Abstract Title Page

**Title:** Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the School Voucher System in Chile

**Authors and Affiliations:**

Luke Keele  
Associate Professor  
Penn State University, University Park, PA 16802  
Email: ljk20@psu.edu

José R. Zubizarreta  
Assistant Professor, Division of Decision, Risk and Operations, and Statistics Department  
Columbia University, New York, NY 10027-6902  
Email: zubizarreta@columbia.edu
**Background / Context:**

Many policy evaluations occur in settings where randomized experiments are difficult or impossible. When randomized interventions are not possible, researchers often choose to conduct an observational study. Cochran and Chambers (1965) defined an observational study as an empirical comparison of treated and control groups where the objective is to elucidate cause-and-effect relationships in contexts where it is not feasible to use controlled experimentation and subjects select their own treatment status. When subjects select their own treatments, differing outcomes may reflect initial differences in treated and control groups rather than treatment effects (Cochran and Chambers 1965; Rubin 1974). Pretreatment differences or selection biases amongst subjects come in two forms: those that have been accurately measured, which are overt biases, and those that are unmeasured but are suspected to exist which are hidden biases. In an observational study of treatment effects, analysts use pretreatment covariates and a statistical adjustment strategy to remove overt biases.

Observational studies often occur in settings where the data have a multilevel or hierarchical structure with covariates observed at two or more levels (Gelman and Hill 2006). Multilevel data structures are common in the social sciences. Educational settings are perhaps the most well-known multilevel data structure where we observe student level measures such as gender, but also school level covariates such as the proportion of female students and the size of the school. A conventional approach to an evaluation problem with multilevel data would use a hierarchical regression model to account for the nested structure of the data and remove overt biases. A hierarchical regression model allows an analyst to fit a regression model using unit level covariates while accounting for unexplained variation among clusters (Raudenbush and Bryk 2002). The cluster level predictors are often referred to as “contextual effects,” and may be interpreted as causal effects under certain assumptions (Gelman 2006). Matching methods are also frequently used to remove overt biases in an observational study. Matched samples are constructed by finding close matches to stochastically balance a large number of covariates (Rosenbaum and Rubin 1983). Matching methods, however, do not typically account for multilevel data structures.

**Purpose / Objective / Research Question / Focus of Study:**

In this paper, we show how to construct matched samples for multilevel data structures. To adapt matching to multilevel data, one might imagine constructing a matched sample that first matches group covariates and then individual level covariates within matched groups. For example, with a school level treatment, a naive approach to this matching problem would be to first match school level covariates and then based on the sample of matched schools match students. We show this naive approach is actually a non-optimal way to construct such a matched sample of clusters and units. Here, we develop an optimal matching method for multilevel data structures. Our method can match at multiple levels which allows us to balance both units and clusters. Counterintuitively our method first matches units and then clusters to create matched pairs of clusters with units matched within these cluster pairs. We apply our matching method to an observational study of whether students that attend private vouchers schools perform better on achievement tests.
Setting:

Chile was the first country to adopt a universal voucher system. Three types of schools were created: public schools, private voucher schools, and private schools. Purely private schools are privately run and do not receive any government funding; funding is entirely financed by tuition paid by parents. Public and private voucher schools receive a direct payment on a per-student basis. Are these private vouchers schools more effective than public schools? Current evidence on the voucher system in Chile is mixed. A number of studies have found that private voucher schools increase test scores by 15% to 60% of a standard deviation (Mizala and Romaguera 2001; Sapelli and Vial 2002; 2005; Anand et al. 2009). Other studies have found either effects that are not statistically detectable (Hsieh and Urquiola 2006; McEwan 2001) or are much smaller (Lara et al. 2011). We conduct an observational study of whether private voucher schools produce students with higher test scores than public schools to demonstrate multilevel matching. In this study, we use the transition from primary school to secondary school to clearly delineate the timing of the treatment.

Population / Participants / Subjects:

In 1988, Chile introduced a national student assessment system known as the SIMCE. The SIMCE is an “educational census.” That is, in the SIMCE, the Ministry of Education collects data to evaluate all students in fourth, eighth, tenth and eleventh grades in language, mathematics and sciences, roughly every two years. SIMCE data are collected from four different sources. First, data are collected from students, which includes test scores that are complemented with other student covariates such as gender. Second, both parents and teachers complete questionnaires. Finally, for schools test scores are aggregated, and a few additional covariates are collected for both primary and secondary schools. Student records can be linked to teacher, parent, and school level covariates. In our study, we use SIMCE data from 2003, 2004, and 2006. The data Chilean form a full panel over time which allows us to carefully separate both school and student level covariates from outcomes.

Significance / Novelty of study:

Our work is novel on two fronts. First, we contribute to the larger literature on the effectiveness of voucher programs. As detailed below, we find that students in vouchers schools do not appear to have higher test scores. Second, this is (to our knowledge) the first paper to detail an optimal method for matching with multilevel data structures. As such, we retain the advantages of matching, but we are also fully faithful to the multilevel data structure which is a key part of many educational interventions.
Intuition may suggest that the best way to match multilevel data is first to match clusters and then within matched clusters to match units. In our case study, this would require first pairing schools and then, within pairs of schools, pairing students. However, this strategy will not always find the largest matched sample that is balanced as two schools that are paired on their school level characteristics may have different student compositions such that when their students are paired it may result in a smaller sample size than optimal. For this reason, the optimal matching strategy needs to contemplate what is optimal both at the student and school levels simultaneously.

Applying Bellman’s principle of optimality (Bellman 1957), the optimal matching strategy is, under the assumption that schools have been matched optimally, first match students and then, considering these optimal student matches, match schools.

To formalize and implement this idea, let $k_t \in \{1, ..., K_t\}$ index the treated schools and $k_c \in \{1, ..., K_c\}$ index the control schools. Let $j_{kt}$ denote treated student $j$ in treated school $k_t$, with $j_{kt} \in \{1, ..., J_{kt}\}$, and $j_{kc}$ denote control student $j$ in control school $k_c$ with $j_{kc} \in \{1, ..., J_{kc}\}$. Let $a_{k_t,k_c} = 1$ if treated school $k_t$ is paired to control school $k_c$ and $a_{k_t,k_c} = 0$ otherwise; similarly let $b_{j_{kt},j_{kc}} = 1$ if treated student $j$ in treated school $k_t$ is paired to control student $j$ in control school $k_c$, and $b_{j_{kt},j_{kc}} = 0$ otherwise.

In the first stage of our matching method, we find the maximum number of pairs of students that satisfies certain covariate balance specifications across all the possible $K_t \times K_c$ pairs of treated and control schools. This is, for all $k_t \in \{1, ..., K_t\}$ and all $k_c \in \{1, ..., K_c\}$ we find

$$n_{k_t,k_c} = \max_b \sum_{j_{kt}=1}^{J_{kt}} \sum_{j_{kc}=1}^{J_{kc}} b_{j_{kt},j_{kc}}$$

subject to the pair matching constraints

$$\sum_{j_{kc}=1}^{J_{kc}} b_{j_{kt},j_{kc}} \leq 1, \quad j_{kt} = 1, ..., J_{kt}$$  \hspace{1cm} (2)

$$\sum_{j_{kt}=1}^{J_{kt}} b_{j_{kt},j_{kc}} \leq 1, \quad j_{kc} = 1, ..., J_{kc}$$  \hspace{1cm} (3)

$$b_{j_{kt},j_{kc}} \in \{0, 1\}$$  \hspace{1cm} (4)

which require matching each treated and control unit at most once, and covariate balancing constraints of the form

$$-\epsilon_q \sum_{j_{kt}=1}^{J_{kt}} \sum_{j_{kc}=1}^{J_{kc}} b_{j_{kt},j_{kc}} (h_q(x_{j_{kt}}) - h_q(x_{j_{kt}})) \leq \epsilon_q \sum_{j_{kt}=1}^{J_{kt}} \sum_{j_{kc}=1}^{J_{kc}} b_{j_{kt},j_{kc}}$$  \hspace{1cm} (5)

where $h_q(\cdot)$ is a suitable transformation of the observed covariate $x$ and $\epsilon_q$ is a scalar indexed by $q = 1, ..., Q$. In other words, in this first stage we find the largest sample of pairs of students...
that is balanced as specified by $h_q(\cdot)$ and $\varepsilon_q$ for $q = 1, \ldots, Q$ across all pairs of treated and control schools.

Then, in the second stage of our matching method, once we know how many balanced pair matches of students can be obtained across all the possible combinations of pairs of schools and collected these values in the matrix $[n_{k_t,k_c}]$, we find the optimal school match that solves

$$
\max_a \sum_{k_t=1}^{K_t} \sum_{k_c=1}^{K_c} n_{k_t,k_c} a_{k_t,k_c}
$$

again subject to pair matching and covariate balancing constraints analogous to (2)-(4) and (5) respectively. In this manner, the multilevel matching problem can be solved optimally by breaking it into a simpler matching subproblems and recursively finding the optimal match.

**Usefulness / Applicability of Method:**

Using the data from Chile as a case study, we demonstrate how to implement the matching algorithm describe above. We find that we are able to produce a match where both student and schools are highly comparable.

**Findings / Results:**

We performed two matches. One with better balance but more limited geographic representation. We also implemented a second match with greater imbalance but more widespread geographic representation. We tested the hypothesis of no effect for private voucher schools in both matches. For the match with greater geographic representation, the approximate one-sided $p$-value is 0.256. In the absence of bias from hidden confounders, the point estimate is $\hat{\tau} = 2.81$ with a 95% confidence interval of -5.68 and 11.34. For the model with better balance but more limited geographic representation, we find the approximate one-sided $p$-value is 0.633. If there are no hidden confounders, the point estimate for the match with greater internal validity is $\hat{\tau} = 0.0743$ with a 95% confidence interval of -8.58 and 9.36. Thus for both matches, we cannot reject the hypothesis that attending a private voucher school has no effect on test scores. In one match, we went to greater lengths to remove overt bias. We also conduct a sensitivity analysis to understand whether our conclusions are a function of hidden bias from unobserved confounders.

**Conclusions:**

Multilevel data structures are common in education settings. As we discussed, in these settings, the optimal matching strategy is to first match individuals and then, considering these optimal individual level matches, match clusters. We implemented this strategy using integer programming to find the largest sample of matched pairs of treated and control individuals within matched pairs of treated and control clusters that is balanced according to specifications given by the user.
Appendices

Appendix A. References


