Optimal full matching under a new constraint on the sharing of controls: Application in pediatric critical care

Simon Dovan Nguyen [1] Ben B. Hansen [1] Mark M. Fredrickson [1] Ryan Barbaro [2]

[1]Department of Statistics [2]Department of Pediatrics

Abstract

Full matching on the propensity score aims to emulate random assignment by placing observations with similar propensity scores into sets with either one treated unit and multiple control units or one control unit and multiple treated units. Sets of the second type, with treatment units forced to share a comparison unit, can be unhelpful from the perspective of statistical efficiency.

In this poster, we introduce an enhancement of Hansen and Klopfer's optimal full matching algorithm that allows the sharing of controls while limiting the number that are permitted to do so. The enhancement is illustrated in a data-scarce pilot study on the effects of extracorporeal membrane oxygenation for treatment of pediatric acute respiratory distress syndrome. In conjunction with existing full matching tools, the enhancement provides additional benefits by prioritizing 1:1 pairs and increasing effective sample size while still maintaining moderate covariate balance.

Issues with optimal full matching

Issue: Full matching's flexibility in matched set structures may harm treatment effect estimation.

Suppose we are interested in estimating the average treatment effect (ATE) with sample size n:

$$\Delta = \frac{1}{n}\sum_{i=1}^{n} y_i(1) - y_i(0) \tag{1}$$

Summing over the b matched sets, we can estimate the average treatment effect with

$$\hat{\Delta} = \sum_{s=1}^{b} \frac{n_s}{n}\left[\sum_{i=1}^{n_s} \frac{Z_{si}Y_{si}}{n_{s1}} - \frac{(1-Z_{si})Y_{si}}{n_{s0}}\right] \tag{2}$$

where n_{s1} (n_{s0}) denotes the number of treatment (control) units in set s with $n_s = n_{s1} + n_{s0}$, $Z_{si} = 1$ if unit i in set s received treatment and 0 if not, and Y_{si} the outcome of unit i in set s.

Denoting the variance of each matched set s as σ_s^2, the variance of $\hat{\Delta}$ can be derived to

$$Var(\hat{\Delta}) = \frac{1}{n^2}\sum_{s=1}^{b} \sigma_s^2 \frac{n_s^2}{n_s - 1}\left(\frac{1}{\theta_s(1-\theta_s)}\right) \quad \text{where } \theta_s = \frac{n_{s1}}{n_s} \tag{3}$$

Objective 1: Assuming $\sigma_s^2 \approx \sigma^2$, equation 3 shows that $Var(\hat{\Delta})$ is small when n is large and $\theta_s \approx \frac{1}{2}$.

As such, promoting pair-like matched sets would be beneficial in decreasing the variance, provided that covariate balance does not increase too seriously.

Objective 2: To equivalently compare the sizes of matched structures into matched pairs, define

$$\text{Effective Sample Size (ESS)} = \sum_{s=1}^{b}\left[(n_{s1}^{-1} + n_{s0}^{-1})/2\right]^{-1} \tag{4}$$

As $Var(\hat{\Delta})$ is inversely proportional to ESS, increasing ESS improves treatment effect estimation.

Existing tools in optimal full matching

Solution: Limit the number of controls shared by one treatment unit.

1. `max.controls`: The maximum number of control units shared by one treatment unit.
2. `min.controls`: The minimum number of control units shared by one treatment unit.
3. `mean.controls`: The average number of controls per treatment.

A new tool in optimal full matching: `shared_treatment_excess`

However, `max.controls` and `min.controls` can be too deterministic on setting bounds on matching configurations. We introduce a new matching restriction, `shared_treatment_excess`, that allows additional slack on the sharing of controls while limiting extreme variations in matching configurations.

Define

$k :=$ the number of treated units sharing a control unit
$v :=$ the number of sets with shared treated units

Then, `shared_treatment_excess` is defined as the difference of the number of treated units sharing a control and the number of sets with shared treated units:

$$\texttt{shared\_treatment\_excess} = k - v \tag{5}$$

`shared_treatment_excess` provides additional slack by allowing the matching algorithm the possibility of either increasing the number of treated units sharing a control and decreasing the number of sets with shared treated units or vice versa.

Extracorporeal membrane oxygenation

Pediatric acute respiratory disease syndrome (PARDS) is a life-threatening illness that prohibits children from breathing, leading to ventilator support. When ventilator support is not sufficient, intensivists may initiate extracorporeal membrane oxygenation (ECMO). Although life saving, ECMO is intense and may lead to long-term harmful complications in a child's functioning and quality of life.

Physicians have no clear guidance on when to apply ECMO. In our study, we evaluate the effect of initiating ECMO early (control group) vs. late (treatment group) on all-cause mortality within 90 days. In order to assess the influence of `shared_treatment_excess`, we compare four models enhanced with `shared_treatment_excess` to four reduced models without it.

Results

1. **Influence on effective sample size**: `shared_treatment_excess` was found to be more influential in increasing ESS than `max.controls` and `min.controls` with the largest increase in ESS in the presence of `max.controls`.
2. **Influence on covariate balance**: As expected, covariate balance worsened in favor of increasing ESS. Interestingly, however, the most imbalanced variables indicated by high standardized differences in the reduced models, were shown to be more well-balanced in the enhanced models with `shared_treatment_excess`.
3. **Influence on matching configurations**: In all models, there is a large increase in 1:1 pair matches with decreases in larger matching configurations.

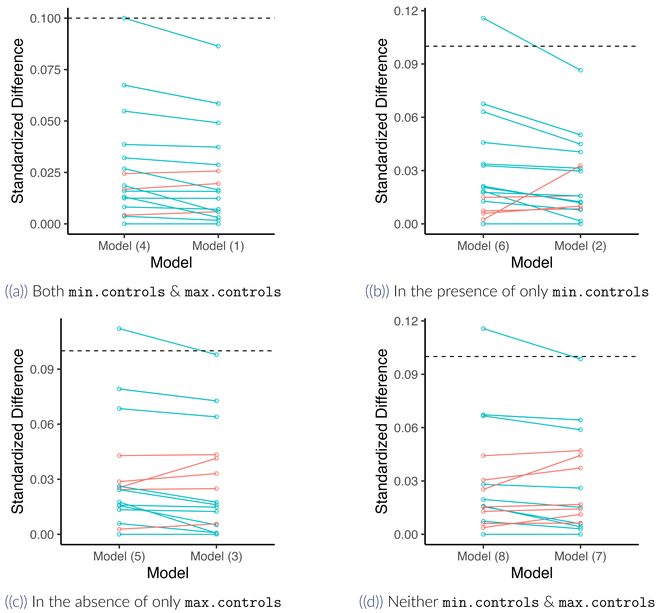
Contribution

This study provides researchers with a new tool in managing the bias-variance tradeoff between effective sample size and covariate balance. Researchers desiring benefits in effective sample sizes and simpler matching configurations at the small expense of covariate balance can utilize `shared_treatment_excess` as an additional tool in optimal full matching.

Figures

Model	ATE	ATE Standard Error	ATE p-value	ESS	Balance p-value
(1) All	0.102	0.0389	0.011	282.695	0.057
(4) `Max` and `Min`	0.099	0.0391	0.012	280.529	0.224
(2) `Max` and `STE`	0.098	0.0389	0.014	282.795	0.050
(6) Only `Max`	0.100	0.0392	0.012	274.472	0.745
(3) `Min` and `STE`	0.107	0.0389	0.008	276.764	0.220
(5) Only `Min`	0.103	0.0391	0.010	274.110	0.681
(7) Only `STE`	0.102	0.0393	0.012	276.364	0.335
(8) Unrestricted	0.101	0.0391	0.012	273.363	0.750

Table 1. Results table. All models indicate that the inclusion `shared_treatment_excess` in the enhanced models lead to a more precise treatment effect estimate at the cost of covariate balance.



((a)) Both `min.controls` & `max.controls`

((b)) In the presence of only `min.controls`

((c)) In the absence of only `max.controls`

((d)) Neither `min.controls` & `max.controls`

Figure 1. Individual match-adjusted covariate standardized difference between treatment and control groups for the enhanced and reduced models. Although `shared_treatment_excess` tends to worsen covariate balance on average, these figures show several cases in which individual covariate balance improves.

References

[1] Mark M. Fredrickson and Ben B. Hansen. Full matching with more precise set size constraints. Joint Statistical Meetings, 2021.

[2] Ben B. Hansen. Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 2004.

[3] Ben B. Hansen and Stephanie Klopfer. Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, 2006.

simondn@umich.edu University of Michigan