Optimal full matching and related designs via network flows

Ben B. Hansen and Stephanie Olsen Klopfer[*]

submitted January 2005; revised October 2005

Abstract

In the matched analysis of an observational study, confounding on covariates $\mathbf{X}$ is addressed by comparing members of a distinguished group ($Z = 1$) to controls ($Z = 0$) only when they belong to the same matched set. The better matchings, therefore, are those whose matched sets exhibit both dispersion in Z and uniformity in $\mathbf{X}$. For dispersion in Z, pair matching is best, creating matched sets that are equally balanced between the groups; but actual data place limits, often severe limits, on matched pairs' uniformity in $\mathbf{X}$. At the other extreme is full matching, the matched sets of which are as uniform in $\mathbf{X}$ as can be, while often so poorly dispersed in Z as to sacrifice efficiency.

This paper presents an algorithm for exploring the intermediate territory. Given requirements on matched sets' uniformity in $\mathbf{X}$ and dispersion in Z, the algorithm first decides the requirements' feasibility. In feasible cases, it furnishes a match that is optimal for $\mathbf{X}$-uniformity among matches with Z-dispersion as stipulated. To illustrate, we describe the algorithm's use in a study comparing womens' to mens' working conditions; and we compare our method to a commonly used alternative, greedy matching, which is neither optimal nor as flexible but is algorithmically much simpler. The comparison finds meaningful advantages, in terms of both bias and efficiency, for our more studied approach.

Key words: matched sampling, minimum cost flow, nonrandom treatment assignment, observational study, quasiexperiment, subclassification

---

1 Introduction

Full matching is a form of subclassification generalizing pair matching and matching with multiple controls. Like those methods, full matching is suited to prospective studies comparing a treatment and a control group, or to retrospective studies comparing cases to controls. Such techniques sort all or part of a sample into non-overlapping matched sets, striving to place similar subjects alongside one another while separating those treatment and control subjects who are too dissimilar to be compared.

Older methodological work suggests, and a growing applied literature confirms, that such sortings of observations can strengthen a subsequent statistical analysis in a number of important ways. On the methodological side, Cochran (1968) documented their utility for reducing bias in estimates of treatment effects in situations where covariance adjustment may not be appropriate, a theme with echoes in later work of Rubin (1976) and his coauthors (Rosenbaum and Rubin 1985b; Rubin and Thomas 1996); related methodological literature promotes the use of matching in combination with covariance adjustment in order to simultaneously secure benefits of both (Rubin 1979; Rosenbaum 2002a; Ho et al. 2004a). Recent applications that use matching to good advantage include Harding's in sociology (2003), Dehejia and Wahba's in labor economics (1999; see also Smith and Todd 2005, Diamond and Sekhon 2005), Ho's in law (2005), and the SUPPORT investigators' in medicine (Connors et al. 1996), among many others.

A common weakness of these studies, one they share with many applications of multivariate-distance matching, is their use of greedy, rather than optimal, matching algorithms, algorithms that match treated and control units in relatively inflexible proportions. Without overshadowing the studies' strengths, this aspect of their approach brings several deleterious side-effects. First, when matching in regions of covariate space where treated subjects are many and controls are few, it lacks a protocol for grouping several treated subjects with a single control; this tempts the analyst to "re-use" controls, creating matched sets that overlap and are not disjoint. Second, in regions of the covariate space where controls are relatively plentiful, it fails to take advantage of their numbers, needlessly leaving some of them unmatched. The first of these side-effects precludes the proper use of existing statistical methods for matched or stratified data, such as permutation inference (Mantel

and Haenszel 1959; Hodges and Lehmann 1963; ch.2 of Rosenbaum 2002b), inference conditional on matched sets (Cox and Hinkley 2000; Hosmer and Lemeshow 2000; Hansen 2004a), or inference from a random effects model (Raudenbush and Bryk 2002; ch.12 of Agresti 2002; Smith 1997); and the second reduces inferential precision.

In a full match, some matched sets may contain one treated subject (or case) alongside one or more controls, while other matched sets may contain multiple treated units (cases) alongside one control. This distinguishes full matching from both pair matching, matching with $k \geq 2$ controls, and matching with a variable number of controls, although full matchings sometimes coincide with matches produced by these simpler methods. Among all methods of matching for two groups, full matching alone is demonstrably optimal as a method of producing similarity within matched sets (Rosenbaum 1991); and it remedies the aforementioned side-effects of greedy matching. Despite these attractions, software for full matching have been lacking.

This paper describes the algorithm beneath our effort to fill that gap; and it demonstrates the advantage of finding matches optimally, by way of a minimum-cost network flow solver. It understands full matching more broadly than as Rosenbaum (1991) originally defined it, permitting the options of restricting matched sets' relative proportions of treatment and control units, and of using only a fraction of the control group; understood in this way, full matching includes matching with $k \geq 1$ or a variable number of controls as special cases, so that an algorithm or program for optimal full matching can also be used to create optimal matchings of these other types. More important benefits, in our view, are those conveyed upon full matching as such, which we illustrate with a small case study (§ 1.1, 5) in which full matching is applied to the analysis of a gender equity study. (This study has no treatment or control group *per se*, but it resembles an observational study in seeking to compare two groups while adjusting for measured covariates.)

1.1 Why use full matching?

The argument for full matching's optimality as a method of subclassification is somewhat abstract, but it has a number of concrete manifestations. The first is its ability to make use of as many observations as are suitable to be included in a comparison.

Table 1: Pair match (solid lines) and match with one to four controls (all lines) for one science department's research-only faculty.

Women		Men	
Subject	$\log_{10}(\text{Grant})$	Subject	$\log_{10}(\text{Grant})$
A	0 —————— R	0
B	0 —————— S	0
C	0 —————— T	0
D	0 —————— U	4.4
E	4.4 —————— V	5.0
F	6.1 ⤙ W	5.7
		X	5.9
		Y	6.0
		Z	6.3

In some studies, one begins with mostly complete data on treated units or cases, but less complete data on a much larger reservoir of potential controls; in such settings, resources may permit additional data collection on no more than one or a small number k of controls per treated subject. In many other studies, however, matching is done after all the data have been collected. In typical cases of this type, full matching closely matches many more subjects than would matching with a fixed number of controls.

Table 1 illustrates this with an example drawn from a study of gender equity to which one of us (B.B.H.) served as a consultant. Men and women professors at a major research university were to be matched on the basis of administrative data, then compared as to the size of their lab and office space assignments. At universities like the one in question, space requirements increase with amount of lab equipment and with the number of students and staff working in a lab, and a professor's total grant funding correlates with all of these. Thus grant funding was used as a proxy for space requirement, and matches were sought between faculty who are as similar as possible in terms of grant totals. (Specifically, $\log(1 + \text{Grant})$.) Since matches were made only within departments and within instructional and research-only career tracks, the study is a good source of small, illustrative matching problems, and that is the spirit in which the data are presented here; the matching actually used took into account several other factors in addition to grant funding.

Given a pre-assembled data base of subjects, one seeks to match as many of them as possible;

4

but pair matchings, one of which is indicated by solid lines in Table 1, necessarily omit three of nine available controls. The pair match of Table 1 is optimal among pair matches, in the sense that it minimizes the net discrepancy in log of Grant Funding between matched counterparts; its net discrepancy is $|0-0|+|0-0|+|0-0|+|0-4.4|+|4.4-5.0|+|6.1-6.0|$, or 5.1. An optimal *match with a varying number of controls* (Ming and Rosenbaum 2001) allows for potentially a different number of controls within each matched set, and is also indicated in the table by the solid and dashed lines taken together. This match remedies pair matching's drawback of needlessly rejecting controls, and it increases net discrepancy only slightly, to 5.9.

With full matching, however, each control can be matched while at the same time net discrepancy is sharply *reduced*, from 5.1 (or 5.9) to 1.5. Apart from its ability to make use of more potential controls, Table 2 shows the better performance of full matching, at the same time illustrating the second important reason to prefer full matching to more restrictive techniques. Considered as groups, the treated and the controls may differ enough that some number of the controls have similar treated counterparts only among a smaller number of treated subjects, while some number of treated subjects have only a smaller number of good potential matches among the controls. Tables 1 and 2 represent this possibility compactly, with U–Z as the controls with fewer treated subjects who are good potential matches for them and A–D as the treated subjects with fewer similarly funded counterparts among the controls. By accommodating this structure, full matching aligns treated and control subjects better than either pair matching or matching with a variable number of controls.

This benefit of full matching is of perhaps greater importance than the example would suggest. Such mismatched concentrations of treated and control group members as given in this example are common in propensity-score matching, since in fitting the propensity score prior to matching, one in effect does one's best to separate treated and controls groups on a linear covariate index; see for instance Heckman et al. (1997) or Dehejia and Wahba (1999).

Table 2: An optimal full match of a science department's research-only faculty. (Any full matching that places E,F, and U–Z into the same sets as does the indicated match is also an optimal, but all other full matchings of these subjects lead to net discrepancies on log(Grant) that exceed the optimal value.)

	Women		Men	
Subject	$\log_{10}(\text{Grant})$	Subject	$\log_{10}(\text{Grant})$
A	0	R	0
B	0	S	0
C	0	T	0
D	0	U	4.4
E	4.4	V	5.0
F	6.1	W	5.7
		X	5.9
		Y	6.0
		Z	6.3

1.2 Recovering the benefits of matching with a fixed number of controls by imposing special restrictions

To obtain optimal full matches, we shall translate matching problems into network flow problems. Simpler algorithms that avoid this detour, such as greedy matching, may appear to enjoy certain advantages over optimal full matching. Consider the following special requirements, for example; one can imagine simple algorithms that accommodate each, if perhaps not optimally. Can optimal full matching do the same?

The statistician may wish to forbid the matching of certain treatment-control pairs, perhaps those that are dissimilar enough that comparing them would undermine the credibility of the analysis. Forbidding matches across levels of a categorical covariate, such as the department in which a researcher is employed, can be accomplished by matching separately within each level of the covariate, but other prohibitions may not be accommodated in this way. An example is *matching within calipers* (Cochran and Rubin 1973); to impose a grant caliper of one (on the log scale) upon the matching problems in Tables 1 and 2 is to insist that only woman-man pairs whose log grant fundings differ by no more than one may be placed in the same matched set.

The statistician may wish to use only a portion of the control group, perhaps because securing control data necessary for the analysis requires follow-up at some cost.

The statistician may wish to place restrictions on the relative proportions with which treated and control subjects are combined into matched sets, perhaps to control the variability of an estimate based on the matching. If the matching of four treated and four control units into a 3:1 and a 1:3 matched set is no better warranted than their placement into a 2:1, a 1:1, and a 1:2 matched set, then the latter, with its three matched comparisons rather than two, will likely lead to a sharper contrast of treatment and control groups. Similarly, when controls are so plentiful and dispersed as to allow that each treated subject be matched to at least two of them, then the statistician might prefer to insist on this rather than allow matchings with just one comparison unit for some treated subjects, or matches in which multiple treated subjects may share a control.

Rosenbaum's (1991) original full-matching algorithm accommodates only the first of the three challenges listed above. That algorithm is implemented by Gu and Rosenbaum (1993), and by Augursky and Kluve (2004), whose findings indicate that the absence of restrictions on treatment and control subjects' relative proportions, in particular, is problematic, so much so that the latter authors reject it in favor of a sub-optimal, greedy approach. In contrast, the algorithm presented here accepts restrictions of each of the three types, singly or in combination. Particularly when combined, such restrictions may render a matching problem impossible to solve; the algorithms of Section 3 also decide whether this is so, furnishing either a match meeting given restrictions or a report that no such match exists. The match, if it exists, will be optimal. The algorithm is relatively fast as well; the time complexity of the algorithm is competitive with that of previously published algorithms for more restrictive matching techniques (§ 4.2).

2 The place of optimal full matching among bipartite matching methods

2.1 A unified setting for optimal matching of two groups

The algorithm we shall present supports one basic operation: given a treatment and a control group, a list of prohibited matches, a numeric discrepancy rating for each permissible match between treatment and control subjects, a target number of potential controls to be matched to the pool

7

of treated units, and restrictions governing the balance between treated and control units within matched sets, either: find a full match of minimum net discrepancy among full matches meeting the restrictions, if such a match exists; or report that the problem is infeasible. Two treatment-control balance restrictions are required: first, an upper limit on the number of controls to be placed in a matched set; second, either a lower limit, one or greater, on the number of controls to be placed in a matched set, or an upper limit of two or greater on the number of treated units that may be placed into a matched set. Such an algorithm suffices to perform pair matching, matching with multiple controls, full matching as originally conceived (*i.e.* full matching with neither omission of potential controls nor restrictions on the degree of treatment-control unit balance within matched sets), and full matching as it is understood in this paper.

To illustrate the use of this operation, we borrow the syntax of OPTMATCH, an add-on package to R that implements it (Hansen 2004b; R Development Core Team 2004). Thus `fullmatch()` is the operation itself, and `max.controls` denotes an upper restriction on the number of controls to be placed in any one matched set. In full matching, any set with two or more controls has exactly one treated subject, so `max.controls` is also an upper limit on a matched set's ratio of controls to treated subjects. Let `min.controls` be the lower limit on that ratio; this is set to a number less than or equal to 1/2 or greater than or equal to 1 according as matched sets with multiple treated subjects are or are not desired. We assume that either `min.controls` or its reciprocal is a positive integer, as is `max.controls`.

Let T and C be the treatment and control groups, respectively, so that $|C|$ potential controls, *i.e.* as many as C has distinct members, are available to be matched to $|T|$ treated units. Assume for the moment that no treated unit is *isolated*, in the sense that every potential matching of it to a control is forbidden.

Finally let M denote the treatment-by-control group *discrepancy matrix*, whose entries $M[t, c]$ are either positive discrepancies between treatment and control observations t and c or ∞, if the matching of t to c is prohibited. The finite elements of M can be computed in any of a number of ways — propensity score distances and distances based on the Mahalanobis metric are common choices, but are by no means the only choices. See Rosenbaum and Rubin (1985a), Rubin (1979),

Olsen (1997), and Rubin and Thomas (2000) for some examples and properties.

Pair matching. When $|C| \geq |T|$,

$$\texttt{fullmatch}(M, \texttt{min.controls} = 1, \texttt{max.controls} = 1, \texttt{omit.fraction} = r),$$

$r := (|C| - |T|)/|C|$, returns either an optimal pair match, if such exists, or an indication of infeasibility, otherwise.

Feasibility requires that there be at least $|T|$ non-isolated controls, but this is not sufficient for feasibility: as a simple example, let M be

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \infty & \infty \\ 0 & \infty & \infty \end{pmatrix}.$$

Matching with multiple controls. Suppose that a match of each treated unit to, say, two to five controls is sought. Our function permits the specification of a desired fraction of the control group to be matched, which defaults to the full proportion of the control group capable of being included in some full match: i.e., the fraction of control units that are not isolated. Thus

$$\texttt{fullmatch}(M, \texttt{min.controls=2}, \texttt{max.controls=5})$$

returns an optimal matching of each treated unit with two to five controls, placing each non-isolated control into some matched set (if there exists such a matching that is consistent with M). The matching problem specified here may be feasible only if the number of (non-isolated) potential controls is from two to five times the number of (non-isolated) treated subjects, but since the problem places non-trivial restrictions on the sizes of matched sets, such a condition is not sufficient for solutions to the problem to exist. When they do not, $\texttt{fullmatch()}$ returns an indication that the problem was infeasible.

Full matching (with and without restrictions on treatment-control balance). The call

$$\texttt{fullmatch}(M, \texttt{min.controls=.25}, \texttt{max.controls=5}, \texttt{omit=.1})$$

generates an optimal full matching of the treated group to all but 1/10 of the potential controls, with matched sets having as few as one control per four treated subjects and as many five controls per treated subject (or an indication that no such match exists, if appropriate).

If not given an `omit` argument, `fullmatch` seeks to place into matched sets each treated unit and each non-isolated control (thereby using all non-isolated controls). If either of `max.controls` or `min.controls` are not specified, they default to $|C|$:1 or 1:$|T|$, respectively, so that in effect no upper (respectively, lower) restriction is placed on matched sets' ratios of controls to treated units. Thus omitting all arguments other than M prompts `fullmatch()` to find an optimal full matching in the original, unrestricted sense of Rosenbaum (1991).

Provided that there are non-isolated units, some full matching in Rosenbaum's (1991) unrestricted sense always exists. Except in this case, however, deciding the existence of matches of a desired type may require elaborate combinatorial calculations. In pair matching, for instance, the fundamental theorem of Hall (1935) states that the existence of a pair match is guaranteed only when for *any* $T' \subseteq T$, the collection of control subjects eligible to be matched to some member of T' is at least as large as T'; proving or refuting this of a matching setup can be quite demanding. It is helpful in practice to have an algorithm that settles such questions automatically.

2.2 A word on greedy matching

These and other calls to `fullmatch()` resolve definitely the question of whether a match of the specified type exists, and if such matches exist, they find one that is optimal, in the sense that its net discrepancy is within a small, specified tolerance of least-possible. In present practice, matches are more commonly found via so-called greedy algorithms, which compare unfavorably to optimal matching on both of these counts.

Greedy matching takes the following approach, or a variation of it. First, treated and control units are put in some order. In this order, treated units are each matched to the control, or the first of the controls, from which they are least discrepant, progressively generating a pair matching. At

each stage, matches are made without attention to how they affect later possibilities for matching; as a consequence, the algorithm may be incapable of finding permissible matchings, even when they exist. For example, suppose that two women and two men scientists are to be matched, with differences in the log of grant funding as discrepancies and with a grant funding caliper of one, on the log scale; suppose further than one man and one woman have the same grant funding, while the other man's log grant funding is .6 less and the remaining woman's is .6 more. The resulting discrepancy matrix is

$$
\begin{array}{ccc}
 & Y & Z \\
A & 0 & .6 \\
B & .6 & \infty\,;
\end{array}
$$

the '∞' reflecting that B's discrepancy with Z exceeds the caliper. If A is matched first, he'll be matched to Y, leaving B without a potential match other than to Z; but that match is forbidden. Greedy matching fails to find a permissible pair matching, even though one exists. (See § 5 for a comparison of greedy and optimal approaches in an actual application.)

The problem may be remedied in part by repeating the procedure after a random reordering of the list of treated units. Still, when this composite algorithm fails to find a suitable matching, one doesn't know whether the algorithm or the problem fed to it is to blame; and when greedy matching does find a permissible match, it offers no assessment of how close to optimal it is.

3 Full matching and equivalent network flow problems

3.1 Full matches

Let sets T and C of treated and control units be given, with a discrepancy $M[t, c] \in [0, \infty]$ for each $(t, c) \in T \times C$, $M[t, c] = \infty$ indicating a forbidden matching. Set $E = \{(t, c) \in T \times C : M[t, c] < \infty\}$; then (T, C, E) is a *bipartite graph* with vertices or nodes $T \cup C$ and edges E. It can be depicted as a plot with a point for each vertex or node and for each edge $(t, c) \in E$, a line connecting t's point to c's. If $F \subseteq E$, then (T, C, F) is a *subgraph* of (T, C, E).

Matches, and subclassifications more generally, can be associated with subgraphs of particular

types. Elements x and y of $T \cup C$ are said to belong to the same subclass if they fall in the same *connected component* of (T, C, F): *i.e.*, if there is a path in (T, C, F) connecting them, a sequence of edges $\mathbf{e}_1, \ldots, \mathbf{e}_k \in F$, $\mathbf{e}_1 = (x, \cdot)$ or $(\cdot, x)$ and $\mathbf{e}_k = (y, \cdot)$ or $(\cdot, y)$, such that for $i = 2, \ldots, k$, $\mathbf{e}_i$ shares a node with $\mathbf{e}_{i-1}$. A given subgraph (T, C, F), then, induces a partition of $T' \cup C'$, where T' is the set of all $t \in T$ such that some $(t, c) \in F$ and C' is the set of all $c \in C$ such that some $(t, c) \in F$. If the components of the partition each contain one member of T and one or more from C, say, then the subgraph (T, C, F) would encode a match with multiple controls. Table 1, for instance, depicts graphs encoding a pair matching and matches with multiple controls.

The graphs in Table 1 also encode full matchings, as does that depicted in Table 2, among many others that are possible. If an additional edge between subjects B and S were added to the graph of Table 2, however, then it would cease to encode a full matching, since the partition of $\{A, \ldots, F\} \cup \{R, \ldots, Z\}$ encoded by it would contain a component with *both* multiple controls $(\{R, S\})$ *and* multiple treated units $(\{A, B, C\})$. In general, a subclassification in which each subclass contains either precisely one treated unit or precisely one control is a full matching (of some subset of T to some subset of C).

The *diameter* of a connected component of (T, C, F) is the smallest number d such that any two members of the component may be connected by a path of d or fewer edges. In their graph-theoretic representations (T, C, F), subclassifications that are and are not full matchings may be neatly distinguished in terms of the diameters of their connected components.

Lemma 1 (Rosenbaum 1991, Lemma 1) *Let (T, C, E) be a bipartite graph and f the subclassification of $T' \cup C'$ ($T' = \{t \in T : (t, c) \in E, \text{ some } c \in C\}$, $C' = \{c \in C : (t, c) \in E, \text{ some } t \in T\}$) corresponding to it. Then f is a full matching if and only if all connected components of (T, C, E) have diameter two or less.*

Finding full matches, then, amounts to finding graphs all connected components of which have diameter one or two.

3.2 Full matches and network flows

This section and the section that follows present and analyze a network-flow algorithm for solving full matching problems that was introduced by the second author in her Ph.D. thesis (Olsen 1997). We begin with a review of the definition of a network flow.

Let (V, E) be a directed graph: V is a set of nodes and $E \subseteq V \times V$ a set of arcs connecting the nodes. If $\mathbf{e} = (w, v) \in E$, then $\mathbf{e}$ is an arc *into v* and *out of w*. For each $v \in V$, a divergence d is given. v is called a *supply, demand,* or *conservation node,* according as $d_v > 0$, $d_v < 0$, or $d_v = 0$. There is at least one supply and at least one demand node, and the sum of all divergences is assumed to be zero. Upper and lower capacities $U_{\mathbf{e}}$ and $L_{\mathbf{e}}$, $0 \leq L_{\mathbf{e}} \leq U_{\mathbf{e}} \leq \infty$, are associated with each arc $\mathbf{e}$. Together, these numbers define a *network*. A *flow* through the network is a function l of E into $[0, \infty)$ such that: at each arc $\mathbf{e}$, $L_{\mathbf{e}} \leq l(\mathbf{e}) \leq U_{\mathbf{e}}$; and at each node v,

$$\sum_{w:(w,v)\in E} l(w, v) - \sum_{x:(v,x)\in E} l(v, x) = d_v. \tag{1}$$

An *integral flow* is a flow l mapping edges to integers. In a given network, there may or may not exist flows or integral flows.

If in addition a cost $c_{\mathbf{e}} \geq 0$ is given for each $\mathbf{e} \in E$, then the cost of a flow l is $\sum_{\mathbf{e} \in E} c_{\mathbf{e}} l(\mathbf{e})$. The *min-cost integral flow problem* is the problem of finding a flow through a given network, with given edge costs, of minimum total flow cost. We shall establish an equivalence between the problems of finding optimal full matches, under given treatment-control balance restrictions, and of finding min-cost integral flows in certain networks.

Let $\mathbf{f}$ denote a full-matching problem (*f.m.p.*) with treatment and control groups $T = T(\mathbf{f})$ and $C = C(\mathbf{f})$, discrepancy matrix $M = M(\mathbf{f})$, treatment-control balance restrictions $\texttt{max.controls} = \texttt{maxc}(\mathbf{f})$ and $\texttt{min.controls} = \texttt{minc}(\mathbf{f})$, and a target number $\mathbf{n_c}(\mathbf{f})$ of controls to be put into matched sets. Say that $\mathbf{f}$ is *well-specified* if $\texttt{minc}(\mathbf{f}) \cdot |T| \leq \mathbf{n_c}(\mathbf{f}) \leq \texttt{maxc}(\mathbf{f}) \cdot |T|$.

Our algorithm begins by representing $\mathbf{f}$ as a network, $\mathbf{nwk}(\mathbf{f})$. $\mathbf{nwk}(\mathbf{f})$ has $(|T|+|C|+2)$ nodes: a node for each member of T and C as well as two additional nodes, " Sink" and " Overflow." For each $t \in T$ and $c \in C$ such that $M[t, c] < \infty$, there is an arc (t, c) out of t and into c, and the

arc has lower and upper capacities $L_{(t,c)} = 0$ and $U_{(t,c)} = 1$, respectively. For each $c \in C$, an arc of lower capacity 0 and upper capacity 1 goes out of c and into `Sink`. For each $t \in T$ and each $c \in C$, there is an arc out of t or c and into `Overflow`, but the capacities of these arcs depend on the treatment-control balance restrictions accompanying the f.m.p. being modeled.

Let $U = \mathtt{maxc(f)}$ be the maximum, and $L = \max(1, \mathtt{minc(f)})$ the minimum, number of controls per matched set permissible. Let $\tilde{U} = \max(1, 1/\mathtt{minc(f)})$ denote the problem's upper limit on the number of treated subjects per matched set. Arcs from members of T to `Overflow` have lower and upper capacities 0 and $U - L$, respectively, and arcs from members of C to `Overflow` have lower and upper capacities 0 and $\tilde{U} - 1$, respectively. When $\mathtt{minc(f)} \geq 1$, $\tilde{U} = 1$ and arcs of the form $(c, \mathtt{Overflow})$, $c \in C$, have zero capacity — in effect, they are not part of the network. Thus in Figure 1 (§ 3.3), which assumes $\mathtt{minc(f)} \geq 1$, these arcs do not appear; Figure 2 (§ 3.4) treats separately the case that $\mathtt{minc(f)}$ is small enough to make $\tilde{U}$ two or greater.

In the network representation of a f.m.p. $\mathbf{f}$, node divergences d_v depend both on treatment-control balance restrictions and on the number $\mathtt{n_c}$ of control subjects to be matched. Members $t \in T$ become supply nodes, with supply $d_t = +U$; members of C become conservation nodes, *i.e.* $d_c = 0$, all $c \in C$; whereas `Sink` and `Overflow` have supplies $-\mathtt{n_c}$ and $\mathtt{n_c} - U \cdot |T|$, respectively. This completes the definition of $\mathbf{nwk(f)}$ in terms of $\mathbf{f}$.

3.3 Properties of f.m.p.'s as network flows

If l is an integral flow for $\mathbf{nwk(f)}$, let $(T, C, \{(t,c) : l(t,c) = 1\})$ be the induced subgraph of the bipartite graph $(T, C, \{(t,c) : M[t,c] < \infty\})$ and let $\mathbf{sc}(l)$ be the subclassification of $T \cup C$ represented by this graph (*cf.* Section 3.1).

It so happens that the network representations of f.m.p.'s operate somewhat differently depending on whether the problem *permits many-one matches, i.e.* matched sets with several treated subjects and one control, which it does if and only if $\mathtt{minc(f)} \leq 1/2$. The two variations on full matching require separate analyses; however, one important property is shared by both.

Lemma 2 *Let $\mathbf{f}$ be a well-specified f.m.p. with restrictions. If there is a full matching f solving $\mathbf{f}$, then there is an integral flow l_f for $\mathbf{nwk(f)}$ such that $f = \mathbf{sc}(l_f)$.*

14

The verification of Lemma 2 is not difficult, but it requires systematic attention to contingencies that need not detain us now. It appears in the Appendix.

Lemma 2 entails that for well-specified $\mathbf{f}$, the existence of an integral flow for $\mathbf{nwk(f)}$ is a necessary condition for existence of a feasible solution to $\mathbf{f}$. Yet it is not a sufficient condition for $\mathbf{f}$ to have a solution, because it is not the case that for any f.m.p. $\mathbf{f}$ and for any integral flow l in $\mathbf{nwk(f)}$, $\mathbf{sc}(l)$ is a full match solving $\mathbf{f}$. As a counterexample, consider the matching problem
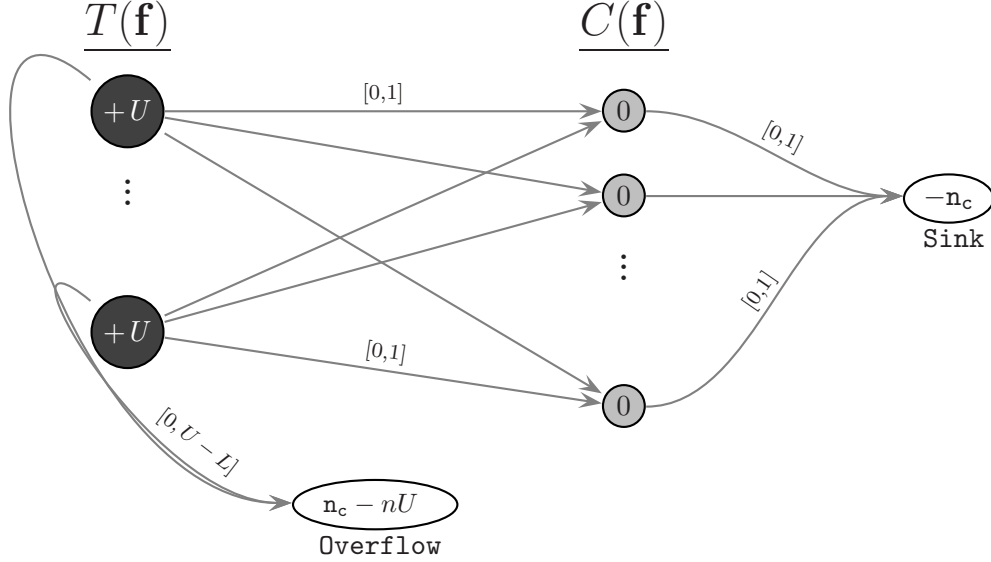
$$\mathbf{f} = (T = \{A, B\}, C = \{C, D\}, \left\{ \begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right\}, \mathtt{maxc} = 2, \mathtt{minc} = 1/2, \mathtt{n_c} = 2),$$

and the flow l through $\mathbf{nwk(f)}$ such that $l(t, \mathtt{Overflow}) = 0$, $t = A$ or B, while $l \equiv 1$ for all other arcs. For such an l, $\mathbf{sc}(l)$ places two treatment and two control units into a single subclass, so it is not a full match. There is a subclass of f.m.p.'s $\mathbf{f}$, however, such that each integral flow l in $\mathbf{nwk(f)}$ induces an admissible solution $\mathbf{sc}(l)$ to $\mathbf{f}$: namely, those problems that *forbid many-one matches*, *i.e.* $\mathbf{f}$ with $\mathtt{minc} \geq 1$. Thus Lemma 2 and the lemma to follow, Lemma 3, together establish that, at least for problems $\mathbf{f}$ that forbid many-one matches, existence of an integral flow for $\mathbf{nwk(f)}$ is necessary and sufficient for existence of a solution to $\mathbf{f}$. For such problems, the question of existence can be decided by any algorithm capable of deciding whether a network admits an integral flow.

Lemma 3 *Let $\mathbf{f}$ be a f.m.p. that forbids many-one matches ($\mathtt{minc}(\mathbf{f}) \geq 1$). If l is an integral flow for $\mathbf{nwk(f)}$, then $\mathbf{sc}(l)$ is a full match placing precisely $\mathtt{n_c}(\mathbf{f})$ members of $C(\mathbf{f})$, and all members of $T(\mathbf{f})$, into matched sets, all of which contain one member of $T(\mathbf{f})$ and between $\mathtt{minc}(\mathbf{f})$ and $\mathtt{maxc}(\mathbf{f})$ members of $C(\mathbf{f})$.*

Proof. Let $\mathbf{f}$ be a f.m.p. that forbids many-one matches, and let l be an integral flow for $\mathbf{nwk(f)}$. To meet its demand, $\mathtt{Sink}$ draws $\mathtt{n_c}(\mathbf{f})$ units of flow, and by the capacity constraints of arcs into $\mathtt{Sink}$, each of these units must come from a distinct control node; so there are at least $\mathtt{n_c}(\mathbf{f})$ control nodes across which l transports at least one unit of flow. But l can transport no flow other than this across control nodes, because there is nowhere else for it to go. (Recall that when many-one matches are forbidden, arcs from control nodes to $\mathtt{Overflow}$ have upper capacity 0.) So

15

Figure 1: Network $\mathbf{nwk}(\mathbf{f})$ for a f.m.p. $\mathbf{f}$ that forbids many-one matches. $U := \mathtt{maxc}(\mathbf{f})$, $L := \mathtt{minc}(\mathbf{f})$, and $n := |T(\mathbf{f})|$. Nodes are labeled with their divergences; arcs, with upper and lower capacity limits. Because their upper capacities are zero, arcs from members of $C(\mathbf{f})$ to $\mathtt{Overflow}$ are not shown.

$\mathbf{sc}(l)$ places precisely $\mathbf{n_c}(\mathbf{f})$ control units into subclasses containing a treated unit. In addition, each of the $\mathbf{n_c}(\mathbf{f})$ control nodes transporting flow receives only one unit of flow, which by integrality of l comes from one treatment node. Should $\mathtt{maxc}(\mathbf{f})$ be greater than one, this treatment node may send flow to other control nodes, but each of these other control nodes itself receives flow from no treatment nodes other than it; this shows that the connected components of the bipartite graph $(T(\mathbf{f}), C(\mathbf{f}), \{(t,c) \in T(\mathbf{f}) \times C(\mathbf{f}) : l(t,c) = 1\})$ have diameter no greater than two. It follows by Lemma 1 that $\mathbf{sc}(l)$ is a full matching. Finally, capacity constraints on the arcs out of treatment nodes are so arranged that l must send a one unit of flow to between $\mathtt{minc}(\mathbf{f})$ and $\mathtt{maxc}(\mathbf{f})$ control nodes. $\square$

3.4 Full matching as a min-cost flow problem

For the case of f.m.p.'s $\mathbf{f}$ that forbid many-one matches, § 3.3 showed that solutions to $\mathbf{f}$ can be found if and only if a corresponding network admits an integral flow. However, for an f.m.p. $\mathbf{f}$ that permits many-one matches, it did not address how to decide whether a matching solving $\mathbf{f}$ exists; nor did it indicate how to find optimal solutions to f.m.p.'s of either type. It so happens that in

addressing the second of these problems, one solves the first as well.

Let $\mathbf{f}$ be a f.m.p. and $\mathbf{nwk}(\mathbf{f})$ the corresponding network. Let $\mathbf{c_f}$ be a function assigning cost to arcs (v, w) in $\mathbf{nwk}(\mathbf{f})$ as follows:

$$c(v, w) = \begin{cases} M(\mathbf{f})[v, w], & v \in T(\mathbf{f}), w \in C(\mathbf{f}); \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

If l is a flow in $\mathbf{nwk}(\mathbf{f})$, write $\mathbf{c_f}(l)$ for the net cost of l, $\sum_{\mathbf{e}} c_{\mathbf{f}}(\mathbf{e})$. Then $\mathbf{mcf}(\mathbf{f}) := (\mathbf{nwk}(\mathbf{f}), \mathbf{c_f})$ encodes the min-cost integral flow problem corresponding to $\mathbf{f}$: determine whether the set $\{c_{\mathbf{f}}(l) : l$ an integral flow in $\mathbf{nwk}(\mathbf{f})\}$ is empty, and if not, find an l minimizing it.

Lemma 4 *Let f be a solution of $\mathbf{f}$ and l an integral flow in $\mathbf{nwk}(\mathbf{f})$ such that $\mathbf{sc}(l) = f$. Then*
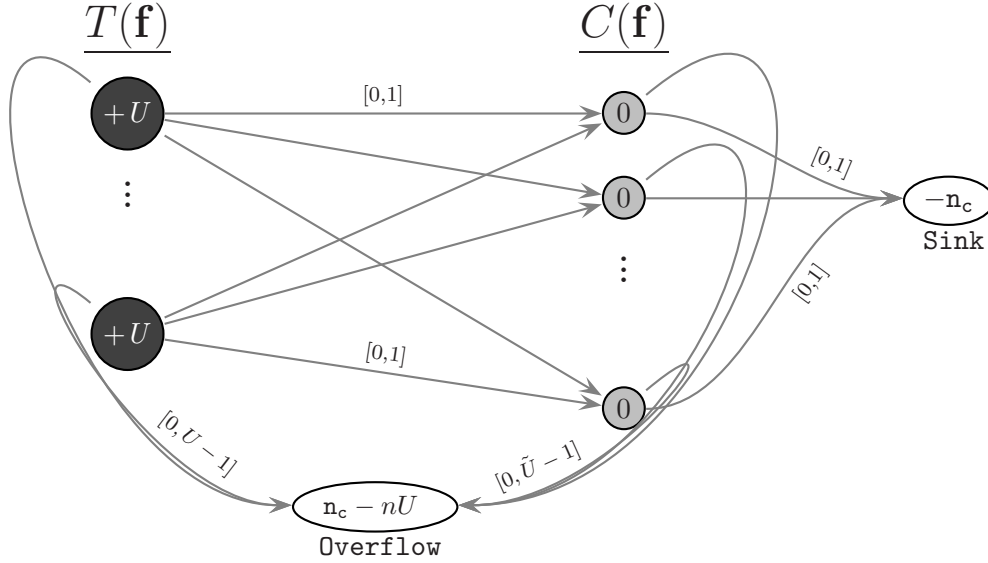
$$\text{Net discrepancy of } f = c_{\mathbf{f}}(l).$$

Proof. By (2), l incurs cost only when it moves flow along arcs from members of $T(\mathbf{f})$ to members of $C(\mathbf{f})$, in which case the costs are proportional to the discrepancy assigned by $\mathbf{f}$ to that pair of units. By $\mathbf{nwk}(\mathbf{f})$'s capacity constraints on these arcs, an integral flow can move either one or no units of flow along such an arc, incurring a cost of 0 or the cost of the arc; and if $\mathbf{sc}(l)$ is known to be a full match, then for any $t \in T(\mathbf{f})$ and $c \in C(\mathbf{f})$ that $\mathbf{sc}(l)$ places in the same subclass, $l(t, c) = 1.$□

In light of Lemmas 2 and 3, Lemma 4 establishes that for f.m.p.'s forbidding many-one matches, finding an optimal solution or establishing that no solution exists can be accomplished by solving the corresponding min-cost integral flow problem. Because Lemma 3 applies only to problems forbidding many-one matches, the same equivalence is not yet assured for all full matching problems with treatment-control balance restrictions. To establish it, we must rule out the possibility of minimum-cost flows l in $\mathbf{mcf}(\mathbf{f})$ which induce subclassifications $\mathbf{sc}(l)$ that are not full matches. This possibility may be excluded for many, but not all, f.m.p.'s that permit many-one matches.

Lemma 5 *Let $\mathbf{f}$ be a f.m.p. such that: (i) $\mathtt{minc}(\mathbf{f}) \leq 1/2$ (many-one matches are permitted); and (ii) all entries of $M(\mathbf{f})$ are strictly positive. If l is a minimum-cost flow for $\mathbf{mcf}(\mathbf{f})$, then $\mathbf{sc}(l)$ is a full match placing all members of $T(\mathbf{f})$ and from $\mathbf{n_c}(\mathbf{f})$ to $|C(\mathbf{f})|$ members of $C(\mathbf{f})$ into matched sets, each containing 1 to $\mathtt{maxc}(\mathbf{f})$ controls and 1 to $1/\mathtt{minc}(\mathbf{f})$ treated units.*

Figure 2: Network $\mathbf{nwk}(\mathbf{f})$ for a f.m.p. $\mathbf{f}$ that permits many-one matches. $U = \mathtt{maxc}(\mathbf{f})$, $\tilde{U} = (\mathtt{minc}(\mathbf{f}))^{-1}$, and $n := |T(\mathbf{f})|$. Nodes are labeled with their divergences; arcs, with upper and lower capacity limits.

Proof. Suppose that l is an integral flow in $\mathbf{nwk}(\mathbf{f})$ such that $\mathbf{sc}(l)$ is not a full matching. We shall show that $\mathbf{c}(l)$ is not least-possible. Lemma 1 entails that the graph $(T, C, \{(t, c) \in T \times C : l(t, c) = 1\})$ has a connected component of diameter greater than two, since $\mathbf{sc}(l)$ is not a full match. So there must be $t_1, t_2 \in T$, $c_1, c_2 \in C$ such that $l(t_1, c_1) = l(t_2, c_1) = l(t_2, c_2) = 1$. Because the arc out of c_1 into $\mathtt{Sink}$ has upper capacity 1, l must send at least one unit of flow along $(c_1, \mathtt{Overflow})$. Let l' be the same as l, except that $l'(t_2, c_1) = 0$, $l'(t_2, \mathtt{Overflow}) = l(t_2, \mathtt{Overflow}) + 1$, and $l'(c_1, \mathtt{Overflow}) = l(c_1, \mathtt{Overflow}) - 1$. These changes do not affect the divergence condition (1). Since l' sends at least one unit of t_2's supply to a member of C, it can send as much as necessary of its remaining supply to $\mathtt{Overflow}$ without violating the capacity constraint on the arc $(t_2, \mathtt{Overflow})$. Finally, $\mathbf{c}(l') = \mathbf{c}(l) - c_{(t_2, c_1)}$, and by assumption $c_{(t_2, c_1)} > 0$, so $c(l)$ is not minimal.

This shows that if l is a min-cost solution to $\mathbf{mcf}(\mathbf{f})$, then $\mathbf{sc}(l)$ is a full matching. In addition,

whenever l solves $\mathbf{mcf}(f)$ and $\mathbf{sc}(l)$ is a full matching, $\mathbf{sc}(l)$ places between $\mathbf{nc}$ and $|T|$ controls into matched sets and respects $\mathbf{f}$'s restrictions on treatment-control balance; the proof of this is routine and is omitted.□

Taken together, these lemmas establish the equivalence of the problem of determining whether a well-specified f.m.p. $\mathbf{f}$ admits a solution and, if it does, producing an optimal solution, with a certain related min-cost flow problem, $\mathbf{mcf}(\mathbf{f})$. Note that we have given a slightly different interpretation to $\mathbf{f}$'s target for the number of controls to be included, $\mathbf{n_c}(\mathbf{f})$, depending on whether $\mathbf{minc}(\mathbf{f}) \geq 1$. If $\mathbf{f}$ forbids many-one matches, then we ask whether a solution matching *precisely* $\mathbf{n_c}(\mathbf{f})$ controls exists, reporting that the problem is infeasible if not. If $\mathbf{f}$ permits many-one matches, then we ask for a solution matching *at least* $\mathbf{n_c}(\mathbf{f})$ control units, reporting infeasibility if and only if this less restrictive class is empty. Of course, for problems asking that all controls be matched, *i.e.* $\mathbf{f}$ with $\mathbf{n_c}(\mathbf{f}) = |C|$, this distinction disappears.

4 Practical considerations

4.1 Incrementing treatment-control discrepancies for stability

Lemma 5 reduces f.m.p.'s to min-cost flow problems only if all discrepancies between treated and control units are strictly positive; but surely problems in which some discrepancies are zero also arise. A simple solution is to modify the initial problem by adding some fixed, positive quantity ϵ to each entry of M: Given a f.m.p. $\mathbf{f}$ and $\epsilon > 0$, let $\mathbf{f}_\epsilon$ be the f.m.p. with $T(\mathbf{f}_\epsilon) = T(\mathbf{f})$, $C(\mathbf{f}_\epsilon) = C(\mathbf{f})$, $M(\mathbf{f}_\epsilon) = M(\mathbf{f}) + \epsilon$, $\mathbf{maxc}(\mathbf{f}_\epsilon) = \mathbf{maxc}(\mathbf{f})$, $\mathbf{minc}(\mathbf{f}_\epsilon) = \mathbf{minc}(\mathbf{f})$, and $\mathbf{n_c}(\mathbf{f}_\epsilon) = \mathbf{n_c}(\mathbf{f})$.

Certainly this change does not affect the answer to the existence question: there exist full matches solving $\mathbf{f}$ if and only if solutions for $\mathbf{f}_\epsilon$ exist. If solutions do exist, then whether $\mathbf{f}$ and $\mathbf{f}_\epsilon$ have the same optimal solutions depends on whether the problems permit many-one matches and on the magnitude of ϵ relative to the entries in M.

For well-specified problems $\mathbf{f}$ that forbid many-one matches, it can be shown that the optimal solution for $\mathbf{f}_\epsilon$ is the same as that for $\mathbf{f}$, whatever be $\epsilon > 0$. The situation is different with problems that permit many-one matches. Let $\mathbf{ms}(f)$ be the number of matched sets in a full matching f and

nc(f) the number of controls matched to some treated unit by f. Unlike with problems forbidding many-one matches, the number of treatment-control discrepancies contributing to a match's net discrepancy, $|T(\mathbf{f})| + \mathbf{nc}(f) - \mathbf{ms}(f)$, may vary, even among matches all of which place the same number of controls into matched sets; this means that optimum solutions to $\mathbf{f}$ and $\mathbf{f}_\epsilon$ may differ. As an example, suppose $T = \{A, B, C\}$ is to be matched with $C = \{X, Y, Z\}$, and suppose that $M[A, Z] = M[B, X] = M[B, Y] = M[C, X] = M[C, Y] = 1$, and that other discrepancies are 0. If f_1 places A, B, and X into one matched set and C, Y, and Z into a second, while f_2 matches A to X, B to Y, and C to Z, then $d(f_1) = 0$ and $d(f_2) = 1$, so f_1 is superior to f_2. Increasing all discrepancies by ϵ, however, $d_\epsilon(f_1) = 4\epsilon$ and $d_\epsilon(f_2) = 1 + 3\epsilon$, so that increments of $\epsilon > 1$ yield f.m.p.'s $\mathbf{f}_\epsilon$ for which f_2 is the optimal solution, not f_1.

Such a shift can be avoided by setting ϵ sufficiently small. However, a larger ϵ is sometimes beneficial: if a many-one and a one-many matched set are replaced with three or more matched sets, then the replacing sets will generally have better treatment-control balance, and they are sure to support at least one additional matched comparison. For this reason, we call ϵ a *stability increment*. In practice, there is a trade-off to be made between, on the one hand, enhancing treatment-control balance — *across* matched sets, homogeneity in their ratios of the number of control to treatment units — and, on the other hand, reducing treatment-control discrepancies — *within* matched sets, dissimilarity among treated and control units. In our R implementation of the algorithm, only a small ϵ is added by default; but the user is free to add a larger stability increment.

4.2 Computational efficiency of full matching via network flows

The min-cost flow problem has been widely studied, and a variety of algorithms for it exist. At their worst, that is on problems specifically designed to exploit their weaknesses, the better algorithms work in pseudopolynomial time, meaning that the number of computations required to execute them is bounded by a polynomial in the number of nodes, N, the number of arcs, A, the greatest arc cost, C, and the largest absolute supply at an arc (Ahuja et al. 1993, p. 340). For example, Bertsekas and Tseng's RELAX-IV algorithm, which is the min-cost flow solver supporting our R function, terminates within $O(NA \log(NC))$ steps, and often much sooner (Bertsekas and Tseng

1994). Now in our networks, $N = |T| + |C| + 2 = n + 2$ and $A \leq |T| \cdot |C| \leq n^2$, so with RELAX-IV a f.m.p. can be solved in no more than $O(n^3 \log(n \max(M[t,c])))$ steps. By comparison, squaring an $n \times n$ matrix in the conventional way involves $O(n^3)$ steps. As n increases, the time required for the most difficult full matching problem with a sample of size n increases faster than does the time required to square a $n \times n$ matrix — but only slightly faster.

For matching with a varying number of controls, *i.e.* for problems that forbid many-one matches, only, Ming and Rosenbaum (2001) describe an optimal matching algorithm with a slightly better time complexity bound, namely $O(n^3)$. We suspect that for many problems, our approach's disadvantage in terms of worst-case complexity is outweighed by its increased flexibility. If run time should become a problem, our procedure's performance on large problems can be substantially improved by imposing a well-chosen caliper r, *i.e.* replacing $\{M[t,c] : t \in T, c \in C\}$ with $\{M[t,c]/[M[t,c] < r] : t \in T, c \in C\}$; the effect of this is to reduce the number of arcs in the network flow problem. (Here, $[condition] = 1$ if *condition* is true, 0 otherwise, and $r/0 = \infty$ for $r \geq 0$.)

5 Example: matching in a gender equity study

Recall the gender equity study described in Section 1.1. As in many studies of people, the conclusions of the study would be of interest to the study subjects themselves. In contrast with typical human subjects, however, ours were themselves scientists, experts in the critique of scientific work; and the results of the study could potentially expose them to moral rebuke. It would be necessary to provide a baseline analysis with as few assumptions and implementation decisions as possible, the justification of which would be transparent and, as nearly as possible, beyond reproach.

To begin, it was decided that for this baseline analysis, men and women would be candidates for comparison only if they held the same rank in the same department, if they differed in total grant funding by a factor of no more than 100, and if they differed in grants awarded to them as principal investigators by no more than a factor of ten. The basis for these decisions will not be reviewed here, except to note that they were relatively stringent, excluding about half of the 886 initial members of the sample, and that they were made after inspection of the covariate characteristics

of the sample, including gender, but before any consideration of outcomes. Every professor with a counterpart meeting these conditions would be included in the comparison. So individualized an inclusion criterion can be met within the framework of full matching, but not with other matched designs.

Within the confines thus set, however, a variety of matchings would still be possible. For permissible pairings (t, c), discrepancies $M[t, c]$ must be chosen; we use the Mahalanobis distance between t and c in terms of logs of the grant-funding variables. Then a method of matching must be chosen. The simplest method guaranteeing a match for each professor with a suitable counterpart would be full matching without restrictions. However, declining to impose restrictions on treatment-control balance might lead to matched sets that vary greatly in their relative proportions of treatment and control subjects, even if a configuration closer to k:1 matching could be achieved with little increase in net discrepancy; this variation makes matched inferences less precise.

To minimize variation in the matched sets' treatment-to-control proportions, for each department and rank combination we first used a line search to determine the maximum feasible value of `min.controls`. That is, we systematically varied the value of `min.controls` in calls to `fullmatch()`, which reports on a matching problem's feasibility by default. With `min.controls` set to the number thus determined, the smallest feasible value of `max.controls` was set for each department and rank in a similar fashion. Such a tuning of matched sets' treatment-control balance can substantially reduce the variance of the estimates that result (Hansen 2004a). It would not have been possible with common greedy approaches to matching, because their failure to produce a matching of a desired type does not prove that such a matching does not exist. OPTMATCH provides convenience functions, `minControlsCap` and `maxControlsCap`, to perform the necessary line searches.

The matching that resulted supported comparisons of lab and office space in 24 departments. These were made non-parametrically, using the aligned rank procedure of Hodges and Lehmann (1963), with space assignment in square feet as the outcome measure. (A description of this method can also be found in Section 2.4.3 of Rosenbaum (2002b).) Only one department showed a significant ($p = .01$) difference between men and women, with women having smaller space assignments than

their matched counterparts.

Using our method, each of the four women of this department was matched to between three and sixteen men (inclusive). A method of matching that does not allow flexibility in the number of controls per treated subject might have matched in pairs, triples, or 3:1 configurations. The R add-on package MATCHIT, either through a call to OPTMATCH or through a greedy algorithm, finds k:1 matches (Ho et al. 2004b). Using the greedy algorithm and the given data for this department, MATCHIT produces the same 1:1 and 3:1 matches as does our optimal approach. With the pair match, however, the difference between men and women falls just short of statistical significance ($p = .06$). In the 3:1 match, full matching's finding of significance is reproduced. However, when tests are inverted to produce a confidence interval for the effect of gender, putatively additive, 3:1 matching's interval runs from 190 to 2120 square feet in favor of men, while restricted full matching's interval is 120 to 1520 square feet in favor of men — about 30% shorter, a noteworthy improvement of precision. Furthermore, without optimal matching it would have been difficult to determine that 3:1 but not 5:1 matching, for instance, could have respected our constraints on which pairs should and should not be placed in the same matched sets. Once that had been determined, optimal full matching with restrictions still showed a substantial advantage over greedy k:1 matching; but in this case that advantage was due to its added flexibility rather than to its ability to find demonstrably optimal matches with certainty.

References

Agresti, A. (2002), *Categorical data analysis*, John Wiley & Sons.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993), *Network flows: theory, algorithms, and applications*, Englewood Cliffs, NJ: Prentice Hall.

Augursky, B. and Kluve, J. (2004), "Assessing the performance of matching algorithms when selection into treatment is strong," Tech. rep., RWI-Essen.

Bertsekas, D. P. and Tseng, P. (1994), "RELAX-IV: a Faster Version of the RELAX

Code for Solving Minimum Cost Flow Problems," Tech. rep., M.I.T., report P-2276, `mit.edu/dimitrib/www/noc.htm`.

Cochran, W. G. (1968), "The Effectiveness of Adjustment By Subclassification in Removing Bias in Observational Studies," *Biometrics*, 24, 295–313.

Cochran, W. G. and Rubin, D. B. (1973), "Controlling Bias in Observational Studies: A Review," *Sankhyā, Series A, Indian Journal of Statistics*, 35, 417–446.

Connors, A. J., Speroff, T., Dawson, N., Thomas, C., Harrell, F. J., Wagner, D., Desbiens, N., Goldman, L., Wu, A., Califf, R., Fulkerson, W. J., Vidaillet, H., Broste, S., Bellamy, P., Lynn, J., and Knaus, W. (1996), "The Effectiveness of Right Hearth Catheterization in the Initial Care of Critically Ill Patients. SUPPORT Investigators." *Journal of the American Medical Association*, 276, 889–97.

Cox, D. R. and Hinkley, D. V. (2000), *Theoretical Statistics*, Chapman & Hall Ltd.

Dehejia, R. and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062.

Diamond, A. and Sekhon, J. S. (2005), "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," Tech. rep., Department of Government, Harvard University.

Gu, X. and Rosenbaum, P. R. (1993), "Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms," *Journal of Computational and Graphical Statistics*, 2, 405–420.

Hall, P. (1935), "On representatives of subsets," *Journal of the London Mathematical Society*, 10, 26–30.

Hansen, B. B. (2004a), "Full matching in an observational study of coaching for the SAT," *Journal of the American Statistical Association*, 99, 609–618.

— (2004b), OPTMATCH, an add-on package for R.

Harding, D. J. (2003), "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on High School Dropout and Teenage Pregnancy," *American Journal of Sociology*, 109, 676–719.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654.

Ho, D. E. (2005), "Why Affirmative Action Does Not Cause Black Students to Fail the Bar," *The Yale Law Journal*, 114, 1997–2004.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2004a), "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," Tech. rep., Department of Government, Harvard University.

— (2004b), MATCHIT: *Matching Software for Causal Inference.*

Hodges, J. L. and Lehmann, E. L. (1963), "Estimates of location based on rank tests (Ref: V42 p1450-1451)," *The Annals of Mathematical Statistics*, 34, 598–611.

Hosmer, D. W. and Lemeshow, S. (2000), *Applied Logistic Regression*, John Wiley & Sons.

Mantel, N. and Haenszel, W. (1959), "Statistical aspects of the analysis of data from restrospective studies of disease," *Journal of the National Cancer Institute*, 22, 719–748.

Ming, K. and Rosenbaum, P. R. (2001), "A note on optimal matching with variable controls using the assignment algorithm," *Journal of Computational and Graphical Statistics.*

Olsen, S. (1997), "Multivariate Matching with Non-Normal Covariates in Observational Studies," Ph.D. thesis, University of Pennsylvania.

R Development Core Team (2004), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-00-3.

Raudenbush, S. W. and Bryk, A. S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications Inc.

Rosenbaum, P. R. (1991), "A Characterization of Optimal Designs for Observational Studies," *Journal of the Royal Statistical Society*, 53, 597– 610.

— (2002a), "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, 17, 286–327.

— (2002b), *Observational Studies*, Springer-Verlag, 2nd ed.

Rosenbaum, P. R. and Rubin, D. (1985a), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.

Rosenbaum, P. R. and Rubin, D. B. (1985b), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.

Rubin, D. B. (1976), "Multivariate Matching Methods That Are Equal Percent Bias Reducing. II: Maximums on Bias Reduction for Fixed Sample Sizes (Corr: V32 P955)," *Biometrics*, 32, 121–132.

— (1979), "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, 318–328.

Rubin, D. B. and Thomas, N. (1996), "Matching Using Estimated Propensity Scores: Relating Theory to Practice," *Biometrics*, 52, 249–64.

— (2000), "Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates," *Journal of the American Statistical Association*, 95, 573–585.

Smith, H. (1997), "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies," *Sociological Methodology*, 27, 325–353.

Smith, J. and Todd, P. (2005), "Does Matching Overcome LaLonde's Critique of Nonexperimental Methods," *Journal of Econometrics*, 125, 305–353.

A   Proof of Lemma 2

For each arc (t, c) of $\mathbf{nwk}(\mathbf{f})$, put $l_f = 1$ if f matches t to c, $l_f = 0$ otherwise. This ensures that the subgraph $(T, C, \{(t, c) : l(t, c) = 1\})$ is that representing f, so that $\mathbf{sc}(l_f) = f$, provided that l_f is a feasible flow in $\mathbf{nwk}(\mathbf{f})$. In $\mathbf{nwk}(\mathbf{f})$, arcs out of treatment and into control nodes have upper and lower capacities 0 and 1, respectively, so l_f conforms to $\mathbf{nwk}(\mathbf{f})$'s rules so far.

Complete the definition of l_f as follows. Write $U = \mathtt{maxc}(\mathbf{f})$, $L = \max(1, \mathtt{minc}(\mathbf{f}))$, and $\tilde{U} = \max(1, 1/\mathtt{minc}(\mathbf{f}))$. For each arc from a treatment node t to $\mathtt{Overflow}$, $l_f = U - $ (number of controls matched by f to t). Then U units flow out of t, as required to balance its supply of $d_t = U$, while the flow through the arc $(t, \mathtt{Overflow})$ is within the permissible range $[0, U - L]$. If c is a control node that is not matched by f to any treatment node, then $l_f(c, \mathtt{Overflow}) = l_f(c, \mathtt{Sink}) = 0$. Place the remaining control nodes into some order; in the event that there are control nodes matched by f to as many as $\tilde{U}$ treatment nodes, see to it that these come before others in this ordering. For arcs $\mathbf{e}$ out of these control nodes, set $l_f(\mathbf{e})$ as follows. To begin, if c is a control node matched to $k \geq 1$ treatment nodes, then set $l_f(c, \mathtt{Overflow}) = k - 1$ and $l_f(c, \mathtt{Sink}) = 1$. Continue setting flows out of control nodes c in this fashion until $\mathbf{n_c}(\mathbf{f})$ units of flow have been directed to $\mathtt{Sink}$. Since $\mathbf{n_c}(\mathbf{f})$ is no larger than than the number of control units matched to some treatment unit by f, $\mathtt{Sink}$ will receive its full $\mathbf{n_c}(\mathbf{f})$ units of flow.

At this point, there may remain control nodes, matched by f to some treatment node, the outflows of which have yet to be specified. Note that, if this is the case, then the remaining control units must be matched by f to fewer than $\tilde{U}$ treated units, since otherwise $|T|$ would have to exceed $\tilde{U} \cdot \mathbf{n_c}(\mathbf{f})$, contrary to the assumption that $\mathbf{f}$ is well-specified. For such $c \in C$ set $l_f(c, \mathtt{Sink}) = 0$ and $l_f(c, \mathtt{Overflow}) = k$, where k is the number of treatment units matched to c by f. In each of these cases, $l_f(c, \mathtt{Overflow})$ and $l_f(c, \mathtt{Sink})$ lie in their permitted capacity ranges of $[0, 1]$ and $[0, \tilde{U} - 1]$, respectively, and flow out of c is the same as flow into it. As $\mathtt{Sink}$ receives $\mathbf{n_c}(\mathbf{f})$ units of flow, its inflow matches the demand on it placed by $\mathbf{nwk}(\mathbf{f})$; and $\mathtt{Overflow}$ receives all remaining flow originating from treatment nodes, as required by $\mathbf{nwk}(\mathbf{f})$. $\square$