	0	.1062
		3	**3**	**1**	**0**	**.1061**
1990	Metro area	1	4	4	3	.1458
		2	3	5	0	.1448
		3	**1**	**7**	**0**	**.1446**
	Univ. region	1	4	0	0	.1238
		2	4	0	0	.1237
		3	**1**	**0**	**0**	**.1233**
1995	Metro area	1	7	1	0	.1903
		2	1	1	0	.1894
		3	**2**	**2**	**0**	**.1894**
	Univ. region	1	2	2	1	.1481
		2	**1**	**1**	**0**	**.1480**
		3	1	1	1	.1481
2000	Metro area	1	5	1	0	.3050
		2	5	1	0	.3044
		3	**5**	**1**	**0**	**.3043**
	Univ. region	1	4	0	0	.2710
		2	1	1	0	.2710
		3	**0**	**0**	**0**	**.2707**
2004	Metro area	1	6	0	0	.2711
		2	0	1	0	.2705
		3	**0**	**1**	**0**	**.2703**
	Univ. region	1	3	0	0	.2161
		2	0	0	0	.2149
		3	**0**	**0**	**0**	**.2149**

Notes: The selected models are emboldened. S = order of the schooling polynomial; R = order of the schooling residual polynomial; P = order of the propensity score polynomial; I = number of interactions between the schooling residual and propensity score terms; I=1 denotes a model with an interaction between the linear terms; I=2 equals I=1 plus an additional interaction between the quadratic terms, etc.

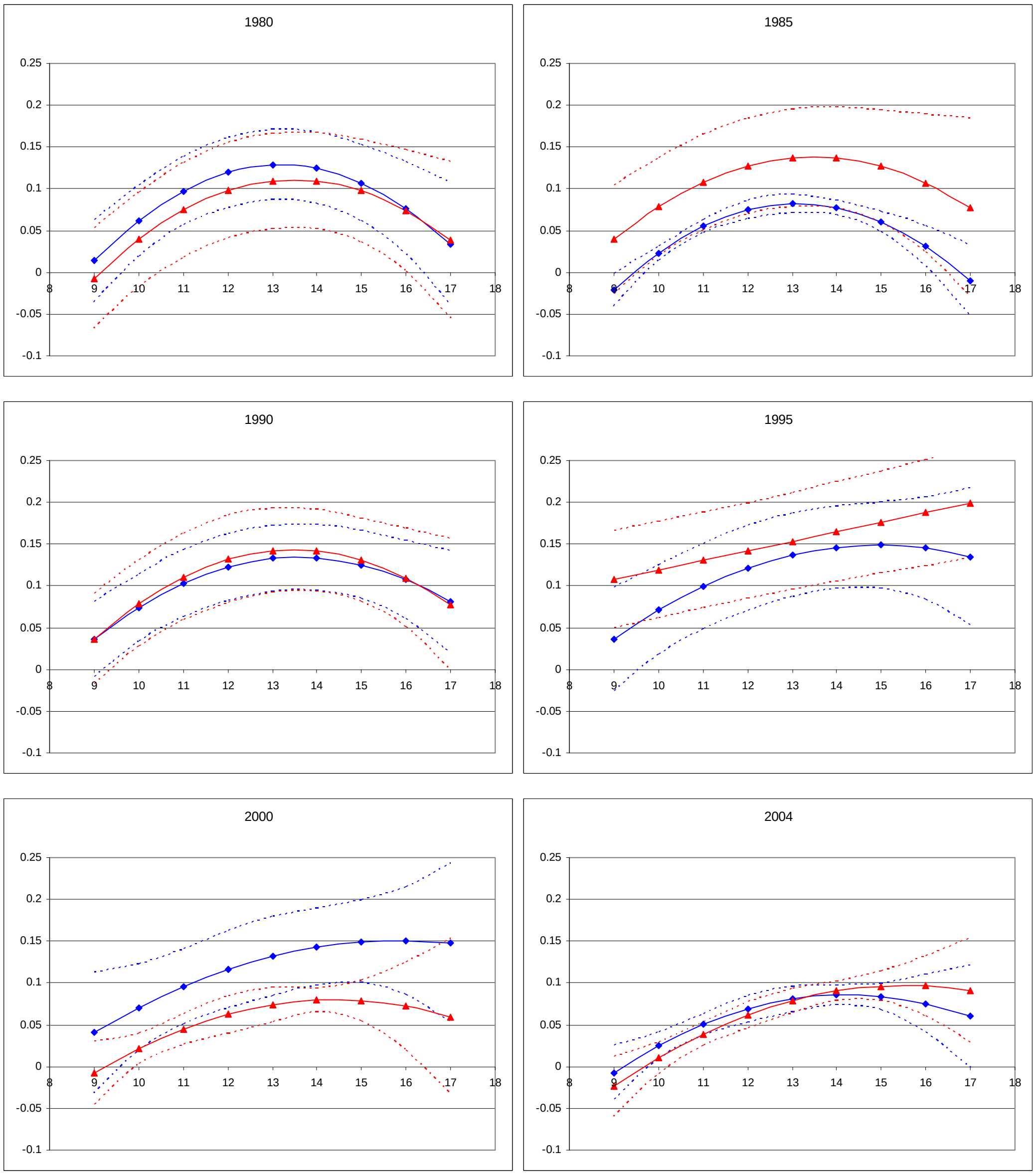
wage equations, suggesting that the linearity of the years of schooling is not supported in our data. The model with a quadratic polynomial of schooling is supported in one case – the 1995 wage equation of university regions – and a cubic polynomial in the remaining eleven cases. The CV results also provide evidence in favour of the inclusion of schooling residual and propensity score terms in most of the wage equations. The inclusion of no selectivity terms at all is supported in only three cases. In eight cases, the CV results suggest a correction for the endogeneity of schooling, and a correction for sample selection is supported in five cases. The inclusion of interaction terms between the schooling residual and propensity score does not get substantiation, implying that these variables are separable in the wage equations.

The fitted CV-minimizing models yield estimates for the marginal and average marginal returns to education. We begin the interpretation of the results by taking a look at the "marginal return curves" in Figure 1. By a simple eyeballing of the curves, three main conclusions can be drawn. First, the marginal return estimates are almost entirely positive for different amounts of schooling attained and remain at a relatively high level during the whole observation period, the point estimates varying between -0.02 and 0.20. Second, the results support the hypothesis that the marginal return is not constant over different amounts of schooling. In fact, the estimated marginal return curves look concave (with the exception of the linearly increasing curve for the university regions in 1995) and indicate that the marginal return increases at low levels of schooling and starts to decrease at around 13–16 years[15]. The results also suggest that the shape of the curves has slightly changed over time, as the decrease in the marginal return at the high levels of schooling is less steep after 1990. Third, the 95 percent confidence intervals for the estimates are substantially large in all cases in which schooling residual terms are included in the wage equation, in comparison with the cases in which they are not. This implies that the correction for the endogeneity of schooling has the price of causing severe inefficiency in the estimation.

Finally, we take a look at the interpretation of the estimation results in terms of our main research question, that is, whether there have been regional differences in the returns. Based on the curves in Figure 1, the marginal returns for the two region types are very close to each other in the cases of 1980, 1990, and 2004, whereas larger differences are estimated for the other years. In 1980 and 1995, the marginal returns are higher for the university regions, while the opposite is true for 2000. A between-year comparison reveals that these differences are consequences of occasional shifts in the marginal return curves: the difference in 1985 occurs after a downward shift in

[15] This kind of concave pattern has also been found in the marginal returns of several other countries by Trostel (2005).

Figure 1
Region-type specific marginal return to education at different amounts of schooling attained. Das-Newey-Vella estimates.



Notes: Δ = Metropolitan area; □ = University region. Dashed lines mark the 95% confidence bounds.

the metropolitan area curve, and the difference in 2000 after a similar shift in the university region curve.

To give a more specific picture of the regional differences, we report the estimated marginal returns and differences in these at 11 and 14 years of schooling in Table 5[16].

---

[16] The point estimate at 11 years of schooling roughly depicts the return for completing the last year at a secondary-level school (general upper secondary or vocational school). Then again, 14 years of

Table 5
Marginal return to education at 11 and 14 years of schooling. Das-Newey-Vella estimates.

	1980		1985		1990		1995		2000		2004	
11 years of schooling												
Metro area	.097	a	.055	a	.103	a	.099	a	.096	a	.050	a
	(.021)		(.004)		(.020)		(.026)		(.023)		(.007)	
Univ. region	.075	b	.108	a	.111	a	.130	a	.045	a	.039	a
	(.029)		(.029)		(.026)		(.029)		(.009)		(.007)	
Difference	.023		-.053	d	-.008		-.031		.051	c	.012	
	(.035)		(.029)		(.034)		(.038)		(.024)		(.009)	
14 years of schooling												
Metro area	.124	a	.078	a	.134	a	.146	a	.143	a	.086	a
	(.022)		(.005)		(.020)		(.025)		(.024)		(.006)	
Univ. region	.109	a	.137	a	.142	a	.165	a	.079	a	.090	a
	(.029)		(.031)		(.025)		(.030)		(.007)		(.006)	
Difference	.015		-.059	d	-.008		-.019		.064	b	-.005	
	(.037)		(.031)		(.032)		(.039)		(.024)		(.008)	

Notes: Standard errors in parentheses. Tests for significance: [a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$; [d] $p < 0.1$.

At the 5 percent level, significant differences between the region types are found only for 2000, in which case the marginal return estimate is higher for the metropolitan area at both 11 and 14 years of schooling by 0.051 and 0.064, respectively. In the case of 1980, the evidence on regional differences is weaker, as the estimated differences are significant only at the 10 percent level. In this case, the marginal return estimates at 11 and 14 years of schooling are higher for the university regions by 0.053 and 0.059, respectively. Relatively large differences are also measured for 1995, but these are not significant due to large standard errors.

Our final inspection concerns the estimates of average marginal return to education and the regional difference in these estimates. In addition to the Das-Newey-Vella estimates, Table 6 shows the average returns estimated by the standard OLS method in which the linearity of schooling and exogenous schooling and location choices are

schooling corresponds to the last year of bachelor studies at a university or university of applied sciences (polytechnic).

assumed. The choice of estimation method seems to affect our conclusions about the average returns: in the cases where schooling residual terms are included in the wage equation, the estimated average return is clearly higher than the corresponding OLS estimate[17]. In some cases, this also affects our conclusion about the regional difference. Whereas the OLS estimates suggest that there have been no significant regional differences in the average returns in 1985 and 2000, the Das-Newey-Vella estimates imply large and significant differences. Naturally, these differences are in the same direction as the differences in the marginal returns reported earlier. For the remaining years, no significant regional differences are found by either method.

Table 6
Average marginal return to education. Comparison of Das-Newey-Vella and OLS estimates.

	1980	1985	1990	1995	2000	2004
Das-Newey-Vella						
Metro area	.079 [a]	.040 [a]	.104 [a]	.115 [a]	.117 [a]	.062 [a]
	(.022)	(.003)	(.020)	(.025)	(.023)	(.005)
Univ. region	.063 [c]	.105 [a]	.114 [a]	.149 [a]	.059 [a]	.065 [a]
	(.029)	(.029)	(.025)	(.029)	(.005)	(.004)
Difference	.016	-.064 [c]	-.010	-.034	.059 [c]	-.004
	(.036)	(.029)	(.032)	(.039)	(.023)	(.006)
OLS						
Metro area	.048 [a]	.048 [a]	.054 [a]	.043 [a]	.060 [a]	.063 [a]
	(.002)	(.002)	(.003)	(.003)	(.004)	(.004)
Univ. region	.046 [a]	.041 [a]	.054 [a]	.050 [a]	.056 [a]	.062 [a]
	(.003)	(.003)	(.004)	(.004)	(.005)	(.004)
Difference	.002	.006	.000	-.006	.004	.001
	(.003)	(.004)	(.005)	(.005)	(.006)	(.005)

Notes: Standard errors in parentheses. Tests for significance: [a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$.

[17] In previous studies, the use of family background instruments has also been found to increase the estimated returns to education in Finland (Uusitalo, 1999; Tokila and Tervo, 2010) as well as in several other countries (Trostel et al., 2002). Kruhse-Lehtonen (2007), in turn, used the region of birth instrument similar to ours, which yielded higher return estimates for most of the studied age-cohort groups, while also smaller estimates for some of the groups.

5 Conclusion

In this paper, we have estimated regional differences in returns to education within a country, Finland, while being the first to simultaneously take into account the endogenous nature of schooling and location choices as well as the possibly nonlinear relationship between wage and the years of schooling. The nonparametric selectivity correction approach of Das et al. (2003) was applied in order to relax the exogeneity and linearity assumptions during the estimation. When comparing the estimates from this procedure to those obtained by the standard OLS method, we found some dissenting evidence on the region-type-specific returns and the differences therein. Thus, we conclude that the regional variation in skill prices may not be adequately captured by methods such as OLS that rely on controlling for observed variables only. Furthermore, we found strong evidence that the marginal return to education has not been constant over different amounts of schooling attained in Finland. The results showed increasing marginal return to post-compulsory-level education and decreasing marginal returns at the higher levels, which is in line with the earlier findings from other countries (Trostel, 2005). However, the decreasing pattern in the marginal returns at the higher levels seemed to have become less prominent towards the end of the observation period, implying that the monetary payoff for completing higher education degrees has improved.

According to the results, the marginal returns to education were relatively similar for the metropolitan area and the group of university regions in most of the observation years. Nevertheless, large and significant differences were found in two cases, 1985 and 2000, indicating lower returns for the metropolitan area in the first case and higher returns in the latter. Although we do not attempt to analyze the exact causes of these estimated regional differences, simple explanations from the supply-demand framework can be offered. A potential explanation for the 1985 result could be that the relatively large supply of educated workforce in the metropolitan area during the early 1980s decreased the marginal returns in this region. Some support for this theory is provided in Table 1, in which the average years of schooling are clearly shown to be larger for the metropolitan area in 1980 and 1985; after this, the gap reduces significantly, suggesting that the convergence of the regional returns may have taken place through the adjustment of regional supplies. Then again, the estimated gap in the regional returns for 2000 might be related to the “information technology boom" that occurred at the time, which may have kept the demand for highly educated workers at a higher level in the metropolitan area. Finally, we note that the regional differences in returns to education seem to have appeared only temporarily; no persistence in these is found.

Acknowledgements

Financial support from the Academy of Finland is gratefully acknowledged (grant number 127049). We also wish to thank Jani-Petri Laamanen and Otto Toivanen, as well as the participants at the Allecon seminar in Jyväskylä (3–4 May 2010) for providing helpful comments.

References

Acemoglu, D., 2002. Technical change, inequality, and the labor market. Journal of Economic Literature 60, 7–72.

Ammermueller, A., Kuckulenz, A., Zwick, T., 2009. Aggregate unemployment decreases individual returns to education. Economics of Education Review 28, 217–226.

Bedi, A.S., Gaston, N., 1999. Using variation in schooling availability to estimate educational returns for Honduras. Economics of Education Review 18, 107–116.

Bennett, R., Glennerster, H., Nevison, D., 1995. Regional rates of return to education and training in Britain. Regional Studies 29, 279–295.

Betts, J.R., 1995. Does school quality matter? Evidence from the National Longitudinal Survey of Youth. The Review of Economics and Statistics 77, 231–250.

Black, D., Kolesnikova, N., Taylor, L., 2009. Earning functions when wages and prices vary by location. Journal of Labour Economics 27, 21–47.

Bourguignon, F., Fournier, M., Gurgand, M., 2007. Selection bias correction based on multinomial logit model: Monte Carlo comparisons. Journal of Economic Surveys 21, 174-205.

Breiman, L., Spector, P., 1992. Submodel selection and evaluation in regression. The X-random case. International Statistical Review 60, 291–319.

Card, D., 2001. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica 69, 1127–60.

Card, D., Krueger, A., 1992. Does school quality matter? Returns to education and the characteristics of public schools in the United States. The Journal of Political Economy 100, 1–40.

Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, L. N., Kenneth Grant, E., Swidinsky, R., and Vanderkamp, J. (Eds.), Aspects of labour market behavior: Essays in honour of John Vanderkamp, Toronto: University of Toronto Press Inc., 201–222.

Dahl, G.B., 2002. Mobility and the return to education: Testing a Roy model with multiple markets. Econometrica 70, 2367–2420.

Das, M., Newey, W.K., Vella, F., 2003. Nonparametric estimation of sample selection models. Review of Economic Studies 33–58.

Falaris, E.M., 1987. A nested logit migration model with selectivity. International Economic Review 28, 429–443.

Griliches, Z., 1977. Estimating the returns to schooling: Some econometric problems. Econometrica 45, 1–22.

Harmon, C., Oosterbeek, H., Walker, I., 2003. The returns to education: Microeconomics. Journal of Economic Surveys 17, 115–156.

Harmon, C., Walker, I., 1999. The marginal and average returns to schooling in the UK. European Economic Review 43, 879–887.

Heckman, J., Layne-Farrar, A., Todd, P., 1996. Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. The Review of Economics and Statistics 78, 562–610.

Kane, T.J., Rouse, C.E., 1993. Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter? NBER Working Paper, Nr. 4268.

Kirjavainen, T., 2009. Essays on the efficiency of school and student achievement. VATT Publications 53, Helsinki: Government Institute for Economic Research.

Kruhse-Lehtonen, U., 2007. Empirical studies on the returns to education in Finland. Publications of the Helsinki School of Economics, A:289.

Lee, L.F., 1983. Generalized econometric models with selectivity. Econometrica 51, 507–512.

Long, M.C., 2008. College quality and early adult outcomes. Economics of Education Review 27, 588–602.

Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67, 565–603.

Sanders, W., 2004. Education and urban household location. Economic Letters 82, 327–331.

Tokila, A., Tervo, H., 2010. Regional differences in returns to education for entrepreneurs versus wage earners. The Annals of Region Science, forthcoming.

Trostel, P., 2005. Nonlinearity in the return to education. Journal of Applied Economics 8, 191–202.

Trostel, P., Walker, I., Woolley, P., 2002. Estimates of the economic return to schooling for 28 countries. Labour Economics 9, 1–16.

Uusitalo, R., 1999. Return to education in Finland. Labour Economics 6, 569–580.

Witten, I.H., Frank, E., 2005. Data mining: Practical machine learning tools and techniques, 2nd ed. Amsterdam: Morgan Kaufmann.

Appendix 1: Variable definitions

- *Monthly wage*. The annual earned income divided by the months worked.
- *Years of schooling*. This variable is based on the highest level of schooling achieved. The variable gets value 9 if the individual has completed only the compulsory level, 12 for the upper secondary level, 14 for the lowest tertiary level, 15 for the bachelor's level, and 17 for the master's level.
- *Job region type*. This variable describes the "rurality" of the sub-region in which the job is located. This variable includes six categories: metropolitan area, university region, regional centre, industrial centre, rural region, sparsely populated region. Only the first two categories are used in this study.
- *Age*. The models include a dummy variable for each value of the individual's age.
- *Tenure*. The number of full months since the beginning of the current employment. This variable is not available for 1980 and 1985. Also a quadratic term of this variable is included in the wage equation to allow a nonlinear relationship between wage and tenure.
- *Married*. This variable gets value 1 if marital status is "married", and 0 otherwise.
- *Swedish speaker*. This variable gets value 1 if the individual's mother tongue is Swedish, and 0 otherwise.
- *Mother's/Father's education*. The highest educational level of mother/father (observed in 1970, 1980, 1990, and 2000). These variables include three categories: compulsory level or no classification available, secondary or the lowest tertiary level, bachelor level or higher.
- *Mother's/Father's socioeconomic status*. The latest socioeconomic class of mother/father (observed in 1970, 1980, or 1990). These variables include six categories: farmer, entrepreneur, high-rank official, low-rank official, worker, no classification available.
- *Region of birth*. The region in which the individual was born. This variable includes 21 categories: Uusimaa, Eastern Uusimaa, Finland Proper or Aland Islands, Satakunta, Tavastia Proper, Pirkanmaa, Päijänne Tavastia, Kymenlaakso, Etelä-Karjala, Southern Savonia, Northern Savonia, North Karelia, Central Finland, Southern Ostrobothnia, Ostrobothnia, Central Ostrobothnia, Northern Ostrobothnia, Kainuu, Lapland, born abroad.

- *Job location*. The region in which the job is located. The wage equations include either three (metropolitan area) or six (university regions) job location categories.