Public and publicly subsidised private primary schools in Spain: new evidence from a quasiexperimental approach

María-Jesús Mancebón[1]
José-María Gómez-Sancho[1]
Mauro Mediavilla[2]
Domingo P. Ximénez-de-Embún[1]

1. Introduction

Studies devoted to evaluating the impact of educational interventions have expanded notably worldwide over the last two decades. There are two factors explaining this. On the one hand, the availability of new, high quality, national and international data. On the other, the development of innovative and sophisticated methods capable of confronting the principal methodological problems facing this kind of studies. These factors have created new opportunities for academics to conduct research that addresses policymakers' concerns about the consequences of actions directed at improving educational outcomes (Murname and Willet, 2011).

One of the subjects that has focused the empirical work of many educational economists has been the public funding of privately run schools. The evaluation of the impact of this policy, widely applied under different guises (vouchers, charter schools, publicly subsidised private schools, busing), has been boosted in the last ten years by the potential that the innovative techniques of causal inference have in the analysis of such a controversial question. These methodologies, clustered under the rubric of Propensity Score Analysis (PSA) by Guo and Fraser (2010), have proved themselves to be extremely useful in the analysis of causal effects in non-experimental studies, that is to say in settings in which the values of all variables (including those describing participation in different potential treatments) are observed rather than assigned by an external agent (Shadish *et al*., 2002).

In this paper we use one of these techniques, propensity score matching (PSM), in order to evaluate the effect of attending a publicly subsidised private school (hereinafter PrSPS) on some of the educational skills promoted by Spanish primary schools. These PrSPS are privately owned and managed but financed by the regional government. Research into this topic is extremely important in Spain, where two models of school management (public and private) coexist and compete for limited public resources. Although advocates of each of these models usually invoke arguments of quality, efficiency or even social equality in order to defend their preferred option, technical studies comparing the performance of public and privately run

[1] University of Zaragoza (Spain)
[2] University of Valencia (Spain) & IEB

schools are so far inconclusive (Toma and Zimmer, 2012). These contradictory results prevent the identification of the optimal model of educational management (public versus private). Our study makes a further contribution to this controversial issue.

The data used in this paper come from a national evaluation project established for the 2006 Spanish Education Act (LOE): the *Evaluación de Diagnóstico* (hereinafter ED), which contains a wealth of information about the socioeconomic context of students and the scores they attain in a standardized external test in the fourth grade. The study specifically concentrates on the Spanish region of Aragon, arguing that the region's funding, governance arrangements and student populations make PrSPS and public schools (hereinafter PuS) more readily comparable than in a larger area. Data were collected in 2010 and were provided by the Local Educational Authority of the region of Aragon, which is responsible for the implementation of the ED in their district.

Our methodological strategy is defined by the sequential application of two methods: propensity score matching (PSM) and hierarchical linear models (HLM). The first of these will allow us to circumscribe a reduced and homogeneous sample of students attending PrSPS and PuS. The application of HLM to this reduced sample of students will allow us to be more accurate in the estimation of the effect of the PrSPS on the educational skills evaluated in the ED. Additionally; we also test the sensitivity of our estimates with respect to unobserved heterogeneity.

The remainder of the paper is organized as follows. The next section briefly describes recent research in this area. In section 3 we provide a description of the data used and the institutional setting. Section 4 expounds the methodological strategy employed to identify the effect of PrSPS on educational achievement. In section 5 we present the results of our estimations. The paper concludes with a summary of our findings in section 6.

2. Background

The origin of research into the effect of school type (private or public) on educational performance is usually attributed to the controversial work of Coleman, Hoffer and Kilgore (1982). This performed a multidimensional comparison of North American public and private schools (Catholic and non-Catholic) from the data supplied by the *High School and Beyond* project. Of all the questions dealt with in the report, those which had greatest media and academic impact were those concerned the comparison of the results obtained by pupils in public and private Catholic schools in standardised tests to evaluate basic cognitive skills (reading, writing and mathematics). Their conclusions, favourable to private schools (PrS, hereinafter), led to a prolific line of research which has lasted until today and which has been directed at overcoming the methodological limitations attributed to the work of Coleman and to testing his results in diverse educational contexts.

The principal deficiencies attributed to the above study were centred on the methodology used to discern the effect of PrS upon the cognitive results of pupils: multiple regression analysis (ordinary least squares, hereinafter OLS) with a set of carefully selected covariates representing the parents' socioeconomic status and other 'background characteristics'[3] of students. Coleman *et al*. (1982) believed that controlling for important pre-existing differences between PuS and PrS' students and their families allowed them to overcome the selection bias that threatened their estimations[4]. Subsequent methodological advances have made clear, however, that the OLS estimate of the parameter (β) of the main predictor (school type), even when incorporating a high number of covariates, will be a biased estimate of the population's average treatment effect (ATE). This is due to the infringement of one of the main assumptions of the OLS method, namely that the residuals are completely unrelated to any predictors included in the regression model.

On the basis of this fact, in recent decades there have emerged a considerable number of studies which have attempted to correct this problem of heterogeneity employing diverse methodological strategies (IV, matching techniques, differences in differences or child fixed effects). The results of this literature are inconclusive concerning the aggregate effect that PrS have on student achievement. While a number of studies in various settings find that PrS outperform traditional PuS (Bedi and Garg, 2000, Morgan, 2001, Anand, Mizala and Repetto, 2009, Lefebvre *et al.* (2011), Kim, 2011), other research has found that student performance attending PrS is not statistically different from that of students in PuS (McEwan, 2001, Bettinger, 2005, Jepsen, 2003, Betts *et al.,* 2006, Witte *et al*, 2007, Chudgar and Quin, 2012). Finally, other studies have concluded that PrS perform worse than PuS (Bifulco and Laad, 2006, Pfeffermann and Landsman, 2011).

In other cases, the effect encountered varies by subject (Altonji *et al*., 2008, Imberman, 2007, Zimmer *et al*., 2012, Davies, 2013), by type of student (Gronberg and Jansen, 2001) or by estimator (Davies, 2013). To these studies must be added those which have shown that the effects of school type vary over time (Sass, 2006, Booker *et al*., 2007, Hanushek *et al*., 2007) and those which s fail to find a consistently positive (or negative) effect of religious schools on overall area-wide educational performance (Allen and Vignoles, 2015).

[3] For a detailed study of the controversy created by the study by Coleman, Hoffer and Kilgore, see number 51(4) of the *Harvard Educational Review* or number 55(2) of *Sociology of Education*.

[4] This bias has its origin in the fact that attendance at a school, whether private or public, is not random but instead is conditioned by characteristics of the family surroundings, which in turn are extremely important in the determination of educational outputs (the family socioeconomic level, for example).

To summarise, the empirical evidence makes clear that the type of influence exercised by the ownership and management of the educational centre on academic results is an open question which requires the performance of additional empirical analyses to those undertaken so far. As Davies states (2013, p.880): "As debates over school choice become increasingly transnational, we need studies from a variety of settings to build a stockpile of international knowledge about school sectors and student achievement". In this context, our study constitutes a contribution aimed at shedding new light on an as yet unclosed debate.

3. Data and institutional background

One of the defining characteristics of the schooling system in Spain is its dual nature, consisting of predominantly public sector provision but with a substantial private sector. The largest segment of the latter are represented by PrSPS, that is to say schools publicly financed by regional education authorities but privately owned and managed[5]. The distribution of students enrolled in primary education among different school types in Spain in 2010 was as follows: 67.44% of students attended a PuS, 28.5% a PrSPS and 4.03% to completely private independent schools (Spanish Ministry of Education, 2013). In this paper, we focus on comparing the performance of PuS and PrSPS in order to evaluate the role of the management model (public versus private) in the promotion of educational skills[6].

Our empirical study employs census data for primary students in the fourth grade in the Spanish region of Aragon in 2010. These data come from the Evaluación de Diagnóstico (ED), a national evaluation of the educational skills of pupils established by the Spanish Education Act (LOE) in 2006 and administered by regional authorities. The ED is involved in the evaluation of several educational competencies that rotate every two years: Spanish; Maths; Science; Digital Skills, Foreign Language, Social Interaction and Citizenship, Arts, Learning by Oneself and Personal Autonomy. In 2010, the second year of the application of the ED, the competencies evaluated were Science and Foreign Language (English). In addition to the assessment of pupils' skills, the ED includes data on a wide range of characteristics. These include basic demographics (gender, age, immigration status, etc.) but also information on children's socioeconomic background (parents' occupational status, parents' years of schooling, household possessions, etc.), data on the academic profile of students (if the child has repeated any academic year, if he or she needs help in doing homework, the daily time devoted to study , etc.) and data on parents' involvement in education, on perceptions of students of themselves

[5] This occurs through the 1985 Right to Education Act (LODE). For a detailed description and historical evolution of the Spanish non-higher education system, see Bernal, 2005.

[6] Green *et al*. (2004) offer an explanation of the principal differences which exist between Spanish PuS and PrSPS.

and the class environment and satisfaction with the school. Table 1 includes the descriptive statistics of all the variables drawn from the ED of 2010, grouped by school type (PuS and PrSPS).

Table 1 shows higher raw results for PrSPS students in the two evaluated educational skills in 2010 (Science and Foreign Language - English)[7]. But raw differences are insufficient to judge the relative quality of PuS versus PrSPS because they do not take into account the differences between students attending each type of school. This is due to the fact that school choice is not exogenous but instead fruit of an individual/family decision, determined by diverse household characteristics such as income and wealth, sociocultural profile etc. (Mancebón and Pérez-Ximénez de Embún, 2011, Burgess and Briggs, 2010, Gallego and Hernando, 2010, Escardibul and Villarroya, 2009 or Tamm, 2008, among others).

This is shown in Table 1, which demonstrates that PrSPS have a much more select student body than PuS. This is true for variables such as parents' occupational status, parents' years of schooling, household possessions, the immigration status of pupils, parents' involvement in education, students' motivation and so on. The differences between PuS and PrSPS are all statistically significant in all these dimensions, a result which mirrors the conclusions reached by other studies which have analysed this topic in Spain with distinct databases. (Mancebón *et al.*, 2012; Calero and Escardibul, 2007; Doncel *et al.*, 2012). In summary, Table 1 underlines the need to apply in our study an estimation strategy which takes into account the differences existing between pupils in PuS and PrSPS and permits the identification of the net effect on school results attributable to school type.

Table 1. Mean and test for equality of means and variances by type of school

Code	Description variable	Mean Total	Mean PuS	Mean PrSPS	Sig. Levene Test for equality of variances	Sig. T-test for equality of means
SCIENCE	Achievement in Sciences	512.37	501.97	526.27	0.00	0.00

[7] The average score of each competence for the total number of schools is 500 and the standard deviation 100, given that as established by the General Report on Diagnostic Evaluation in Aragon 2010 "the evaluation of each competence in Aragon as a whole is established at the level of the average scores transformed into a reference value which has been fixed at 500, with a standard deviation of 100". Here, the approach of the Spanish Diagnostic Evaluation is similar to that of the evaluations of the PISA Project of the OECD. In Table 1 the average score differs from 500 due to the elimination from the sample of private schools without public financing and of those situated in municipalities in which there exists no choice between public and private schools.

FOREIGN LANGUAGE (ENGLISH)	Achievement in foreign language (English)	513.02	499.18	531.50	0.00	0.00
JobMum1	Mother white collar highly skilled	0.29	0.24	0.37	0.00	0.00
JobMum2	Mother white collar low skilled	0.41	0.42	0.39	0.00	0.03
JobMum3	Mother blue collar high skilled	0.04	0.04	0.03	0.00	0.00
JobMum4	Mother blue collar low skilled	0.26	0.30	0.21	0.00	0.00
JobDad1	Father white collar high skilled	0.39	0.31	0.49	0.00	0.00
JobDad2	Father white collar low skilled	0.25	0.26	0.23	0.00	0.01
JobDad3	Father blue collar high skilled	0.30	0.35	0.23	0.00	0.00
JobDad4	Father blue collar low skilled	0.06	0.07	0.05	0.00	0.00
EducationMum	Education mother (years)	11.45	10.78	12.34	0.00	0.00
EducationDad	Education father (years)	11.45	10.78	12.34	0.00	0.00
ZoneGeo1	Student born in Spain	0.87	0.84	0.91	0.00	0.00
ZoneGeo2	Student born in Africa	0.01	0.01	0.00	0.00	0.00
ZoneGeo3	Student born in Asia	0.01	0.01	0.01	0.40	0.68
ZoneGeo4	Student born in Europe	0.05	0.06	0.03	0.00	0.00
ZoneGeo5	Student born in Latin America	0.05	0.06	0.04	0.00	0.00
ZoneGeo6	Student born in an Arab country	0.01	0.02	0.01	0.00	0.00
More5years	More than 5 years living in Spain	0.94	0.93	0.95	0.00	0.00
Gender	Gender (female=1, male=0)	0.49	0.49	0.48	0.27	0.58
Repeater	Student has repeated one or more academic years (repeater=1, non-repeater =0)	0.08	0.09	0.06	0.00	0.00
NumBooks	More than 100 books at home	0.54	0.50	0.60	0.00	0.00
UseBooks	Student uses books frequently	0.72	0.70	0.75	0.00	0.00
Room	Student has own room to study	0.95	0.94	0.96	0.00	0.00
Internet	Internet at home	0.86	0.84	0.88	0.00	0.00
NumTVs	Number of TVs at home	2.11	2.08	2.15	0.01	0.00
NumPCs	Number of computers at home	1.55	1.49	1.63	0.11	0.00
NumTvPay	Number of pay TVs at home	0.44	0.43	0.46	0.00	0.07
NumVideoGames	Number of video games at home	1.73	1.66	1.82	0.16	0.00
NumMP4	Number of MP$ at home	1.01	0.93	1.11	0.00	0.00
StudTim0	Less than 2 hours study every day	0.37	0.37	0.35	0.00	0.13
StudTim1	2 hours study every day	0.16	0.15	0.17	0.00	0.05
StudTim2	More than 2 hours study every day	0.48	0.48	0.48	1.00	1.00
Needhelp	Student needs help in homework	0.22	0.22	0.22	0.28	0.59
RevPar0	Parents do not check either diary or homework	0.21	0.23	0.19	0.00	0.00
RevPar1	Parents check diary but not homework	0.10	0.07	0.13	0.00	0.00
RevPar2	Parents check homework but not diary	0.16	0.20	0.12	0.00	0.00
RevPar3	Parents check both diary and homework	0.53	0.50	0.57	0.00	0.00
RevTeacher	Private tutoring	0.09	0.08	0.09	0.15	0.47
Attitude	Student always finishes homework	0.93	0.92	0.94	0.00	0.01
Aptitude	Student answers homework correctly	0.85	0.84	0.87	0.00	0.00
N		6724	3845	2879		

Source: Authors' calculations, from ED 2010 (Aragon Local Education Authority).

4. Research strategy

When evaluating the impact of PrSPS on students' educational outcomes it is important to take into consideration certain restrictions of the methodological type which affect our study.

Firstly, the data from ED 2010 have a hierarchical structure, due to the fact that the sample selection of individuals occurs at two levels (students and schools). Thus, data are nested.

Consequently, some of the characteristics of students attending the same school are correlated, violating the hypothesis of independence of the observations upon which traditional regression models are based. The application of OLS to these data structures produces an underestimation of the true standard errors, leading to spurious results (Hox, 1995).

Secondly, when assessing the impact of school type, it should be noted that in Spain the distribution of students is not random but, as noted above, schools are chosen by families[8]. Among other influences, family socioeconomic characteristics constitute one of the main determinants of the selection pattern (Escardíbul and Villarroya, 2009; Mancebón and Pérez-Ximénez de Embún, 2014). The "school type" predictor is therefore an endogenous variable. This gives rise to correlations between this predictor and the residuals of the regressions, creating biased parameter estimates. This is what Heckman (1979) defined as sample selection bias.

The characteristics described immediately above are the basic factors conditioning the empirical strategy employed here. This strategy takes material shape in a two-level analysis. Firstly, a Propensity Score Matching (PSM) analysis was conducted, in order to define a homogenous student subsample in terms of the observable characteristics which may simultaneously influence the selection of school type and educational scores. In other words, the PSM analysis allows for the creation of a subsample that is not affected by sample selection bias in observables. In order to control for the impact of unobservable variables on results a sensitivity analysis is employed.

Secondly, a multilevel equation model (HLM) is estimated to allow for the hierarchical structure of the data supplied by ED 2010. This model permits differentiation between those influences acting on the student and those acting on the school. It is expected that this empirical strategy will lead to more accurate estimations and reduced bias.

In the two following sections we synthetically explain the methodological bases of the two techniques to be used in the empirical studies.

4.1. Propensity Score Matching

Selection bias is a methodological problem inherent to all the impact evaluations that use data from administrative records (such as ED). This problem is related to the bias associated to the

[8] LODE (1985): Organic Law 8/1985, 3 July, regulating Education. Official Spanish State Bulletin 159.

pre-treatment differences between treated and non-treated individuals (Caliendo and Kopeinig, 2008) and arises where the assignment of participants to evaluated treatment is not random. This situation is widespread in educational research and is the main econometric problem encountered when trying to measure the effect of privately run schools on the academic performance of children (Lefebvre *et al*., 2011).

The search for research designs and analytic strategies to confront this problem had led to innovative methods originated in the econometrics and statistical fields (for an extensive review of these approaches, see Guo and Fraser, 2010). One of the most popular methods to cope with the possible occurrence of selection bias in observational studies is propensity score matching (PSM) which is erected upon the Neyman-Rubin counterfactual conceptual framework (Neyman, 1923, Rubin, 1974 and 1978).

As is well known, a counterfactual is a theoretical construction that refers to a potential outcome, that is to say to what would have happened to a treated individual if (s)he had not received treatment, *ceteris paribus*. From a theoretical point of view, the counterfactual renders the process of causal inference trivial, because if the value of the counterfactual is established for each person treated, the individual treatment effect (ITE) for this person can be easily evaluated by comparing their real and potential outcomes. The average of these ITEs across all participants in the evaluated treatment would allow the estimation of the average treatment effect (ATT) for the individuals treated. Finally, the application of a statistical test, such as the t-test, would permit evaluation of whether the ATT is extrapolated to the population from which the participants have been sampled (Murname and Willett, 2011).

The fundamental evaluation problem from an empirical point of view arises because the counterfactual is by nature non-observable. At this point the challenge is to find an analytical strategy to proxy the counterfactuals. The contribution of Rosenbaum and Rubin (1983) to this task is extremely valuable. In particular, these authors propose a semiparametric methodology, known as Propensity Score Matching (hereinafter PSM), to deal with the possible occurrence of selection bias (Caliendo and Kopeinig, 2008).

The purpose of PSM is to proxy a credible value of the counterfactual for each of the individuals belonging to the treatment group (hereinafter TG). To accomplish this aim it is necessary to take into account that the only information available regarding the performance achieved without treatment is that corresponding to non-treated individuals (hereinafter control group or CG). Given this consideration, the problem to overcome is to find a procedure that allows the TG and CG to be balanced in all the characteristics relevant to the production of outcomes. The principal advantage of the PSM resides in its capacity to perform matchings between the individuals from the TG and CG when the number of covariates (X) is numerous.

The matchings proposed by Rosenbaum and Rubin (1983) are not performed upon the original variables, but instead upon a single magnitude, the propensity score (ps, hereinafter), which synthesizes all the information contained in the X control variables which simultaneously influence participation in the treatment evaluated and in the outcomes under study. The ps is calculated via a logistic regression model or a similar tool and represents the conditional probability of participating in the evaluated intervention of each individual in the sample, given their observable characteristics X, that is to say:

$$ps = P(W = 1 \mid X) \qquad (1)$$

This magnitude has a very special value in the correction of the selection bias, since the matchings performed upon the base of the ps permit the delimitation of a subsample formed by all the individuals from the TG and those from the CG which are similar in all the observable characteristics that may affect the outcome evaluated. This avoids the differences between the results from each group being contaminated by the differences in the observable characteristics of the members of each group (unconfoundedness).

The key to PSM functioning lies in the creation of good matching, namely in finding the CG individuals having a ps similar to that of the TG individuals. In other words, finding $\forall$ i $\in$ W=1one (some) j $\in$ W=0 such that Pi(W=1) $\approx$ Pj(W=1). This requires that P(W=1 | X)< 1 and P(W=1 | X)>0 $\forall$ X. The fulfilment of these two relationships ensures that the two groups (TG and CG) contain similar individuals regarding observable characteristics (known as the common support assumption). In formal terms, the challenge of this technique resides in finding $\forall i \in W = 1$ those $j \in W = 0$, such that $ps_i \approx ps_j$, where W = 1 indicates participation in the treatment evaluated.

The matching process may be conducted using different algorithms. Guo and Fraser (2010) offer a detailed description of the topic. Several of these algorithms will be used in our empirical work to test the sensitivity of our estimates to the algorithm employed.

Having selected the subsample of comparable individuals, the following step in the PSM is to calculate the estimator of the average treatment effect for treated individuals (ATT or average treatment effect for treated), which is then defined as:

$$\widehat{ATT_{match}} = E\left(\widehat{Y_{match,1}} \middle| W_{match} = 1\right) - E\left(\widehat{Y_{match,0}} \middle| W_{match} = 0\right) \qquad (2)$$

where the match subindex indicates that the estimations refer to the subsample delimited via the PSM, W is the indicator of the group of individuals compared (treatment group:W=1 and control group W=0) and Y indicates the outcome of each group.

In this way an estimation is obtained of the effect of the intervention W upon the outcomes (Y), liberated from the problem of selection bias in observables.

4.2. Hierarchical Linear Models

As just indicated, the application of the PSM permits availability of debugged estimations of the ATT with regard to the observable variables (X) which distinguish the members of the TG and the CG and which are potentially important in the determination of outcomes (Y).

However, the potential influences on the educational results include, usually, more variables than those which simultaneously influence participation in a concrete educational intervention, that is to say those considered in the construction of the ps. Given this situation, the calculation of the net effect of an intervention, such as W, in the educational context requires the contrast of the influence of those other factors (X) which are potentially important in the determination of Y. For this it is fundamental to realize a post-matching analysis. This will provide greater precision in the estimation of the effect of the treatment. Three types of influence deserve attention: the characteristics of the schools at which individuals are educated, the attributes of the students not incorporated into the calculation of the propensity score (those contemporary to the receipt of the treatment) and the differences between the individuals from the TG and CG in unobservable variables.

The testing of the importance of the first two characteristics (the characteristics of schools and characteristics of the pupils not taken into account in the calculation of the propensity score) can be performed via a regression model on the matched sample. In effect, insofar as the subsample delimited via the PSM is not affected by the problem of selection bias in observables which affected the original sample, the regression analysis is now pertinent when identifying the effect of the intervention W upon the results.

With regard to the evaluation of the importance which the unobservable factors may have upon the results obtained, the analysis requires from the performance a sensitivity analysis such as that proposed by Altonji *et al.* (2008) and Rosenbaum, (2002). In section 5.1.3 we shall explain and apply its approach. In what remains of the current section, the foundations of the regression model to be applied in our study will be expounded. On this point, we should underline that our selection of the ideal regression model to conduct our estimates was conditioned by peculiarities of the ED data.

Of all the available regression models, the HLM adapt best to these peculiarities. Their main advantage is that they permit differentiation between those influences acting at the student level (first level of analysis) and those acting at the class and school level (second and third levels).

They are, therefore, models which are especially appropriate for working with data nested at various levels, such as those supplied by almost all educational databases, including ED[9].

These models permit the analysis of variables acting at different levels (individuals, classes and schools, for example) and they allow the identification of the proportion of the total variance of an outcome attributed to each of the specified levels. In analytical terms, the level 1 (student) equation is determined as follows:

$$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} a_{pjk} + e_{ijk} \; con \; e_{ijk} \sim N(0, \sigma^2) \qquad (3)$$

where Y_{ijk} is the expected result from the individual i in the class j in the school k; a_{pjk} is a level 1 explanatory variable for the individual i in the class j in the school k, π_{pjk} are the level 1 coefficients (p=0,1,…,P) and e_{ijk} is the level 1 random effect which is assumed to follow a normal distribution. At level 2 (classes), the π coefficients are treated as variables to be estimated, and thus we have:

$$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk} \qquad (4)$$

where β_{pqk} (q=0,1,…,Q_p) are the level 2 coefficients, X_{qjk} is a level 2 predictor and r_{pjk} is a random effect. It is assumed that for each unit j the vector $(r_{0jk}, r_{1jk}, \ldots, r_{pjk})$' is distributed according to a normal distribution in which each element has an average of zero and a covariance matrix T_π with a maximum dimension of (P+1)x(P+1). Each of the level 2 coefficients, β_{pqk}, are converted into the variables to be explained at level 3 (school):

$$\beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} + u_{pqk} \qquad (5)$$

where γ_{pqs} (s=0,1,…,S_{pq}) are the level 3 coefficients, W_{sk} is a level 3 predictor and u_{pqk} is a level 3 random effect. It is assumed that the vector of random effects is distributed as a normal distribution in which each element has a mean of zero and a covariance matrix T_β with the following maximum dimension: $\sum_{p=0}^{P}(Q_p + 1) \times \sum_{p=0}^{P}(Q_p + 1)$.

5. Empirical results

[9] Bryk and Raudenbusch (1988) recommend the use of this type of general model when analysing the effects of schools on educational outcomes. There exist multiple applications of this methodology to the educational context. Among these are Willms (2006), Somers *et al*. (2004) and Mancebón *et al*. (2012), the last of these being applied to Spanish data from PISA 2006.

This section presents the principal results obtained from the empirical analysis performed. Firstly, the estimations obtained from the application of the PSM are commented upon. Next, we expound the principal contributions to these estimations offered by the application of the HLM.

5.1. PSM Results

5.1.1. Estimation of the Propensity Score Model

The strategy for estimating the PSM takes concrete shape, as explained in the previous section, by finding a group of students from PrSPS which is comparable with the students who attend a PuS in all those variables which can potentially condition the choice of school and the scores obtained of good marks in the competences evaluated in the ED. The selection equation must first be estimated, that is to say, the equation which permits the propensity score (ps) to be predicted must be constructed and, secondly, the sample of pupils belonging to the TG and CG in this indicator must be balanced. The estimation of the selection equation is of decisive importance, since it affects both balance on propensity scores and the final estimate of the treatment effect. A crucial point in the specification of this equation is that only the variables which affect both school choice and academic performance must be included. In addition, only those regressors which are potential predictors of educational outcomes and which occur prior to the choice of school (or were stable between the time of the choice of school and the time of the outcome assessment) should be included as explanatory variables (Caliendo and Kopeining, 2008).

Econometric literature offers various methods of estimation of the conditional probability of receiving a treatment (in our case, of attending a PrSPS): logistic regression, probit model, linear-probability function and discriminant analysis (Murname and Willett, 2011). In our study we have opted to use a logistic regression model, and specifically a generalized boosted model (GBM). The key feature and advantage of this model is that the analyst does not need to specify functional forms of the predictor variables. In addition, GBM permits nonlinear and interaction effects to be captured (McCaffrey, 2004). Finally, the data-adaptive algorithm on which this method is grounded leads to estimations of the ps that balance the observable covariates of the TG and CG[10]. For this reason GBM constitutes a highly suitable model to be used in the context of the PSM (Chowa *et al*., 2013).

[10] This is due to the fact that the adjustment it supplies is that which minimises the average standardized absolute mean difference (ASAM) between the individuals from the TG and CG.

A very important issue that should not be overlooked in interpreting the results of a GBM is that it does not provide estimated regression coefficients such as βs. Instead it provides *influence*, which is the percentage of log likelihood explained by each input variable[11].

Bearing in mind these premises, the specification of the GBM which was applied in our study included in the estimation those variables from the databases which, in the light of the previous empirical evidence regarding the determinants of school choice and the determinants of school results, could simultaneously affect the choice of PrSPS and academic performance. That is to say, when specifying the selection equation no consideration is taken of either the variables which can potentially contribute to explaining the differences in the cognitive competences evaluated in the diagnostic test, but which do not influence the choice of school (study habits, for example), nor those which could be determinants of that choice but do not influence the competences cited (the distance to the centre, for example).

Table 2 presents the results of the estimation of the selection equation[12]. It can be observed that the variables which capture the greatest degree of influence in the probability of attending a PrSPS are the years of study of the mother and father (16 and 21%, respectively), followed by the variables which proxy the degree of possessions in the household (number of TVs, PCs, video game consoles, MP4s, study room). The influence of the employment of the parents is also important. The dummies which approximate the employment of the mother account for 5.7% and those of the father 10.6%. Although the R^2 obtained is low, in these models the percentage of correct predictions of the model estimated is more important; in our case this reaches practically 70%, which the literature considers to be a fairly high degree of reliability. The final part of the table shows various parameters used in the estimation of the GBM models.

Figures 1 and 2a show the distribution of the predictions of the propensity scores estimated for the individuals from PuS and PrSPS. It can be clearly observed, in both the boxplot and the distribution graph, that there is a very broad area of common support. In other words, individuals in the TG have individuals in the CG with whom they can be compared, as their ps scores are the same.

Table 2: Results of the GBM (Dependent variable P (W=1)

Variable	Influence
JobMum2	1.76
JobMum3	1.23
JobMum4	2.73
JobDad2	2.84

[11] See Guo and Fraser (2010), page 144.

[12] In the estimations those individuals with data missing from the variables have been eliminated, following a case-wise deletion procedure.

JobDad3	6.76
JobDad4	1.01
EducationMum	16.02
EducationDad	21.08
ZoneGeo1	0.98
ZoneGeo2	0.76
ZoneGeo3	2.08
ZoneGeo4	2.07
ZoneGeo5	0.57
NumBooks	1.91
Room	5.99
NumTVs	6.57
NumPCs	4.51
NumTvPay	3.58
NumVideoGames	9.36
NumMP4	8.20
Best num iterations	16453.00
Train R^2	0.084
Test R^2	0.045
% correct prediction	68.4%
Train fraction	0.5
Bag	0.5
Shrinkage factor	0.0005
Distribution	Logistic
Max num interactions	4
Max num iterations	20000
Seed	0

Figure 1. Boxplot ps scores

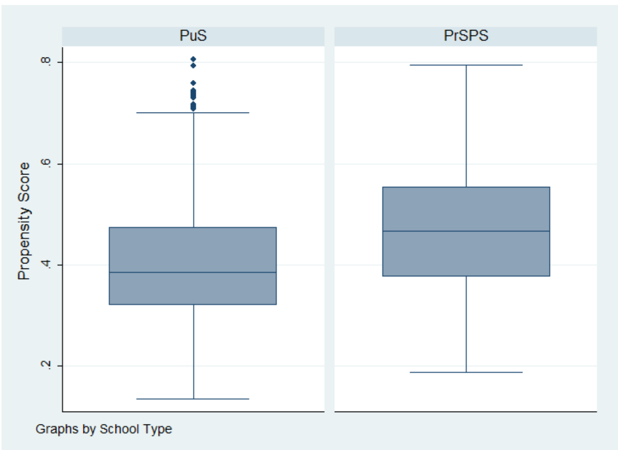
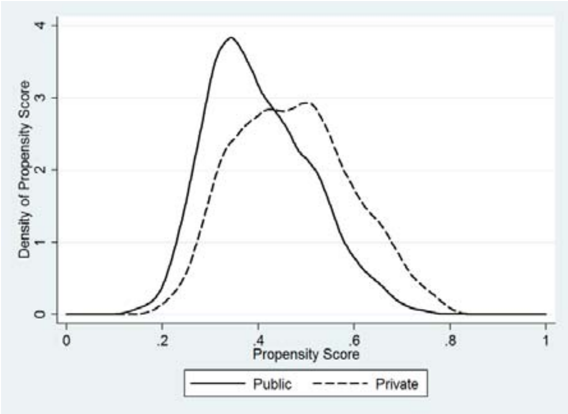
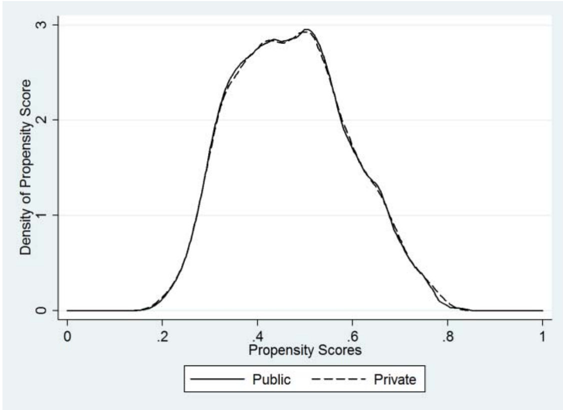


Figure 2. Propensity score distribution by school type

a. Full sample



b. Matched sample



5.1.2. Matching and resampling estimation

After estimating the ps the matching process is then undertaken. Various algorithms can be found in the literature regarding the performance of this process: greedy matching, optimal matching and fine balance (Guo and Fraser, 2010). The present study uses the first of these, which may be applied via a range of variants (Smith and Tood, 2005). The two most commonly used algorithms are nearest neighbour matching (hereinafter NNM), which allows for diverse variants, and methods based on kernel functions (hereinafter KM). The first of these matches each individual from the TG with that from the CG having the most similar ps value. KM constructs matches using all the individuals in the potential control sample in such a way that it gathers more information from those who are closer matches and less from distant observations. In so doing, KM uses comparatively more information than other matching algorithms (Guo and Fraser, 2010, chapter 7). The present study applied these two algorithms, as well as several of the options permitted by NNM (with and without replacement, with caliper and without caliper, 1 to 1, 1 to 2 and 1 to 3). The KM, in turn, was applied with different bandwidths. This was done in an attempt to test the sensitivity of the matching to the different estimation methods.

The analysis led us to opt for the Epanechnikov kernel type KM with a bandwidth of 0.03, since it best equates the individuals from the TG and the CG. The sample was only reduced by 9 individuals from the CG who were not paired with any individual from the TG. The remaining individuals from the CG were weighted on the basis of the number of times that they were matched with individuals from the TG. These weights were required to be used in the subsequent analyses.

Figure 2b displays the distribution of the ps in the original sample and in the matched subsample. In the latter, it is observable that there is an almost perfect overlap in the distribution for public schools and PrSPS. Figure 3 shows the matching used between students from PuS and PrSPS[13].

Figure 3. Propensity score matching blocks

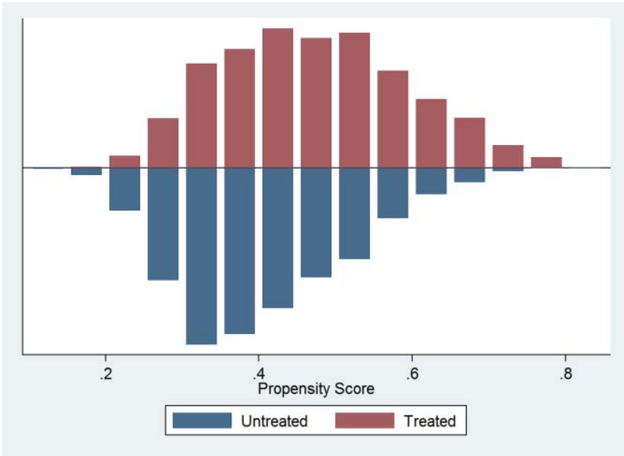


Table 3 compares the scores in Science and Foreign Language (English) for the unmatched and matched samples (ATT). It shows that in the two samples PrSPS have a positive and statistically significant effect on both Sciences and Foreign Language (English) scores for students evaluated in ED 2010.

Table 3. Two-group t-test

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Sciences	Unmatched	526.27	501.97	24.30	2.43	9.99
	ATT	526.27	519.15	7.11	2.72	2.62
Foreign Language (English)	Unmatched	531.50	499.18	32.32	2.40	13.49
	ATT	531.50	518.97	12.53	2.68	4.68

5.1.3. Sensitivity analysis: selection on unobservables

The matching method approach is based on the conditional independence (CIA), which states that the researcher should observe all variables simultaneously influencing the participation

[13] The Annex presents other results obtained from the implementation of the PSM. In particular, Table A.1. shows the differences in averages in the ps and the covariates for the complete sample and the paired sample, and similarly the reduction in the bias achieved. Figure A.1. shows graphically the pre- and post-matching bias for each of the variables. Figure A.2. depicts the distribution of the variables used in the PSM by type of school for the complete sample (figures on the left) and the matched sample (figures on the right).

decision and outcome variables. If there unobserved variables, which simultaneously affect assignment into treatment and the outcome variable, a “hidden bias” might arise to which matching estimators are not robust.

The idea is scrutinize the estimated management effects to see whether they are sensitive to selection bias due to correlation between unobserved factors and a person treatment status. Although we have many background variables, treatment effect in nonexperimental studies may be contaminated with selection bias due to unobserved factors like motivation, ability, preferences, etc. The purpose of sensitivity analysis is to ask whether inferences about the management effects may be altered by factors not observed in the data (Aakvik, 2001; Hujer et al., 2004; Caliendo et al., 2005; Altonji *et al*., 2008).

In order to estimate the extent to which such "selection on unobservables" may bias our qualitative and quantitative inferences about the effects of the typology of the school, we present the results from using Rosenbaum’s (2002) procedure for bounding the treatment effect estimates in Table 4.

There we give the results of the p-value from Wilcoxon sign-rank tests for the averaged treatment effect on the treated while setting the level of hidden bias to a certain value γ, which reflects our assumption about unmeasured heterogeneity or endogeneity in treatment assignment expressed in terms of the odds ratio of differential treatment assignment due to an unobserved covariate[14]. At each γ we calculate a hypothetical significance level “p-value critical”, which represents the bound on the significance level of the treatment effect in the case of endogenous self-selection into treatment status.

Table 4 shows that robustness to hidden bias varies across the two variables. The finding of a positive effect of management on Science is the least robust to the possible presence of selection bias. The critical level of $\boldsymbol{\gamma}$ at which we would have to question our conclusion of a positive effect is between 1.12 and 1.15, i.e. is attained if an unobserved covariate caused the odds ratio of treatment assignment to differ between treatment and control cases by a factor of about 1.15. For Foreign Language model it would require a hidden bias of $\boldsymbol{\gamma}$ between 1.30 and 1.33 to render spurious the conclusion of a positive benefit effect on publicly subsidised private school.

[14] For a mathematical demonstration, see DiPrete and Gangl (2004).

Table 4 – Rosenbaum Bounds for Management Benefit Treatment Effects

Variable	γ*	P_value critical
Sciences	1	9.7e-07
(N=2879 matched pairs)	1.03	.000023
	1.06	.000329
	1.09	.002888
	1.12	.016511
	1.15	**.064211**
	1.18	.177506
	1.21	.365214
	1.24	.587657
Foreign Language (English)	1	2.8e-15
(N = 2879 matched pairs)	1.03	5.1e-13
	1.06	5.1e-11
	1.09	3.0e-09
	1.12	1.0e-07
	1.15	2.3e-06
	1.18	.000033
	1.21	.00032
	1.24	.002172
	1.27	.010623
	1.30	.038508
	1.33	**.10649**

* gamma: log odds of differential assignment due to unobserved factors

A critical value of 1.15 suggests that individuals with the same X-vector differ in their odds of participation by a factor of 1.15 or 15%. It is important to note that these are worst-case scenarios. Hence, a critical value of 1.15 does not mean that unobserved heterogeneity exists and that there is no effect of treatment on the outcome variable. This result only states that the confidence interval for the effect would include zero if an unobserved variable caused the odds ratio of treatment assignment to differ between treatment and comparison groups by 1.15. Additionally, this variable's effect on the outcome would have to be so strong that it almost perfectly determines the outcome in each pair of matched cases in the data. However, even if there is unobserved heterogeneity to a degree of 15% in the group of Science, inference about the treatment effect would not be changed.

To repeat, the Rosenbaum bounds are in this sense a “worst-case” scenario. Nonetheless, they convey important information about the level of uncertainty contained in matching estimators by showing just how large the influence of a confounding variable must be to undermine the conclusions of a matching analysis. As a conclusion, this analysis allows confirmed that only a large amount of unobserved heterogeneity would alter the inference about the estimated effects. Even so, always is necessary has some caution when interpreting the results.

5.2. Postmatching analysis: HLM results

The analysis performed in the two previous sections has permitted the delimitation of a subsample unaffected by the problem of selection bias which affected the initial sample. From this, it has been possible to perform an initial estimation of the impact of attending a PrSPS upon the educational achievements of pupils (Table 3). This estimation has taken into consideration the educational results of the individuals belonging to the TG and CG, once the effect of the covariates which jointly determine PrSPS attendance and children's cognitive development has been discounted. However, students' cognitive skills are also influenced by other variables which do not affect their participation in the treatment and which, as a result, have not been included in the estimation of the ps. This is why a more precise estimation of the effect than that offered by the comparison of the scores for the unmatched and matched samples (ATT) would require a refined approach which took into account the variables which may influence the evaluated skills but have not been included in the PSM. Consequently, a post-matching analysis was undertaken, the results of which are shown below.

As previously explained, the regression model best fitting the data supplied by the ED 2010 is a hierarchical lineal model (HLM). This type of model, as we said earlier, permits the identification of the proportion of the total variance of the outcomes obtained by students which may be attributed to the different estimation levels. In our case, level 1 is represented by the student, level 2 is represented by the class and level three is represented by the school.

The appropriateness of applying an HLM is justified empirically by the intra-class correlation (ICC) values of the null model of Science and Foreign Language (English) performance (the two being the dependent variables of the regression). Tables 5 and 6 show, respectively, these values for an HLM at two levels and three levels[15].

As observed, in the model at three levels the ICC for the class level is 12.3% for Sciences and 4.0% for Foreign Language (English). For level 3 (school) these values are 18.9% and 32.9% respectively. These results show that the class level explains a minimal percentage of the variance of the results in Foreign Language (English), although this level does explain a higher percentage of the results in Sciences. Consequently, it was decided to apply a two-level model for achievement in a Foreign Language (English) and a three-level model for Sciences.

The models were estimated by imposing fixed effects on the parameters (with the exception of the independent term), rejecting the null hypothesis that there existed statistically significant random effects. The dependent variables in the regression are the marks achieved by fourth grade pupils from Aragon in the ED tests for 2010 in Science and Foreign Language (English). The predictors of the regression and the results of the HLM are listed in Table 7, grouped by

[15] The intra-class correlation (ICC) is the proportion of the total variance explained by the differences between classes (level 2) and between schools (level 3). If the ICC were zero, a hierarchical model would not be necessary, since in this case the total variance of the scores would not be explained by the differences existing between students attending different classes or schools.

levels. The left-hand side of the table presents the results from the two-level model for Foreign Language (English). The right-hand side offers the results for the three-level model, more adequate for the estimation of the determinants of the outcomes in Science.

The predictor which has greatest interest in our study is attendance at a PrSPS. It can be observed that this effect is positive and significant for Science, while for Foreign Language (English) it is not statistically significant. The coefficient estimated in Science is 22 points, which indicates that a pupil whose remaining characteristics are identical has a score in this competence 22 points higher in a PrSPS than in a PuS.

With regard to the effects of the covariates included in the regression the following can be observed. The size of the municipality in which the school is located and attendance at a school in the city of Zaragoza (capital of the Autonomous Community of Aragon) have a significant effect upon competence in English. The net effect of attendance at a school situated in the city of Zaragoza is +15.16[16].This result may be explained by the greater effort which in recent years has been made in bilingual programmes, which have been concentrated especially in the city of Zaragoza.

The results also underline the non-existence of peer effects for fourth year primary pupils[17]. Only the average years of study of mothers at school level show positive and significant effects upon competence in English. The variables at pupil level show results common in the literature regarding the determinants of educational performance. Girls obtain better results in the competence of Foreign Language (English), while boys stand out in Sciences. The occupation and level of education of the parents have the expected effect. Greater occupational and educational level (in this case the relevant factor is the mother) mean better results at school in both competences. In the case of the variable which approximates the effect of immigration (residence in Spain of over 5 years) the effect is that expected in the scientific competences (positive and significant), while it is not significant in competences in English. Another variable which presents the expected effect is the number of books existing in the household: homes which state they have over 100 books affect positively the acquisition of educational competencies. To this result must be added a positive and statistically significant effect of variable “use of books by the pupil”: pupils who state they frequently read books present better academic results.

---

[16] This value has been calculated as: (Population of Zaragoza × $\hat{\beta}$ (Municipality size)+ $\hat{\beta}$ (Zaragoza city).

[17] These effects have been shown to be important for high school students. See, for instance, Schneeweis and Winter-Ebmer (2007) or Schindler (2007), both of them using PISA data.

Table 5. HLM regression: random effects (3-levels)

	Science		Foreign language (English)	
	Null model	Complete model	Null model	Complete model
Schools	1805.91	1639.54	3128.13	2061.81
Classes	1169.27	949.38	379.75	439.40
Students	6554.70	4528.35	5993.29	4010.39
Total	9529.88	7117.27	9501.16	6511.61
ICC(school)	18.9%		32.9%	
ICC(class)	12.3%		4.0%	
% of total variance explained by variables		25.3%		31.5%
% of level 1 (students) variance explained by variables		30.9%		33.1%
% of level 2 (classes) variance explained by variables		18.8%		-15.7%
% of level 3 (schools) variance explained by variables		9.2%		34.1%

Table 6. HLM regression: random effects (2 levels)

	Science		English	
	Null model	Complete model	Null model	Complete model
Schools	2661.88	2393.96	3373.56	2172.77
Students	7470.67	5354.93	6466.98	4466.59
Total	10132.55	7748.90	9840.55	6639.36
ICC (schools)	26.3%		34.3%	
% of total variance explained by variables		23.5%		32.5%
% of level 1 (students) variance explained by variables		28.3%		30.9%
% of level 2 (schools) variance explained by variables		10.1%		35.6%

Of the different items used in the ED to approximate family wealth only the number of televisions in the home demonstrates a significant influence on results (negative influence).

The effect shown by the time of dedication to school tasks out of school negatively influences performance. Children who declare they dedicate over two hours daily to these tasks display worse results than those who dedicate less than two hours. Homework does not appear to constitute a good strategy for stimulating the capacities of 10-year-old children. A–possible interpretation of this effect could be that children who dedicate more time to schoolwork outside the classroom are those who have greater learning difficulties. An identical interpretation is merited by the results which present the variables of “help with study” and “revision of tasks by parents or private teachers”.

Attitude, approximated by the variable “I do the tasks”, shows a positive effect in both competences but is not significant in English. The case of aptitude, approximated by the variable “my homework is correct when we correct it in class”, displays in turn a positive effect in the results.

Additionally, the regression incorporates information regarding three factors extracted from a principal components analysis applied to the data concerning the school environment. These data proceed from the answers supplied by the pupils evaluated. This analysis permitted us to identify three factors that we term RELCEN, SELFCONF and PERCAMB.The first factor contains information regarding the evaluation the child makes of his or her school (if the centre has cultural and sports activities, if the pupil uses the school’s library, if the installations are well cared for, etc.). Factor 2 synthesizes the information offered by variables related to the self-perception of the pupils’ academic capacity (if pupils understand what they read, if they express themselves well, if they write correctly, if they are good at languages, etc.). Factor 3, finally, reflects the subjective perceptions of the school environment (if there is a good atmosphere in the pupil’s class, if his/her classmates help each other, if the pupil has a good relationship with his/her teachers, if the teachers stimulate their pupils, etc).

The results vary depending on the competence evaluated. While in Foreign Language (English) factor 1 presents a positive and significant effect, in Sciences the effect is negative but not significant. The other two factors influence in a statistically significant way the two competencies: self-confidence (factor 2) positively, while the perception of the school atmosphere (factor 3) does so negatively.

Table 7. Estimation of fixed effects with robust standard errors in the HLM

Two-level model Foreign Language (English)		Three-level model Sciences	
School variables (Level 2)		**School variables (Level 3)**	
Variable	**Coefficient**	**Variable**	**Coefficient**
Intercept	976.69*** (192.4)	Intercept	517.63*** (172.9)
SCHTYPE (publicly subsidised private primary school)	7.57 (7.6)	SCHTYPE (publicly subsidised private primary school)	22.6*** (7.8)
TERUEL province(1) *Ref: Huesca province*	11.97 (17.2)	TERUEL province(1) *Ref: Huesca province*	2.77 (12.0)
ZARAGOZA province(1) *Ref: Huesca province*	12.27 (14.9)	ZARAGOZA province(1) *Ref: Huesca province*	-10.44 (16.1)
Municipality size (number inhabitants)	0.00*** (0.0)	Municipality size (number inhabitants)	0.00 (0.00)
ZARAGOZA city	-813.51*** (315.8)	ZARAGOZA city	-41.17 (290.4)
		Class variables (Level 2)	
PCTGIRLS (Percentage girls at school)	7.52 (36.1)	PCTGIRLS (Percentage girls in class)	-32.34 (43.9)
PCTREP (Percentage repeaters at school)	54.86 (74.3)	PCTREP (Percentage repeaters in class)	6.8 (53.9)
PCTMORE5YEARS (Percentage of pupils living over 5 years in Spain at school at school)	51.11 (43.2)	PCTMORE5YEARS (Percentage of pupils living over than 5 years in Spain at school in class)	-21.14 (45.3)
PCTJOBMUM1 (Percentage of students whose mother is White Collar High Skilled at school)	49.56 (49.2)	PCTJOBMUM1 (Percentage of students whose mother is white collar highly skilled in class)	10.62 (23.6)
PCTJOBMUM2 (Percentage of students whose mother is white collar low skilled at school)	-30.19 (37.6)	PCTJOBMUM2 (Percentage of students whose mother is white collar low skilled at class)	-4.01 (25.8)
PCTJOBMUM3 (Percentage of students whose mother is blue collar highly skilled at school)	-86.35 (90.0)	PCTJOBMUM3 (Percentage of students whose mother is blue collar highly skilled in class)	-85.6 (64.4)
AVEDUCATIONMUM (Average number of years of mothers education at school)	8.43** (3.9)	AVEDUCATIONMUM (Average number of years of mothers' education ins class)	1.19 (3.2)
Student variables (Level 1)		**Student variables (Level 1)**	
GENDER (Female =1)	20.3*** (2.4)	GENDER (Female =1)	-11.19*** (2.4)
REPEATER (If student has repeated any year = 1)	-39.73*** (6.1)	REPEATER (If student has repeated any year = 1)	-28.72*** (6.2)
JOBMUM1 (Student's mother is white collar highly skilled)	11.13*** (3.9)	JOBMUM1 (Student's mother is white collar highly skilled)	11.52*** (4.3)
JOBMUM2 (Student's mother is white collar low skilled)	0.88 (3.2)	JOBMUM2 (Student's mother is white collar low skilled)	1.15 (3.5)
JOBMUM3 (Student's mother is blue collar highly skilled)	2.05 (9.0)	JOBMUM3 (Student's mother is blue collar highly skilled)	0.2 (7)

Table 7. Estimation of fixed effects with robust standard errors in the HLM (cont)

Two-level model Foreign Language (English)		**Three-level model Sciences**	
Student variables (Level 1)		**Student variables (Level 1)**	
Variable	**Coefficient**	**Variable**	**Coefficient**
JOBDAD1 (Pupil's father is white collar highly skilled)	9.19* (5.4)	JOBDAD1 (Pupil's father is white collar highly skilled)	1.6 (6.5)
JOBDAD2 (Student's father is white collar low skilled)	1.71 (5.6)	JOBDAD2 (Student's father is white collar low skilled)	-2.01 (6.6)
JOBDAD3 (Student's father is blue collar highly skilled)	3.4 (5.1)	JOBDAD3 (Student's father is blue collar highly skilled)	1.24 (6.2)
EDUCATIONMUM (Mother's years of education)	1.42*** (0.3)	EDUCATIONMUM (Mother's years of education)	1.47*** (0.3)
MORE5YEARS (Over 5 years living in Spain)	-8.94 (6.1)	MORE5YEARS (Over 5 years living in Spain)	18.74*** (6.6)
NUMBOOKS (Over 100 books at home)	7.22** (3.1)	NUMBOOKS (Over 100 books at home)	13.26*** (2.8)
USEBOOKS (Uses books frequently)	13.75*** (3.0)	USEBOOKS (Uses books frequently)	13.76*** (3.2)
NUMTVs (Number of TVs at home)	-5.24*** (1.6)	NUMTVs (Number of TVs at home)	-6.42*** (1.7)
STUDTIM1 (2 hours of homework every day)	-3.54 (3.4)	STUDTIM1 (2 hours studying every day)	-10.48*** (3.5)
STUDTIM2 (Over 2 hours studying every day)	-11.68*** (2.6)	STUDTIM2 (Over 2 hours studying every day)	-12.77*** (2.6)
NEEDHELP (Student needs help in homework)	-23.15*** (3.1)	NEEDHELP (Student needs help in homework)	-27.68*** (3.5)
REVPAR1 (Parents check diary but not homework)	-2.87 (4.8)	REVPAR1 (Parents check diary but not homework)	-11.38** (5.2)
REVPAR2 (Parents check homework but not diary)	-0.59 (3.8)	REVPAR2 (Parents check homework but not diary)	-10.11** (4.2)
REVPAR3 (Parents check both homework and diary)	-10.92*** (3.1)	REVPAR3 (Parents check both homework and diary)	-17.28*** (3.6)
REVTEACHER (Private tutoring)	-18.3*** (4.7)	REVTEACHER (Private tutoring)	-19.97*** (4.6)
ATTITUDE (Student always finishes homework)	9.87 (6.6)	ATTITUDE (Student always finishes homework)	17.51** (7.4)
APTITUDE (Student answers homework correctly)	11.79*** (4.5)	APTITUDE (Student answer homework correctly)	17.49*** (4.2)
RELCEN (Factor 1)	4.67*** (1.0)	RELCEN (Factor 1)	-0.12 (1.4)
SELFCONF (Factor 2)	21.89*** (1.4)	SELFCONF (Factor 2)	18.87*** (1.5)
PERCAMB (Factor 3)	-4.09*** (1.3)	PERCAMB (Factor 3)	-8.27*** (1.3)

(1) This variable proxies the location of the school in the region of Aragon. This region consists of three provinces: Zaragoza, Teruel and Huesca (this last being the category of reference)

6. Conclusions

The analysis performed in this study has underlined the existence of a certain advantage of the publicly subsidised private school (PrSPS) of Aragon compared to the public schools (PuS) in some educational competencies, in particular those related to the dominance of abilities in solving problems and questions related to scientific skills. Even having taken into consideration the differences in the sociocultural background of the pupils attending the two types of school (differences which favour the PrSPS), attendance at a PrSPS favours the obtaining of better results by pupils in the Diagnostic Evaluation undertaken in 2010 in the Autonomous Community of Aragon.

In the case of competencies in Foreign Language (English), the second competence evaluated in the 2010 edition of the ED, the study performed does not permit the establishment of statistically significant relationships between school type and the skills acquired by Aragonese pupils in that cognitive dimension.

These results simply evidence the difficulty of establishing a clear causal effect between the school management model and academic achievements. In effect, we began our study by underlining the lack of consensus existing in the literature on the differential quality of the PuS and PrSPS, finding studies with contradictory conclusions. Our study is a new contribution to this field; it has shown that in certain educational competencies (Sciences) PrSPS present advantages, while in others (English) the contributions of this type of school are similar to those of PuS. Our results, in line with those obtained by Zimmer *et al*. (2012) and Imberman (2007), point towards a possible specialisation of primary schools in certain cognitive skills.

Acknowledgments

The authors received financial support from the Spanish Government, Ministerio de Economía y Competitividad ry of Science and Innovation, Project (EDU2013-42480- R). Research support from Fundación Ramón Areces is also acknowledged by Mauro Mediavilla y Domingo Pérez-Ximénez de Embún.

References:

Aakvik, A. (2001): “Bounding a matching estimator: the case of a Norwegian training program”, *Oxford Bulletin of Economics and Statistics*, 63 (1), 0305-9049.

Allen, R. and Vignoles, A. (2015): “Can school competition improve standards?. The case of faith schools in England”, *Empirical Economics*, DOI 10.1007/s00181-015-0949-4.

Altonji, J. G., Elder, T.E. and Taber, C. R. (2008): "Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization", *American Economic Review*, 98 (2), pp. 345-350.

Anand, P., Mizala, A. and 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*, 28, (3), pp. 370-381.

Angrist, J. D. and Pischke, J. S. (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. USA.

Batlagi, D.H. (1999): *Econometrics*, 2nd edition, Springer: Berlin.

Bedi, A.S., & Garg, A. (2000). The effectiveness of private versus public schools: the case of Indonesia. *Journal of Development Economics*, 61, 463- 494.

Bernal, J.L. (2005): "Parental choice, social class and market forces: the consequences of privatization of public services in education", *Journal of Education Policy*, 20(6), pp. 779-792.

Bettinger, E.P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review*, 24, 133-147.

Betts, J.R., Rice, L. A., Zau, A.C., Tang, E. and Koedel, C.R. (2006): *Does school choice work?. Effects on student integration and achievement*, San Francisco CA. Public Policy Institute of California.

Bifulco, R. and Ladd, H.F. (2006): "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina", *Education Finance and Policy*, 1(1), pp. 50-90.

Booker, K, Gilpatric, S, Gronberg, T.J. and Jansen, D.W. (2007): "The impact of charter school attendance on student performance", *Journal of Public Economics*, 91, 849–876

Brewer, D.J. and McEwan, P.J. (2010): *Economics of Education*. Elsevier. Academic Press.

Bryk A.S. and Raudenbusch S.W. (1988): "Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Msodel", American Journal of Education, 97 (1), pp. 65-108.

Burgess, S. and Briggs, A. (2010): "School assignment, school choice and social mobility". *Economics of Education Review*, 29, pp. 639–649.

Calero, J. and Escardibul, J.O. (2007): "Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003". *Hacienda Pública Española. Revista de Economía*, 183 (4), pp. 33-66.

Caliendo, M., Hujer, R. and Thomsem, S. (2005): "The Employment Effects of Job Creation Schemes in Germany: A Microeconometric Evaluatio". *IZA Discussion Paper*, 1512.

Caliendo, M. and Kopeinig, S. (2008): "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, 22(1), pp. 31-72.

Campbell, D. T. (1957): "Factors relevant to the validity of experiments in social settings", *Psychological Bulletin*, 54(4), pp. 297-312.

Chowa, G., Masa, R. D., Wretman, C.J. and Ansong, D. (2013): "The impact of household possessions on youth's academic achievement in the Ghana Youthsave experiment: A propensity score analysis", *Economics of Education Review*, 33, pp. 69-81

Chudgar and Quin (2012): "Relationship between private schooling and achievement: Results from rural and urban India", *Economics of Education Review*, 31 (4), pp. 376-390.

Coleman, J., Hoffer, T. and Kilgore, S. (1982): *Secondary school achievement. Public, catholic and private schools compared*. New York: Basic Books, Inc. Publishers.

Cordero, J.M., Crespo, E. and Pedraja, F. (2013): "Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España", Revista de Educación, 362, en prensa. DOI:10.4438/1988-592X-RE-2011-362-161.

Crespo, E. and Santín, D. (2013): "Does school ownership matter? An unbiased efficiency comparison for regions of Spain", *Journal of Productivity Analysis*, DOI 10.1007/s11123-013-0338.

Davies, S. (2013): "Are there Catholic school effects in Ontario, Canada?", *European Sociological Review*, 29 (4), pp. 871–883.

DiPrete, T. and Gangl, M. (2004): "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments", *WZB Discussion Paper*, SP I 2004-101

Doncel, L.M., Sainz, J. and Sanz, I. (2012): "An estimation of the advantage of charter over public schools", *Kyklos*, 65 (4), pp. 442-463.

Escardíbul, J.O., and Villarroya A. (2009): "The inequalities in school choice in Spain in accordance to PISA data". *Journal of Education Policy*, 24, (6), pp. 673-695.

Figel, J. (2010): *Informe Educación y Atención a la Primera Infancia en Europa: un medio para reducir las desigualdades sociales y culturales*. Agencia Ejecutiva en el ámbito Educativo, Audiovisual y Cultural (EACEA P9 Eurydice). Ed. Ministerio de Educación y Ciencia. Also available at http://www.eurydice.org.

Gallego, F.A. and Hernando, A. (2010): "School Choice in Chile: Looking at the Demand Side", Working Document no. 356, Universidad Pontificia Católica de Chile

Greene, K.V. and Kang, B. (2004): "The effect of public and private competition on high school outputs in New York State". *Economics of Education Review*, 23, pp. 497-506.

Gronberg, T.J. and Jansen, D. (2001): *Navigating newly chartered waters. An analysis of charter school performance*. Austin, TX. Texas Public Policy Foundation.

Guo, S. and Fraser, M.W. (2010): *Propensity Score Analysis*. Statistical Methods and Applications. SAGE publications. London.

Hanushek, E.A., Kain, J.F., Rivkin, S.G. y Branch, F. (2007): "Charter school quality and parental decision making with school choice", *Journal of Public Economics*, 91, 823-848.

Hanushek, E., Machin, S. and Woessmann, L. (2011): *Handbook of the Economics of Education (4).* Elsevier. Amsterdam.

Heckman, J.J. (1979): "Sample section bias as a specification error", *Econometrica, Journal of the Econometric Society* (47), pp. 153-161.

Heckman, J. J. (2005): "The scientific model of causality*", Sociological Methodology*, 35, pp. 1–97.

Holland, P. (1986): "Statistics and causal inference (with discussion)", *Journal of the American Statistical Association*, 81, pp. 945-970.

Hox, J. (1995): *Applied Multilevel Analysis*. TT-Publikaties. Amsterdam.

Hsieh, C. T. and Urquiola, M. (2006): "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program", *Journal of Public Economics*, 90, (8–9), pp. 1477-1503.

Hujer, R., Caliendo, M. and Thomsen, S. (2004): “New Evidence on the Effects of Job Creation Schemes in Germany – A Matching Approach with Threefold Heterogeneity”, *Research in Economics*, 58 (4), pp. 257-302

Imberman, S. (2007): Achievement and behaviour in charter schools: drawing a more complete picture. Working Paper number 142. Center for the Study of Privatization in Education. New York.

Jepsen, C. (2003): “The effects of Catholic primary schools”, *Journal of Human Resources*, 38, 928-941.

Kim, Y.J. (2011): “Catholic schools or school quality? The effects of Catholic schools on labor market outcomes”, *Economics of Education Review*, 30 (3), pp. 546-558.

Lefebvre, P., Merrigan, P. and Verstraete, M. (2011): “Public subsidies to private schools do make a difference for achievement in mathematics: Longitudinal evidence from Canada”, *Economics of Education Review*, 30 (1), pp. 79-98.

Mancebón, M.J. and Muñíz, M.A. (2008): “Public High Schools in Spain. Disentangling managerial and program efficiencies”, *Journal of the Operational Research Society*, 59, pp. 892- 901.

Mancebón, M.J. and Pérez-Ximénez de Embún, D. (2011): “Equality of school choice: a study applied to the Spanish region of Aragon”, *Education Economics*, 22 (1), pp. 90-111.

Mancebón, M.J., Calero, J., Choi, A. and Pérez-Ximénez de Embún, D. (2012): “The Efficiency of Public and Publicly-Subsidized High Schools in Spain. Evidence from PISA-2006”, *Journal of the Operational Research Society*, 63, pp. 1516-1533.

McCaffrey, D.F., Ridgeway, G. and Morral, A.R. (2004): “Propensity score estimation with boosted regression for evaluating causal effects in observational studies”, *Psychological Methods*, 9, pp. 403-425.

McEwan, P.J. (2001): “The effectiveness of Public, Catholic and Non-Religious Private schools in Chile’s voucher system”, *Education Economics*, 9 (2), 103-128.

Morgan, S. L. and Winship, C. (2008): Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press. UK.

Murname, R.J. and Willett, J.B. (2011): *Methods matter*. Oxford University Press. New York.

Neyman, J.S. (1923): “Statistical problems in agricultural experiments”, *Journal of the Royal Statistical Society*, Series B, 2, pp. 107-180.

Perelman, S. and Santin, D. (2008): “Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results”. *Education Economics*, 19 (1), pp. 29-49.

Pfeffermann, D. and Landsman, V. (2011): “Are private schools better than public schools?. Appraisal for Ireland by methods of observational studies”, *The Annals of Applied Statistics*, 5(3), 1726-1751.

Reynolds, C.L. (2012): “Where to attend? Estimating the effects of beginning college at a two-year institution”, *Economics of Education Review*, 31 (4), pp. 345-362.

Rosenbaum, P. R. and Rubin, D.B. (1983): “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, 70, pp. 41-55.

Rosenbaum, P. R. (2002): *Observational Studies*. Springer, New York.

Rubin, D. B. (1974): “Estimating causal effects of treatments in randomized and non-randomized studies”, *Journal of Educational Psychology*, 66, pp. 688-701.

Rubin, D. B. (1990): "Bayesian inference for causal effects: the role of randomization", *Journal of Statistical Planning and Inference*, 25, pp. 279-292.

Rubin, D. B. (1978): "Formal models of statistical inference for causal effects: the role of randomization", *Annals of Statistics*, 6, pp. 34-58.

Salinas, J. and Santín, D. (2012): "Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en PISA 2006", *Revista de Educación*. 358, pp. 382-405.

Sass, T.R. (2006): "Charter schools and student achievement in Florida", *Journal of Education Finance and Policy*, 1, 91-122.

Schneeweis, N. and Winter-Ebmer, R. (2007): "Peer effects in Austrian schools", *Empirical Economics*, 32, pp. 387–409.

Schindler, B. (2007): "School composition effects in Denmark: quantile regression evidence from PISA 2000", *Empirical Economics*, 33, pp. 359–388.

Shadish, W. R., Cook, T.D. and Campbell, D. T. (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company. Boston.

Smith, H. L. and Tood, P. E. (2005): "Does matching overcome LaLonde's critique of non-experimental estimators?", *Journal of Econometrics*, 125, pp. 305-353.

Somers M.A., McEwan P.J. and Willms J.D. (2004): "How Effective Are Private Schools in Latin America?", *Comparative Education Review*, 48, (1), pp. 48-69.

Tamm, M. (2008): "Does money buy higher schooling?: Evidence from secondary school track choice in Germany", *Economics of Education Review*, 27(5), pp. 536-545.

Toma, E. y Zimmer, R. (2012): "Two decades of charter schools: Expectations, reality, and the future", *Economics of Education Review*, 31 (2), pp. 209-210.

Willms J.D. (2006): *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. UIS Working Paper, no. 5. UNESCO Institute for Statistics: Montreal.

Witte, J.F., Weimer, D. Shober, A. and Schlomer, P. (2007): "The performance of charter schools in Wisconsin", *Journal of Policy Analysis and Management*, 26, pp. 574-575.

Zimmer, R. Gill, B., Booker, K., Lavertu, S. and Witte, J. (2012): " Examining charter student achievement effects across seven states", *Economics of Education Review*, 31 (2), 213-224.

Appendix:

Table A.1. Average differences based on school type for the variables in the pre- and post-matching samples and bias reduction

Variable		Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
Propensity score	Unmatched	0.47	0.40	60.8		24.87	0.00
	Matched	0.47	0.47	0.9	98.5	0.32	0.75
JobMum1	Unmatched	0.37	0.24	29.7		12.15	0.00
	Matched	0.37	0.37	0.7	97.6	0.26	0.80
JobMum2	Unmatched	0.39	0.42	-5.4		-2.20	0.03
	Matched	0.39	0.40	-2.0	62.5	-0.77	0.44
JobMum3	Unmatched	0.03	0.04	-7.6		-3.06	0.00
	Matched	0.03	0.03	-1.7	77.5	-0.71	0.48
JobMum4	Unmatched	0.21	0.30	-21.7		-8.74	0.00
	Matched	0.21	0.20	2.3	89.6	0.93	0.36
JobDad1	Unmatched	0.49	0.31	36.7		14.98	0.00
	Matched	0.49	0.50	-1.2	96.8	-0.44	0.66
JobDad2	Unmatched	0.23	0.26	-6.5		-2.61	0.01
	Matched	0.23	0.24	-2.0	68.8	-0.77	0.44
JobDad3	Unmatched	0.23	0.35	-27.5		-11.07	0.00
	Matched	0.23	0.21	3.4	87.8	1.38	0.17
JobDad4	Unmatched	0.05	0.07	-10.7		-4.29	0.00
	Matched	0.05	0.05	-0.3	97.2	-0.13	0.90
YearsMum	Unmatched	12.34	10.78	33.9		13.66	0.00
	Matched	12.34	12.49	-3.2	90.7	-1.26	0.21
YearsDad	Unmatched	12.34	10.78	33.5		13.53	0.00
	Matched	12.34	12.47	-2.9	91.4	-1.13	0.26
ZonaGeo1	Unmatched	0.91	0.84	21.0		8.36	0.00
	Matched	0.91	0.91	0.9	95.8	0.38	0.70
ZonaGeo2	Unmatched	0.00	0.01	-9.3		-3.65	0.00
	Matched	0.00	0.00	1.0	89.4	0.60	0.55
ZonaGeo3	Unmatched	0.01	0.01	-1.0		-0.42	0.68
	Matched	0.01	0.00	2.1	-101.6	0.89	0.37
ZonaGeo4	Unmatched	0.03	0.06	-12.8		-5.11	0.00
	Matched	0.03	0.04	-2.5	80.9	-1.06	0.29
ZonaGeo5	Unmatched	0.04	0.06	-10.9		-4.36	0.00
	Matched	0.04	0.04	0.2	97.8	0.10	0.92
ZonaGeo6	Unmatched	0.01	0.02	-8.1		-3.19	0.00
	Matched	0.01	0.01	-0.7	91.8	-0.30	0.76
NumBooks	Unmatched	0.60	0.50	18.7		7.56	0.00
	Matched	0.60	0.61	-2.0	89.4	-0.76	0.45
Own bedroom	Unmatched	0.96	0.94	8.1		3.23	0.00
	Matched	0.96	0.96	-1.2	84.9	-0.51	0.61
Internet	Unmatched	0.88	0.84	12.7		5.06	0.00
	Matched	0.88	0.89	-3.4	73.4	-1.40	0.16
NumTVs	Unmatched	2.15	2.08	9.8		3.96	0.00
	Matched	2.15	2.15	0.8	92.1	0.30	0.77
NumPCs	Unmatched	1.63	1.49	17.2		6.97	0.00
	Matched	1.63	1.66	-4.1	76.2	-1.56	0.12
NumTvPag	Unmatched	0.46	0.43	4.5		1.82	0.07
	Matched	0.46	0.48	-3.4	24.1	-1.24	0.21
NumConso	Unmatched	1.82	1.66	16.3		6.62	0.00
	Matched	1.82	1.84	-1.9	88.6	-0.71	0.48
NumMP4	Unmatched	1.11	0.93	18.3		7.47	0.00
	Matched	1.11	1.14	-3.0	83.4	-1.11	0.27
Abs(bias)	Unmatched			17.7		617.20	0.00
	Matched			1.9		31.47	0.09

Figure A1. Pre- and post-matching bias between PuS and PrSPS

Standardised

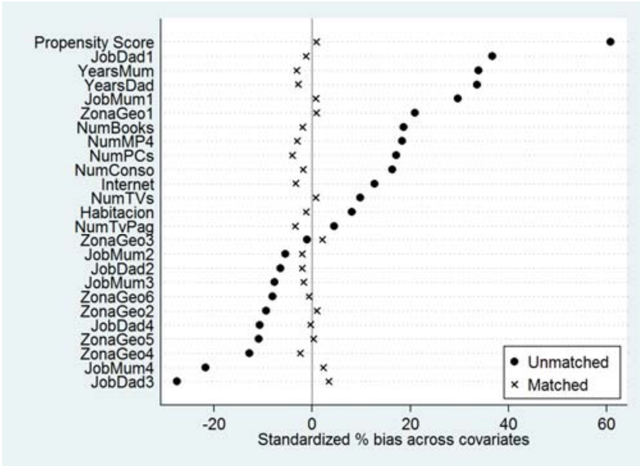


Figure A.2. Distribution of the variables in the unmatched and matched samples

Mother's education (years)

Full sample

Matched sample

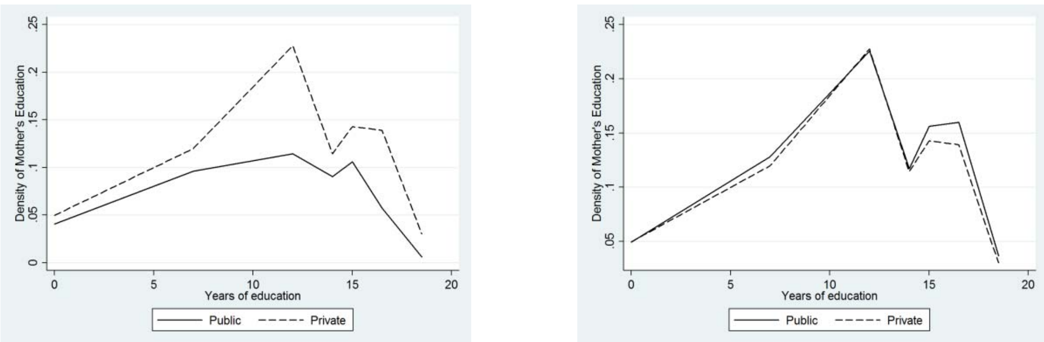
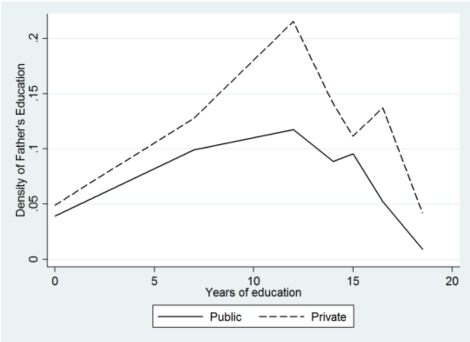


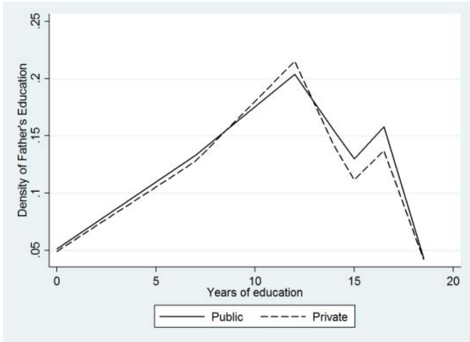
Figure A.2. Distribution of the variables in the unmatched and matched samples (cont.)

Father's education (years)

Full sample

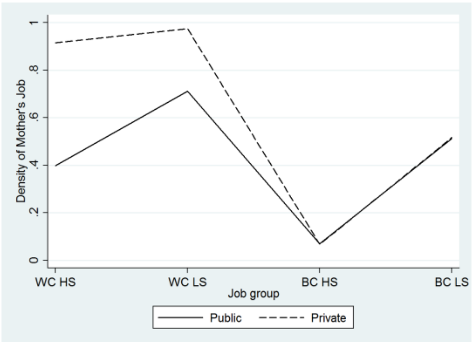


Matched sample

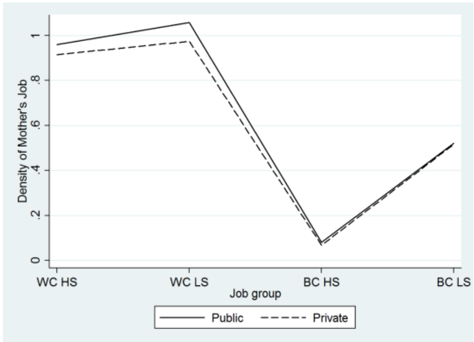


Mother's job

Full sample

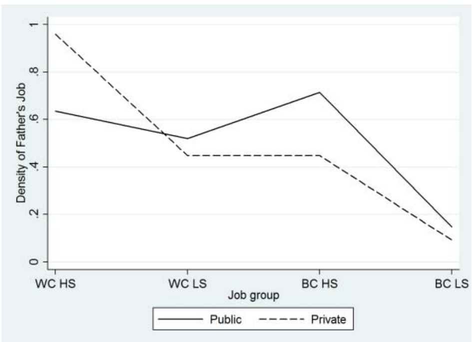


Matched sample



Father's job

Full sample



Matched sample

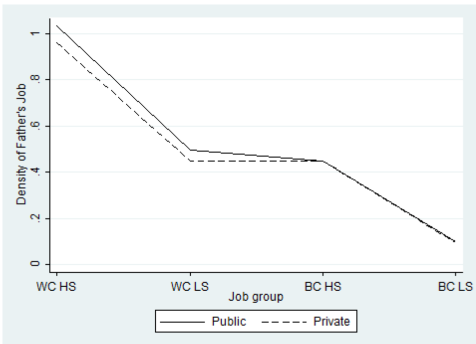
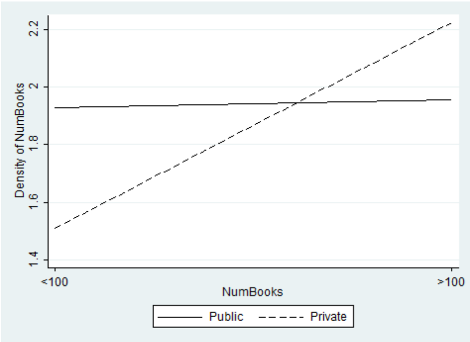


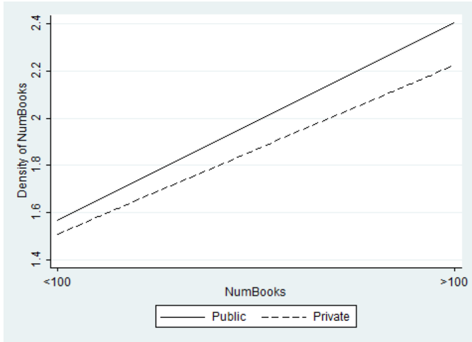
Figure A.2. Distribution of the variables in the unmatched and matched samples (cont.)

Number of books at home

Full sample

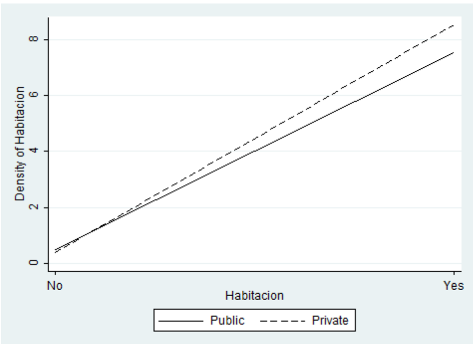


Matched sample

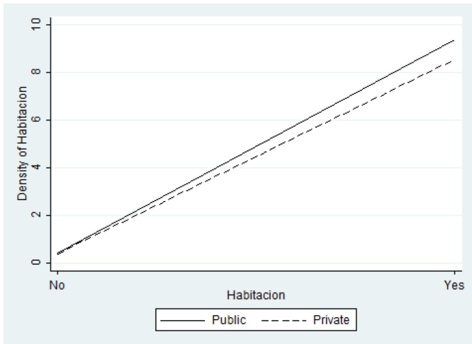


Own bedroom to study

Full sample

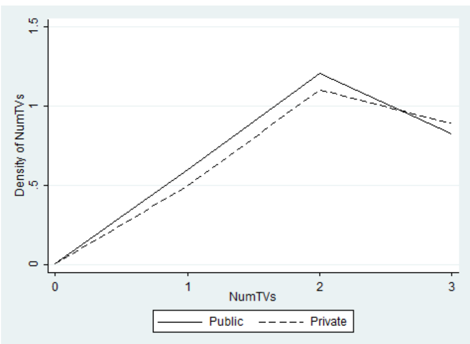


Matched sample



Number of TV sets at home

Full sample



Matched sample

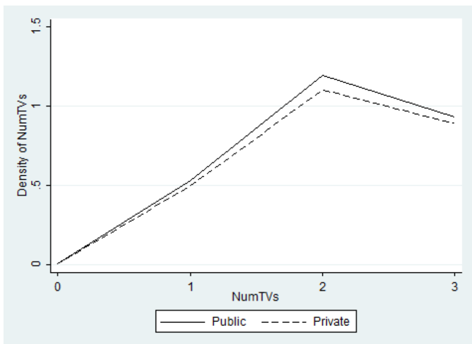
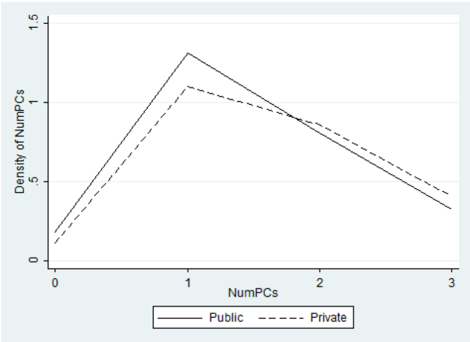


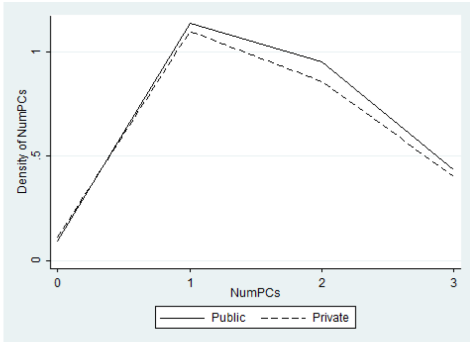
Figure A.2. Distribution of the variables in the unmatched and matched samples (cont.)

Number of PCs at home

Full sample

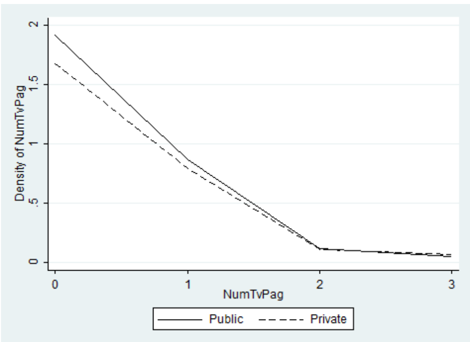


Matched sample



Number of pay TVs at home

Full sample



Matched sample