University of Konstanz
Department of Economics

How Far is the East? Educational Performance in Eastern Europe

Alina Botezat and Ruben R. Seiberlich

Working Paper Series
2011-15

How Far is the East? Educational Performance in Eastern Europe

Alina Botezat
Romanian Academy - 'Gh.Zane' Institute for Economic and Social Research

Ruben R. Seiberlich*
University of Konstanz

This version: May 22, 2011

Abstract

When the Soviet Union collapsed a transition process started in Eastern Europe. This included a number of reforms to adapt the educational system to the new requirements of the job market. To assess the educational systems in Eastern Europe, this paper takes a look at the gap in PISA test scores between Finland, the best performing country, and seven Eastern European countries, as well as, between Eastern European countries. The methodology applied in this paper is a semiparametric version of the threefold Blinder-Oaxaca decomposition, an approach which is not yet used in the research regarding the differences in school outcomes.

JEL classification: J24, I21, C14
Keywords: PISA, test score gap, decomposition, semiparametric, propensity score matching

*Corresponding author. Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone ++49-7531-88-5111, Fax -4450, email: Ruben.Seiberlich@uni-konstanz.de. Ruben Seiberlich gratefully acknowledges support received from the State of Baden-Württemberg through its LGFG-scholarship. Alina Botezat acknowledges support received within the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU- 6/1.5/S/25. This paper benefited from helpful ideas and discussions with Benjamin Hess, Elisabeta Jaba, Fabian Krüger and especially our supervisor Winfried Pohlmeier. Moreover, we are very grateful to Ronald Oaxaca for his comments and we thank the seminar participants at the University of Konstanz and the University of Iaşi, as well as the session participants at the Annual Meeting of the Irish Society of New Economists 2010, Dublin, and at the Annual Meetings of the Society of Labor Economists (SOLE) 2011, Vancouver for their helpful comments. All remaining errors are ours.

1 Introduction

Over the past twenty years the countries of Eastern Europe have gone through periods of transition and structural changes which also affected the educational system. Most Eastern European countries have performed reforms to adapt the educational system to the new requirements of the job market. One aim of these reforms was to improve the quality of schooling, which is the driving force for the economic growth Hanushek and Kimko (2000). The success of these reforms in education can be assessed by analyzing the results of international standardized test scores such as PISA (Programme for International Student Assessment), TIMSS (Third International Mathematics and Science Study), or PIRLS (Progress in International Reading Literacy Study). The results from PISA 2006, for example, show that there is a high variation in performance of the Eastern European countries. Many of the Eastern European countries are still in a transition process and have not yet overcome the initial disadvantages compared to Western countries. Most of them perform statistically significantly below the OECD average and only Estonia, Slovenia and Czech Republic perform in the upper part of the distribution OECD (2007).

The first aim of this paper is to analyze the PISA test score gaps between Finland and seven Eastern European countries (Estonia, Czech Republic, Hungary, Romania, Bulgaria, Latvia and Slovakia). Using data from the 2006 survey, we choose Finland as benchmark for our analysis. It is the best performing country in the PISA study and is considered to have the most effective and equitable school system Ammermüller (2007). Understanding the test score differences is of huge importance since it allows improving the school systems and, therefore, directly provides relevant information for educational policies.

The second aim is to disentangle the PISA test score gap between countries which had similar educational systems 20 years ago. Estonia as well as Latvia belonged to the Soviet Union until 1991, the Czech and Slovak Republic together formed Czechoslovakia until the end of 1992. The precondition that two countries belonged to the same country forms a natural experiment, that reveals how two countries, which start from more or less the same point, develop over the subsequent years.

To achieve the two aims, we disentangle the effects that explain the gaps in order to show which factors contribute to the differences in school performance. This enables us to answer the following questions: Could institutional reforms improve the school performance and overcome the disadvantages resulting from an unfavorable social background? Given the same socio-economic background, which educational system manages to increase the returns to these individual characteristics? Will a student perform better if, given his socio-cultural background, he would attend the school system of a country that on average performs better than that of his home country? In other words, how would school performance change if students from Eastern Europe participated in the Finnish school system?

This paper contributes to the previous literature in several ways: First of all, it makes an original contribution by introducing a semiparametric method into the educational literature, which is commonly used in explaining the gender differences in wages, but not in the research regarding the decomposition of differences school outcomes. This is important in its own right since recent papers have demonstrated that the functional form assumptions of the parametric Blinder (1973) and Oaxaca (1973) decomposition can give misleading results (Barsky, Bound, Charles, and Lupton (2002), Mora (2008)). The methodology applied here is a semiparametric version of the threefold Blinder-Oaxaca decomposition which disentangles the effects in an endowment, return and an interaction effect between these two. The method is based on an approach proposed by Frölich (2007), who uses propensity score matching to compute the counterfactual outcome. To account for differences in the common support we further follow the procedure developed by Ñopo (2008). Furthermore, this is the first paper that decomposes the differences in PISA test scores between the best performing country in the study and several Eastern European countries as well as between some Eastern European countries.

The reminder of the paper is the following: The next section provides a general overview of the educational systems in Eastern Europe. The second section focuses on the identification strategy used to decompose the gap in school performance. The third section presents the PISA study 2006 and describes the data. Section 4 discusses the estimation results and section 5 takes a closer look at the unexplained part of the total gap. The last section concludes.

2 Overview of the Educational Systems in Eastern Europe

According to Cerych (1997) and Radó (2001) the following issues of the educational reforms in Eastern Europe can be identified: In all countries, a depolitisation of education took place, implying the end of ideological control and orientation of the system. Furthermore, educational change led to the decentralization and liberalization in educational management by breaking down the state monopoly. Moreover, the pupils or their parents respectively have now freedom of choice concerning their educational path. Another issue of the reforms was redefining the quality in education. During communism, the most important indicators for quality was the participation rates and the achievement of the most talented students Radó (2001). The new reforms instead focused on the quality of curricula and exams, on the improvement of the in-service training of the teachers, as well as on the specific learning interests of pupils. Aside from the problem of financing of education, another issue of the reforms was that of equity. In the old system, the tracking of students in vocational, technical and general schools, as well as the institutionalized segregation of disabled and minority children (Roma, for example) deepened the inequalities in the education system, which were then reflected on the labor market and in society.

Even if these countries started reforms at the same time, their subsequent evolution was different, depending, especially, on the development and the speed of economic reforms. Looking, for example, at countries such as Estonia, Czech Republic and Hungary, that went through a process of rapid privatization Bjørnskov and Potrafke (2011), they are also those countries that are among Eastern European countries the best performing in PISA test scores. Thus, with few exceptions, we cannot speak of continuity in educational reforms as long as they depend on factors outside the system itself. Only in the case of Hungary and Estonia there were undivided educational policies, due to measures taken before 1989. In Hungary, the decentralization of the educational system had already started in 1985 Radó (2001). The Estonian schools already won a degree of autonomy regarding the content of curricula during the Soviet period when textbooks were predominantly written by Estonian authors Kitsing (2008).

As culture has an impact on institutions and economic outcomes Guiso, Sapienza, and Zingales (2006), we should also mention the heterogeneity in the cultural and historical backgrounds in order to explain the different evolution of the educational reforms in Eastern Europe. Our sample consists of Romania and Bulgaria from South-Eastern Europe, Hungary, Czech and Slovak Republic from the Visegrad group and Estonia and Latvia from the Baltic region. Romania and Bulgaria are traditionally Orthodox Christian with some Islamic influence. Economically and educationally, they mainly developed after World War II. In comparison, Hungary, Czech and Slovak Republic are historically of Roman Christian culture and developed economically and educationally much earlier. The two Baltic Republics, on the other hand, are historically influenced by Germany and Russia. Since their independence in the early 1990's, they have close contact with Scandinavian countries. Their current educational systems therefore combine Nordic and Central European characteristics Cerych (1997).

Generally, previous empirical research on the school performance of the Eastern European countries is quite limited, providing mixed results and inconclusive evidence. One reason was the lack of reliable data that can objectively describe the educational process in these countries. Before 1989, data reported on human capital stock (years of schooling, for example) were overestimated Beirne and Campos (2007) and, after 1989, the participation at the international standardized tests (TIMSS, PIRLS, PISA) was not the same for all countries. Estonia, for example, participated for the first time in the PISA Study in 2006. The existence of such comparative data and of cross-national individual-level survey has allowed in the last years the extension of research, promising to answer key questions concerning the quality of the educational system in Eastern Europe.

For the transition period, the paper by Ammermüller, Heijke, and Wössmann (2005) provides evidence regarding the production of school quality in Eastern European countries. Even if these countries faced similar characteristics in the economic and political development, the impact of individual factors, school resources and institutional settings on school performance shows different patterns. Using TIMSS data from 1995, the authors show that the student background has a lower impact in those countries which perform worse (Lithuania, Latvia and Romania) and which adopted reforms regarding the school system later than the other countries. The largest effects are obtained in Czech Republic and Hungary. The impact of school resources and teacher characteristics on school performance is low in magnitude and does not necessarily indicate a particular pattern. Only in some cases (Romania, Czech Republic and Hungary), a better training and a richer experience of the teachers can positively influence the test scores. The most favorable institutional setting is in Czech Republic, although the results show that the variation in test scores cannot be explained by institutional differences between countries. All in all, Ammermüller, Heijke, and Wössmann (2005) show substantial effects of student background on educational performance and much lower impact of resources and institutional settings.

One of the efforts of the educational reforms in Eastern Europe was to adjust the educational systems to the new labor market conditions. In this sense, the literature has also attempted to estimate returns to schooling in different transition economies. Flabbi, Paternostro, and Tiongson (2008) analyze a sample of eight countries (Bulgaria, Czech Republic, Hungary, Latvia, Poland, Russia, Slovak Republic and Slovenia) over the early transition period up to 2002. Using data from the ISSP (International Social Survey Programme), they test if the transition to a market economy leads to higher returns of schooling and find a weak empirical evidence in this sense. Large cross-country variation could be identified only in the levels of returns to schooling. They classify the countries in the following three groups. Hungary and Poland exhibit high levels of returns to schooling, Bulgaria, Latvia, Slovenia as well as Russia 'medium' returns of schooling and Czech and Slovak Republic low returns.

3 Identification Strategy

Understanding the differences in school achievement is one of the central themes in economics of education. The analysis of disparities in school performance are focused either on the gender gap in different subjects Fryer and Levitt (2010), Niederle and Vesterlund (2010), on the differences between countries McEwan and Marshall (2004), Ammermüller (2007), and between different subgroups Card and Rothstein (2007), Patacchini and Zenou (2009), Krieg and Storer (2006), Duncan and Sandy (2007), Schneeweis (2010).

All of these studies use a parametric approach and most of them used the Blinder-Oaxaca decomposition or a modified parametric version of it. The traditional Blinder-Oaxaca decomposition determines the source of the differences at the means and breaks down a gap into two parts by estimating one counterfactual. The first part, the characteristics effect, can be explained by the differences in the characteristics of individuals and the second part, commonly known as the unexplained gap, is a structure effect, which reflects the differences in slope coefficients. A comprehensive overview of the Blinder-Oaxaca decomposition is provided by Fortin, Lemieux, and Firpo (2010). The main disadvantages of the Blinder-Oaxaca decomposition are the ignorance of the common-support problems and the functional form assumptions. To avoid these drawbacks, we apply a semiparametric method, which makes it possible to identify the counterfactual outcome for every individual separately, allowing for arbitrary individual effect heterogeneity Heckman, LaLonde, and Smith (1999), Imbens (2004). Moreover, the counterfactual outcomes are only computed for those individuals who are actually comparable.

This semiparametric method identifies the counterfactual outcome for each individual as it is done in the evaluation literature. There, the interest usually lies in the estimation of the effect of a program. To isolate the true effect of the program on a particular individual, the observed outcome has to be compared to the outcome that would have resulted had the individual not been treated (not participated in the program). To estimate this counterfactual outcome, information on the non-participants is used. One possibility is to match treatment with comparison units that are similar in terms of their observable characteristics. Generally, matching directly on the vector of characteristics would be computationally demanding and, due to the curse of dimensionality, it would become hard to find good matches if the number of covariates is large.

To overcome this problem, Rosenbaum and Rubin (1983) demonstrate that matching can be done on a single-index variable, namely the propensity score. Frölich (2007) is the first who uses such a matching procedure outside the treatment evaluation literature. He shows that mean independence is sufficient for consistency of propensity score matching and uses it to decompose the gender wage gap between those who are on the common support analogously to the Blinder-Oaxaca decomposition into a characteristics and return effect. In this paper, we will extend this procedure to estimate a threefold decomposition.

To obtain the propensity score, we estimate the probability that an individual belongs to the better performing country ($D = 1$) by a logit regression, i.e.

$$p = \Pr[D = 1|X = x] = F(x'\beta) \tag{1}$$

where $F(x'\beta)$ represents the cumulative logistic distribution. In the next step, the density of this propensity score is estimated using a Gaussian Kernel estimator. Kernel matching then uses all members of one group to generate a match for each observation in the other group. Thereby, the contribution of each member is determined by the bandwidth and is smaller, the poorer the match is. The bandwidths are selected by leave-one-out cross-validation and are chosen to minimize the least-squares criterion. Let $f_1(p)$ be the distribution of the propensity score $p = p(X)$ among those from country $D = 1$ (the better performing country), $f_0(p)$ the distribution among those pupils from country $D = 0$ (the worse performing country) and $f_d^S(p)$, for $d = 0, 1$, the density of p in the subpopulation of those from country $D = d$ belonging to the common support, S, of the two countries.[1] In such a way, the test score gap for the common support subpopulation

$$\Delta_S = E_S[Y^1|D = 1] - E_S[Y^0|D = 0] \tag{2}$$

where Y^d indicates the outcome of those from country $D = d$, for $d = 0, 1$, can be expressed as

$$\Delta_S = \int_S E[Y|p(x) = p, D = 1] f_1^S(p) \, \mathrm{dp} - \int_S E[Y|p(x) = p, D = 0] f_0^S(p) \tag{3}$$

Since the decomposition is only for the common support subpopulation, the overlap assumption is always fulfilled and the only identifying assumption is mean independence given x, e.g. if

[1] $f_d^S(p) = \frac{f_d(p)}{\mu_{S|D=d}}$ is scaled such that the integral integrates to one. Thereby, $\mu_{S|D=d}$ is the empirical probability of being on the common support conditional on being from country d.

$E[Y|D=0,X=x]=E[Y|D=1,X=x]$ holds Frölich (2007) shows that the counterfactual outcomes is identified by estimating

$$E_S[Y^1|D=0]=\int_S E[Y|p(x)=p,D=1]f_0^S(p)\ \mathrm{dp} \text{ and} \tag{4}$$

$$E_S[Y^0|D=1]=\int_S E[Y|p(x)=p,D=0]f_1^S(p)\ \mathrm{dp} \tag{5}$$

where the counterfactual outcome for $p(x)=p$ can be estimated by the Nadaraya-Watson estimator

$$\hat{E}\left[Y|p(x)=p,D=d\right]=\frac{\sum_i^{n_d}K\left(\frac{p-p_i^d}{h}\right)Y_i^d}{\sum_i^{n_d}K\left(\frac{p-p_i^d}{h}\right)},\quad \text{for } d=0,1 \tag{6}$$

Thereby, K is the kernel function, h the bandwidth, n_d the number of observations and p^d the propensity score of those in country d.[2] The first counterfactual $E_S[Y^1|D=0]$ gives the expected outcome those from country $D=0$ would have in country $D=1$.[3] In section 4, we will argue why the assumption of mean independence given x is reasonable in our analysis.

In order to disentangle the effects of the gap, we extend the procedure applied by Frölich (2007) by decomposing the gap into three parts:

$$\begin{aligned}
&\int_S E[Y|p(x)=p,D=1]f_1^S(p)\ \mathrm{dp}-\int_S E[Y|p(x)=p,D=0]f_0^S(p)\ \mathrm{dp}\\
&=\underbrace{\int_S E[Y|p(x)=p,D=0]\left[f_1^S(p)-f_0^S(p)\right]\ \mathrm{dp}}_{\Delta_c}\\
&+\underbrace{\int_S \left[E[Y|p(x)=p,D=1]-E[Y|p(x)=p,D=0]\right]f_0^S(p)\ \mathrm{dp}}_{\Delta_r}\\
&+\underbrace{\int_S \left[E[Y|p(x)=p,D=1]-E[Y|p(x)=p,D=0]\right]\left[f_1^S(p)-f_0^S(p)\right]\ \mathrm{dp}}_{\Delta_{cr}}
\end{aligned} \tag{7}$$

In terms of the Blinder-Oaxaca decomposition, the first term can be attributed to differences in the distributions of individual characteristics (over the common support) and is, therefore, the characteristics effect (Δ_c). It captures the difference of the test scores that would vanish if the characteristics of the students from the worse performing country would follow the same distribution as those of the students from the better performing country. The second summand is the part of the gap (over the common support) that can be explained by those factors, other

[2] The bandwidth h is chosen by leave-one-out cross validation. Following Frölich (2004) we choose as bandwidth search grid $0.01\sqrt{1.2}^{g-2}$ for $g=1,...,59$ and ∞.

[3] Note that the problem of self-selection does not occur in our context as the treatment is the attendance of a school system in another country. Since we only use natives and second generation immigrants, this cannot be influenced by the individuals.

than individual characteristics, that determine the school performance (institutional aspects of the school system, resources, cultural factors). It would vanish if the students from the worse performing country would attend the school system of the better performing country and thus is analogous to the return effect (Δ_r) in the Blinder-Oaxaca decomposition. The term in the last brackets (Δ_{cr}) is the interaction effect between the characteristics and the return effect, reflecting the fact that the gap could also be determined by the simultaneous existence of differences in the distributions of individual characteristics and in the returns.

We decide to apply the threefold decomposition, used for the first time in decomposing the gap in test score by Ammermüller (2007), for the following reason. When we have to decompose a gap in test score, we should take into account that individuals can be better endowed with characteristics that, at the same time, are better rewarded by their school systems than by the other school system. In our case, the interaction term (if positive) expresses how much better the students from the worse performing country would score on average if the students from the better performing country did not have the advantage of being better endowed with those characteristics that are also better rewarded in terms of test scores in their country, or less endowed with those characteristics that are better rewarded in the worse performing country.

Since, in this study, we want to identify the PISA test score gap of all individuals from two country and not just the gap of those who are on the common support, we also need to account for individuals whose distributions of the propensity score do not overlap.

To account for the fact that some of the observations in the sample cannot be matched, Ñopo (2008) suggests first decomposing the observed average test score of both countries into a part that is due to those who are on the common support (S) and into a part resulting from those out of the common support ($\bar{S}$), i.e.

$$E\left[Y^d|D=d\right] = (1-\mu_{\bar{S}|D=d}) \int\limits_S E[Y|p(x)=p, D=d] f_d^S(p)\,\mathrm{dp}$$

$$+\,\mu_{\bar{S}|D=d} \int\limits_{\bar{S}} E[Y|p(x)=p, D=d] f_d^{\bar{S}}(p)\,\mathrm{dp} \tag{8}$$

for $d = 0, 1$ and where $\mu_{\bar{S}|D=d}$ is the empirical probability of being unmatched conditional on being from country d. In order to obtain expressions involving expected values conditional on the respective partitioned domains, the integrals are rescaled. If we use these expressions to calculate the PISA score gap Δ, we get

$$\Delta = \underbrace{\mu_{\bar{S}|D=1}\left[\int\limits_{\bar{S}} E[Y|p(x)=p,D=1] f_1^{\bar{S}}(p)\ \mathrm{dp} - \int\limits_{S} E[Y|p(x)=p,D=1] f_1^{S}(p)\ \mathrm{dp}\right]}_{\Delta_1}$$
$$+ \left[\int\limits_{S} E[Y|p(x)=p,D=1] f_1^{S}(p)\ \mathrm{dp} - \int\limits_{S} E[Y|p(x)=p,D=0] f_0^{S}(p)\ \mathrm{dp}\right]$$
$$+ \underbrace{\mu_{\bar{S}|D=0}\left[\int\limits_{S} E[Y|p(x)=p,D=0] f_0^{S}(p)\ \mathrm{dp} - \int\limits_{\bar{S}} E[Y|p(x)=p,D=0] f_0^{\bar{S}}(p)\ \mathrm{dp}\right]}_{\Delta_0} \tag{9}$$

Like in (7), the second part of this expression, which represents the gap for the common support subpopulation, can again be disentangled into three parts. Therefore, the whole gap Δ is disentangled into five parts, i.e.

$$\Delta = \Delta_1 + \Delta_c + \Delta_r + \Delta_{cr} + \Delta_0 \tag{10}$$

Additionally to the characteristics, return and interaction effect, we get two terms which make up the part of the gap that can be explained by differences between two groups of pupils from the countries, those whose characteristics can be matched to pupils from the other country and those who cannot. Using this method for decomposing PISA test score gaps, it is ensured that we look separately at those individuals who actually have comparable background characteristics and those who do not. Each of the five components of the total gap can be interpreted as follows:

- Δ_1 represents the gap in test scores that can be explained by differences between those students from the better performing country who can be matched to students from the other country and those who remain out of the common support, weighted by the empirical fraction of those who are out of common support from the better performing country. A positive value of the gap indicates that the students from the better performing country, who are out-of-support, perform better than their counterparts, who are on the common support.

- Δ_0 describes the gap in test scores that can be explained by differences between those students from the worse performing country who cannot be matched and those who are on the common support, weighted by the empirical fraction of those who are out of common support from the worse performing country.

- Δ_c, Δ_r and Δ_{cr} represent the characteristics, return and interaction effect computed only for the common support subpopulation and were described above.

Compared to the parametric Blinder-Oaxaca decomposition, the approach applied here differs in the following aspects. The regression function is not specified as linear and the counterfactual scores are simulated only for the common support subpopulation. In the case of Blinder-Oaxaca, the regressions are performed for all students, extrapolating the validity of the counterfactuals also for those individuals out of the common support.

Since we estimate the effects for every individual, another advantage of the semiparamteric approach is that we can decompose the test score gap for the common support subpopulation at different quantiles τ. Therefore, we estimate the three parts as follows:

$$\Delta_c^\tau = F^{-1}_{y^0|D=1,S}(\tau) - F^{-1}_{y^0|D=0,S}(\tau) \tag{11}$$

$$\Delta_r^\tau = F^{-1}_{y^1|D=0,S}(\tau) - F^{-1}_{y^0|D=0,S}(\tau) \tag{12}$$

$$\Delta_{cr}^\tau = F^{-1}_{y^1|D=1,S}(\tau) - F^{-1}_{y^0|D=1,S}(\tau) - F^{-1}_{y^1|D=0,S}(\tau) + F^{-1}_{y^0|D=0,S}(\tau) \tag{13}$$

$$\Delta_S^\tau = \Delta_c^\tau + \Delta_r^\tau + \Delta_{cr}^\tau = F^{-1}_{y^1|D=1,S}(\tau) - F^{-1}_{y^0|D=0,S}(\tau) \tag{14}$$

The adjusted quantiles are obtained indirectly via inverting the adjusted distribution functions.

4 Data

The following analysis is based on data from PISA 2006. PISA assesses the achievement of 15-year-olds in mathematics, reading and science literacy. Apart from test scores, data on pupils' social and cultural background were collected as well as information about the school environment of students OECD (2007).

The data contain information on more than 35 000 students and more than 2000 schools. For comparison reasons, the scores have been standardized to a mean of 500 and a standard deviation of 100. Our sample consists of data from Finland and seven Eastern European countries: Estonia, Czech Republic, Hungary, Romania, Bulgaria, Latvia and Slovakia. A general description of the variables used in this study is given in Table (A.1). Since the performance of the immigrants from the first-generation could also reflect the influence of other school systems than the one they currently attend, we decide to drop these students from the samples. Moreover, the share of first generation immigrants is quite different for the countries in our sample.

[Table (A.1) about here]

Having to deal with a high volume of data, the problem of missing data in PISA study is inevitable. As Ammermüller (2007) noted, dropping individuals with missing information could lead to an upward bias in test scores, since the missing data are not missing at random, being predominant among students who have low test scores. One solution to overcoming this problem is to predict the values of these data using the complete information available from all students. Thus, we decide to impute all the missing values by applying a method suggested by Wössmann, Lüdemann, Schütz, and West (2009).

Table (A.2) presents the weighted means and standard deviations for the variables used in our study.

[Table (A.2) about here]

To apply propensity score matching, we only use data at the individual level. These include measures for the students' characteristics (age, gender and grade) and for family background (number of books at home, parents' education). Furthermore, these variables are commonly used to measure the (in)equality of educational opportunities (Wössmann (2008), Schütz, Ursprung, and Wössmann (2008), Martins and Veiga (2010)). From these indicators, the number of books is the most important measure of family background, which best predicts the student performance (Wössmann (2003), Fuchs and Wössmann (2007), Wössmann (2008)). Therefore, we expect that the differences in these observable characteristics of the students from different countries explain the gap in their school performance being confident that the mean independence assumption holds in our analysis.

According to Schütz, Ursprung, and Wössmann (2008), there is a high variation across school systems as to what extent they achieve equal opportunities for children from different family backgrounds. Their results, as well as those from PISA OECD (2007), confirm the fact that from Europe, Finland has the most equitable school system. Using these variables that describe the family background and which are important determinants of school performance to match the Finnish students with those from countries from Eastern Europe, we can regard the differences in school performance in such matched groups as unexplained by individual and background characteristics. In this way, we intend to measure precisely how much of the total gap to Finland can be explained by differences in the distributions of observable individual characteristics and how much can be explained by other factors, such as school resources and different institutional features of the school system.

We decide not to include school variables in obtaining the propensity score for the following two main reasons. First of all, the matches become poor when including school or country variables. Secondly, by matching on individual characteristics, the resulting estimate of the counterfactual outcome represents the conditional probability that an individual with propensity score p would, for example, attend a comprehensive school, a private school or a class with less than 20 pupils. To our understanding, this is of interest and not what an individual would have in the other country if he/she would attend exactly the school type and class design he/she attends in his/her home country.

The descriptive statistics reported in Table (A.2) show some differences in observable characteristics between students from different countries. Looking at the grade in which students are, reveals the fact that in Finland, Estonia and Latvia most of the students from the sample are in the 9^{th} grade and the percentage of those being in the 10^{th} or 11^{th} class is below 3 percent. In Czech and Slovak Republic most of the students are in the 10^{th} grade and in Estonia, more than a quarter of the students are in 8^{th} grade, while in Bulgaria, Czech and Slovak Republic the percentage is below 7. Regarding the number of books, more than a third of students from Bulgaria and Romania have less than 50 books at home, while the corresponding percentage in the other countries is between 16 and 20. In all countries, the parents are well educated, but some differences can still be noticed. In Finland, the majority of the parents have a tertiary education whereas the majority in the Eastern European countries have upper secondary education. Among the Eastern European countries, the parents in Czech and Slovak Republic are best educated. In both countries, more than 75 percent of the students have parents who completed upper secondary education.

As mentioned before, we mainly analyze the gap in school performance between Finland and countries from Eastern Europe in this paper. According to data from Table (A.2), the range of differences in test scores is very large: between 152 points (Finland - Romania in reading) and 32 points (Finland - Estonia in science). Also the spread of the test scores in countries from Eastern Europe is very different: higher in Bulgaria and in Czech Republic, lower in Estonia, Latvia and Romania.

Regarding the differences in the supports, we present in Table (A.3) the fractions of individuals, whose combinations of age, gender, grade, number of books and parents' education cannot be matched.

[Table (A.3) about here]

It can be seen that for 14 of the 18 cases less than 3 percent of the students are out of the common support. Nevertheless, for the decompositions between Finland and Hungary, Bulgaria, Czech as well as Slovak Republic, 15.39% to 54.84% of the students cannot be matched with pupils from Finland.[4] In all country comparisons at least 97.74% of Finnish students are on the common support and therefore as comparison units available.

5 Estimation Results

To estimate the different components of the PISA test score gap, we include all individual variables explained above. Since the estimation results are similar for math and science from the point of view of the magnitude and sign effects, we only report the science results. All of our decompositions are formulated from the point of view of the worse performing country ($D = 0$).

Results for the Decompositions of the Science score gaps between Finland and Eastern European countries

Table (A.4) shows the results of the semiparametric decompositions for the science PISA test scores between Finland and seven Eastern European countries.

[Table (A.4) about here]

The first striking result is that, for all seven countries, the return effect is significantly positive and the effect with the largest magnitude. This indicates that, given their average

[4]The reason for these shares is that, in these countries, the fraction of those in grade 10 or 11 is relatively high, whereas this number is nearly zero in Finland. This leads to an almost perfect predictor in the logit regression and, therefore, to more propensity scores out of the common support.

characteristics, the students from each of the seven Eastern European countries would have on average higher test scores in science if they attended the Finnish school system. Particularly the pupils from Bulgaria and Romania would profit from a such school system; making it possible to increase their score in science by more than 100 points on average.

The characteristics effect is smaller in magnitude and only significant for six country comparisons. The characteristics effect is not significantly different from zero between Finland and Czech Republic. It is positive for four countries. This reveals that the Finnish students tend to have slightly more favorable characteristics than the students from Eastern European countries on average. Only when we compare Finland with Hungary and Slovak Republic are the students' characteristics from these two worse performing countries actually more advantageous than the characteristics of the Finnish students. In these cases, the interaction effect is also high relative to the corresponding total gap, showing that the gap would be smaller if the Finnish students had not the advantage of being better endowed with those characteristics which are also better rewarded by the Finnish school system compared to the other school system.

Analyzing the values of Δ_0, we notice that the differences in the average test scores between those who are on the common support and those who are out of the common support is significantly different from zero only for the comparison of Finland with the Czech Republic and Hungary. In these cases, this effect matters even more than the differences in the characteristics for explaining the total gap. The negative values of Δ_0 show that the students from the worse performing countries who are out of the common support score significantly higher in the PISA science test score than those who are on the common support. This negative effect can again be explained by the fact that, in these countries, we obtained a sizeable share of pupils in grade 10 or 11. As described before, those pupils are more likely to be out of common support but, since they are older (e.g. had more schooling), their scores are on average higher in the PISA tests. The magnitude of Δ_1 on the other hand is negligible and insignificant in all cases.

The result of the comparison between Finland and the Czech Republic also illustrates the importance of controlling for those who are out of the common support. If we would have looked at the common support subpopulation only, the gap in the average science test score would be 67.86 instead of 51.35 and therefore over 30% larger than the actual gap of the whole sample.

All in all and under our identifying assumption, our estimation results suggest that the higher average score in science in Finland is mainly due to the fact that the Finnish school system is more efficient in transforming the given inputs into PISA test score points.

Results for the Decompositions of the Reading score gaps between Finland and Eastern European countries

Table (A.5) contains the results for the PISA reading scores.

[Table (A.5) about here]

It can be seen that, except for the comparison of Finland and Latvia, the gaps for reading scores are larger than the gaps for the science results. Moreover, the results yield more or less the same interpretation as the results for the PISA science test scores. Again, all return effects are significantly positive and by far the largest in magnitude in explaining the majority part of the gap. Like in the decomposition of the science test scores, the characteristics effect are only negative for Hungary and Slovak Republic. For the other five countries, the characteristics effect is positive indicating that, on average, the Finnish students are slightly better endowed with more favorable characteristics or less endowed with less favorable characteristics. For the reading scores, four of the interaction effects are significantly different from zero. All of them are positive and suggest that the Finnish students have a slight advantage due to the fact that they are better endowed with those characteristics that also yield a higher return in Finland. The values for Δ_1 are again all around zero and insignificantly different from zero. Like for the PISA science test score gap, the values of Δ_0 are significantly negative for Hungary as well as the Czech and Slovak Republic, indicating again that the students from those countries, who cannot be matched with Finnish students, score on average higher in the PISA reading test than their counterparts who can be matched. As before, we attribute this to the fact that those who are out of the common support obtained more schooling by the time of the test.

Results for the Decompositions of the Science and Reading score gaps among Eastern European countries

As indicated before, the results from PISA study show that there is a significant variation in the performance, not only between Finland and Eastern European countries, but also between countries from Eastern Europe, which shared the same educational system for decades. We refer here to Czech and Slovak Republic as well as Estonia and Latvia. Since each pair of countries also share a common history with respect to their religion, culture and the influence of other countries, we expect them to be more similar than the students in the previous decompositions. Given these considerations, it is interesting to have a look at the gap of each of these two pair of countries that were more common twenty years ago but have developed differently since the early 1990's, in order to explain their different evolution over time in terms of test scores at PISA study. The decomposition results are presented in Tables (A.6) and (A.7).

[Table (A.6) about here]

[Table (A.7) about here]

As we expected, the characteristics effects are not significantly different from zero in both cases, indicating that the distributions of the individual propensity scores are very similar in each of these two pair of countries.

In both cases, almost the whole gap can be explained by the return effect. While for the decomposition gap between Estonia and Latvia the interaction effect is very small and insignificant, in case of Czech and Slovak Republic the magnitude of this effect is not negligible and works in favor of Slovak students. Even if the students from Czech Republic have the advantage of higher returns, they are less endowed with those characteristics that are better rewarded by their school system than by the Slovak system or more endowed with those characteristics that are better rewarded by the Slovak school system, as reflected by the negative values of the interaction effects.

Due to the similar distribution of the propensity score, the matching procedure works well for all individuals from the two compared countries. As can be seen in Table for more than 99.73% comparison units are available. This result also explains the fact that Δ_1 and Δ_0 have no contribution in explaining the total gap.

Results for the Semiparametric Decompositions of the Science and Reading score gaps at different quantiles

In this part we will focus on the gaps of the PISA test score at different quantiles instead of the mean.

Table (A.8) presents the results for the science test scores, showing that the distributions of the gaps are quite different for various country comparisons.

[Table (A.8) about here]

The course of the total gap shows different patterns. For some countries' comparisons (Finland-Estonia, Finland-Hungary and Finland-Romania), it increases in the lower part of the distribution up to 25$^{\text{th}}$ or 50$^{\text{th}}$ percentile and then decreases more or less for the rest part of the distribution. When we compare Finland-Czech Republic and Finland-Bulgaria the total gap in science is the highest for the first percentile (p_5) and then decreases steadily until the last percentile (p_{95}). The opposite pattern is found in the case of Czech-Slovak Republic, where the total gap is the smallest at the first percentile (p_5) and then increases steadily along the whole distribution. The decomposition results for Finland-Estonia and Finland-Latvia show that the total gap is relatively stable along the distribution.

Looking at the evolution of the characteristics effect at different quantiles, it can be noticed that, in all nine cases, the effect is negative for at least the last two quantiles. This reveals that for the high performing students, the characteristics of the students from worse performing countries are in fact more advantageous than those of the students from the better performing country. From the median to the lowest part of the distribution, the characteristics effect increases steadily in most cases (except for Finland-Czech Republic and Finland-Hungary), showing that the weaker students perform, the more disadvantageous are their characteristics relative to those of the students from Finland, at the respective percentile.

In all nine decompositions, the return effect decreases over the whole distribution. Except for the decomposition between Finland and Romania, the return effect even becomes negative at the 95% quantile. These results reveal the fact that, for weaker students, the respective school system in Eastern Europe is less able to convert the endowments of the students into a good performance compared to the school system of the better performing country. In other words, these students, given their characteristics, would perform better in science if they would attend the other school system. For high performing students this is different. For them, their school systems better succeed in transforming the students' characteristics into a good educational achievement, than could the reference school systems of the better performing countries. These results show that those countries who are better performing on average (e.g. in our decompositions Finland, Estonia and Czech) are better because they are better able to convert the given characteristics of the poor performing students into relatively high scores and not because they get the best out of the very good performing students.

The interaction effect increases steadily, being positive from about the 75 percent quantile upwards. The magnitude of this effect is very high, having a great influence on the total gap size at each quantile. At the lower part of the test score distribution, we observe the same pattern: The characteristics and return effects are generally positive, while the interaction effect is always negative. This result shows that the total gap at these quantiles would have been higher if the students from the better performing country would not be less endowed with those characteristics that are better rewarded by their school system as by the less performing system. Looking at the upper quantile, we notice that even if the characteristics and return effects are negative, the total gap turns positive due to the magnitude of the interaction effects. It thus shows, that, even if the characteristics of the students from the worse performing countries are more advantageous and the school system, that they attend, succeeds in transforming these characteristics into a good school performance, the advantage of the students from the better performing country is much higher, as reflected by the sign and magnitude of the interaction effect. These students are better endowed with those characteristics that are also better rewarded by their school system, generating higher test scores.

Table (A.9) displays the decomposition of the gaps at different quantiles for the reading test score.

[Table (A.9) about here]

Compared to the results presented above for the science test score gaps, the decompositions at different quantiles for the reading score reveal significant differences. On the one hand, the total gap in reading is smaller than the total gap in science at the upper part, and much higher at the lower part of the distribution. On the other hand, the differences in gaps between the two extremes of the distribution (p_5-p_{95}) are higher for the reading as for the science test score. Thus, these results show that there is a higher heterogeneity in students' performance not only between students from different countries at the respective percentile, but also along the same distribution of the reading test scores.

Compared to the science test scores gap, the course of the total gap in reading test scores is the same for all country comparisons between Finland and the seven Eastern European countries. For all these countries, the gap is decreasing from the first percentile to the end of upper part of the distribution. This result gives further insight as to why the Finnish students perform best in the PISA 2006 study. The Finnish school achieves that the poor performing students perform much better than the poor performing students of the other countries. If we look at the comparison between Estonia and Latvia, as well as, Czech and Slovak Republic, we find this pattern only for the first pair. In the case of Czech and Slovak Republic, the total gap is smallest in the lower part of the distribution and then increases steadily.

The characteristics effect decreases steadily from the first to the last percentile. As in the case of the science score gaps, the effect is negative only at the upper part of the distribution. The course of the return effect for reading is similar to the return effect in science test score, with the distinction that it remains positive, not only when we compare the test scores between Finland and Romania, but also in the case of Finland-Bulgaria. The Romanian and Bulgarian school systems are less efficient in transforming the advantage that the best performing students are better endowed into higher test scores. Compared to the other Eastern European countries, which achieve higher returns in the upper part of the distribution than Finland, they are still less efficient than the Finnish school system. Like in the science test score results, the interaction effect increases steadily, being here positive from about the 75 percent quantile upwards.

6 Understanding the unexplained part of test score gap

As our results show, the unexplained part is the largest in explaining the total test score gap. Therefore, we show that the individual characteristics we use to estimate the propensity score are not the main driving force of the differences in PISA test scores, e.g. evidence is provided that those from the better performing country are not just better endowed with more favorable characteristics. Even though this result contributes to understanding the differences in school performance, it does not provide information about what makes the difference. One possibility is to include further institutional and country specific variables, but, as this would add almost perfect predictors to the estimation of the propensity score, the matching procedure would no longer appropriate since the common support area would shrink.

We do not think that controlling for these covariates would be appropriate, not only from a methodical standpoint. Let's consider that enough comparison units exists. If we would then control for a certain school type, for example, those comparison units who are in the same school type would be more likely to form the counterfactual outcome. Nevertheless, it might be that, given his/her characteristics, this individual would not participate in this school type in the other country but would be much more likely to visit another school type. If we match only on the basis of individual characteristics, this conditional probability is reflected in the obtained counterfactual outcome.

If our identifying assumption about the mean independence holds, there are no other individual confounders than those used to construct the propensity score. In this case, the unexplained part is not due to unobserved individual characteristics and can be attributed to the different school system. This section, therefore, is intended to provide a better insight into the unexplained part of the gap.

According to the economic literature on educational achievement, other factors that influence the difference in test scores are institutional features of the school system, such as different accountability measures, school autonomy, school tracking, and school inputs such as class size, shortage of materials, school location, teacher education and instruction time. Thus, in order to understand the magnitude of the unexplained part of the test score gap we should take a closer look at different characteristics of the school systems considered in this study. This comparative and descriptive approach also aims to provide an explicit view why we did not include such variables in our matching procedure. We refer here to different aspects regarding the design of the school systems, such as, the existence or nonexistence of certain educational policies, the timing of implementation of certain measures designed to have long term effects and other issues, more or less adjacent to the educational act, which generally affects the quality of certain features of the educational system.

As mentioned before, in the case of Eastern European countries, the most important institutional features that have proven to affect school outcomes (Wössmann, Lüdemann, Schütz, and West (2009), Hanushek and Wössmann (2010)), were also those that have undergone numerous reforms in educational policies over the past 20 years. One of these institutional features is school autonomy. Wössmann, Lüdemann, Schütz, and West (2009) show that students perform better in schools that have autonomy in hiring teachers, in budget allocation as well as in choosing school textbooks and teaching methods. In Table (A.10), we look whether the schools have autonomy with respect to hiring teachers, operating expenditures, content of the compulsory minimum curriculum, the curricular content of optional subjects, the choice of teaching method, the choice of school textbooks, criteria for internal assessment of pupils and the decision whether pupils should repeat a class. As can be seen, the four best performing countries in our sample are also those who have at least limited autonomy in all eight categories. By comparison, we see that in Bulgaria and Romania, the school autonomy had not even been implemented in 2005, the time when the PISA data collection for 2006 took place which might explain a part of the large unexplained gap between Finland and these countries.

[Table (A.10) about here]

If we look at Estonia and Latvia, the unexplained gap can further be due to the fact that in Estonia, for example, the autonomy in choosing school textbooks has a long tradition. Even before 1990, Estonia refused to use teaching materials from other countries as the other Baltic states did, preferring to teach from textbooks written by Estonian authors. We, therefore, could presume that increasing the decision-making power regarding hiring teachers, operating expenditures with salaries and choice of textbooks will have a positive impact on the school performance of students from the worse performing countries and especially for those from Bulgaria and Romania. However, this follows only from the descriptive part. As Wössmann, Lüdemann, Schütz, and West (2009) show, a positive association between school autonomy and student achievement is not automatically guaranteed, especially then when a high degree of asymmetric information on school's interests generates incentives for opportunistic behavior. We expect to find such behavior in countries (like in Bulgaria and Romania), where the quality of institutions (including schools) is affected by corruption which, in turn, determines an unproductive use of schooling Rogers (2008). According to Wössmann, Lüdemann, Schütz, and West

(2009), the opportunistic behavior within schools can be controlled in those systems, where external exams are implemented. Thus, monitoring schools and student performance through examinations organized by external authorities, the positive impact of school autonomy on student achievement counteracts the negative effect of the opportunistic behavior.

Table (A.11) shows measures of the accountability of the school system for the countries in our sample and reveals that the three least successful countries are those in which external evaluation was implemented after 2000.

[Table (A.11) about here]

Another impact of the implementation of external examination is the indirect pressure exerted on teachers and schools, whose performance is also evaluated in terms of results at these tests. This creates a huge demand for investment in teacher training, especially in those countries where the course contents must be changed and adapted to the new requirements. Table (A.10) shows that the in-service training for teachers is compulsory in all countries except the Slovak Republic. Therefore, we cannot state any association between the in-service training for teachers and the PISA test scores. Moreover, the effectiveness of these programs depends entirely on their quality and on the way they cover the demand of training needed. However, most of these in-service trainings in countries from Eastern Europe are still old-fashioned, being organized within schools which are not fully reformed and which still feature many traits of the old system Radó (2001).

Moreover, there are some differences between the unexplained gaps and some different quantitative patterns can be obtained. When we compare Estonia and Latvia, Czech and Slovak Republic, as well as, Finland with Hungary or Czech Republic, the unexplained gap, both in science and reading scores, is higher than the whole gap. Similarly, this is the case if we compare Finland and Slovak Republic in reading scores. If the unexplained part is larger than the whole gap, it can either reflect the fact that the students from the worse performing country, who cannot be matched, obtain higher scores than their counterparts who are on the common support (as in the case of Finland - Czech Republic, for example). In this case, Δ_0 is negative and the differences in distributions of the individual characteristics literally do not explain any of the total gap in the test scores. Alternatively, such a result can reflect more advantageous individual and background characteristics of the students from the worse performing country compared to Finnish students or it can be an interaction of these two effects. This is the case for the comparison of Finland and Hungary, for example. Here Δ_c and Δ_0 are both significantly negative and, therefore, the total gap is smaller than the unexplained part. In all these cases, the students from the worse performing country who belong to the common support subpopulation would, on average, even outperform the students from the better performing country if they would attend their school system.

Looking at the magnitude of the unexplained gap along the percentiles, we notice the same pattern for science and reading scores: This part of the gap is higher at the lower end of the test score distributions and then decreases steadily towards the upper end of the test scores distributions. Comparing to the total gap, in the case of worse performing students, the unexplained part of the gap surpasses the total gap at the respective quantile. Among the best performing students, the unexplained gap decreases and it is even smaller in absolute terms than the corresponding total gap, thus showing that, at the upper percentiles of the test scores, the differences in distributions of the individual and background characteristics explain the total gap more than the unexplained gap.

7 Conclusion

The countries from Eastern Europe have gone through periods of transition, characterized by a continuous reforming of the educational system, over the last 20 years . The aim of these reforms was to adapt the school system to the educational standards of western Europe, as well as, to the new labor market requirements.

This paper analyzes the differences in PISA test scores between these countries and Finland, which is considered to have the most efficient and equitable educational system in Europe, as well as the differences between countries, which had similar educational systems 20 years ago. Understanding the test score differences is of huge importance, since it allows improving the school systems and, therefore, directly provides relevant information with respect to educational policies. Thus, we contribute to the literature by showing where the difference in PISA test scores between Eastern European countries and Finland, the best performing country in the PISA 2006 study, as well as among Eastern European countries come from.

Moreover, we contribute to the literature by introducing a semiparametric matching procedure into the educational literature. To decompose the gap in test scores, we use a method which is based on an approach proposed by Frölich (2007), who uses propensity score matching to compute the counterfactual outcome. To account for differences in the supports we further follow the procedure developed by Ñopo (2008). Such a method relaxes the functional form assumptions of the usual Blinder-Oaxaca decomposition, often used to decompose the gap in school performance, and illustrates the need to look at those individuals who have actually comparable background characteristics and those who do not separately. In this paper we extend this procedure to estimate a threefold decomposition.

Applying this method in decomposing the gaps in PISA test scores provides interesting insights. We provide evidence that only a small part of the gap can be attributed to the fact that the Finnish students are better endowed with more favorable individual characteristics. The main part of the gap still remains after controlling for the individual background. If our identifying assumption of mean independence holds, this remaining part of the gap is an estimate of the potential gain the Eastern European students could obtain if they would participate in the Finnish school system, or alternatively of their school system could approach the Finnish one. The students from South Eastern Europe are those who have the largest potential outcome increase if they would attend a school system similar to the Finnish one.

Estonia, the country which already adapted its school system to the Finnish school system more than any other country in our sample, would gain the least if they would adapt further to the Finnish school system. Estonia not only performs best out of our Eastern European countries but also among the best of all participating countries.

If we take a closer look at the distribution of the test score gaps in the science and reading test score gap between each pair of the countries, we find that only for the comparison of the science scores in Finland and Romania does the gap increase from the 5% quantile to the 95% quantile. If we look at the science test score gap between Finland and the Baltic Republics, we notice that the gap is more or less constant throughout the distribution. The striking finding is that in 11 out of 14 comparisons, the gap decreases from the 5% quantile to the 95% quantile. This result clearly shows that the Finnish students are not, on average, the best performing country because they manage to get the high performers to score better than the high performers of the other countries but because their school system manages to have their weakest students perform better than the weak students of the other countries. This clearly points to very important policy advice to the Eastern European educational system; namely, that they should not focus on the elite but rather on the weakest. Especially protruding is this observation, if we look at the distribution of the reading test scores where in all seven cases, the gap is decreasing from the 5% to the 95% quantile.

Moreover, our paper exploits the fact that some Eastern European countries have had a very similar school system 20 years ago. Estonia and Latvia both belonged to the Soviet Union and share a similar history and the Czech and Slovak Republic composed Czechoslovakia until 1992. This provides us with a situation similar to a natural experiment. In both cases, the countries started from a very similar point but then developed differently over the past years. For these countries, we do not find any differences in individual characteristics such that the whole gap is due to the differences in the educational systems. For both countries we even find that the students from the country that score worse, on average, would even outperform the students from the better performing country if all would attend the same school system. This result holds independent of the subject.

If we look at the distribution of the gap, we find that it stays constant if we compare the science test score between Estonia and Latvia, whereas it is decreasing if we focus on the reading score. This indicates that, concerning reading, the poor performing Estonian students perform better than the poor performing Latvia students, whereas the better performing students score closer together in the two countries. Compared to that, we find that the gap is increasing over the distribution for both subjects if we compare the Czech Republic to Slovakia. Hence, this is the only pair of countries where the better performing country scores higher due to the fact that the best performing students perform much better than the best performing students of the worse performing country and the weak performing students from the better performing country perform only slightly better than the weak performing students from the other country.

References

Ñopo, H. (2008): "Matching as a Tool to Decompose Wage Gaps," *Review of Economics and Statistics*, 90(2), 290–299.

Ammermüller, A. (2007): "PISA : what makes the difference?," *Empirical Economics*, 33(2), 263–287.

Ammermüller, A., H. Heijke, and L. Wössmann (2005): "Schooling quality in Eastern Europe: Educational production during transition," *Economics of Education Review*, 24(5), 579–599.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002): "Accounting for the Black-White Wealth Gap: A Nonparametric Approach," *Journal of the American Statistical Association*, 97, 663–673.

Beirne, J., and N. F. Campos (2007): "Educational inputs and outcomes before the transition from communism," *The Economics of Transition*, 15(1), 57–76.

Bjørnskov, C., and N. Potrafke (2011): "Politics and Privatization in Central and Eastern Europe: A Panel Data Analysis," *The Economics of Transition*, 19(2), 201–230.

Blinder, A. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 8(4), 436–455.

Card, D., and J. Rothstein (2007): "Racial segregation and the black-white test score gap," *Journal of Public Economics*, 91(11-12), 2158–2184.

Cerych, L. (1997): "Educational Reforms in Central and Eastern Europe: processes and outcomes," *European Journal of Education*, 32(1), 75–96.

Duncan, K. C., and J. Sandy (2007): "Explaining the Performance Gap Between Public and Private School Students," *Eastern Economic Journal*, 33(2), 177–191.

Eurydice (2007): "School Autonomy in Europe. Policies and Measures," *European Comission*, Brussels.

——— (2008): "Levels of Autonomy and Responsibilities of Teachers in Europe," *European Comission*, Brussels.

——— (2009): "National Testing of Pupils in Europe: Objectives, Organisation and Use of Results," *European Comission*, Brussels.

Flabbi, L., S. Paternostro, and E. R. Tiongson (2008): "Returns to education in the economic transition: A systematic assessment using comparable data," *Economics of Education Review*, 27(6), 724–740.

Fortin, N., T. Lemieux, and S. Firpo (2010): *Decomposition Methods in Economics,* Handbook of Labour Economics, Vol. 4A, 1-102, O. Ashenfelter and D. Card (eds.), Amsterdam: North-Holland.

Frölich, M. (2004): "Finite-sample properties of propensity-score matching and weighting estimators," *Review of Economics and Statistics*, 86(1), 77–90.

Frölich, M. (2007): "Propensity score matching without conditional independence assumption–with an application to the gender wage gap in the United Kingdom," *Econometrics Journal*, 10(2), 359–407.

Fryer, R. G., and S. D. Levitt (2010): "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2(2), 210–40.

Fuchs, T., and L. Wössmann (2007): "What accounts for international differences in student performance? A re-examination using PISA data," *Empirical Economics*, 32(2), 433–464.

Guiso, L., P. Sapienza, and L. Zingales (2006): "Does culture affect economic outcomes?," *The Journal of Economic Perspectives*, 20(2), 23–48.

Hanushek, E. A., and D. D. Kimko (2000): "Schooling, Labor-Force Quality, and the Growth of Nations," *The American Economic Review*, 90(5), 1184–1208.

Hanushek, E. A., and L. Wössmann (2010): *The Economics of International Differences in Educational Achievement,* Handbook of the Economics of Education, Vol. 3, E.A. Hanushek, S. Machin and L. Woessmann (eds.), Amsterdam: North Holland.

Heckman, J. J., R. LaLonde, and J. Smith (1999): *The Economics and Econometrics of Active Labor Market Programs,* Handbook of Labour Economics, Vol. 3, 1865-2097, O. Ashenfelter and D. Card (eds.), Amsterdam: North-Holland.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4–29.

Kitsing, M. (2008): "PISA 2006–Estonian Results," *Ministry of Education and Research of Estonia*, Tartu.

Krieg, J. M., and P. Storer (2006): "How Much Do Students Matter? Applying the Oaxaca Decomposition to Explain Determinants of Adequate Yearly Progress," *Contemporary Economic Policy*, 24(4), 563–581.

Martins, L., and P. Veiga (2010): "Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities?," *Economics of Education Review*, 29(6), 1016 – 1033.

McEwan, P., and J. Marshall (2004): "Why does academic achievement vary across countries? Evidence from Cuba and Mexico," *Education Economics*, 12(3), 205–217.

Mora, R. (2008): "A nonparametric decomposition of the Mexican American average wage gap," *Journal of Applied Econometrics*, 23(4), 463–485.

NIEDERLE, M., AND L. VESTERLUND (2010): "Explaining the Gender Gap in Math Test Scores: The Role of Competition," *The Journal of Economic Perspectives*, 24(2), 129–144.

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14(3), 693–709.

OECD (2007): "PISA 2006 science competencies for tomorrow's world," *OECD*, Paris.

PATACCHINI, E., AND Y. ZENOU (2009): "On the sources of the black-white test score gap in Europe," *Economics Letters*, 102(1), 49–52.

RADÓ, P. (2001): *Transition in Education.* Open Society Institute, Institute for Educational Policy.

ROGERS, M. L. (2008): "Directly unproductive schooling: How country characteristics affect the impact of schooling on growth," *European Economic Review*, 52(2), 356–385.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

SCHNEEWEIS, N. (2010): "Educational institutions and the integration of migrants," *Journal of Population Economics*, forthcoming.

SCHÜTZ, G., H. W. URSPRUNG, AND L. WÖSSMANN (2008): "Education Policy and Equality of Opportunity," *Kyklos*, 61(2), 279–308.

WÖSSMANN, L. (2003): "Schooling Resources, Educational Institutions and Student Performance: the International Evidence," *Oxford Bulletin of Economics and Statistics*, 65(2), 117–170.

——— (2008): "How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States," *Zeitschrift für Betriebswirtschaft*, 78 (1), 45–70.

WÖSSMANN, L., E. LÜDEMANN, G. SCHÜTZ, AND M. R. WEST (2009): *School Accountability, Autonomy and Choice around the World*, no. 13. Cheltenham: Edward Elgar.

Appendix

Table A.1: Variables' description

Variable	Min	Max	Description	Notation
Test scores				
Reading score	5.67	781.96	mean of five plausible values for reading	scorereading
Math score	40.61	819.05	mean of five plausible values for math	scoremath
Science score	93.56	820.52	mean of five plausible values for science	scorescience
Student Background				
Student's sex	0	1	1 for male	Dmale
Student's age	182.04	195.96	Student's age in month	age_m
7^{th} grade	0	1	1 for 7^{th} grade of students	grade7
8^{th} grade	0	1	1 for 8^{th} grade of students	grade8
9^{th} grade	0	1	1 for 9^{th} grade of students	grade9
10^{th} or 11^{th}	0	1	1 for 10^{th} or 11^{th} grade of students	grade1011
Books Cat.1	0	1	1 if less than 10 books at home	book1
Books Cat.2	0	1	1 if 11-25 books	book2
Books Cat.3	0	1	1 if 26-100 books	book3
Books Cat.4	0	1	1 if 101-200 books	book4
Books Cat.5	0	1	1 if 201-500 books	book5
Books Cat.6	0	1	1 if more than 500 books	book6
Mother's no sec. Ed.	0	1	1 if completed at most ISCED 1	m_nosec
Mother's lower sec. Ed.	0	1	1 if completed ISCED 2	m_lower_sec
Mother's upper sec. Ed.	0	1	1 if completed ISCED 3A,3B,3C or 4	m_upper_sec
Mother's tertiary Ed.	0	1	1 if completed ISCED 5B or higher	m_ter
Father's no sec. Ed.	0	1	1 if completed at most ISCED 1	f_nosec
Father's lower sec. Ed.	0	1	1 if completed ISCED 2	f_lower_sec
Father's upper sec. Ed.	0	1	1 if completed ISCED 3A,3B,3C or 4	f_upper_sec
Father's tertiary Ed.	0	1	1 if completed ISCED 5B or higher	f_ter

Source: PISA 2006 data, own calculations.

Table A.2: Weighted means and standard deviations

	Finland		Estonia		Czech R.		Hungary	
Variable	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.
scorereading	548.50	75.86	506.68	80.18	484.87	105.76	483.31	90.05
scoremath	550.26	75.15	516.48	75.80	511.14	98.41	491.81	86.84
scorescience	565.56	80.66	533.55	79.67	514.21	94.84	504.71	84.70
Dmale	0.49	0.50	0.51	0.50	0.56	0.50	0.52	0.50
age_m	187.77	3.40	189.66	3.44	190.53	3.43	188.49	3.43
grade7	0.00	0.03	0.03	0.17	0.01	0.07	0.02	0.14
grade8	0.11	0.32	0.25	0.44	0.03	0.18	0.05	0.23
grade9	0.89	0.32	0.70	0.46	0.44	0.50	0.66	0.47
grade1011	0.00	0.01	0.02	0.13	0.52	0.50	0.27	0.44
book1	0.05	0.22	0.05	0.21	0.05	0.23	0.06	0.24
book2	0.11	0.31	0.11	0.31	0.10	0.30	0.11	0.31
book3	0.34	0.48	0.29	0.45	0.35	0.48	0.28	0.45
book4	0.24	0.43	0.24	0.43	0.23	0.42	0.21	0.41
book5	0.19	0.39	0.20	0.40	0.17	0.38	0.18	0.39
book6	0.07	0.26	0.12	0.32	0.09	0.29	0.16	0.37
m_nosec	0.04	0.19	0.00	0.03	0.01	0.08	0.01	0.09
m_lower_sec	0.06	0.24	0.03	0.17	0.03	0.17	0.14	0.35
m_upper_sec	0.33	0.47	0.62	0.49	0.77	0.42	0.57	0.50
m_ter	0.58	0.49	0.35	0.48	0.20	0.40	0.28	0.45
f_nosec	0.06	0.24	0.00	0.06	0.00	0.05	0.01	0.07
f_lower_sec	0.09	0.28	0.04	0.21	0.02	0.14	0.09	0.29
f_upper_sec	0.42	0.49	0.77	0.42	0.76	0.43	0.68	0.47
f_ter	0.43	0.50	0.18	0.39	0.22	0.41	0.22	0.42
Number of obs.	4609		4703		5813		4395	

Source: PISA 2006 data, own calculations.

Table A.2 - continued: Weighted means and standard deviations

	Latvia		Slovakia		Bulgaria		Romania	
Variable	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.
scorereading	481.09	84.09	467.35	99.67	405.52	110.95	396.02	86.35
scoremath	487.98	77.43	493.07	89.85	415.70	95.60	414.92	79.31
scorescience	491.63	80.02	489.41	89.49	436.63	102.63	418.37	77.68
Dmale	0.48	0.50	0.51	0.50	0.52	0.50	0.50	0.50
age_m	189.61	3.42	188.64	3.38	188.87	3.42	188.91	3.31
grade7	0.02	0.15	0.01	0.08	0.00	0.06	0.01	0.08
grade8	0.16	0.37	0.02	0.14	0.07	0.25	0.14	0.34
grade9	0.79	0.41	0.38	0.49	0.74	0.44	0.83	0.38
grade1011	0.03	0.17	0.59	0.49	0.19	0.39	0.03	0.17
book1	0.06	0.24	0.08	0.27	0.20	0.40	0.15	0.36
book2	0.12	0.33	0.12	0.32	0.16	0.37	0.21	0.41
book3	0.31	0.46	0.40	0.49	0.28	0.45	0.33	0.47
book4	0.24	0.43	0.23	0.42	0.16	0.37	0.15	0.36
book5	0.17	0.38	0.12	0.33	0.12	0.33	0.10	0.30
book6	0.10	0.30	0.05	0.22	0.08	0.27	0.06	0.24
m_nosec	0.00	0.05	0.00	0.04	0.02	0.12	0.03	0.16
m_lower_sec	0.02	0.14	0.04	0.21	0.11	0.31	0.09	0.29
m_upper_sec	0.69	0.46	0.81	0.40	0.59	0.49	0.54	0.50
m_ter	0.29	0.45	0.15	0.36	0.29	0.45	0.34	0.48
f_nosec	0.00	0.06	0.00	0.05	0.01	0.12	0.03	0.16
f_lower_sec	0.03	0.18	0.03	0.16	0.08	0.27	0.07	0.25
f_upper_sec	0.77	0.42	0.82	0.39	0.74	0.44	0.60	0.49
f_ter	0.20	0.40	0.16	0.36	0.16	0.37	0.30	0.46
Number of obs.	4542		4675		4255		5102	

Source: PISA 2006 data, own calculations.

Table A.3: The share of observations out of the common support

Country 1	**Finland**	**Finland**	**Finland**	**Finland**	**Finland**	**Finland**	**Finland**	**Estonia**	**Czech**
Country 2	**Estonia**	**Czech R.**	**Hungary**	**Latvia**	**Slovakia**	**Bulgaria**	**Romania**	**Latvia**	**Slovakia**
Country 1	0.72%	0.15%	0.03%	0.52%	1.06%	1.54%	1.13%	0.04%	0.77%
Country 2	1.07%	2.03%	0.88%	1.50%	0.39%	0.39%	0.52%	0.00%	0.15%

Source: PISA 2006 data, own calculations.

Table A.4: Semiparametric Decomposition Result for Science

Countries	Δ_1	Δ_c	Δ_r	Δ_{cr}	Δ_0	Δ
FIN-EST	−0.11	3.81	30.78	−2.31	−0.15	32.02
	(0.15)	(1.72)	(2.08)	(2.12)	(0.19)	(1.73)
FIN-CZE	−0.01	6.05	55.85	−9.73	−0.80	51.35
	(0.09)	(3.19)	(2.92)	(3.82)	(0.34)	(1.85)
FIN-HUN	−0.03	1.06	59.24	0.84	−0.27	60.85
	(0.06)	(2.05)	(2.17)	(2.38)	(0.13)	(1.82)
FIN-LTV	0.00	−0.41	69.36	5.38	−0.39	73.94
	(0.14)	(2.30)	(2.18)	(2.63)	(0.41)	(1.84)
FIN-SLK	−0.13	7.15	63.55	5.69	−0.11	76.15
	(0.24)	(4.70)	(2.31)	(5.02)	(0.15)	(1.89)
FIN-BUL	0.04	29.84	115.34	−16.64	0.36	128.94
	(0.19)	(1.88)	(2.25)	(2.10)	(0.17)	(2.02)
FIN-ROM	0.29	12.02	134.06	0.58	0.25	147.20
	(0.15)	(1.55)	(2.04)	(1.73)	(0.11)	(1.92)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.5: Semiparametric Decomposition Result for Reading

Countries	Δ_1	Δ_c	Δ_r	Δ_{cr}	Δ_0	Δ
FIN-EST	−0.15	−1.27	41.37	4.92	−0.04	44.82
	(0.16)	(2.40)	(2.06)	(2.71)	(0.19)	(1.70)
FIN-CZE	−0.02	7.16	63.48	−6.09	−0.91	63.64
	(0.09)	(3.69)	(2.96)	(4.08)	(0.43)	(2.03)
FIN-HUN	−0.02	0.02	61.67	3.77	−0.26	65.19
	(0.05)	(2.46)	(2.17)	(2.73)	(0.16)	(1.84)
FIN-LTV	−0.07	−2.07	62.96	7.01	−0.42	67.41
	(0.11)	(2.51)	(2.39)	(2.93)	(0.48)	(1.82)
FIN-SLK	−0.12	4.41	67.82	9.19	−0.15	81.15
	(0.17)	(4.96)	(2.33)	(5.16)	(0.16)	(1.92)
FIN-BUL	0.02	30.37	127.39	−15.15	0.35	142.98
	(0.16)	(2.03)	(2.46)	(2.29)	(0.17)	(2.14)
FIN-ROM	0.20	9.95	137.99	3.98	0.38	152.49
	(0.13)	(1.73)	(2.05)	(1.84)	(0.15)	(1.95)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.6: Semiparametric decompositions for science between Eastern European countries

Countries	Δ_1	Δ_c	Δ_r	Δ_{cr}	Δ_0	Δ
EST-LTV	0.02	7.10	40.68	−5.88	0.00	41.92
	(0.04)	(0.96)	(1.78)	(1.04)	(0.10)	(1.90)
CZE-SLK	0.59	18.84	27.12	−22.00	0.24	24.80
	(0.25)	(1.16)	(2.12)	(1.74)	(0.28)	(1.96)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.7: Semiparametric decompositions for reading between Eastern European countries

Countries	Δ_1	Δ_c	Δ_r	Δ_{cr}	Δ_0	Δ
EST-LTV	−0.03	7.02	22.20	−6.60	0.00	22.59
	(0.06)	(0.95)	(1.84)	(1.07)	(0.09)	(1.95)
CZE-SLK	0.40	20.71	24.28	−28.09	0.21	17.51
	(0.21)	(1.30)	(2.38)	(1.98)	(0.26)	(2.30)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.8: Semiparametric decomposition at different quantiles for science score

		5 % Quantile	**25 % Quantile**	**50 % Quantile**	**75 % Quantile**	**95 % Quantile**
FIN-EST	Δ_c	117.32	51.77	3.10	−44.21	−113.41
		(7.91)	(1.76)	(2.56)	(2.81)	(3.90)
	Δ_r	157.14	80.93	29.42	−22.61	−88.06
		(4.65)	(2.92)	(2.52)	(2.65)	(3.80)
	Δ_{cr}	−245.46	−100.25	1.89	98.89	232.05
		(8.82)	(2.86)	(3.19)	(3.53)	(4.79)
	Δ_s	**29.00**	**32.45**	**34.41**	**32.08**	**30.59**
		(4.60)	(2.53)	(2.34)	(2.29)	(3.52)
FIN-CZE	Δ_c	30.50	64.58	18.40	−35.28	−83.58
		(21.13)	(6.80)	(2.17)	(3.09)	(6.44)
	Δ_r	187.92	112.49	52.09	−5.02	−50.87
		(5.34)	(3.67)	(3.08)	(4.13)	(14.64)
	Δ_{cr}	−148.96	−109.09	−17.25	79.93	161.12
		(21.42)	(7.42)	(3.22)	(4.67)	(15.72)
	Δ_s	**69.47**	**67.98**	**53.24**	**39.63**	**26.67**
		(4.46)	(2.75)	(2.48)	(2.63)	(3.05)
FIN-HUN	Δ_c	101.99	50.29	2.88	−42.28	−104.27
		(13.43)	(3.10)	(3.55)	(3.24)	(3.98)
	Δ_r	183.73	107.88	56.22	7.65	−56.37
		(3.72)	(3.15)	(3.07)	(2.97)	(5.69)
	Δ_{cr}	−221.94	−91.69	3.66	92.54	211.92
		(14.05)	(3.86)	(4.15)	(3.99)	(6.59)
	Δ_s	**63.78**	**66.49**	**62.76**	**57.91**	**51.29**
		(3.52)	(2.67)	(2.43)	(2.10)	(3.08)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.8 - continued: Semiparametric decomposition at different quantiles for science score

		5 % Quantile	25 % Quantile	50 % Quantile	75 % Quantile	95 % Quantile
FIN-LTV	Δ_c	77.47	54.45	3.50	−43.58	−112.03
		(18.42)	(2.52)	(2.26)	(2.70)	(3.70)
	Δ_r	189.51	117.58	68.90	17.41	−45.18
		(4.60)	(3.42)	(3.02)	(3.18)	(3.24)
	Δ_{cr}	−196.21	−96.32	3.14	100.95	227.99
		(18.93)	(3.74)	(3.32)	(3.66)	(4.67)
	Δ_s	**70.77**	**75.72**	**75.53**	**74.78**	**70.77**
		(4.30)	(2.75)	(2.40)	(2.50)	(3.03)
FIN-SLK	Δ_c	34.87	57.02	25.80	−28.16	−99.18
		(37.21)	(2.01)	(3.75)	(4.29)	(5.65)
	Δ_r	198.20	115.14	56.38	4.11	−51.74
		(4.45)	(3.21)	(3.09)	(4.91)	(3.06)
	Δ_{cr}	−147.48	−88.61	−3.57	91.75	212.18
		(37.41)	(3.35)	(4.53)	(6.01)	(6.05)
	Δ_s	**85.60**	**83.55**	**78.61**	**67.70**	**61.26**
		(4.30)	(2.49)	(2.55)	(2.93)	(2.88)
FIN-BUL	Δ_c	156.14	92.13	37.44	−20.41	−112.79
		(4.16)	(2.98)	(4.27)	(4.93)	(5.81)
	Δ_r	240.06	184.95	119.97	57.19	−34.31
		(7.59)	(3.49)	(3.43)	(3.44)	(3.14)
	Δ_{cr}	−243.55	−126.67	−23.41	76.42	231.49
		(7.89)	(4.07)	(4.78)	(5.39)	(6.19)
	Δ_s	**152.65**	**150.41**	**134.00**	**113.20**	**84.39**
		(4.98)	(2.49)	(3.06)	(3.08)	(3.30)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.8 - continued: Semiparametric decomposition at different quantiles for science score

		5 % Quantile	25 % Quantile	50 % Quantile	75 % Quantile	95 % Quantile
FIN-ROM	Δ_c	87.54	60.61	20.72	−23.20	−99.21
		(5.61)	(2.94)	(2.81)	(3.19)	(5.26)
	Δ_r	217.21	179.03	134.12	102.59	27.46
		(5.65)	(4.02)	(3.64)	(2.77)	(3.55)
	Δ_{cr}	−175.13	−91.18	−3.59	72.51	215.44
		(7.61)	(4.41)	(3.89)	(3.51)	(5.73)
	Δ_s	**129.61**	**148.45**	**151.25**	**151.90**	**143.69**
		(4.08)	(2.80)	(2.93)	(2.62)	(3.63)
EST-LTV	Δ_c	114.11	47.64	3.41	−30.49	−94.36
		(3.40)	(2.55)	(2.97)	(2.29)	(5.29)
	Δ_r	143.85	85.17	38.74	6.55	−63.15
		(4.07)	(3.12)	(3.59)	(2.85)	(3.36)
	Δ_{cr}	−217.12	−89.74	−0.66	66.65	197.79
		(4.96)	(3.46)	(4.23)	(3.23)	(6.12)
	Δ_s	**40.84**	**43.08**	**41.50**	**42.71**	**40.28**
		(4.56)	(3.05)	(2.40)	(2.38)	(3.44)
CZE-SLK	Δ_c	127.24	55.25	9.63	−22.18	−69.66
		(3.69)	(2.29)	(2.64)	(4.67)	(7.32)
	Δ_r	129.61	81.58	29.93	−23.20	−71.73
		(12.81)	(3.95)	(3.09)	(3.34)	(3.92)
	Δ_{cr}	−240.99	−120.51	−14.75	73.27	175.71
		(12.83)	(4.47)	(3.80)	(5.26)	(8.16)
	Δ_s	**15.85**	**16.32**	**24.80**	**27.88**	**34.32**
		(4.33)	(2.87)	(2.51)	(3.17)	(3.16)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.9: Semiparametric decomposition at different quantiles for reading score

		5 % Quantile	**25 % Quantile**	**50 % Quantile**	**75 % Quantile**	**95 % Quantile**
FIN-EST	Δ_c	92.44	48.98	−2.42	−46.97	−108.72
		(18.89)	(1.83)	(1.83)	(2.66)	(3.63)
	Δ_r	168.57	86.52	37.42	−11.02	−67.92
		(5.90)	(3.23)	(2.60)	(2.58)	(2.74)
	Δ_{cr}	−210.48	−88.50	9.45	99.79	218.51
		(19.46)	(3.08)	(2.59)	(3.33)	(4.36)
	Δ_s	**50.53**	**47.00**	**44.45**	**41.80**	**41.87**
		(5.37)	(2.71)	(2.04)	(2.17)	(2.83)
FIN-CZE	Δ_c	52.82	69.21	15.54	−45.52	−84.31
		(28.27)	(7.03)	(2.54)	(3.17)	(6.65)
	Δ_r	217.89	127.96	57.37	−7.00	−69.76
		(7.05)	(4.03)	(3.25)	(3.69)	(10.64)
	Δ_{cr}	−155.77	−108.38	−11.02	92.24	178.31
		(28.56)	(7.55)	(3.28)	(4.41)	(12.32)
	Δ_s	**114.94**	**88.79**	**61.89**	**39.72**	**24.24**
		(5.63)	(3.18)	(2.54)	(2.51)	(3.12)
FIN-HUN	Δ_c	118.43	50.38	−5.32	−43.24	−101.90
		(19.96)	(3.98)	(4.01)	(2.64)	(3.82)
	Δ_r	200.82	112.62	50.54	5.24	−49.11
		(5.34)	(3.60)	(3.08)	(2.76)	(3.82)
	Δ_{cr}	−228.85	−88.09	15.75	92.03	202.81
		(20.25)	(4.68)	(4.56)	(3.45)	(5.16)
	Δ_s	**90.40**	**74.92**	**60.97**	**54.02**	**51.80**
		(4.80)	(3.07)	(2.32)	(2.10)	(2.87)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.9 - continued: Semiparametric decomposition at different quantiles for reading score

		5 % Quantile	25 % Quantile	50 % Quantile	75 % Quantile	95 % Quantile
FIN-LTV	Δ_c	82.42	49.90	-0.49	-45.28	-105.79
		(17.44)	(3.19)	(2.31)	(2.56)	(3.94)
	Δ_r	187.21	108.66	57.93	14.53	-42.97
		(5.07)	(3.45)	(3.03)	(2.94)	(5.21)
	Δ_{cr}	-187.75	-82.81	8.66	91.75	206.61
		(18.13)	(4.18)	(3.03)	(3.25)	(6.44)
	Δ_s	**81.88**	**75.75**	**66.09**	**61.00**	**57.85**
		(3.74)	(2.83)	(2.43)	(2.54)	(2.83)
FIN-SLK	Δ_c	50.16	62.87	17.59	-39.50	-103.60
		(48.28)	(2.11)	(4.70)	(4.19)	(5.93)
	Δ_r	234.65	124.57	52.91	1.00	-53.52
		(4.26)	(3.20)	(3.11)	(4.44)	(3.67)
	Δ_{cr}	-157.48	-88.79	7.03	101.06	206.04
		(48.48)	(3.21)	(5.22)	(5.86)	(6.21)
	Δ_s	**127.33**	**98.66**	**77.54**	**62.56**	**48.92**
		(3.75)	(2.74)	(2.38)	(2.23)	(3.71)
FIN-BUL	Δ_c	177.02	93.69	30.23	-26.25	-115.16
		(4.02)	(3.21)	(4.57)	(5.23)	(5.86)
	Δ_r	270.69	199.21	121.36	66.72	-25.33
		(7.88)	(4.04)	(3.70)	(3.13)	(3.30)
	Δ_{cr}	-255.13	-121.01	-9.30	75.05	225.16
		(7.87)	(4.45)	(5.19)	(5.62)	(6.26)
	Δ_s	**192.58**	**171.89**	**142.29**	**115.51**	**84.67**
		(4.23)	(2.94)	(2.97)	(2.62)	(3.71)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.9 - continued: Semiparametric decomposition at different quantiles for reading score

		5 % **Quantile**	**25 %** **Quantile**	**50 %** **Quantile**	**75 %** **Quantile**	**95 %** **Quantile**
FIN-ROM	Δ_c	115.40	63.15	11.29	−39.36	−104.71
		(7.02)	(3.15)	(2.72)	(3.58)	(5.38)
	Δ_r	242.09	182.64	135.87	96.48	30.58
		(5.91)	(4.10)	(3.11)	(3.37)	(3.58)
	Δ_{cr}	−192.27	−82.78	5.37	85.50	209.61
		(8.76)	(4.47)	(3.57)	(3.91)	(5.69)
	Δ_s	**165.22**	**163.02**	**152.54**	**142.61**	**135.49**
		(3.89)	(2.97)	(2.26)	(3.15)	(3.77)
EST-LTV	Δ_c	132.95	49.00	−6.11	−38.28	−84.65
		(3.33)	(2.51)	(2.44)	(2.51)	(5.53)
	Δ_r	152.20	67.67	11.81	−24.11	−82.58
		(3.92)	(2.96)	(3.19)	(3.16)	(3.24)
	Δ_{cr}	−253.95	−87.96	15.88	81.25	183.05
		(5.67)	(3.24)	(3.62)	(3.69)	(6.25)
	Δ_s	**31.20**	**28.71**	**21.58**	**18.85**	**15.81**
		(5.65)	(3.12)	(2.57)	(2.52)	(2.72)
CZE-SLK	Δ_c	160.54	59.06	6.18	−23.96	−79.36
		(3.78)	(2.72)	(2.93)	(5.98)	(8.07)
	Δ_r	164.16	79.47	19.17	−35.87	−88.02
		(12.04)	(3.98)	(3.22)	(3.06)	(4.54)
	Δ_{cr}	−310.91	−128.36	−9.64	82.48	191.45
		(12.61)	(4.65)	(4.04)	(6.49)	(8.79)
	Δ_s	**13.79**	**10.17**	**15.72**	**22.65**	**24.07**
		(5.92)	(3.64)	(3.08)	(2.61)	(3.63)

Source: PISA 2006 data, own calculations. The country which has worse performance is always the reference country. Standard errors are in brackets and simulated with 1000 bootstrap replications.

Table A.10: Autonomy and in-service teacher training

	Finland	**Estonia**	**Czech R.**	**Hungary**
Dates of major reforms that have increased the school autonomy [a]	progressively implemented	1992	1990	1993
Dates of major reforms that have increased the curricular autonomy [b]	progressively implemented	1990	2004	1993
School autonomy regarding				
hiring teachers[a]	delegation [c] possible	full	full	limited
operating expenditures (salaries) [a]	delegation possible	full	limited	limited
the content of the compulsory minimum curriculum [b]	limited	full	full	limited
the curricular content of optional subjects [b]	limited	full	full	limited
the choice of teaching methods [b]	full	full	full	full
the choice of school textbooks [b]	delegation possible	limited [d]	full	full
the criteria for the internal assessment of pupils [b]	full	full	full	full
decisions as to whether pupils should repeat a year [b]	full	full	full	full
In-service teacher training [a]				
Status of in-service teacher training	compulsory	compulsory	compulsory	compulsory
Amount of time per year (in hours)	18	32	no time indication	17
Incentives	no incentives	promotion	salary increase	salary increase

[a] *Source:* Eurydice (2007)
[b] *Source:* Eurydice (2008)
[c] *Delegation* means that bodies or local authorities may delegate their decision-making power to schools.
[d] Teachers choose the textbooks from a predetermined list.

Table A.10 - continued: Autonomy and in-service teacher training

	Latvia	Slovak R.	Bulgaria	Romania
Dates of major reforms that have increased the school autonomy [a]	1991	1990	2008	2006
Dates of major reforms that have increased the curricular autonomy [b]	1993	1990	-	1998
School autonomy regarding				
hiring teachers [a]	full	full	full	no autonomy
operating expenditures (salaries) [a]	full	limited	no autonomy	no autonomy
the content of the compulsory minimum curriculum [b]	no autonomy	no autonomy	no autonomy	no autonomy
the curricular content of optional subjects [b]	limited	full	limited	full
the choice of teaching methods [b]	full	full	full	full
the choice of school textbooks [b]	limited [c]	limited	full	limited
the criteria for the internal assessment of pupils [b]	limited	full	full	full
decisions as to whether pupils should repeat a year [b]	no autonomy	full	full	full
In-service teacher training [a]				
Status of in-service teacher training	compulsory	optional	compulsory	compulsory
Amount of time per year (in hours)	36	no time indication	no time indication	19
Incentives	no incentives	salary increase	salary increase promotions	promotion

[a] *Source:* Eurydice (2007)
[b] *Source:* Eurydice (2008)
[c] Teachers choose the textbooks from a predetermined list.

Table A.11: Accountability

	External exams [a]			Tracking [c]
	for promotion, streaming	for monitoring schools	School years in which they take place	Number of school programs [b]
Finland	-	X (1998 [d])	6,9	1
Estonia	X (1997)	X (1997)	9	1
Czech R.	-	-	-	6
Hungary	-	X (1986)	4,6,8	4
Latvia	X (1994)	X (1994)	3,6,9	2
Slovak R.	-	X (2003)	9	6
Bulgaria	-	X (2006)	4,5,6	4
Romania	X (2007)	X (1995)	7,8	3

[a] *Source:* Eurydice (2009)
[b] *Source:* PISA 2006 Database
[c] We use the school programs, in which the students are enrolled, as proxies for school types.
[d] Year of first full implementation of external exams (national tests).