An International Comparison of the Effect of Private Education Spending on Student Academic Performance: Evidence from the Programme for International Student Assessment (PISA), 2006

Deockhyun Ryu	Changhui Kang†
Department of Economics	Department of Economics
Chung-Ang University	Chung-Ang University
Seoul, South Korea	Seoul, South Korea
Email: dhryu@cau.ac.kr	Email: ckang@cau.ac.kr
Fax: +82-2-812-9718	Fax: +82-2-812-9718

Abstract

In order to shed light on the effectiveness of private education, this paper undertakes an international study on impacts of private education spending on student academic performance, using the OECD's PISA 2006 data. To circumvent endogeneity of the education spending, it relies on propensity-score matching and nonparametric bounding methods. The primary result from both methods is that the effect of private education spending remains only modest for all the countries examined in the PISA 2006 data. In terms of an elasticity, the largest magnitudes of the effect imply that a 10 percent increase in spending leads to no more than a 0.6 percent average improvement in test scores. Private education spending seems to have either little average impact on student academic performance, or a weak positive effect at best.

JEL *Classification* : I20, C30

Keywords : Private Education Spending, Test Scores, PISA, Matching, Nonparametric Bounds

† Corresponding Address: Department of Economics, Chung-Ang University, 221 Heukseok-Dong Dongjak-Gu, Seoul 156-756, South Korea; E-mail: ckang@cau.ac.kr, Phone: +82-2-820-5862, Fax: +82-2-812-9718.

An International Comparison of the Effect of Private Education Spending on Student Academic Performance: Evidence from the Programme for International Student Assessment (PISA), 2006

Abstract

In order to shed light on the effectiveness of private education, this paper undertakes an international study on impacts of private education spending on student academic performance, using the OECD's PISA 2006 data. To circumvent endogeneity of the education spending, it relies on propensity-score matching and nonparametric bounding methods. The primary result from both methods is that the effect of private education spending remains only modest for all the countries examined in the PISA 2006 data. In terms of an elasticity, the largest magnitudes of the effect imply that a 10 percent increase in spending leads to no more than a 0.6 percent average improvement in test scores. Private education spending seems to have either little average impact on student academic performance, or a weak positive effect at best.

JEL *Classification* : I20, C30

Keywords : Private Education Spending, Test Scores, PISA, Matching, Nonparametric Bounds

1 Introduction

Along with the debate on the effectiveness of public education expenditures, causal impacts of private education are a longstanding issue in economics of education research. Studies focusing on private school attendance (e.g., Catholic schools) in western countries are a traditional path taken by researchers. They do not, however, agree on the true impacts of private school attendance. Evans and Schwab (1995) and Neal (1997) report the merits of Catholic secondary schooling; in contrast, Altonji et al. (2005), Figlio and Stone (1999) and Goldhaber (1996) find no significant gaps between public and private schools. Research drawing on private school voucher experiments in the U.S. and some Latin American countries reveals recent evidence on impacts of private school attendance (Anand et al., 2009; Angrist et al., 2002; Krueger and Zhu, 2004). They generally show a slightly positive effect of private school attendance on student academic performance (McEwan, 2004).

Another useful line of research on private education is investigating causal effects of private tutoring expenditures in East Asian countries and some others. This is a small but growing literature exploring impacts of private tutoring on student educational outcomes. The studies, however, have yet to show unambiguous evidence on causal impacts of private tutoring. Dang (2007), Dang and Rogers (2008) and Ono (2007) support strong effects of private tutoring in Vietnam and Japan; in contrast, Briggs (2001), Gurun and Millimet (2008), Kang (2007) and Ryu and Kang (2009) document negligible impacts of tutoring and coaching on education outcomes in the U.S., Turkey and South Korea, respectively.

Although informative for furthering our understanding about the effectiveness of private education investments, studies on private tutoring vary greatly in the empirical strategies and analyzed countries.[1] Hence, it is currently hard to determine whether the conflicting empirical results of the impact of private tutoring arise because of variations in estimation methods or those in analyzed countries. In order to improve the knowledge on the effectiveness of private tutoring and private education investments in general, it is necessary to apply an identical empirical method to different countries using a standardized international data set.

Given such limitations of studies on private tutoring, the current paper undertakes an international study on impacts of private education spending on student academic performance, using the OECD's Programme for International Student Assessment (PISA) 2006. The PISA 2006 data set offers an excellent instrument for such an undertaking. PISA is an internationally standardized assessment of reading, math and science literacy that is developed by the OECD and partner countries in order to measure the degree of academic preparedness of students at age 15. A unique feature of the PISA 2006 data, which is rarely available in other international tests, is that a question is asked as to how much parents privately and directly paid to educational providers for services on top of tax-based public education expenditures. As such private payments cover tuition fees to the child's school including private schools and cram schools and other miscellaneous costs directly paid to the school, we interpret the amount

[1] Using data from Vietnam, Dang (2007) relies on a joint Tobit-ordered probit model that involves a fairly complicated likelihood function whereby an identification of the effect of private tutoring comes from an instrumental variable (IV) of tutoring fees charged by the schools in the commune. As the author admits, such a variable is likely to proxy for local living standards, which can be directly correlated with a student's educational outcome. Ono (2007) examines the effect of *ronin*—spending additional years upon graduation of high school to enter prestigious colleges in Japan—on college quality, employing the average quality of colleges within the respondent's prefecture of origin as an IV for the *ronin* status. As in Dang (2007), such a variable can also be related with a respondent's college quality as a proxy for unobservable student and family characteristics. Briggs (2001) addresses endogeneity of coaching for standardized admissions tests (SAT or ACT) in the U.S. by means of Heckman's selection correction methods but it is unclear where an identification comes from besides functional form assumptions. For data from South Korea, Kang (2007) employs a nonparametric bounding method to circumvent the endogeneity problem.

of the answered private payments as private education spending for a student in addition to public education spending. In the PISA 2006 data such information is available for a total of 16 countries/economies: Bulgaria, Colombia, Germany, Denmark, Hong Kong - China, Croatia, Iceland, Italy, S. Korea, Luxembourg, Macao - China, New Zealand, Poland, Portugal, Qatar and Turkey.

Using the PISA 2006 data, the current paper contributes to the literature by exploring the causal effect of private education spending on student academic performance for each sampled country and comparing it internationally across different countries. With the PISA 2000 data, Vandenberghe and Robin (2004) undertook similar research, examining the effectiveness of private (as opposed to public) education on student achievement across countries. They, however, relied on a simple classification of private vs. public education. The current paper improves upon them by exploring impacts of private educational *expenditures*—a more detailed measure of educational investments—across countries.

It is well known that educational expenditures in general are not exogenously and randomly given to a student; there is little doubt that private education spending is endogenous in an estimation of its impact on educational outcomes. To circumvent endogeneity of the private education spending, we apply two estimation methods that gain growing popularity for causal estimations in the recent program evaluation literature. One is a propensity-score matching method; the other is a nonparametric bounding method (Blundell and Costa-Dias, 2009; Imbens and Wooldridge, 2008; Heckman et al., 1999).[2] The results from both methods suggest that the effects of private education spending do not seem large for all the countries examined in the PISA 2006 data. In terms of an elasticity, the largest magnitudes of the effect drawn from the matching method imply that a 10 percent increase in spending leads to no more than a 0.6 percent average improvement in test scores. Moreover, some countries reveal even negative impacts of an increase in private education spending on test scores. For all the countries examined by the bounding method, there is also no compelling evidence that private education spending strongly improves a student's test performance. In a majority of the analysis samples, the elasticities based on the upper bounds of the average treatment effect (ATE) suggest that a

[2] An instrumental variables (IV) method can be another proper method of estimation, provided that a strong and exogenous instrument is available in the data. Unfortunately, however, we fail to find a good IV in the PISA 2006 data. In nonparametric bounding analysis we employ a dummy variable indicating whether or not the education level of a student's mother is above high school as an IV for private education spending. But such an IV is not exogenous but monotone in the sense that will be defined in section 2.2.

10 percent increase in spending is unlikely to raise a student's test score by more than 1 percent. In addition, lower bounds of the ATE fail to be greater than zero in every analysis sample. Taken altogether, a conservative interpretation of the results of the matching and bounding methods would be that private education spending has either little average impact on student academic performance, or a weak positive effect at best.

The rest of the paper is organized as follows. Section 2 outlines the empirical strategy of the paper. Data are discussed in section 3; empirical results are shown in section 4. Section 5 concludes.

2 Empirical Framework

To begin with, let us consider an empirical model expressed by:

$$y_i = \beta_0 + \beta_1 ln(S_i) + \beta_2 X_i + u_i$$

where y_i is a test score of student i, S_i is private educational spending for i, X_i is a vector of i's characteristics, and u_i is the error term. Running an OLS regression of y_i against $ln(S_i)$ and X_i is a simple method of estimating the causal parameter β_1 using cross-sectional data. To the extent that S_i is not exogenously and randomly determined for student i, however, an OLS method fails to yield an unbiased estimate for β_1. For example, if educationally-motivated families spend more on private education and these families are not appropriately measured, the estimate for β_1 is likely to be biased upward. On the other hand, if parents determine the expenditure on private tutoring according to the quality of the child, then the estimate can also be biased: if parents tend to spend more on weak children and less on strong children, the estimate will be biased downward; if parents spend less on weak children and more on strong children, the estimate will be biased upward.

In the absence of a randomized experiment for private education spending, a causal estimation in principle calls for an extra variable for a student that strongly affects the parents' decision to invest in her education, while being independent of educational outcomes of the student when S_i is controlled for. It seems, however, fairly difficult if not impossible for studies to find such a variable in household data. In the face of difficulties in causal estimations, the current paper employs two estimation methods that address the endogeneity of private education spending. The first method is a propensity-score matching method; the second method is a nonparametric bounding method

To explain the empirical methods in further detail, let us first define y as a normalized—with mean zero and variance one—test score of a student for each country. Let T be a treatment indicator of the amount of private spending on education, which was reported in the PISA 2006 data in the form of a categorical variable with 5 discrete ranges. (The details on the classifications are given in section 3.) T is equal to zero if the amount of spending reported by the parents is in the lowest category (category 1); one if it is in the second lowest category (category 2); and two if it is in the other three categories (categories 3, 4 and 5). Four cut-off values of the amount of spending that separate the five categories differ by the country. (See Appendix Table 1 for the cut-off values (in US$) for each country.) Each student receives treatment $t \in T = \{0, 1, 2\}$.

2.1 A Matching Estimator

As introduced in many survey papers (e.g., Blundell and Costa-Dias, 2009; Imbens and Wooldridge, 2008; Smith and Todd, 2005), a matching method attempts to yield a counterfactual observation(s) of the treated observation based on observational characteristics. Since there can be more than one observational characteristics for which the treated observation and a counterfactual observation(s) are matched, studies rely on the propensity score which is a probability of treatment predicted by observational characteristics (Rosenbaum and Rubin, 1983). By averaging the difference between the treated outcome and counterfactual outcomes for all the treated observations, the matching method attempts to yield causal estimates for the average treatment effects on the treated (ATT) and the average treatment effect (ATE). In the current paper, the treatment is a varying level of private education spending and the outcome of interest is a test score given a certain level of spending.

A key assumption employed in the matching literature is that treatment assignment and potential outcomes are independent conditional on a set of the individual's observable characteristics X (the Conditional Independence Assumption, CIA). In the context of the current paper, the CIA states that for a student who shows the same observable characteristics (X), a level of private education spending is independent of the potential test score under each level of private education spending. However, to the extent that there may exist unobservable factors that can affect parents' decision over the level of private education spending, the CIA may be problematic in the current evaluation problem. In fact, the same problem can give rise to any evaluation problem for that matter. Nonetheless, the matching literature argues that matching estimators in general have low bias if the data include a rich set of variables related to treatment assignment and potential outcomes (Smith and Todd, 2005). Provided that the PISA 2006 data offers fairly rich information on treatment assignment, a propensity-score matching method is likely to yield good estimates for the causal impact of private education spending.

While a typical presentation of a matching method is based on a binary treatment, the current study deals with three discrete levels of education spending in the analysis. In order to consider more than two levels of the treatment, we rely on Lechner (2001) who has developed general propensity-score matching methods for more than two mutually exclusive treatments; Larsson (2003), Lechner (2002) and Ryu and Kang (2009) are real-data examples of the method. Hence, what follows heavily draws on the preceding papers.

Given that there are three different levels of spending or treatments ($\{0, 1, 2\}$), we denote potential outcomes by $\{y^0, y^1, y^2\}$. For each student, only one outcome can be observable in the data and the others are counterfactuals. Our evaluation problem is to estimate the average impact of treatment m compared to treatment l for combinations of $m, l \in \{0, 1, 2\}$ $(m > l)$. More formally, the outcome of interest is:

$$\theta_0^{m,l} = E(y^m - y^l | T = m) = E(y^m | T = m) - E(y^l | T = m).$$

Here, $\theta_0^{m,l}$ is the multiple-treatment version of the average treatment effect on the treated (ATT), which denotes the expected average treatment effect of treatment m relative to treatment l for participants in treatment m. To the extent that $E(y^m|T = m)$ is easily constructed from the data, matching methods attempt to construct the unobservable counterfactual $E(y^l|T = m)$ under the CIA, which is formally expressed by: with II meaning independence,

$$y^0, y^1, y^2 \quad \amalg \quad T \ \mid \ X$$

Extending Rosenbaum and Rubin (1983) into the multiple treatments framework, Lechner

(2001) shows that it is not necessary for matching to condition on multidimensional X but only to condition on the participation probability conditional on X (the propensity score). Hence,

$$E(y^l|T=m) = E_{P^{m|ml}_{(X)}}[E(y^l|P^{m|ml}(X),\ T=l) \mid T=m]$$

where $P^{m|ml}(x) = P^{m|ml}(T=m|T=l \text{ or } T=m,\ X=x)$. We employ a matching protocol for the estimation of $\theta_0^{m,l}$ suggested in Lechner (2001, Table 1). It is described in the appendix.

Along with estimates for ATT, we report the average treatment effect of treatment m relative to treatment l for a participant drawn randomly from the population (ATE). According to Lechner (2001), the ATE is calculated as follows:

$$\widehat{\gamma}_N^{m,l} = \sum_{j=0}^{2} \left[\left(\widehat{E}(y^m|T=j) - \widehat{E}(y^l|T=j) \right) \cdot P(T=j) \right]$$

Such an ATE is comparable to the average treatment effects drawn from a bounding method of the next section.

2.2 A Nonparametric Bounding Method

Our second method of estimation is a nonparametric bounding method that was first introduced in economics by Manski (1990) and further developed in Manski (1997), Manski and Pepper (2000).[3] The basic idea of the bounding method is that instead of obtaining point estimates (as in instrumental variables (IV) or matching methods) that often rely on strong assumptions, one may calculate lower and upper bounds of the ATE under a few weaker assumptions. Provided that the bounds are sufficiently narrow and informative to locate the causal effect, we interpret that the magnitude of the true effect is somewhere between the estimated lower and upper bounds.

As in the matching method, each student receives treatment $t \in T = \{0, 1, 2\}$. Define the response function $y(\cdot) : T \rightarrow Y$ to map treatments into outcomes. The realized outcome $y \equiv y(z)$ is the level of y for a student who actually receives treatment z. The latent outcome $y(t)$ $(t \neq z)$ describes what level of performance the student would have achieved had he or she received treatment t. Of primary interest are the average causal effects of increasing private

[3]Some recent examples of this method include Blundell et al. (2007), Gonzalez (2005), Kang (2007), Kreider and Pepper (2007), Lechner (1999), Pepper (2000) and Ryu and Kang (2009) among others.

education spending on the test score: $E[y(1) - y(0)]$, $E[y(2) - y(1)]$, and $E[y(2) - y(0)]$.

In order to set up bounds for the treatment effects, we first decompose $E[y(t)]$ by

$$E[y(t)] = E[y|z = t]Pr(z = t) + E[y(t)|z \neq t]Pr(z \neq t) \tag{1}$$

To make bounds analysis feasible, let us suppose that y is bounded by $[K_0,\, K_1]$. We set K_0 equal to -5 and K_1 equal to 5.[4] Since the unobservable counterfactual $E[y(t)|z \neq t]$ is bounded by $[K_0,\, K_1]$, we have the worst-case (WC) bounds of $E[y(t)]$ given by

$$\begin{aligned} &E[y|z = t]Pr(z = t) + K_0 Pr(z \neq t) \\ &\leq E[y(t)] \leq \\ &E[y|z = t]Pr(z = t) + K_1 Pr(z \neq t) \end{aligned} \tag{2}$$

In order to further tighten the bounds of $E[y(t)]$, assumptions are introduced below individually as well as jointly. An assumption is monotone treatment response (MTR), which is specified as follows:

$$t_l < t_m \longrightarrow y(t_l) \leq y(t_m) \tag{3}$$

This assumption is drawn from a theory that there will be non-negative impacts of increased educational spending on a student's academic performance. A majority of empirical studies support the validity of such an assumption. Although the exact magnitude of a positive effect of educational spending is hardly agreed, it is rare that studies find strong *negative* impacts of monetary educational investments on performance (see Hanushek (1997, 2003); an exception is a study of Leuven et al. (2007)).[5]

Another assumption to be employed is monotone treatment selection (MTS), which is specified by:

$$t_l < t_m \longrightarrow E[y(t)|z = t_l] \leq E[y(t)|z = t_m] \tag{4}$$

[4] Alternative values of K_0 and K_1 do not affect the results of the paper qualitatively. Bounds based on alternative values of K_0 and K_1 are available upon request.

[5] The MTR bounds of $E[y(t)]$ can be expressed by

$$\begin{aligned} &E[y|z \leq t]Pr(z \leq t) + K_0 Pr(z > t) \\ &\leq E[y(t)] \leq \\ &E[y|z \geq t]Pr(z \geq t) + K_1 Pr(z < t) \end{aligned}$$

This assumption supposes that sorting into treatment is not exogenous but monotone in the sense that the average latent outcome $y(t)$ is greater for those students whose parents spend a large amount of money on private education ($z = t_m$) than for those whose parents spend a small amount ($z = t_l$, $t_l < t_m$). For instance, high income parents are more likely to spend a large amount of money on their child's education than low-income parents, while children of high income parents tend to be more academically able and smarter than those of low-income parents (see, e.g., Haveman and Wolfe (1995)). While it specifies a source of endogeneity in a conventional OLS method of examining the impacts of private education spending, the MTS assumption can make an important contribution to tightening the bounds of the true effect in combination with MTR.[6]

Joining MTR with MTS assumptions, we can obtain the MTR+MTS bounds of $E[y(t)]$ given by

$$\begin{aligned}&\textstyle\sum_{k<t} E(y|z=k)Pr(z=k) + E(y|z=t)Pr(z \geq t)\\ &\leq E[y(t)] \leq\\ &\textstyle\sum_{k>t} E(y|z=k)Pr(z=k) + E(y|z=t)Pr(z \leq t)\end{aligned} \tag{5}$$

Manski and Pepper (2000, p.1004) proposes a test of the joint MTR+MTS hypotheses, which states that $E[y|z=u]$ must be a weakly increasing function of u. Under MTR+MTS, it should be satisfied that

$$\begin{aligned} u' \leq u \quad \Rightarrow \quad & E[y|z=u'] = E[y(u')|z=u'] \\ & \overset{by\ MTR}{\leq} \quad E[y(u)|z=u'] \overset{by\ MTS}{\leq} \quad E[y(u)|z=u] = E[y|z=u] \end{aligned} \tag{6}$$

After applying such a test to each sampled country, we exclude 8 countries—Germany, Denmark, Croatia, Iceland, Italy, Luxembourg, Macao - China, Qatar—from the bounding analysis. The sample mean $\widehat{E}(y|z=u)$ fails to be a weakly increasing function of u, hence either MTR or MTS assumption does not hold for these countries.

As can be seen shortly in Table 5, MTR+MTS bounds of the effect of private education spending on test scores seem to be tight enough to give meaningful interpretations of

[6]The MTS assumption yields the bounds of $E[y(t)]$ given by

$$\begin{aligned}&E[y|z=t]Pr(z \geq t) + K_0 Pr(z < t)\\ &\leq E[y(t)] \leq\\ &E[y|z=t]Pr(z \leq t) + K_1 Pr(z > t)\end{aligned}$$

the estimated bounds for all the countries examined. Nonetheless, one can further tighten MTR+MTS bounds by introducing a monotone instrumental variable (MIV) υ that satisfies mean monotonicity—$E[y(t)|\upsilon = u_1] \leq E[y(t)|\upsilon = u_2]$ if $u_1 < u_2$—as proposed in Manski and Pepper (2000). Mean monotonicity supposes that the expected test score of students with $\upsilon = u_1$ is less than or equal to that of students with $\upsilon = u_2$ for any given level of private spending. For an MIV, we propose to use a dummy variable (H) indicating whether or not the education level of a student's mother is above high school. A rationale is that given a certain level of private education spending, students with high-educated mothers are more likely to perform better than those with low-educated mothers thanks to, for example, better innate ability, emotional support for education, more involvement in study, etc.[7]

Combining MIV with MTR+MTS, the MIV+MTR+MTS bounds of $E[y(t)]$ are given by

$$\begin{aligned}
&\textstyle\sum_{u \in H} Pr(H = u) \cdot \\
&\quad \left\{\sup_{u_1 \leq u} [\textstyle\sum_{k<t} E(y|H = u_1, z = k) Pr(z = k|H = u_1) + E(y|H = u_1, z = t) Pr(z \geq t|H = u_1)]\right\} \\
&\leq E[y(t)] \leq \\
&\textstyle\sum_{u \in H} Pr(H = u) \cdot \\
&\quad \left\{\inf_{u_2 \geq u} [\textstyle\sum_{k>t} E(y|H = u_2, z = k) Pr(z = k|H = u_2) + E(y|H = u_2, z = t) Pr(z \leq t|H = u_2)]\right\}
\end{aligned} \tag{7}$$

We calculate conditional expectations, $E[y(t)|\cdot]$, nonparametrically by relying on local linear regression (Fan, 1992) in which the control variable is a PISA index of economic, social and cultural status of the student's family, and $\widehat{E}[y(t)|\cdot]$ is evaluated at its mean value. Given the bounds of $E[y(t)]$ under varying assumptions, the lower bound (LB) of an average treatment effect (ATE), $E[y(t_m) - y(t_l)]$ ($t_m > t_l$), is calculated by the difference between the lower bound of $E[y(t_m)]$ and the upper bound of $E[y(t_l)]$; the upper bound (UB) of the ATE is obtained by the difference between the upper bound of $E[y(t_m)]$ and the lower bound of $E[y(t_l)]$. Along with the bounds of $E[y(t)]$ and ATEs are calculated bootstrap 5th and 95th percentiles of the lower and upper bound, respectively. The interval between these percentiles shows a conservative 90% confidence interval for the estimated bounds. The number of the bootstrap samples is 100.

In section 4 we report estimated bounds under the MTR+MTS assumption for each country,

[7] Under mean monotonicity, MIV bounds are expressed by

$$\begin{aligned}
&\textstyle\sum_{u \in H} Pr(H = u) \left\{\sup_{u_1 \leq u} [E(y|H = u_1, z = t) Pr(z = t|H = u_1) + K_0 Pr(z \neq t|H = u_1)]\right\} \\
&\leq E[y(t)] \leq \\
&\textstyle\sum_{u \in H} Pr(H = u) \left\{\inf_{u_2 \geq u} [E(y|H = u_2, z = t) Pr(z = t|H = u_2) + K_1 Pr(z \neq t|H = u_2)]\right\}
\end{aligned}$$

while those under each individual assumption are suppressed. In addition, in Appendix Table 2, we show estimated MIV+MTR+MTS bounds for each of the 6 countries excluding Turkey (Grades 9 and 10). The MIV+MTR+MTS bounds of Turkey are dropped, since the estimated LBs of ATEs generally exceed its UBs probably due to a failure of mean-monotonicity of the MIV.

3 The Data - PISA 2006

For empirical analysis we employ data from the OECD's Programme for International Student Assessment (PISA), 2006. PISA is an internationally standardized assessment developed by the OECD and partner countries in order to measure the degree of academic preparedness of students at age 15. The PISA 2006 study is the third cycle of the survey in which a total of 56 countries and regions including all 30 OECD member countries and 16 partner countries take part. For each country the target population is 15-year-old students—students aged from 15 years 3 months to 16 years 2 months—who are enrolled in a formal educational institution regardless of the grade at the time of the survey.[8]

The PISA 2006 sampling procedure consists of two stages. In the first stage each country drew a stratified random sample of schools (usually 150 schools per country) in which 15-year-old students were enrolled. The second stage randomly sampled 35 of the 15-year-old students in each of these schools. This sampling procedure typically yields a sample of 4,500 to 10,000 tested students in each country.

The performance tests, which were paper and pencil tests, were conducted for the students in the sample. The subjects were reading, math and science literacy. We employ a total score summed over reading, math and science tests as a measure of student performance, since the spending on education is an overall spending for unspecified subjects as explained shortly.[9] In addition to the tests, a wealth of background information on the students was collected by individual student questionnaires. The study also conducted interviews for the principal of each school in order to gather information on school characteristics.

[8] A more detailed description of the general characteristics of the PISA 2006 study is found in OECD (2006). For more technical details interested readers are referred to OECD (2008). Important documents can be found at the PISA web site, www.pisa.oecd.org.

[9] Missing test scores of a subject is not a concern in the PISA 2006, because the number of raw test score observations is identical for three subjects within each country.

A unique feature of the PISA 2006 study is that in order to add to details on family background information on tested students, a parent questionnaire was administered for a total of 16 countries/economies: Bulgaria, Colombia, Germany, Denmark, Hong Kong - China, Croatia, Iceland, Italy, S. Korea, Luxembourg, Macao - China, New Zealand, Poland, Portugal, Qatar and Turkey. In that parent questionnaire, a question is asked as to how much parents paid to educational providers for services in the last twelve months. Such payments include any tuition fees paid to the child's school, any other fees paid to individual teachers in the school or to other teachers for any tutoring for the child, as well as any fees for cram school, but exclude the costs of goods like sports equipment, school uniforms, computers or textbooks that are not included in a general fee. Although not explicit in the question, such payments do not contain public education spending paid in the form of taxes, since it is not a part of parents' direct payments made to schools. Yet, they include tuition fees if the student attend a private school or cram schools. Hence, we interpret the amount of private payments reported in the answer of the question as total spending on a student's private education, using it as our key explanatory variable.

An answer to this question consists of five discrete categories that divides non-negative amounts into five mutually exclusive areas. Four cut-off values that separate these five areas differ by the country. (See Appendix Table 1 for the cut-off values (in US$) for each country.) For example, for South Korea, 1.5 million Won ($ 1,465), 3 million Won ($ 2,929), 4.5 million Won ($ 4,394) and 6 million Won ($ 5,857) are used in sequence for cut-off values. For each country we aggregate the five areas into three by pooling three highest categories into one. If necessary, we convert the categorical variable into a continuous variable by taking the median value of each category. We use purchasing power parity (PPP) exchange rates for 2006 to convert different currencies into US dollars.

For both matching and bounding analyses that follow, Portugal is ruled out, because the range of education expenditures defined by the categorical variable is too wide and crude to be informative for this country. For Portugal, 83.9 percent of respondents choose an expenditure category that covers Euro 20 ($ 25) to 4,000 ($ 5,000). As a result, the average expenditure calculated on the basis of a continuous variable amounts to $ 2,933 (see Table 1), which is the largest among 16 countries in the PISA 2006 sample.

The matching methods are applied to the remaining 15 countries except for Portugal. Among the 15 countries, the bounding analysis is applied to the 7 countries alone—Bulgaria, Colombia, Hong Kong - China, S. Korea, New Zealand, Poland, and Turkey—because the MTR+MTS assumption fails to hold, hence the bounding analysis is not suitable for the other 8 countries. Namely, sample mean $\widehat{E}(y|z = u)$ fails to weakly increase over u for the these 8 countries.

Finally, to make the analysis sample of a country as homogenous as possible, we restrict each country's sample to a grade that has the largest number of students. A sample of grade 9 students is employed for Bulgaria, Germany, Denmark, Croatia, Luxembourg and Poland ; a sample of grade 10 students is used for Colombia, Hong Kong, Iceland, Italy, S. Korea, Macao and Qatar; a sample of grade 11 students for New Zealand; two separate samples of grade 9 and 10 students are exploited for Turkey, since each grade contains a sufficiently large number of students for analysis.

4 Estimation Results

4.1 Descriptive Statistics and OLS Results

A left panel of Table 1 shows descriptive statistics of the total test score and private spending on education for the 16 countries to which a parent questionnaire was administered. As explained in section 3, the sample size of the raw data ranges from 3,789 (Iceland) to 21,773 (Italy) for different countries. The mean of the total test score summed over reading, math and science tests varies from 979.3 (Qatar) to 1,626.5 (Hong Kong).

A unique piece of information that can be drawn from the PISA 2006 data is the average level of private educational spending for 15-year-old children in each country. If we exclude Portugal as an exceptional case as discussed earlier, parents of east Asian countries such as Hong Kong ($2,375), South Korea ($2,828) and Macao ($1,706) spend the greatest amount of private money on children's education; those of New Zealand, Qatar and Turkey expend a medium amount; and those of Denmark, Bulgaria and Poland the smallest amounts among parents in the 15 countries. Such patterns of private education spending generally agree with the findings of Baker et al. (2001) and Bray (1999).

A right panel of Table 1 reports OLS results of an empirical model for each country that is expressed by:

$$y_i = \beta_0 + \beta_1 ln(S_i) + \beta_2 X_i + u_i$$

where y_i is a normalized test score of student i, S_i is a continuous variable for private educational spending for i, X_i is a vector of i's characteristics[10], and u_i is an error term. The right panel of the table shows the OLS estimate of β_1, its standard error and T-value for each country. In the last column we convert $\widehat{\beta}_1$ into an elasticity showing a percent change in test score due to a 10 percent increase in spending, which is evaluated at the mean test score of each country.

For all the countries, a 10 percent greater private spending on education does not seem to be strongly associated with a higher test score, although statistical significance varies by the country. A 10 percent greater spending is associated with a maximum of 0.015 SD higher (S. Korea) and a minimum of 0.01 SD lower test score (Croatia and Italy). If the estimates are converted into the elasticities, the effect of a 10 percent greater private education spending ranges from even 0.189 percent lower (Italy) to 0.231 percent higher test scores (S. Korea). The simple average of the elasticities across the 16 analysis samples is little different from zero.

While they can be somewhat informative, such estimated elasticities drawn from OLS estimates are not causal estimates for the effect of an increase in spending due to the well-known endogeneity problem. We below rely on matching and bounding methods to draw causal estimates.

4.2 Matching Results

Table 2 shows the estimation results of the matching methods, using the sample for each country. The estimates of average treatment effects on the treated ($\widehat{\theta}_N^{m,l}$, ATT) are reported in columns (1) to (3); average treatment effects ($\widehat{\gamma}_N^{m,l}$, ATE) in columns (4) to (6); and the elasticities of ATE that show a percent change in test score due to a 10 percent increase in spending in columns (7) to (9). The elasticity $\widehat{\eta}_N^{m,l}$ for each country is calculated by the following formula:

$$\widehat{\eta}_N^{m,l} = 10 \times \frac{\widehat{\gamma}_N^{m,l} \times SD \times \widehat{E}(S_i|t=m)}{S^{ml} \times \widehat{E}(Y_i|t=m)} \tag{8}$$

where SD is a standard deviation of the raw total test score before normalization (Y_i); $\widehat{\gamma}_N^{m,l}$ is an estimated ATE of $E[y^m - y^l]$; and $S^{ml} \equiv \widehat{E}(S_i|t=m) - \widehat{E}(S_i|t=l)$ ($m > l$; $m, l = 0, 1, 2$).

The matching results suggest that an increase in private education spending does not yield

[10] They include i's sex, birth year, number of books at home, PISA index of home educational resources, PISA index of economic, social and cultural status, whether the test language is the language i speaks at home, whether i's school is a private or public school, and a type of community in which i's school is located.

strong positive causal impacts on the test scores. While most of ATE's for a majority of countries fail to be statistically significantly different from zero effects, Colombia, Germany, Denmark and Italy see some of the ATEs significantly different from zero. Nonetheless, the sizes of the effects for these countries do not seem to be large. For instance, Denmark's $\widehat{\gamma}_N^{1,0}$ and Germany's $\widehat{\gamma}_N^{2,1}$ are, respectively, equal to 0.204 and 0.244, and are significantly different from zero at the 5 percent level of significance. If both of the ATEs are converted into the elasticities, however, they imply that a 10 percent increase in education spending raises the test score by no more than 0.38 and 0.60 percent, respectively. Moreover, such elasticities are among the highest in the 16 analysis samples. Another notable case is Italy whose $\widehat{\gamma}_N^{1,0}$ and $\widehat{\gamma}_N^{2,0}$ are equal to negative 0.062, both being significantly different from zero. The converted elasticities suggest that a 10 percent higher education spending leads to even a decrease in test score by 0.17 and 0.13 percent, respectively.

To the extent that the cut-off values of different categories of spending levels differ by the country, elasticities rather than ATTs or ATEs are a better measure of comparison across countries. The simple average, standard deviation, maximum and minimum of the elasticities across the 16 analysis samples are presented at the bottom of Table 2. The average of elasticity $\widehat{\eta}_N^{1,0}$ shows a 0.053 percent improvement in test score due to a 10 percent increase in spending. The maximum $\widehat{\eta}_N^{1,0}$ (0.454) is attained by Colombia, and the minimum $\widehat{\eta}_N^{1,0}$ (-0.169) is given by Italy. As for elasticity $\widehat{\eta}_N^{2,1}$, the average elasticity is 0.061; the maximum is 0.599 (Germany); and the minimum is -0.041 (Hong Kong). As for elasticity $\widehat{\eta}_N^{2,0}$, the average elasticity is 0.050; the maximum is 0.428 (Denmark); and the minimum is -0.167 (Turkey, Grade 9). In sum, the effects of private education spending do not seem large for all the countries examined in the PISA 2006 data. The largest magnitudes of the elasticities imply that a 10 percent increase in spending leads to no more than a 0.6 percent improvement in test scores. There are some countries that reveal even negative impacts of an increase in private education spending on test scores.

The primary finding that private education spending does not yield a large improvement in academic performance hold if the overall sample is disaggregated by sex of the student. Tables 3 and 4 present the estimates of ATEs and elasticities for the boys and girls sample, respectively, of each country.

For boys, most of ATE's for a majority of countries are not significantly different from zero effects, although some of the ATEs are different from zero for Colombia, Denmark and Italy. As is seen earlier, however, the sizes of ATEs for these countries do not seem large. If we focus on elasticities for comparisons across countries, the effects of private education spending do not seem strong for all the countries examined. The average of elasticity $\widehat{\eta}_N^{1,0}$ across the 16 analysis samples is 0.095; the maximum is 0.628 (Denmark); and the minimum is −0.151 (Italy). As for elasticity $\widehat{\eta}_N^{2,1}$, the average is 0.036; the maximum is 0.385 (Germany); and the minimum is −0.151 (Denmark). As for elasticity $\widehat{\eta}_N^{2,0}$, the average is 0.056; the maximum is 0.4 (Colombia); and the minimum is −0.115 (Turkey, Grade 9).

Weak impacts of increased private education spending reiterate for the girls samples of the countries. Overall, the averages of the elasticities are slightly above zero effects for girls as well. In addition, the largest magnitudes of elasticities imply that a 10 percent rise in spending leads to no more than a 0.75 percent improvement in test scores. Furthermore, some countries even show negative impacts of an increase in private education spending on test scores.

4.3 Bounding Results

Relying on the bounding method for causal estimation, Table 5 shows MTR+MTS bounds of the effect of increased private education spending for each of seven countries (eight analysis samples). A lower bound (LB), an upper bound (UB), the 5th percentile of the LB, and the 95th percentile of the UB are presented in columns (1) to (4), respectively. In columns (5) and (6) we convert the UB estimate and its 95th percentile into the elasticity representing a percent change in test score due to a 10 percent increase in spending. The elasticity is calculated by equation (8).

For all the eight analysis samples examined by the bounding method, there is no compelling evidence that private education spending strongly improves a student's test performance. Such a result largely reiterates the findings of the matching method. In a majority of the analysis samples, the elasticities imply that a 10 percent increase in spending is unlikely to raise a student's test score by more than 1 percent. Moreover, LBs of the effect of spending fail to be greater than zero in every analysis sample. For South Korea and Turkey's grade 9 samples, however, the UBs of the effect seem relatively large, implying that a 10 percent increase in spending raises test score at the largest by about 0.52 to 0.92 percent.[11] Nonetheless, the

[11]Employing a data set of South Korea (the Korean Education and Employment Panel, KEEP), Kang (2007)

corresponding LBs do not exceed zero, which suggests that the true effect of private education spending is not far from zero.

The primary finding that private education spending lead to at best a small improvement in performance hold when the overall sample is disaggregated by sex of the student. Table 6 reports the elasticities of the MTR+MTS bounds of the average effect of private education spending for boys and girls separately. For boys and girls alike, a 10 percent increase in spending is unlikely to raise a student's test score by more than 1 percent in a majority of analysis samples; there is little evidence that the impacts of private education spending are strong and different between boys and girls. An exception appears to be South Korea where the UBs of the effects seem relatively large. The elasticities of the UBs range from 0.44 to 1.18 for boys, and from 0.60 to 0.73 for girls. However, the LBs fail to be different from zero, implying the true impacts of the education spending are unlikely to be strong.

Tighter MIV+MTR+MTS bounds of the effect of private education spending shown in Appendix Table 2 draw a largely similar picture. In a vast majority of the samples, a 10 percent increase in spending raises a student's test score at the largest by less than 1 percent. LBs of the effect of spending fail to exceed zero. Suppressed results of MIV+MTR+MTS bounds for separate boys and girls samples largely repeat the findings of Table 6. Taken altogether, it is difficult to find compelling evidence that private education spending strongly improves student performance. A conservation interpretation of the empirical results of the matching and bounding methods would be that private education spending has either little average impact on student academic performance, or a weak positive effect at best.

5 Concluding Remarks

Using the PISA 2006 data, this paper undertakes an international study on impacts of private education spending on student academic performance. Its primary result drawing from the matching and bounding methods is that the effects of private education spending remain only modest for all the countries examined in the PISA 2006 data. In terms of an elasticity, the largest magnitudes of the effects imply that a 10 percent increase in spending leads to no more

implements a similar bounding method in order to measure the bounds of the causal effect of private tutoring expenditures on test scores. While lower limits of his MIV+MTR+MTS bounds fail to exceed zero, their upper limits suggest that a 10 percent increase in spending raises a student's test score by 0.53 to 0.76 percent (Table 6). Such amounts are fairly close to the elasticities (0.52 to 0.92) for S. Korea obtained by the current study.

than a 0.6 percent average improvement in test scores. We fail to find compelling evidence that an increase in private education spending greatly raises student academic performance in the countries examined.

Although weak impacts of private education spending may be in contrast to what is expected by some researchers, such a finding is aligned with recent studies on the effectiveness of monetary education expenditures on academic achievement. As for the impacts of public education spending, Guryan (2003) shows that a 10 percent increase in spending raises test scores by 0.77 to 1.15 percent. Leuven et al. (2007) reveals that roughly a 10 percent increase in school budget *reduces* language and arithmetic scores of eight graders by 0.1 SD, i.e., approximately 1.5 percent in elasticity. Relying on randomization lotteries in the Chicago Public Schools in the U.S., Cullen et al. (2005, 2006) report the absence of a positive impact of public school choice and high-quality peers on traditional student outcomes.

As for private education investments, McEwan (2004) summarizes that there are small achievement gains for students who attend private school using vouchers but such gains are confined only to African American students. However, he adds that such results are highly sensitive to analytical assumptions and gains are not evident for other racial and ethnic groups. Using data on private tutoring expenditures for high school students in South Korea, Kang (2007) reports that the causal effects of private tutoring are at most modest: no greater than a 0.76 percent increase in test scores due to a 10 percent increase in private tutoring expenditure.

Overall, the conclusion of the current paper lends further support for Hanushek's skeptical views on the effectiveness of school resources and monetary education investments on student outcomes (Hanushek, 1986; 1997; 2003). Private education spending is no more effective for student performance than public school expenditures are. Two questions naturally arise from such a conclusion. The first is why private education spending does not lead to an improvement in performance; the second is why then parents spend private money for their children's education with no compelling evidence on its positive impacts. For the first question, Kang (2007) offers two potential explanations in the context of private tutoring in South Korea: (1) poor overall quality of teachers in the private tutoring sector, and (2) peer pressure among parents for private tutoring. The peer pressure can also be a potential answer to the second question, since it may lead parents to spend too much for children's private tutoring. These explanations, however, are obviously confined to South Korea. Finding factors that explain the conclusion of this paper for other countries would be a topic for further research.

Appendix: A propensity-score matching protocol

Step 1 Estimate the conditional probabilities on the subsample of participants in m and l to obtain propensity score $\widehat{P}^{m|ml}(x)$.

Step 2 For a given value of m and l, the following steps are performed:

1. Choose one observation in the subsample defined by participation in m and delete it from that subsample.
2. Find an observation in the subsample of participants in l that is as close to the one chosen in step 2.(1) in terms of $\widehat{P}^{m|ml}(x)$. Do not remove that observation to be used again.
3. Repeat (1) and (2) until no participant is left in subsample m.
4. Using the matched comparison group formed in (3), compute the sample mean $\widehat{E}(y^l|T = m)$ as an estimate for $E(y^l|T = m)$.

Step 3 Repeat step 2 for all combinations of m and l.

Step 4 Compute the estimate of $\theta_0^{m,l}$ by $\widehat{\theta}_N^{m,l} = \widehat{E}(y^m|T = m) - \widehat{E}(y^l|T = m)$. Obtain the standard error of $\widehat{\theta}_N^{m,l}$ by generating 100 bootstrap samples.[12]

[12] A STATA software called PSMATCH2 that was developed by Leuven and Sianesi (2003) is employed for matching estimations.

References

Altonji, J.G., Elder, T.E., Taber, C.R., 2005a. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, Journal of Political Economy 113 (1), 151-184.

Anand, P., Mizala, A., Repetto, A., 2009. Using school scholarships to estimate the effect of private education on the academic achievement of low-income students in Chile, Economics of Education Review, forthcoming.

Angrist, J.D., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment, American Economic Review 92 (5), 1535-1558.

Baker, D.P., Akiba, M., LeTendre, G.K., Wiseman, A.W., 2001. Worldwide shadow education: Outside-school learning, institutional quality of schooling, and cross-national mathematics achievement, Educational Evaluation and Policy Analysis 23 (1), 1-17.

Blundell, R., Costa-Dias, M., 2009. Alternative Approaches to Evaluation in Empirical Microeconomics, Journal of Human Resources, forthcoming.

Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds, Econometrica 75 (2), 323-363.

Bray, M., 1999. The shadow education system: Private tutoring and its implications for planners, UNESCO International Institute for Educational Planning, Paris.

Briggs, D.C., 2001. The Effect of Admissions Test Preparation: Evidence from NELS:88, Chance 14 (1), 10-18.

Cullen, J.B., Jacob, B.A., Levitt, S., 2005. The impact of school choice on student outcomes: an analysis of the Chicago Public Schools, Journal of Public Economics 89 (5-6), 729-760.

Cullen, J.B., Jacob, B.A., Levitt, S., 2006. The Effect of School Choice on Participants: Evidence from Randomized Lotteries, Econometrica 74 (5), 1191-1230.

Dang, H., 2007. The determinants and impact of private tutoring classes in Vietnam, Economics of Education Review, 26 (6), 683-698.

Dang, H., Rogers, F.H., 2008. How to interpret the growing phenomenon of private tutoring: human capital deepening, inequality increasing, or waste of resources?, Policy Research Working Paper Series of The World Bank, No. 4530

Evans, W.N., Schwab, R.M., 1995. Finishing high school and starting college: Do Catholic schools make a difference?, Quarterly Journal of Economics 110 (4), 941-974.

Fan, J., 1992. Design Adaptive Nonparametric Regression, Journal of of the American Statistical Association 87, 998-1004.

Figlio, D.N., Stone, J.A., 1999. Are private schools really better?, In: Polachek, S. (Eds.), Research in Labor Economics, Vol.18, JAI Press, Stamford, Connecticut, pp. 115-140.

Goldhaber, D.D., 1996. Public and private high schools: Is school choice an answer to the productivity problem?, Economics of Education Review 15 (2), 93-109.

Gonzalez, L., 2005. Nonparametric Bounds on the Returns to Language Skills, Journal of Applied Econometrics 20 (6), 771-795.

Gurun, A., Millimet, D.L., 2008. Does Private Tutoring Payoff?, IZA Discussion Paper No. 3637, The Institute for the Study of Labor (IZA).

Guryan, J., 2003. Does Money Matter? Estimates from Education Fianance Reform in Massachusetts, Mimeo.

Hanushek, E.A., 1986. The economics of schooling: Production and efficiency in public schools, Journal of Economic Literature 24 (3), 1141-1177.

Hanushek, E.A., 1997. Assessing the effects of school resources on student performance: An update, Educational Evaluation and Policy Analysis 19 (2), 141-164.

Hanushek, E.A., 2003. The failure of input-based schooling policies, Economic Journal 113 (485), F64-F98.

Haveman, R., Wolfe, B., 1995. The Determinants of Children's Attainments: A Review of Methods and Findings, Journal of Economic Literature 33 (4), 1829-1878.

Heckman, J., LaLonde, R., Smith, J., 1999. The Economics and Econometrics of Active Labour Market Programme, in Ashenfelter, O. and Card, D. (eds.), The Handbook of Labor Economics, Volume III.

Imbens, G.W., Wooldridge, J.M., 2008. Recent Developments in the Econometrics of Program Evaluation, IZA Discussion Paper No. 3640, The Institute for the Study of Labor (IZA).

Kang, C., 2007. The Effect of Private Tutoring Expenditures on Academic Performance: Evidence from a Nonparametric Bounding Method, Working Paper, Department of Economics, National University of Singapore.

Kreider, B., Pepper, J.V., 2007. Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors, Journal of the American Statistical Association 102 (478), 432-441.

Krueger, A., Zhu, P., 2004. Another look at the New York City school voucher experiment, American Behavioral Scientist 47, 699-717.

Larsson, L., 2003. Evaluation of Swedish youth labour market programmes, Journal of Human Resources 38 (4), 891-927.

Lechner, J., 1999. Nonparametric bounds on employment and income effects of continuous vocational training in East Germany, Econometrics Journal 2 (1), 1-28.

Lechner, M., 2001. Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption, in Lechner, M., Pfeiffer, F. (eds), Econometric Evaluation of Labour Market Policies, Heidelberg: Physica/Springer, pp. 43-58.

Lechner, M., 2002. Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies, Review of Economics and Statistics 84 (2), 205-220.

Leuven, E., Lindahl, M., Oosterbeek, H., Webbink, D., 2007. The Effect of Extra Funding for Disadvantaged Pupils on Achievement, Review of Economics and Statistics 89 (4), 721-736.

Leuven, E., Sianesi, 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, available at "http://ideas.repec.org/c/boc/bocode/s432001.html".

Manski, C.F., 1990. Nonparametric Bounds on Treatment Effects, American Economic Review 80 (2), 319-323.

Manski, C.F., 1997. Monotone Treatment Response, Econometrica 65(6), 1311-1334.

Manski, C.F., Pepper, J.V., 2000. Monotone Instrumental Variables, with an Application to the Returns to Schooling, Econometrica 68 (4), 997-1012.

McEwan, P.J., 2004. The Potential Impact of Vouchers, Peabody Journal of Education 79 (3), 57-80.

Neal, D., 1997. The effects of Catholic secondary schooling on educational achievement, Journal of Labor Economics 15 (1), 98-123.

OECD (Organization for Economic Co-Operation and Development), 2006. Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006, OECD Publications, Paris.

OECD (Organization for Economic Co-Operation and Development), 2008. PISA 2006 Technical Reports, OECD Publications, Paris.

Ono, H., 2007. Does examination hell pay off ? A cost-benefit analysis of "ronin" and college education in Japan, Economics of Education Review, 26 (3), 271-284.

Pepper, J.V., 2000. The Intergenerational Transmission Of Welfare Receipt: A Nonparametric Bounds Analysis, Review of Economics and Statistics 82 (3), 472-488.

Rosenbaum, P.R., Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika 70 (1), 41-55.

Ryu, D., Kang, C., 2009. The Effect of Private Tutoring Expenditures on Academic Performance of Middle School Students: Evidence from the Korea Education Longitudinal Study, Unpublished manuscript.

Smith, A.S., Todd, P.E., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators?, Journal of Econometrics 125 (1-2), 305-353.

Vandenberghe,V., Robin, S., 2004. Evaluating the effectiveness of private education across countries: a comparison of methods, Labour Economics 11 (4), 487-506.

Table 1: Descriptive Statistics of the Sample and OLS Estimates of the Effect of Private Spending on Test Scores

	Descriptive statistics of the raw data			OLS results Dependent variable: Normalized test score					
Country:	Total sample	Total score Mean (SD)	Private Spending ($) Mean (SD)	Grade	Sample used	Estimate	S.E.	T-value	Elasticity evaluated at the mean test score
Bulgaria*	4,498	1,250.5 (305.5)	241.7 (243.6)	9	2,983	0.020	0.013	1.52	0.049
Colombia*	4,478	1,141.5 (242.9)	166.7 (191.9)	10	1,608	0.050	0.018	2.75	0.105
Germany	4,891	1,513.1 (295.3)	774.8 (617.1)	9	1,822	-0.018	0.010	-1.84	-0.036
Denmark	4,532	1,504.3 (252.5)	471.3 (726.8)	9	1,910	-0.045	0.019	-2.31	-0.075
Hong Kong-China*	4,645	1,626.5 (253.5)	2,375.1 (2,395)	10	2,809	0.029	0.018	1.64	0.045
Croatia	5,213	1,436.8 (244.0)	442.8 (380.5)	9	3,707	-0.096	0.019	-5.17	-0.163
Iceland	3,789	1,479.5 (267.3)	393.3 (510.8)	10	2,240	-0.037	0.016	-2.27	-0.067
Italy	21,773	1,406.0 (275.9)	272.5 (163.9)	10	13,066	-0.096	0.009	-10.18	-0.189
S. Korea*	5,176	1,624.2 (257.7)	2,827.9 (1,810)	10	4,905	0.146	0.017	8.65	0.231
Luxembourg	4,567	1,456.2 (277.0)	675.2 (722.3)	9	1,468	-0.047	0.036	-1.33	-0.077
Macao-China	4,760	1,526.4 (219.9)	1,705.8 (1,106)	10	1,684	-0.053	0.035	-1.53	-0.076
New Zealand*	4,823	1,572.7 (290.5)	1,364.1 (1,497)	11	2,800	-0.002	0.013	-0.18	-0.004
Poland*	5,547	1,500.0 (260.8)	238.2 (233.3)	9	4,825	0.072	0.014	5.14	0.125
Portugal	5,109	1,410.9 (263.1)	2,932.9 (2,128)						
Qatar	6,265	979.3 (263.0)	1,339.4 (1,335)	10	1,968	-0.057	0.016	-3.55	-0.154
Turkey*	4,942	1,295.1 (249.7)	965.8 (1,667)	9	1,892	0.116	0.024	4.80	0.224
				10	2,478	0.033	0.016	2.01	0.063
				Average		0.001			0.000
				S.D.		0.069			0.124
				Max		0.146			0.231
				Min		-0.096			-0.189

Note: * indicates that a bounding method is applied to that country.

Table 2: Average Treatment Effects on the Treated (ATT) and Average Treatment Effects (ATE) of the Matching Methods: The Whole Sample

	Average Treatment Effects on the Treated (ATT)			Average Treatment Effects (ATE)			Elasticities of ATE		
Country:	$\theta_N^{1,0}$	$\theta_N^{2,1}$	$\theta_N^{2,0}$	$\gamma_N^{1,0}$	$\gamma_N^{2,1}$	$\gamma_N^{2,0}$	$\eta_N^{1,0}$	$\eta_N^{2,1}$	$\eta_N^{2,0}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bulgaria	0.046	0.083*	0.176*	0.031	-0.013	0.050*	0.090	-0.039	0.120
	(0.049)	(0.048)	(0.064)	(0.027)	(0.022)	(0.030)			
Colombia	0.117	0.143	0.341*	0.203*	-0.013	0.162*	0.454	-0.030	0.298
	(0.084)	(0.090)	(0.153)	(0.067)	(0.044)	(0.069)			
Germany	-0.061	-0.310*	-0.306*	-0.049	0.244*	0.030	-0.094	0.599	0.059
	(0.061)	(0.146)	(0.158)	(0.049)	(0.097)	(0.049)			
Denmark	-0.009	-0.034	0.416	0.204*	0.045	0.261*	0.375	0.092	0.428
	(0.117)	(0.172)	(0.300)	(0.098)	(0.043)	(0.131)			
Hong Kong	0.036	-0.014	0.128	0.082	-0.025	0.083*	0.186	-0.041	0.125
-China	(0.065)	(0.057)	(0.079)	(0.057)	(0.029)	(0.049)			
Croatia	-0.082*	-0.063	-0.269*	0.024	0.007	0.019	0.058	0.019	0.037
	(0.045)	(0.067)	(0.074)	(0.025)	(0.017)	(0.036)			
Iceland	-0.127*	0.044	-0.073	0.000	-0.001	-0.015	0.000	-0.002	-0.026
	(0.073)	(0.089)	(0.078)	(0.037)	(0.026)	(0.046)			
Italy	-0.112*	-0.038	-0.176*	-0.062*	0.024*	-0.062*	-0.169	0.078	-0.131
	(0.028)	(0.026)	(0.028)	(0.017)	(0.011)	(0.015)			
S. Korea	0.050	0.171*	0.217*	0.020	0.009	0.037	0.049	0.025	0.066
	(0.048)	(0.045)	(0.042)	(0.031)	(0.021)	(0.026)			
Luxembourg	0.027	-0.021	-0.217*	-0.027	0.045	0.029	-0.060	0.114	0.057
	(0.070)	(0.073)	(0.090)	(0.045)	(0.031)	(0.046)			
Macao-China	0.076	-0.031	-0.151	0.026	0.024	0.032	0.052	0.055	0.049
	(0.082)	(0.075)	(0.098)	(0.050)	(0.037)	(0.055)			
New Zealand	-0.015	0.136	0.020	-0.007	0.015	0.018	-0.013	0.033	0.030
	(0.052)	(0.127)	(0.109)	(0.038)	(0.050)	(0.038)			
Poland	0.058	0.097*	0.106*	0.001	0.013	0.029	0.002	0.028	0.051
	(0.049)	(0.043)	(0.063)	(0.031)	(0.015)	(0.030)			
Qatar	-0.140*	-0.097	-0.282*	-0.042	0.028	-0.066	-0.126	0.089	-0.160
	(0.061)	(0.119)	(0.150)	(0.058)	(0.036)	(0.050)			
Turkey	0.067	0.230*	0.432*	-0.006	-0.009	-0.098*	-0.016	-0.016	-0.167
(Grade 9)	(0.081)	(0.087)	(0.096)	(0.034)	(0.024)	(0.046)			
Turkey	0.016	0.101	0.149*	0.021	-0.012	-0.020	0.062	-0.024	-0.040
(Grade 10)	(0.047)	(0.068)	(0.067)	(0.028)	(0.016)	(0.031)			
Average	-0.003	0.025	0.032	0.026	0.024	0.031	0.053	0.061	0.050
S.D.	0.076	0.126	0.242	0.075	0.060	0.084	0.161	0.147	0.149
Max	0.117	0.230	0.432	0.204	0.244	0.261	0.454	0.599	0.428
Min	-0.140	-0.310	-0.306	-0.062	-0.025	-0.098	-0.169	-0.041	-0.167

Notes: Bootstrap standard errors are reported in parentheses. * indicates that the estimate is significant at the 0.05 level. The elasticities are percent changes in test score due to a 10 percent increase in spending. See text for details.

Table 3: Average Treatment Effects (ATE) of the Matching Methods: A Boys Sample

		Average Treatment Effects (ATE)			Elasticities of ATE		
Country:	N	$\gamma_N^{1,0}$	$\gamma_N^{2,1}$	$\gamma_N^{2,0}$	$\eta_N^{1,0}$	$\eta_N^{2,1}$	$\eta_N^{2,0}$
		(1)	(2)	(3)	(4)	(5)	(6)
Bulgaria	1,533	-0.038	-0.021	0.040	-0.109	-0.063	0.096
		(0.045)	(0.032)	(0.041)			
Colombia	674	0.123	0.014	0.217*	0.276	0.031	0.400
		(0.070)	(0.071)	(0.073)			
Germany	917	0.021	0.157	0.053	0.041	0.385	0.104
		(0.089)	(0.120)	(0.093)			
Denmark	827	0.341*	-0.073	0.170	0.628	-0.151	0.279
		(0.153)	(0.088)	(0.150)			
Hong Kong	1,330	0.056	0.034	0.101	0.127	0.055	0.153
-China		(0.116)	(0.050)	(0.107)			
Croatia	1,860	0.029	0.004	0.033	0.069	0.010	0.064
		(0.039)	(0.023)	(0.051)			
Iceland	1,044	0.052	0.026	-0.038	0.112	0.062	-0.068
		(0.052)	(0.036)	(0.067)			
Italy	6,037	-0.055*	0.030	-0.050	-0.151	0.094	-0.106
		(0.028)	(0.018)	(0.027)			
S. Korea	2,464	-0.009	0.044	0.018	-0.022	0.116	0.033
		(0.040)	(0.028)	(0.038)			
Luxembourg	768	-0.036	0.024	-0.038	-0.079	0.060	-0.076
		(0.065)	(0.047)	(0.076)			
Macao-China	731	0.028	0.015	0.024	0.055	0.034	0.038
		(0.070)	(0.063)	(0.065)			
New Zealand	1,286	0.026	-0.055	-0.012	0.048	-0.117	-0.021
		(0.091)	(0.115)	(0.075)			
Poland	2,349	0.041	-0.003	-0.007	0.105	-0.006	-0.012
		(0.047)	(0.029)	(0.042)			
Qatar	850	0.081	0.043	0.023	0.244	0.138	0.055
		(0.078)	(0.048)	(0.069)			
Turkey	1,014	0.033	-0.034	-0.067	0.087	-0.063	-0.115
(Grade 9)		(0.067)	(0.045)	(0.076)			
Turkey	1,302	0.030	-0.007	0.040	0.089	-0.014	0.078
(Grade 10)		(0.044)	(0.029)	(0.052)			
Average		0.045	0.012	0.032	0.095	0.036	0.056
S.D.		0.088	0.050	0.075	0.176	0.119	0.133
Max		0.341	0.157	0.217	0.628	0.385	0.400
Min		-0.055	-0.073	-0.067	-0.151	-0.151	-0.115

Notes: Bootstrap standard errors are reported in parentheses. * indicates that the estimate is significant at the 0.05 level. The elasticities are percent changes in test score due to a 10 percent increase in spending. See text for details.

Table 4: Average Treatment Effects (ATE) of the Matching Methods: A Girls Sample

Country:	N	Average Treatment Effects (ATE) $\gamma_N^{1,0}$ (1)	$\gamma_N^{2,1}$ (2)	$\gamma_N^{2,0}$ (3)	Elasticities of ATE $\eta_N^{1,0}$ (4)	$\eta_N^{2,1}$ (5)	$\eta_N^{2,0}$ (6)
Bulgaria	1,450	0.056	-0.086*	0.037	0.161	-0.258	0.088
		(0.037)	(0.031)	(0.040)			
Colombia	934	0.183*	0.010	0.121	0.410	0.022	0.222
		(0.081)	(0.062)	(0.084)			
Germany	905	-0.037	0.305*	-0.003	-0.072	0.748	-0.006
		(0.059)	(0.116)	(0.065)			
Denmark	1,083	-0.118	0.204*	0.045	-0.217	0.420	0.074
		(0.086)	(0.052)	(0.152)			
Hong Kong	1,479	-0.040	0.045	-0.010	-0.091	0.073	-0.016
-China		(0.078)	(0.040)	(0.064)			
Croatia	1,847	-0.033	-0.018	0.004	-0.078	-0.051	0.008
		(0.033)	(0.027)	(0.040)			
Iceland	1,196	0.044	0.001	-0.035	0.093	0.002	-0.063
		(0.041)	(0.028)	(0.051)			
Italy	7,029	-0.048*	0.013	-0.048*	-0.132	0.041	-0.102
		(0.021)	(0.014)	(0.019)			
S. Korea	2,441	0.054	0.000	0.036	0.130	0.000	0.063
		(0.040)	(0.027)	(0.034)			
Luxembourg	700	0.020	0.066	0.021	0.044	0.166	0.043
		(0.048)	(0.047)	(0.053)			
Macao-China	953	0.024	0.060	-0.034	0.047	0.139	-0.053
		(0.062)	(0.049)	(0.063)			
New Zealand	1,514	0.062	-0.026	0.085	0.117	-0.056	0.144
		(0.059)	(0.065)	(0.058)			
Poland	2,476	0.057	-0.012	0.038	0.148	-0.025	0.067
		(0.041)	(0.027)	(0.033)			
Qatar	1,118	-0.191*	0.072	-0.087*	-0.577	0.229	-0.212
		(0.060)	(0.047)	(0.049)			
Turkey	878	0.043	-0.005	0.045	0.114	-0.009	0.077
(Grade 9)		(0.061)	(0.045)	(0.077)			
Turkey	1,176	0.038	-0.037	-0.035	0.111	-0.077	-0.067
(Grade 10)		(0.046)	(0.029)	(0.045)			
Average		0.007	0.037	0.011	0.013	0.085	0.017
S.D.		0.083	0.093	0.051	0.209	0.224	0.101
Max		0.183	0.305	0.121	0.410	0.748	0.222
Min		-0.191	-0.086	-0.087	-0.577	-0.258	-0.212

Notes: Bootstrap standard errors are reported in parentheses. * indicates that the estimate is significant at the 0.05 level. The elasticities are percent changes in test score due to a 10 percent increase in spending. See text for details.

Table 5: MTR+MTS Bounds of the Effect of Private Spending on Test Scores: The Whole Sample

	Estimated Bounds				Elasticity	
	(1)	(2)	(3)	(4)	(5)	(6)
	LB	UB	LB 5 pctile	UB 95 pctile	UB	UB 95 pctile
Bulgaria						
$E[y(1) - y(0)]$	0.000	0.091	0.000	0.201	0.262	0.579
$E[y(2) - y(1)]$	0.000	0.229	0.000	0.345	0.691	1.041
$E[y(2) - y(0)]$	0.000	0.255	0.000	0.396	0.608	0.946
Colombia						
$E[y(1) - y(0)]$	0.000	0.180	0.000	0.360	0.402	0.805
$E[y(2) - y(1)]$	0.000	0.274	0.000	0.400	0.620	0.906
$E[y(2) - y(0)]$	0.000	0.354	0.000	0.549	0.652	1.010
Hong Kong-China						
$E[y(1) - y(0)]$	0.000	0.004	0.000	0.111	0.008	0.251
$E[y(2) - y(1)]$	0.004	0.083	0.000	0.189	0.135	0.306
$E[y(2) - y(0)]$	0.000	0.033	0.000	0.164	0.049	0.249
S. Korea						
$E[y(1) - y(0)]$	0.000	0.215	0.000	0.312	0.522	0.756
$E[y(2) - y(1)]$	0.000	0.347	0.000	0.445	0.919	1.179
$E[y(2) - y(0)]$	0.000	0.403	0.000	0.514	0.713	0.909
New Zealand						
$E[y(1) - y(0)]$	0.000	0.031	0.000	0.107	0.058	0.202
$E[y(2) - y(1)]$	0.000	0.159	0.000	0.305	0.338	0.651
$E[y(2) - y(0)]$	0.000	0.171	0.000	0.330	0.291	0.561
Poland						
$E[y(1) - y(0)]$	0.000	0.167	0.000	0.270	0.435	0.702
$E[y(2) - y(1)]$	0.000	0.095	0.000	0.173	0.204	0.373
$E[y(2) - y(0)]$	0.000	0.197	0.000	0.303	0.349	0.537
Turkey (Grade 9)						
$E[y(1) - y(0)]$	0.000	0.347	0.000	0.510	0.927	1.363
$E[y(2) - y(1)]$	0.000	0.360	0.000	0.565	0.657	1.033
$E[y(2) - y(0)]$	0.000	0.502	0.000	0.728	0.861	1.247
Turkey (Grade 10)						
$E[y(1) - y(0)]$	0.000	0.049	0.000	0.138	0.143	0.404
$E[y(2) - y(1)]$	0.000	0.084	0.000	0.181	0.173	0.375
$E[y(2) - y(0)]$	0.000	0.101	0.000	0.201	0.195	0.389

Table 6: MTR+MTS Bounds of the Effect of Private Spending on Test Scores by Sex

	Elasticity					
	Boys			Girls		
	(1)	(2)	(3)	(4)	(5)	(6)
	LB	UB	UB 95 pctile	LB	UB	UB 95 pctile
Bulgaria						
$E[y(1) - y(0)]$	0.000	0.389	0.815	0.000	0.014	0.287
$E[y(2) - y(1)]$	0.000	0.730	1.284	0.043	0.591	0.971
$E[y(2) - y(0)]$	0.000	0.711	1.188	0.000	0.382	0.740
Colombia						
$E[y(1) - y(0)]$	0.000	0.600	1.050	0.000	0.302	0.740
$E[y(2) - y(1)]$	0.000	0.633	1.055	0.000	0.629	1.013
$E[y(2) - y(0)]$	0.000	0.793	1.232	0.000	0.594	1.057
Hong Kong-China						
$E[y(1) - y(0)]$	0.000	0.000	0.240	0.000	0.158	0.533
$E[y(2) - y(1)]$	0.007	0.007	0.151	0.004	0.235	0.462
$E[y(2) - y(0)]$	0.000	0.000	0.164	0.000	0.180	0.468
S. Korea						
$E[y(1) - y(0)]$	0.000	0.438	0.715	0.000	0.597	0.881
$E[y(2) - y(1)]$	0.006	1.182	1.515	0.000	0.730	1.028
$E[y(2) - y(0)]$	0.000	0.778	1.002	0.000	0.677	0.906
New Zealand						
$E[y(1) - y(0)]$	0.000	0.000	0.159	0.000	0.069	0.245
$E[y(2) - y(1)]$	0.017	0.420	0.813	0.000	0.358	0.720
$E[y(2) - y(0)]$	0.000	0.291	0.633	0.000	0.313	0.620
Poland						
$E[y(1) - y(0)]$	0.031	0.663	1.040	0.000	0.230	0.559
$E[y(2) - y(1)]$	0.000	0.107	0.311	0.000	0.329	0.556
$E[y(2) - y(0)]$	0.000	0.422	0.664	0.000	0.306	0.557
Turkey (Grade 9)						
$E[y(1) - y(0)]$	0.000	0.349	0.932	0.000	0.330	0.882
$E[y(2) - y(1)]$	0.000	0.463	0.846	0.000	0.247	0.452
$E[y(2) - y(0)]$	0.000	0.607	1.039	0.000	0.382	0.655
Turkey (Grade 10)						
$E[y(1) - y(0)]$	0.000	0.066	0.195	0.000	0.041	0.122
$E[y(2) - y(1)]$	0.000	0.065	0.135	0.000	0.131	0.271
$E[y(2) - y(0)]$	0.000	0.091	0.177	0.000	0.141	0.272

Appendix Table 1: Cut-off Values of Private Education Spending (in US$)

	Cut-off order			
Country:	1	2	3	4
Bulgaria	63.7	318.5	636.9	955.4
Colombia	21.5	215.4	430.9	646.3
Germany	12.5	1,500.0	3,000.0	4,500.0
Denmark	83.3	666.7	1,333.3	2,000.0
Hong Kong-China	642.7	1,285.3	6,426.7	12,853.5
Croatia	336.1	840.3	1,344.5	1,848.7
Iceland	159.7	798.7	1,437.7	2,076.7
Italy	125.0	250.0	375.0	500.0
S. Korea	1,464.7	2,929.3	4,394.0	5,858.7
Luxembourg	125.0	1,000.0	2,000.0	3,000.0
Macao-China	1,248.4	2,496.9	3,745.3	4,993.8
New Zealand	140.8	2,112.7	4,225.4	6,338.0
Poland	92.6	185.2	370.4	771.6
Portugal	25.0	5,000.0	10,000.0	15,000.0
Qatar	274.7	1,648.4	3,022.0	4,395.6
Turkey	447.8	895.5	3,731.3	11,194.0

Appendix Table 2: MIV+MTR+MTS Bounds of the Effect of Private Spending on Test Scores

	Estimated Bounds				Elasticity	
	(1)	(2)	(3)	(4)	(5)	(6)
	LB	UB	LB 5 pctile	UB 95 pctile	UB	UB 95 pctile
Bulgaria						
$E[y(1)-y(0)]$	0.000	0.053	0.000	0.166	0.153	0.479
$E[y(2)-y(1)]$	0.005	0.202	0.000	0.337	0.608	1.015
$E[y(2)-y(0)]$	0.000	0.209	0.000	0.374	0.498	0.893
Colombia						
$E[y(1)-y(0)]$	0.000	0.167	0.000	0.326	0.373	0.729
$E[y(2)-y(1)]$	0.000	0.211	0.000	0.389	0.478	0.880
$E[y(2)-y(0)]$	0.000	0.287	0.000	0.515	0.528	0.947
Hong Kong-China						
$E[y(1)-y(0)]$	0.000	0.009	0.000	0.144	0.021	0.328
$E[y(2)-y(1)]$	0.005	0.080	0.000	0.175	0.131	0.285
$E[y(2)-y(0)]$	0.000	0.037	0.000	0.187	0.055	0.283
S. Korea						
$E[y(1)-y(0)]$	0.000	0.237	0.000	0.338	0.575	0.818
$E[y(2)-y(1)]$	0.000	0.355	0.000	0.461	0.943	1.223
$E[y(2)-y(0)]$	0.000	0.425	0.000	0.537	0.752	0.949
New Zealand						
$E[y(1)-y(0)]$	0.000	0.022	0.000	0.093	0.041	0.176
$E[y(2)-y(1)]$	0.000	0.141	0.000	0.273	0.302	0.583
$E[y(2)-y(0)]$	0.000	0.152	0.000	0.306	0.259	0.521
Poland						
$E[y(1)-y(0)]$	0.000	0.201	0.000	0.313	0.522	0.815
$E[y(2)-y(1)]$	0.000	0.093	0.000	0.202	0.200	0.436
$E[y(2)-y(0)]$	0.000	0.222	0.000	0.363	0.393	0.644