Supplementary Materials for "Multivariate Continuous Blocking to Improve Political Science Experiments"*

Ryan T. Moore†

7 March 2013 at 09:47

1 Univariate Examples and Definitions

Consider a small political field experiment in which you will assign a campaign message in four precincts, two predominantly Democratic and two predominantly Republican. To prevent your preconceptions from influencing the experiment, you decide to randomize the two available campaign messages (say, A and B) to the precincts. You know that precinct partisanship is likely to strongly affect voters' responses to the campaign messages. You randomly select two precincts to assign message A to, and both Democratic precincts are chosen. What do you do? If you proceed, you will be unable to untangle the effects of partisanship from that of the campaign messages. Better design could have prevented this problem. By simply sorting the precincts by partisanship then randomizing one Democratic and one Republican precinct to each message, the experiment could generate more useful information. This pre-randomization sorting of units, *blocking*, can easily incorporate many variables, whether discrete or continuous, and confers several statistical and political ad-

*Thanks to Keith Schnakenberg and Jack Frasier for research assistance; thanks to Michael L. Anderson, Jake Bowers, Thad Dunning, Jeff Gill, Donald P. Green, Jens Hainmueller, Gary King, Kevin Quinn, Jas Sekhon, Laura Stoker, and anonymous reviewers for suggestions on previous versions; thanks to Daniel Rubenson, Peter John Loewen, Larry Schweinhart, and Zongping Xiang for generously sharing data and answering questions. The replication archive is available as Moore (2012b).

†Department of Political Science, Washington University in St. Louis, 241 Seigle Hall, Campus Box 1063, One Brookings Drive, St. Louis MO 63130. tel: 314.935.3461; fax: 801.922.7901; `rtm` (at) `wustl` (dot) `edu`; `http://rtm.wustl.edu`.

Unit	X	T	Y_T	Y_C
1	2	?	_	_
2	2	?	_	_
3	4	?	_	_
4	3	?	_	_
5	3	?	_	_
6	4	?	_	_
7	10	?	_	_

$\longrightarrow$

	Unit	X	T	Y_T	Y_C
block 1	1	2	1	1	_
	2	2	0	_	3
block 2	3	4	1	4	_
	6	4	0	_	6
block 3	4	3	1	9	_
	5	3	0	_	7
	7	10	?	_	_

Table 1: Typical Blocking Problem and Solution. Prior to treatment assignment (left table), units may be blocked to any other unit using covariate X. In right table, units are exact-blocked to other units, then treatment is randomized within blocks.

vantages in real experiments. Here, we demonstrate some of these advantages, providing additional examples to accompany our primary paper, Moore (2012a).

The timing of random assignment differentiates experimental blocking from a similar procedure, observational matching.[1] Consider the simple univariate case in Table 1. The columns of the table represent unit identification numbers, values of the background covariate X, an as-yet unknown vector of treatment assignments (1 for treatment, 0 for control), and the unobserved potential outcomes for each unit (Y_T for outcomes under treatment, Y_C for those under control). The fact that only one of treatment or control can be assigned to each unit, and thus only one of Y_T or Y_C can be observed for each unit is known as the "fundamental problem of causal inference" (Holland, 1986). On the Table's left, the treatment has not yet been assigned.

Blocking on X would pair Unit 1 to Unit 2, Unit 3 to Unit 6, and Unit 4 to Unit 5.

[1]While some literature uses "matching" to denote the experimental procedure, "blocking" draws a clearer distinction between pre- and post-assignment procedures. In several statistical schools, such as under the Neyman-Rubin Causal Model (Rubin, 1974, 1978), "matching" usually refers explicitly to the practice of pairing observational treatment units with controls. We advocate using "blocking" to denote the pre-assignment procedure, and "matching" to refer to the post-assignment procedure of grouping already-treated units with similar control units. In some fields, pre-assignment blocking is considered a special case of "matching", which may occur among all units before assignment ("nonbipartite matching") or only after treatment between already-treated units and control units ("bipartite matching"). Using "matching" to represent pre-assignment procedures appears especially common in biostatistics and public health, while the experimental design literature in statistics prefers "blocking". Other authors use "stratification" to denote a similar set of design procedures (Donner, 1998), though this can create confusion with "stratification" from survey sampling, where researchers usually do not control the assignment mechanism.

Unit	X	T	Y_T	Y_C
1	2	1	1	_
2	2	1	3	_
3	4	1	4	_
4	3	0	_	9
5	3	0	_	7
6	4	0	_	6
7	10	0	_	8

$\longrightarrow$

	Unit	X	T	Y_T	Y_C
match 1 {	1	2	1	1	_
	4	3	0	_	9
match 2 {	2	2	1	3	_
	5	3	0	_	7
match 3 {	3	4	1	4	_
	6	4	0	_	6
	7	10	0	_	8

Table 2: Typical "Matching" Problem and Solution. After treatment assignment, treatment units are matched to control units using covariate X. Bracketed rows denote reasonable matched pairs. Match 3 is an "exact match", but match 1 and match 2 are inexact.

Then, we perform random allocation within each block. The right side of Table 1 shows the blocks bracketed, and column T shows one possible randomization. Within each blocked pair, one unit is assigned to treatment, one to control. One estimate of the average treatment effect using exact one-to-one blocking simply takes the mean of the differences between the treatment outcome and the control outcome within each block. Estimates of the treatment effect depend on decisions like whether to exclude or include Unit 7, whose covariate value (and perhaps potential outcomes) differs greatly from those of Units 1 through 6. If X measures family sizes, for example, Unit 7's value is well outside the $\{2, 3, 4\}$ set defined by the other units, and may represent a family with significantly different preferences and constraints than the others.

By contrast, the matching problem occurs after treatment has been assigned. For example, an observational researcher might encounter the first three units of the left panel of Table 1 already treated and the last four untreated. Table 2 displays such a scenario; the left side of Table 2 is similar to that of Table 1, but the researcher has not been able to control the treatment assigned to the units. Observational matching then might create the three pairs in the right panel of Table 2. Here an advantage of a well-designed experiment over observational research is evident – these observational units could not produce the well-balanced experiment shown in the right panel of Table 1.

Balance

Complete randomization balances experimental data in expectation, but blocking before assignment often constitutes an improvement. Assume that each unit has a single observed covariate value, X_i. Then $\bar{X}_T = \frac{1}{n}\sum_{i\in\{I_i=1,T_i=1\}} X_i$ and $\bar{X}_C = \frac{1}{n}\sum_{i\in\{I_i=1,T_i=0\}} X_i$ are the simple means of the covariate in the treatment and control groups, respectively. Although covariate balance denotes identical multivariate distributions, we take a conventional first step and focus on univariate mean balance. Let the covariate imbalance between treatment and control groups be $M \equiv |\bar{X}_T - \bar{X}_C|$; higher values indicate more imbalance.

Given a set of units with background measures, an experimenter may choose to block or not to block prior to random assignment. If the experimenter blocks, then a particular blocking algorithm must be selected, yielding particular blocks. After a particular blocking has been implemented, randomization assigns treatment and control to one unit in each block. Given the obtained blocking, the particular assignment obtained may create more or less covariate imbalance.

In Table 1, the exact blocking promotes perfect covariate balance, and all randomizations will yield the same value on the covariate imbalance measure, $M = 0$. However, for different arrangements of units into blocks, and of randomizations after blocking, this will not always be the case. For example, a suboptimal blocking might pair units 1 and 4, 2 and 5, and 3 and 6. In this case, the measure of covariate imbalance would be greater than zero, indicating worse imbalance: $M \equiv \left|\bar{X}_T - \bar{X}_C\right| = \left|\frac{\sum_{i=1}^{3} X_i}{3} - \frac{\sum_{i=4}^{6} X_i}{3}\right| = \frac{2}{3}$.

Assume there are 20 units taking unique X values from 1 to 20 such that the unit numbered 1 has covariate value 1, the unit numbered 2 has covariate value 2, etc. (i.e., $X_i = i$). Under an unblocked design, the worst possible random assignment of 10 units to treatment and 10 units to control would assign the set of units including $X_i \in \{1, ..., 10\}$ to (say) control, and the set including $X_i \in \{11, ..., 20\}$ to treatment. Then, the covariate imbalance measure would be $\left|\bar{X}_T - \bar{X}_C\right| = \left|\frac{\sum_{i=11}^{20} X_i}{10} - \frac{\sum_{i=1}^{10} X_i}{10}\right| = 10$.

If instead we block on X_i, we may do very well, we may do very poorly. The worst possible pairwise blocking will create pairs $(1, 20), (2, 19), \ldots, (10, 11)$. The worst possible randomization then might assign the low value in each pair control and the high value in each pair treatment, replicating the pairs and assignments in the worst randomization under an unblocked design above. Combining the worst possible blocking with the worst possible randomization yields an imbalance measure of $M = 10$, and does no worse than the worst possible unblocked design. The best possible randomization for this poor blocking would alternate, roughly assigning the low value of one pair to control, the low value of the next pair to treatment, etc. This combination of the worst possible blocking and the best possible randomization yields an imbalance measure of $M = 0$.

On the other hand, the best possible blocking would create pairs $(1, 2), (3, 4), \ldots$. Conditional on this blocking, the best possible randomization would roughly alternate, assigning the low value of one pair to control, the low value of the next pair to treatment, etc., producing $M = 0$. The worst possible assignment randomization would assign the low value in each pair control and the high value in each pair treatment. The worst assignment would then yield a covariate imbalance measure of only $M = 1$. Blocking and randomization are thus nearly symmetric tools for achieving covariate balance: the best blocking and the worst randomization yield about the same level of imbalance as the worst blocking and the best randomization. If the randomization is unlucky, good blocking successfully limits imbalance, and even bad blocking does at least as well as no blocking at all.

Combining the best possible blocking with the best possible randomization yields an imbalance measure of 0. Thus, even under the best blocking, we slightly prefer fortuitous randomizations that create balance to randomizations that limit balance. Under an unblocked design, the best possible randomization produces the same level of imbalance as under the best blocked design. However, this best randomization is much rarer in unblocked designs. There are $\frac{20!}{10!10!} = 184{,}756$ different combinations of 10 treatment and 10 control units possible. However, under no blocking, fewer than 3% of these combinations yield a covariate imbalance measure of $M = 0$, compared to 25% of the assignments under the best

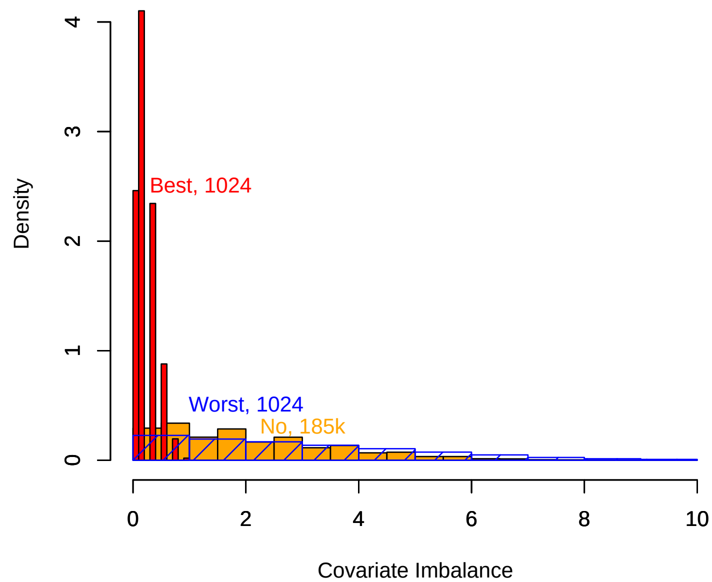

Figure 1: Distributions of Imbalance for Best, Worst, and No Blocking. For each blocking, a histogram represents all possible levels of covariate imbalance. The worst blocking does no worse than no blocking. The best blocking strongly limits the likely imbalance. Numbers in Figure indicate how many possible randomizations exist under that blocking.

blocking. Only about 25% of the unblocked assignments yield $M < 1$, the worst value under the best blocking. Every randomization under the best blocking produces better balance than the average unblocked design.

Figure 1 shows the distribution of all possible levels of imbalance in this 20-unit experiment. As blocking procedures improve upon the worst possible one, the blue histogram will become more like the red one and even under the worst possible assignment randomization, blocking will yield better covariate balance on average than the unblocked design. As Figure 1 shows, reasonable blocking can powerfully limit covariate imbalance, while even poor blocking is not worse than an unblocked design.

Efficiency

Assume that the treatment effect for each unit in Table 1 is exactly zero, and thus that the missing potential outcome for each unit is equal to the observed one. Then, the causal effect for each unit is zero, and the true average causal effect is zero. Considering every possible permutation of three treatment, three control units, the standard deviation of the possible estimates of that mean is about 2.43 under random allocation, but only about 1.23 under exact blocking. This represents a substantial improvement of about 49%.

Monte Carlo results confirm the improved efficiency of blocking over random allocation. In a series of simulations varying the number of experimental units (50 or 500), covariates (2 or 6), the linearity of the outcome model in the covariates, the size of the treatment effect relative to the model coefficients, and the adjustment set (none, half, or all of the covariates), blocking produces more efficient estimates of the treatment effect. Comparing the simple unadjusted difference in mean treatment effect estimates from one set of 3200 random assignments, the difference between the 95th and 5th percentiles of the sampling distribution averages 76% larger under complete randomization than under blocking. For regression-adjusted estimates, the difference is 14% larger on average under complete randomization than under blocking. We show similar results in the applications of our article.

A Block-Level Effect

In Table 1, the administrator of Unit 1 could derive a local causal estimate of $Y_{1T} - Y_{2C} = 1 - 3 = -2$ from the block containing Units 1 and 2. The administrator would then expect that this estimate has as little error as that from any other block, since all included units are exact-blocked.

Limited Resources

The data from Table 2 provides an example of how blocking can inform resource allocations in a study. If the first three units were forced to be paired with the next three and resources only existed to implement treatment in one block, then exact blocked Units 3 and 6 would provide an estimate of the treatment effect with less error due to covariate imbalance than either of the other blocks, in both of which $M = 1$.

2 Multivariate Examples

2.1 Robust Estimates of the Scaling Matrix

2.1.1 Blocks from Moore (2012a) Example

Tables 3, 4, and 5 display the blocks described in Section 4.2 of Moore (2012a).

	Unit 1	Unit 2	Distance
1	15	4	0.18
2	9	7	0.37
3	13	5	0.37
4	18	10	0.43
5	17	3	0.44
6	14	11	0.50
7	20	1	0.65
8	16	2	0.95
9	12	6	1.32
10	19	8	2.12

Table 3: Blocks from Data excluding Outliers, using Non-Outlier-Resistant Covariate Weighting Matrix.

2.1.2 Additional Example

Consider four data points that lie in a near-rectangle: $A = (1, 1)$, $B = (2, 1)$, $C = (1, 4.1)$, and $D = (2, 4)$. In Mahalanobis distance, the closest unit to A is B, by a small margin. See

	Unit 1	Unit 2	Distance
1	15	4	0.18
2	13	5	0.37
3	9	7	0.42
4	18	10	0.46
5	17	3	0.50
6	14	11	0.59
7	20	1	0.76
8	16	2	1.03
9	12	6	1.56
10	19	8	2.18
11	22	21	13.43

Table 4: Blocks from Data including Outliers, using Outlier-Resistant Covariate Weighting Matrix. All pairings identical to no outliers, non-resistant weighting above.

Table 6 for the exact values. Adding an outlier at $(1, 100)$ means the y-variation in the first four points becomes meaningless, and C is considered much closer to A. The outlier can also leverage the covariation to make D closest to A. As this extreme outlier is unlikely to yield a useful comparison to any of the other units, we consider this influence undesirable.

In contrast, the MCD and MVE estimates of the covariance matrix both generate the original multivariate distances whether or not either of the outliers is included.

2.2 Increased Efficiency in the Presence of Causal Heterogeneity

In our paper, we show the increase in efficiency due to blocking for a constant treatment effect. Here, we provide another example, demonstrating the same type of advantage in the presence of unit heterogeneity in the causal effect that correlates with a covariate.

We assume a small experiment of twelve units, with a single covariate drawn from a standard normal distribution for half the units and a normal centered at 10 for the other half. We assume the outcome under control is just the covariate value, and the outcome under treatment is twice the covariate value. Thus, the unit heterogeneity is large here, as true treatment effects average around 1 for the units with smaller covariate values, versus around 10 for the units with larger covariate values.

	Unit 1	Unit 2	Distance
1	17	15	0.12
2	9	7	0.24
3	19	13	0.25
4	14	11	0.28
5	10	1	0.33
6	12	3	0.45
7	4	2	0.58
8	20	18	0.69
9	8	5	1.45
10	22	6	2.24
11	21	16	3.64

Table 5: Blocks from Data including Outliers, using Non-Outlier-Resistant Covariate Weighting Matrix. Only two pairings are identical to no outliers, non-resistant weighting above.

		Distance from A		
Scale Matx	Points	B	C	D
Cov	A-D	**1.73**	1.76	2.45
Cov	A-$D + (1, 100)$	2.00	**0.08**	2.03
Cov	A-$D + (20, 100)$	1.72	1.01	**0.74**
MCD	A-D	**1.73**	1.76	2.45
MCD	A-$D + (1, 100)$	**1.73**	1.76	2.45
MCD	A-$D + (20, 100)$	**1.73**	1.76	2.45
MVE	A-D	**1.73**	1.76	2.45
MVE	A-$D + (1, 100)$	**1.73**	1.76	2.45
MVE	A-$D + (20, 100)$	**1.73**	1.76	2.45

Table 6: Example of consistent blocks in the presence of outliers under resistant estimates of the covariate scaling matrix. **Bold** values denote unit closest to A.

We perform 1000 completely randomized, randomly allocated, and pair-blocked assignments, and show the overall differences in means in Figure 2. Given the large treatment heterogeneity, the blocked estimates gather much closer to the true sample treatment effect than the completely randomized or randomly allocated estimates. As expected, across a large number of randomizations, all three methods are unbiased on average.

Figure 2: Example of increased efficiency from blocking in the presence of treatment effect heterogeneity correlated with a covariate. Vertical dashed line represents true average treatment effect.

References

Donner, Allan. 1998. “Some Aspects of the Design and Analysis of Cluster Randomization Trials.” *Applied Statistics* 47:95–113.

Holland, Paul. 1986. “Statistics and Causal Inference.” *The Journal of the American Statistical Association* 81 (396):945–960.

Moore, Ryan T. 2012a. “Multivariate Continuous Blocking to Improve Political Science Experiments.” *Political Analysis* 20:460–479.

Moore, Ryan T. 2012b. “Replication data for: Multivariate Continuous Blocking to Improve Political Science Experiments.” http://hdl.handle.net/1902.1/18341. IQSS Dataverse Network [Distributor] V1 [Version].

Rubin, Donald B. 1974. “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of Educational Psychology* 66:688–701.

Rubin, Donald B. 1978. “Bayesian inference for causal effects: The role of randomization.” *The Annals of Statistics* 6:34–58.