

Simultaneous Sensitivity Analysis in Stata: `arsimsens` and `pairsimsens`

Daniel Lempert

lemperds@potsdam.edu

Department of Politics

SUNY, Potsdam

Potsdam, NY 13676, USA

Abstract

A simultaneous sensitivity analysis assesses how sensitive an inference of a non-zero treatment effect is to an unobserved confounder with a specified relationship to the treatment and response. Gastwirth et al. (1998) develops a method of simultaneous sensitivity analysis that can be used after 1:1 matching; Small et al. (2009) modifies the method so that it can be applied after 1: k and full matching. This paper describes the commands `pairsimsens` and `arsimsens`, which implement, respectively, the analyses of Gastwirth et al. (1998) and Small et al. (2009) in Stata. The `.ado` and `.hlp` files for the software presented in the paper are provided in a `.zip` file in the supplementary materials.

Keywords: Simultaneous Sensitivity Analysis, Stata, Aligned-Rank Test, McNemar Test, Wilcoxon Signed-Rank Test

1. Introduction

Unlike in a randomized experiment, in an observational study subjects are not assigned to treatments randomly. This presents a challenge for drawing inferences about whether a treatment has an effect on subjects' responses. Particularly problematic are unobserved confounders: unmeasured covariates that may impact both whether a subject is treated and what his response is after treatment. A sensitivity analysis may be used to assess how inferences would change if there existed a confounder with a hypothesized relationship to treatment and/or outcome. Foundational work in sensitivity analysis is reviewed in Rosenbaum (2002, 105-170).

A *primal* sensitivity analysis considers a confounder that has an analyst-specified association with assignment to treatment, assuming a near-perfect relationship with the response (e.g., Rosenbaum, 2010, 65-94). A *dual* sensitivity considers a confounder that has an analyst-specified association with the response, and assumes a near-perfect relationship with assignment to treatment. I discuss below the *simultaneous* sensitivity analysis, which allows the researcher to specify a confounder's relationship to both treatment and response (Gastwirth et al., 1998). Note that the primal and dual sensitivity analyses are special cases of the simultaneous sensitivity analysis. Primal sensitivity analysis for matched pairs is implemented in Stata by Gangl (2004) (for continuous responses) and by Subramanian and Overby (2014) (for binary responses).¹

1. For a recent overview of software available for other, less closely related, types of sensitivity analysis, see Liu et al. (2013, 572).

Stratification of subjects via matching on covariates is a central tool of observational studies; subjects may be stratified into pairs via 1:1 matching, or into larger strata via 1:k or full matching (Hansen, 2004; Hansen and Klopfer, 2006; Rosenbaum, 1991; Rosenbaum, 2010, 163-186). Gastwirth et al. (1998) develops a method of simultaneous sensitivity analysis that can be used after pair matching. Small et al. (2009) significantly modifies the method so that it can be applied after 1:k matching (matching with a fixed number of controls) and full matching (in which strata consist of one treated subject, one control subject, or both). After sketching relevant results in Gastwirth et al. (1998) and Small et al. (2009), I introduce the Stata commands `pairsimsens` and `arsimsens`, which allow the user to assess how sensitive an inference of a non-zero treatment effect is, after pair or full matching.

2. Methods

In this section, I review randomization inference and introduce the notation and assumptions that are common to the sensitivity analyses presented in Gastwirth et al. (1998) and Small et al. (2009). Then, I describe the results used in `pairsimsens` and `arsimsens` from Gastwirth et al. (1998), and consider the applications to the McNemar Test (for binary responses) and Wilcoxon’s signed rank test (for ordinal or continuous responses) after pair matching. Last, I describe the central result in Small et al. (2009), and the application to the aligned-rank test after full matching.

2.1 Randomization Inference and Simultaneous Sensitivity Analysis: Notation and Assumptions

Matching on observed covariates \mathbf{x} constructs I matched sets, $i = 1, \dots, I$ each containing $n_i \geq 2$ units, of which m_i are treated and $n_i - m_i$ are controls. In each set, $m_i = 1$ or $m_i - n_i = 1$, or both. Let $Z_{ij} = 1$ if unit j in set i is treated and $Z_{ij} = 0$ if it is control. r_{Cij} is the response that unit j in set i would exhibit if placed in the control group—call r_{Cij} “response under control.” r_{Tij} is the response that would be exhibited if unit j in set i were exposed to treatment. Only one of r_{Cij} or r_{Tij} is observed for a given unit: call this R_{ij} —response under treatment assignment actually received (e.g., Rosenbaum, 2010, 21-63). Let \mathbf{r}_{Ci} be the n_i -dimensional vector of responses under control in group i , and $\mathbf{r}_C = [\mathbf{r}_{C1}^T, \dots, \mathbf{r}_{CI}^T]^T$. Define \mathbf{R}_i , \mathbf{R} , \mathbf{Z}_i , and \mathbf{Z} analogously.

Fisher’s (1935) null hypothesis of no treatment effect states that $r_{Tij} = r_{Cij}$, $\forall i, j$ (implying $r_{Cij} = R_{ij}$). If treatment assignment within sets is random, this hypothesis can be assessed via randomization inference. In an observational study, there is no random assignment; however, by matching on all relevant pre-treatment covariates \mathbf{x} , we assume that the probability of assignment to treatment within matched groups is equal. (Rosenbaum, 2010, 65-90). Randomization inference can then proceed by “shuffling” treatment assignment, taking every permutation of each \mathbf{Z}_i while holding r_{Cij} fixed, and calculating a test statistic $t(\mathbf{Z}, \mathbf{r}_C)$, for each of these equiprobable shuffles, thereby determining a null distribution of the test statistic. The null hypothesis is tested by calculating a significance level: the proportion of the time $t(\mathbf{Z}, \mathbf{r}_C)$ is greater than or equal to the observed test statistic. For larger data sets, calculating $t(\mathbf{Z}, \mathbf{r}_C)$ for every possible treatment assignment is impractic-

cal, so computational shortcuts to calculate significance levels have been developed (e.g., Rosenbaum, 2010, 21-63).

But randomization inference requires that all relevant pre-treatment covariates have been taken into account when constructing the matched sets. A sensitivity analysis is useful when this requirement may not hold. In general, a sensitivity analysis allows the analyst to describe what the null distribution could look like—and thus, how inference should change—in the presence of an unmeasured pre-treatment covariate, or confounder. Assumptions and some notation for simultaneous sensitivity analysis follow, drawing very heavily on the exposition in Gastwirth et al. (1998) and Small et al. (2009), to which I refer the reader for important detail and more general results.

Consider a binary confounder u_{ij} .² Assume that u_{ij} is the relevant confounder, in that, for every subject, treatment assignment is conditionally independent of response under control, given the observed covariates and u_{ij} . Let \mathbf{u}_i be the n_i -dimensional vector of u_{ij} in group i , and $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_J^T]^T$.

In a simultaneous sensitivity analysis, two sensitivity parameters are specified. The sensitivity parameter γ describes the strength of the relationship between u_{ij} and assignment to treatment, and the sensitivity parameter δ describes the relationship between u_{ij} and response under control, as follows. Assume the following holds in the population before matching: letting $\beta_i(z)$ be an unknown function and $\alpha_i(u_{ij})$ be a normalizing constant,

$$\Pr(Z_{ij} = z | \mathbf{x}_{ij}, u_{ij}) = \exp\{\alpha_i(u_{ij}) + \beta_i(z) + \gamma z u_{ij}\}. \quad (1)$$

Since the treatment considered is binary, $\alpha_i(u_{ij})$ can be computed; letting $\beta_i = \beta_i(1) - \beta_i(0)$ be an unknown number, (1) may be written as,

$$\Pr(Z_{ij} = 1 | \mathbf{x}_{ij}, u_{ij}) = \frac{\exp(\beta_i + \gamma u_{ij})}{1 + \exp(\beta_i + \gamma u_{ij})}, \quad (2)$$

and $\Pr(Z_{ij} = 0 | \mathbf{x}_{ij}, u_{ij}) = 1 - \Pr(Z_{ij} = 1 | \mathbf{x}_{ij}, u_{ij})$.

Assume further that in the population, letting $\zeta_i(u_{ij})$ be a normalizing constant and $\kappa_i(r)$ an unknown function,

$$\Pr(r_{Cij} = r | \mathbf{x}_{ij}, u_{ij}) = \exp\{\zeta_i(u_{ij}) + \kappa_i(r) + \delta r u_{ij}\}. \quad (3)$$

Any number of outcome models, including the logit, multinomial logit, Poisson, normal, and gamma can be written in the form of (3) (see Gastwirth et al., 1998, 909). Note that β_i and κ_i vary with i because they are functions of \mathbf{x}_i . An implication of this model is that $\Gamma = e^\gamma$ is, for any pair of observations with the same values of the observed covariates, but with different values of the confounder, the maximum ratio of the odds that one is treated, to the odds that the other is treated (e.g., Rosenbaum, 2002, 106-109). Also, if responses

2. This assumption is not as limiting as it might seem at first glance. For $0 \leq u_{ij} \leq 1$, Wang and Krieger (2006) show that for matched pairs, $n_i = 2$, the values of u_{ij} that maximize the null distribution of the test statistic are $u_{ij} = 0$ or 1 . Thus, for matched pairs, assuming that u_{ij} is binary is just as conservative as assuming $0 \leq u_{ij} \leq 1$. The restriction that $0 \leq u_{ij} \leq 1$ is just a restriction on the scale of the unobserved covariate, a restriction needed if the numerical scale on the sensitivity parameters γ and δ is to have any meaning (Rosenbaum, 2002, 109). Small et al. (2009, 208-209) show a result related to that of Wang and Krieger (2006): that for matched sets with three subjects, $n_i = 3$, the values of u_{ij} that maximize the null distribution of the test statistics are $u_{ij} = 0$ or 1 .

are binary, $\Delta = e^\delta$ is, for any pair of observations with the same values of the observed covariates, but with different values of the confounder, the maximum ratio of that odds that one has higher response, to the odds that the other has the higher response. If responses are not binary, and one unit has response r and the other $r^* > r$, the maximum ratio is $e^{(r^*-r)\delta}$ (see the discussion in Gastwirth et al., 1998, 909-910, 916). Note that this is the model for the population before matching; in the sample after matching—as detailed in Sections 2.2, 2.3, 3.4 and 3.9—inference for continuous responses will be based not on the responses themselves, but ranks associated with the responses, adjusted to lie in $[0, 2]$.

I next describe simultaneous sensitivity analysis for matched pairs, then for matching with multiple controls or full matching. To simplify discussion, below (and in the software) I restrict attention to $\Gamma \geq 1$ and $\Delta \geq 1$ —that is, a confounder (weakly) positively associated with both treatment and response.³

2.2 Simultaneous Sensitivity Analysis for Matched Pairs

Consider now the case where all strata consist of matched pairs (i.e., $n_i = 2$ and $m_i = 1$, $\forall i$), such that $\mathbf{x}_{ij} = \mathbf{x}_{ik}$, $\forall i$. Following Gastwirth et al. (1998), but slightly modifying notation so as to be consistent with Small et al. (2009), let $r_{Ci(j)}$ and $Z_{i(j)}$ denote the order statistics for the responses under control and treatments in pair i , so that $r_{Ci(2)} \geq r_{Ci(1)}$ and $Z_{i(2)} \geq Z_{i(1)}$; let $\tilde{\mathbf{r}}_C$ and $\tilde{\mathbf{Z}}$ be the $I \times 2$ matrices consisting of the $r_{Ci(j)}$ and $Z_{i(j)}$. Define A_i so that $A_i = 1$ if the treated observation in matched pair i has the (strictly) higher response, and $A_i = 0$ otherwise; let $q_i \geq 0$ be any function of $(\tilde{\mathbf{r}}_C, \tilde{\mathbf{Z}})$. Let p_i be the probability that in pair i , the treated subject has the higher response.⁴ Last, let $\chi(E) = 1$ if event E occurs, and $= 0$ if it does not, and \mathbf{B} be the set that consists of the 2^I distinct vectors of length I that have coordinates 0 or 1.

Consider test statistics of the form $T = t(\mathbf{Z}, \mathbf{r}_C) = \sum A_i q_i$. (As relevant below, McNemar’s statistic and the Wilcoxon signed rank statistic are of this form.) For such statistics, Gastwirth et al. (1998) shows that the probability of obtaining test statistic $T \geq s$, under the null, given confounding, is:

$$\text{pr}(T \geq s | \tilde{\mathbf{Z}}, \tilde{\mathbf{r}}_C) = \sum_{b \in \mathbf{B}} \chi \left(\sum_{i=1}^I b_i q_i \geq s \right) \prod_{i=1}^I p_i^{b_i} (1 - p_i)^{1-b_i}, \quad (4)$$

for any number s . Given a binary confounder and treatment dose, with specified $\gamma \geq 0$ and $\delta \geq 0$, Gastwirth et al. (1998) shows that (4) is maximized when $p_i = \bar{\pi}_i \bar{\theta}_i + (1 - \bar{\pi}_i)(1 - \bar{\theta}_i)$, with

$$\bar{\pi}_i \equiv \frac{1}{1 + \exp\{-\gamma(z_{i(2)} - z_{i(1)})\}} = \frac{\Gamma}{\Gamma + 1}, \quad \bar{\theta}_i \equiv \frac{1}{1 + \exp\{-\delta(r_{Ci(2)} - r_{Ci(1)})\}}. \quad (5)$$

3. This is justified as follows. The ultimate quantity of interest will be the maximum probability of obtaining the observed test statistic, under the null, for specified Γ and Δ . If one of Γ and Δ is less than 1, this probability is the same as for $\Gamma = \Delta = 1$. If both Γ and Δ are less than 1, this probability is the same as for $1/\Gamma$ and $1/\Delta$ (see Gastwirth et al., 1998, 913).

4. This probability is calculated as follows. Let $\pi_i \equiv \frac{1}{1 + \exp\{-\gamma(z_{i(2)} - z_{i(1)})\}}$ and $\theta_i \equiv \frac{1}{1 + \exp\{-\delta(r_{Ci(2)} - r_{Ci(1)})\}}$. Then, $p_i = \text{pr}(A_i = 1 | \tilde{\mathbf{Z}}, \tilde{\mathbf{r}}_C) = \pi_i \theta_i + (1 - \pi_i)(1 - \theta_i)$.

In large samples, (4) is approximated by:

$$1 - \Phi \left(\frac{s - \sum p_i q_i}{\sqrt{\sum p_i (1 - p_i) q_i^2}} \right). \quad (6)$$

2.3 Simultaneous Sensitivity Analysis after Matching with Multiple Controls or Full Matching

Now consider matched sets constructed after matching with multiple controls or full matching—strata may consist of more than two subjects, but may not contain more than one of *both* treated and control observations. As above, the ultimate quantity of interest is the maximum probability of obtaining the observed test statistic $T = t(\mathbf{Z}, \mathbf{r}_C) = \mathbf{Z}^T \mathbf{q}$ under the null, for specified γ and δ . The key step in approximating this probability is finding, for each group i , the values of u_{ij} that maximize μ_i , the expectation of the null distribution (Gastwirth et al., 2000; Small et al., 2009). A formal presentation of the approximation, from Small et al. (2009, 205-206), follows.

For any vector \mathbf{w} , define $\text{Orb}(\mathbf{w})$ as the set containing every vector that can be obtained by permuting the coordinates of \mathbf{w} . Let $\vec{\mathbf{r}}_{C_i}$ be the vector of \mathbf{r}_{C_i} 's coordinates arranged in increasing order. In set i , for $k = 0, \dots, n_i$, define $\tilde{\mathbf{u}}_k$ as the vector with k zeros followed by $n_i - k$ ones. Small et al. (2009, 205-206), relying in part on Gastwirth et al. (2000), show

$$\mu_{ik} = \sum_{\mathbf{z}_i \in \text{Orb}(\mathbf{Z}_i)} \sum_{\mathbf{r}_i \in \text{Orb}(\vec{\mathbf{r}}_{C_i})} \mathbf{z}_i^T \mathbf{q}_i(\mathbf{r}, \mathbf{m}) \frac{\exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{z}_i)}{\sum_{\mathbf{b}_i \in \text{Orb}(\mathbf{Z}_i)} \exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{b}_i)} \frac{\exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{r}_i)}{\sum_{\mathbf{w}_i \in \text{Orb}(\vec{\mathbf{r}}_{C_i})} \exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{w}_i)}$$

and

$$\sigma_{ik}^2 = \sum_{\mathbf{z}_i \in \text{Orb}(\mathbf{Z}_i)} \sum_{\mathbf{r}_i \in \text{Orb}(\vec{\mathbf{r}}_{C_i})} \{ \mathbf{z}_i^T \mathbf{q}_i(\mathbf{r}, \mathbf{m}) - \mu_{ik} \}^2 \frac{\exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{z}_i)}{\sum_{\mathbf{b}_i \in \text{Orb}(\mathbf{Z}_i)} \exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{b}_i)} \frac{\exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{r}_i)}{\sum_{\mathbf{w}_i \in \text{Orb}(\vec{\mathbf{r}}_{C_i})} \exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{w}_i)},$$

with $\mathbf{r} = [\vec{\mathbf{r}}_{C_1}, \dots, \vec{\mathbf{r}}_{C_{i-1}}, \mathbf{r}_i, \vec{\mathbf{r}}_{C_{i+1}}, \dots, \vec{\mathbf{r}}_{C_I}]^T$. Call

$$\mu_{i \max} = \max_{k \in \{0, 1, \dots, n_i\}} \mu_{ik}$$

$$A_i = \{k : \mu_{ik} = \mu_{i \max}\}$$

$$\sigma_{i \max}^2 = \max_{k \in A_i} \sigma_{ik}^2.$$

Then, the maximum probability of obtaining test statistic $T \geq s$, under the null, given confounding, is approximated by:

$$\Pr(T \geq s \mid \vec{\mathbf{r}}_C, \mathbf{m}, \mathbf{X}, \mathbf{u}) = 1 - \Phi \left(\frac{s - \sum_{i=1}^I \mu_{i \max}}{\sqrt{\sum_{i=1}^I \sigma_{i \max}^2}} \right). \quad (7)$$

2.4 Minimum Hodges-Lehmann Point Estimate

Informally, the Hodges-Lehmann point estimate $\hat{\tau}$ is the quantity τ that, when subtracted from the observed responses of treated subjects, gives the test statistic that would be obtained if the null hypothesis were true. The formal definition is more complicated, since there may be no such τ , or there may be infinitely many such τ . Thus, to be more precise, let \bar{t} be the expectation of the test statistic $t(\mathbf{Z}, \mathbf{R})$ if the null hypothesis is true. The Hodges-Lehmann estimate is defined as,

$$\hat{\tau} \equiv \frac{\inf\{\tau : \bar{t} > t(\mathbf{Z}, \mathbf{R} - \tau\mathbf{Z})\} + \sup\{\tau : \bar{t} < t(\mathbf{Z}, \mathbf{R} - \tau\mathbf{Z})\}}{2}. \quad (8)$$

Essentially, if there is no exact solution, $\hat{\tau}$ is the average of the smallest τ that is too large, and the largest τ that is too small (Rosenbaum, 2002, 47-50).

The minimum Hodges-Lehmann point estimate $\hat{\tau}_{\min}$ is the value of τ that, when subtracted from the observed responses of treated subjects, gives the *maximum* possible expectation of the test statistic, if the null hypothesis holds, given a confounder with specified Γ and Δ .⁵ Note that $\hat{\tau}_{\min}$ for $\Gamma = \Delta = 1$ is just $\hat{\tau}$ (Rosenbaum, 2002, 147-148).

3. The Commands: Description and Applications

In this section, I describe the commands `pairsimsens` and `arsimsens`, explaining the syntax, options, output, and saved results. I also illustrate applications, by replicating selected results from Gastwirth et al. (1998) and Small et al. (2009).

3.1 Preliminaries for `pairsimsens`

Each stratum must consist of exactly two subjects: one treated and one control. Data should be organized such that each subject is an observation. (If data is organized such that each observation is a stratum, see the Stata command `reshape`.) `pairsimsens` requires Stata 9 or newer.

3.2 Syntax for `pairsimsens`

```
pairsimsens response [if] [in], treat(treatment) str(stratum) [gam(#) del(#)
mcnemar wsignedrank hlestimate]
```

3.3 Options for `pairsimsens`

`treat(treatment)` takes as its input the name of the treatment variable, which should be a 0/1 variable, such that 1 indicates that an observation is treated, and 0 indicates that an observation is control. Note that this option is required.

`str(stratum)` takes as its input the name of the variable that indicates the stratum to which an observation belongs. This should be an integer variable, and there should be exactly two observations in each stratum: one treated and one control. This option is also

5. This is guaranteed to be the minimum estimate for statistics where \bar{t} is a decreasing function of τ ; all statistics considered here satisfy this criterion.

required.

`gam(#)` takes as its input the value of Γ for the hypothesized unobserved confounder. Recall that $\Gamma = e^\gamma$. If this option is not specified, the default value of 1, indicating no confounding relationship, is used.

`del(#)` takes as its input the value of Δ for the hypothesized unobserved confounder. Recall that $\Delta = e^\delta$. If this option is not specified, the default value of 1, indicating no confounding relationship, is used.

`mcnemar` specifies that McNemar's test statistic is to be used in conducting the sensitivity analysis. If this option is specified, the response should be dichotomous. Specifically, it should be a 0/1 variable, such that 1 indicates the positive outcome.

`wsignedrank` specifies that the Wilcoxon signed ranked test is to be used in conducting the sensitivity analysis. This option should be specified for ordinal or continuous responses.

`hleestimate` returns the minimum Hodges-Lehmann point estimate, given Γ and Δ . If this option is specified, `wsignedrank` should also be specified.

3.4 Output and Applications for `pairsimsens`

Gastwirth et al. (1998, 911-912) describes a case-referent (Rosenbaum, 2002, 7-8) study that assessed whether a certain brand of tampon caused toxic shock syndrome. Both exposure to treatment and response are binary, so McNemar's (1947) test is appropriate to test the null hypothesis of no treatment effect (Rosenbaum, 2002, 31). Ties (concordant pairs of responses) are discarded. McNemar's test statistic is simply the number of pairs where the treated observation has the positive outcome; i.e., $T = t(\mathbf{Z}, \mathbf{r}_C) = \sum A_i q_i$, with $q_i = 1, \forall i$.

The data for this example is reproduced in `gkr98-tab1.dta` (all `.dta` and `.do` files are included as supplementary material). In the example, for all seven of the pairs where outcomes differed and exactly one subject was exposed to the brand of tampon in question, the subject who had toxic shock was exposed. The output for `pairsimsens`, with the `mcnemar` option specified, is shown below. The specified values of Γ and Δ , and the number of discordant matched pairs in the data are displayed, along with the (exact) maximum probability of obtaining at least the observed test statistic (here, 7), under the null.

```
. pairsimsens resp, treat(treat) str(strat) g(1) d(1) mcnemar
```

```
SIMULTANEOUS SENSITIVITY ANALYSIS FOR MCNEMAR TEST
```

```
Gamma: 1
Delta: 1
Discordant Matched Pairs: 7
max[p(t>=7|u)]: .0078125
```

Were this a randomized experiment, the probability that exposure to treatment is without effect would be .0078. To evaluate sensitivity of this result to an unobserved confounder (and to display the results compactly), run the file `gkr98-tab1-rep.do`. Replicating Table 1 in Gastwirth et al. (1998, 911), this gives:

| Gamma | Delta | | | | |
|-----------|----------|-----------|-----------|-----------|-----------|
| | 1 | 2 | 3 | 4 | 1.701e+38 |
| 1 | .0078125 | .0078125 | .0078125 | .0078125 | .0078125 |
| 2 | .0078125 | .016334 | .02298357 | .0279936 | .05852766 |
| 3 | .0078125 | .02298357 | .0372529 | .04902228 | .13348389 |
| 4 | .0078125 | .0279936 | .04902228 | .06722989 | .2097152 |
| 1.701e+38 | .0078125 | .05852766 | .13348389 | .2097152 | 1 |

As Gastwirth et al. (1998, 912) explains, the inference of a non-zero effect is not sensitive to (for example) a confounder that doubled the odds of both exposure and toxic shock: for $\Gamma = \Delta = 2$, the maximum p is .016. However, a confounder that quadrupled the odds of both exposure and toxic shock would cast the inference into doubt: for $\Gamma = \Delta = 4$, the maximum p is .067.

Gastwirth et al. (1998, 915-918) also describes a study in which lead levels in the blood (measured in $\mu g/dl$) of matched pairs of children were compared. In each pair there was one treated child, whose parent worked in a certain battery manufacturing factory, and one control child, whose parents worked elsewhere. With matched pairs, where treatment is binary, and response is continuous, the Wilcoxon (1945) signed-rank test is appropriate to test the null hypothesis (Rosenbaum, 2002, 32). The test statistic is the ranks of the absolute differences in responses for each pair, summed over all pairs in which the treated observation has the higher response. That is, $T = t(\mathbf{Z}, \mathbf{r}_C) = \sum A_i q_i$, with q_i the rank of $r_{Ci(2)} - r_{Ci(1)}$ from 1 to I .

Gastwirth et al. (1998, 915-916) recommends, and `pairsimsens` implements, the following modification. In the definition of $\bar{\theta}_i$, given in (5), replace $r_{Ci(2)} - r_{Ci(1)}$ with $2q_i/I$. Note that approximately half of all pairs have $2q_i/I$ between 0 and 1, and approximately half have $2q_i/I$ between 1 and 2. This allows Δ to be conveniently interpreted as the maximum odds that, for the pair with the median value of $r_{Ci(2)} - r_{Ci(1)}$, the observation that has the higher value of the unobserved confounder also has the higher response (for other advantages of this approach, see Gastwirth et al., 1998, 915-916).

The data for the example is reproduced in `gkr98-tab4.dta`. The output for `pairsimsens`, with the `wsignedrank` and the `hlestimate` options specified, is shown below. In addition to the specified Γ and Δ , the output gives $\sum p_i q_i$ (“Expectation”), $\sum p_i(1 - p_i)q_i^2$ (“Variance”), $\frac{T - \sum p_i q_i}{\sqrt{\sum p_i(1 - p_i)q_i^2}}$ (“Deviate”), the (approximated) maximum probability of obtaining at least the observed test statistic (here, 499) under the null (“`max[p(t>=499|u)]`”), and the minimum Hodges-Lehmann point estimate (“`min[H-L Point Estimate]`”).

```
. pairsimsens resp, t(treat) s(strat) g(1) d(1) hle ws
```

SIMULTANEOUS SENSITIVITY ANALYSIS FOR WILCOXON S-R TEST

Gamma: 1
Delta: 1
Expectation: 264
Variance: 2858.625
Deviate: 4.395307544374739
max[p(t>=499|u)]: 5.53080153576e-06
min[H-L Point Estimate]: 15.5

Assuming no confounding, the standardized deviate is 4.40, the associated p -value is 5.531×10^{-6} , and the Hodges-Lehmann point estimate is 15.5. To evaluate the impact of an unobserved confounder on the inference of a non-zero effect, and on the minimum Hodges-Lehmann point estimate, run `gkr98-tab4-5-rep.do`. This gives, for selected values of Γ and Δ , the minimum standardized deviate (replicating Table 4 in Gastwirth et al. (1998, 916)), the associated p -values, and the minimum Hodges-Lehmann point estimates (replicating Table 5 in Gastwirth et al. (1998, 918)):

. table Gamma Delta, contents(mean dev)

| Gamma | Delta | | | |
|-----------|-----------|-----------|-----------|-----------|
| | 1 | 3 | 5 | 1.701e+38 |
| 1 | 4.3953075 | 4.3953075 | 4.3953075 | 4.3953075 |
| 3 | 4.3953075 | 3.0825899 | 2.7865236 | 2.2244857 |
| 5 | 4.3953075 | 2.6927099 | 2.2980715 | 1.4805171 |
| 1.701e+38 | 4.3953075 | 1.9127413 | 1.2050303 | |

. table Gamma Delta, contents(mean pval)

| Gamma | Delta | | | |
|-----------|-----------|-----------|-----------|-----------|
| | 1 | 3 | 5 | 1.701e+38 |
| 1 | 5.531e-06 | 5.531e-06 | 5.531e-06 | 5.531e-06 |
| 3 | 5.531e-06 | .00102604 | .00266384 | .01305789 |
| 5 | 5.531e-06 | .0035437 | .01077886 | .06936765 |
| 1.701e+38 | 5.531e-06 | .02789059 | .11409581 | |

. table Gamma Delta, contents(mean hle)

| Gamma | Delta | | | |
|-----------|-------|------|------|-----------|
| | 1 | 3 | 5 | 1.701e+38 |
| 1 | 15.5 | 15.5 | 15.5 | 15.5 |
| 3 | 15.5 | 15.5 | 15.5 | 15.5 |
| 5 | 15.5 | 15.5 | 15.5 | 15.5 |
| 1.701e+38 | 15.5 | 15.5 | 15.5 | 15.5 |

| Gamma | 1 | 3 | 5 | 1.701e+38 |
|-----------|------|------|------|-----------|
| 1 | 15.5 | 15.5 | 15.5 | 15.5 |
| 3 | 15.5 | 11 | 10 | 8.5 |
| 5 | 15.5 | 10 | 8.5 | 5.5 |
| 1.701e+38 | 15.5 | 7 | 4 | |

As Gastwirth et al. (1998, 916-918) explains, the inference of a non-zero effect, in this example, is quite robust, even in the presence of substantial unobserved confounding. Even a confounder that quintupled both the odds of exposure to treatment and the odds the exposed subject would have the higher response in a pair, would not affect the inference of a non-zero effect (for $\Gamma = \Delta = 5$, $p = .011$). There would, however, be a substantial decrease in the estimated minimum effect size: for such a confounder, the minimum Hodges-Lehmann estimate is $8.5 \mu\text{g}/\text{dl}$, compared to the estimate of $15.5 \mu\text{g}/\text{dl}$ without confounding. Nor does a confounder that almost perfectly predicts which observation in a pair has the higher response threaten inference, if it only triples the odds of exposure (for $\Delta = \infty$ and $\Gamma = 3$, $p < .05$). But a confounder that near-perfectly predicts which of a pair has the higher response and quintuples the odds of exposure would threaten inference (for $\Delta = \infty$ and $\Gamma = 5$, $p > .05$)—although in this case, the minimum Hodges-Lehmann estimate is $5.5 \mu\text{g}/\text{dl}$, it is not significantly different from 0.

3.5 Saved Results for `pairsimsens`

The following saved results are available after `pairsimsens`.

After both `mcnemar` and `wsignedrank`:

`r(dev)`, the standardized deviate.

`r(del)`, the value of Δ specified.

`r(gam)`, the value of Γ specified.

`r(tsobs)`, the observed test statistic.

`r(pval)`, the maximum probability that the null holds.

After `mcnemar` only:

`r(apval)`, the maximum probability that the null holds, approximated based on the standardized deviate. (Only for purposes of comparison; in applications, `r(pval)` should be used.)

$\mathbf{r}(\mathbf{p})$, p_i , with $p_i = \bar{\pi}_i\bar{\theta}_i + (1 - \bar{\pi}_i)(1 - \bar{\theta}_i)$. (Note that p_i is constant across i for the McNemar Test.)

$\mathbf{r}(\mathbf{dmp})$, the number of discordant matched pairs.

After wsignedrank only:

$\mathbf{r}(\mathbf{h1r})$, the minimum Hodges-Lehmann point estimate, rounded to three decimal places.

$\mathbf{r}(\mathbf{h1e})$, the minimum Hodges-Lehmann point estimate. (Because `pairsimsens` calculates the two terms in the numerator of (8) with error on the order of 10^{-6} , $\mathbf{r}(\mathbf{h1r})$, not $\mathbf{r}(\mathbf{h1e})$, is reported in the program's output.)

$\mathbf{r}(\mathbf{sig2})$, $\sum p_i(1 - p_i)q_i^2$, with $p_i = \bar{\pi}_i\bar{\theta}_i + (1 - \bar{\pi}_i)(1 - \bar{\theta}_i)$.

$\mathbf{r}(\mathbf{mu})$, $\sum p_i q_i$, with $p_i = \bar{\pi}_i\bar{\theta}_i + (1 - \bar{\pi}_i)(1 - \bar{\theta}_i)$.

$\mathbf{r}(\mathbf{retmat})$, an $I \times 5$ matrix, where the first column is the stratum number, the second is the adjusted rank $2q_i/I$, the third is $\bar{\theta}_i$, the fourth is $\bar{\pi}_i$, and the fifth is p_i , with $p_i = \bar{\pi}_i\bar{\theta}_i + (1 - \bar{\pi}_i)(1 - \bar{\theta}_i)$.

3.6 Preliminaries for `arsimsens`

Data should be organized such that each observation is a subject; each stratum should have exactly one control observation, exactly one treated observation, or both; responses should be ordinal or continuous. `arsimsens` requires Stata 9 or newer. A computational shortcut used by `arsimsens` is described in Appendix A.

3.7 Syntax for `arsimsens`

```
arsimsens response [if] [in], treat(treatment) str(stratum) [gam(#) del(#)
verbose hlestimate]
```

3.8 Options for `arsimsens`

`treat(treatment)` takes as its input the name of the treatment variable, which should be a 0/1 variable, such that 1 indicates that an observation is treated, and 0 indicates that an observation is control. Note that this option is required.

`str(stratum)` takes as its input the name of the variable that indicates the stratum to which an observation belongs. This should be an integer variable, and there should not be more than one of *both* treated and control observations in a single stratum. This option is also required.

`gam(#)` takes as its input the value of Γ for the hypothesized unobserved confounder. Recall that $\Gamma = e^\gamma$. If this option is not specified, the default value of 1, indicating no confounding

relationship, is used.

`del(#)` takes as its input the value of Δ for the hypothesized unobserved confounder. Recall that $\Delta = e^\delta$. If this option is not specified, the default value of 1, indicating no confounding relationship, is used.

`verbose` shows the progress of the program’s execution, by displaying the stratum for which $\mu_{i \max}$ and $\sigma_{i \max}^2$ is being calculated.

`hleestimate` gives the minimum Hodges-Lehmann point estimate, for the specified Γ and Δ .

3.9 Output and Applications for `arsimsens`

Small et al. (2009, 206-208) considers the effect of military service on smoking. The treatment is a binary indicator of veteran status, and the outcome is “pack-years:” the number of years that the subject has smoked cigarettes, multiplied by the number of packs per day smoked. After full matching on a set of covariates, subjects are stratified into groups of two, three, or four. When some strata contain more than two subjects, with a binary treatment dose, and a continuous outcome, the Hodges-Lehmann (1962) aligned-rank test is an appropriate test of the null hypothesis (Lehmann 1975; Rosenbaum 2002, 139). Hodges and Lehmann (1962) define the aligned rank as the rank of the aligned responses, where the aligned response is each observation’s response minus the mean response in its stratum. The test statistic $\mathbf{Z}^T \mathbf{q}$ is the sum of the aligned ranks for the treated observations.

Small et al. (2009, 208) recommends, and `arsimsens` implements, the following modification (see also Gastwirth et al., 1998, 915-916): replace each r_{Cij} with its aligned rank, divided by half the maximum of the aligned ranks. Then, let $\mathbf{q}(\mathbf{r}_C) = \mathbf{r}_C$. This modification means that approximately half of all observations have r_{Cij} between 0 and 1, and approximately half have r_{Cij} between 1 and 2. In turn, this allows Δ to be conveniently interpreted as the maximum odds that, for the “average pair” of observations in the same stratum, the observation that has the higher value of the unobserved confounder also has the higher response; the *average pair* of observations is average in the sense of having an average difference in responses: one observation has aligned rank at the 25th percentile and the other at the 75th.

The data for the example is reproduced in `Smalletal.dta`. The output for `arsimsens` is shown below. In addition to the specified Γ and Δ , the output gives $\sum_{i=1}^I \mu_{i \max}$ (“Expectation”), $\sum_{i=1}^I \sigma_{i \max}^2$ (“Variance”), $\frac{T - \sum_{i=1}^I \mu_{i \max}}{\sqrt{\sum_{i=1}^I \sigma_{i \max}^2}}$ (“Deviate”), the maximum probability of obtaining at least the observed test statistic (here, 1347.3) under the null (“max [p(t>=1347.3|u)]”), and the minimum Hodges-Lehmann point estimate (“min[H-L Point Estimate]”).

```
. arsimens resp, t(tr) s(strat) gam(1) del(1) hle
```

```
SIMULTANEOUS SENSITIVITY ANALYSIS FOR ALIGNED RANK TEST
```

```

Gamma: 1
Delta: 1
Expectation: 1273.335325877381
Variance: 347.4356434880523
Deviate: 3.968060923335258
max[p(t>=1347.29849595064|u)]: .0000362299060822
min[H-L Point Estimate]: 2.857

```

In the absence of confounding, the probability of obtaining at least the observed test statistic, given the null, is $p = .00004$, and the Hodges-Lehmann point estimate is 2.857. To evaluate the impact of an unobserved confounder on the inference of a non-zero effect, and on the minimum Hodges-Lehmann point estimate, run `smalletal-rep.do`. This gives, for specified values of Γ and Δ , the maximum probability that the null holds (replicating Table 5 in Small et al. (2009, 208)), as well as the minimum Hodges-Lehmann point estimates:

```
. table Gamma Delta, contents(mean pval)
```

| Gamma | Delta | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 1 | 1.1 | 1.2 | 1.5 | 2.5 | 1.701e+38 |
| 1 | .00003623 | .00003623 | .00003623 | .00003623 | .00003623 | .00003623 |
| 1.1 | .00003623 | .00004983 | .00006622 | .0001316 | .00046241 | .00363386 |
| 1.2 | .00003623 | .00006628 | .00011257 | .0003906 | .0032933 | .06463972 |
| 1.5 | .00003623 | .00013385 | .00040034 | .00428191 | .11298419 | .92932353 |
| 2.5 | .00003623 | .00054009 | .00418381 | .14281653 | .98001349 | 1 |
| 1.701e+38 | .00003623 | .01397183 | .27702887 | .99979383 | 1 | 1 |

```
. table Gamma Delta, contents(mean hle)
```

| Gamma | Delta | | | | | |
|-----------|-------|-------|-------|-------|-------|-----------|
| | 1 | 1.1 | 1.2 | 1.5 | 2.5 | 1.701e+38 |
| 1 | 2.857 | 2.857 | 2.857 | 2.857 | 2.857 | 2.857 |
| 1.1 | 2.857 | 2.667 | 2.6 | 2.5 | 2.333 | 1.6 |
| 1.2 | 2.857 | 2.6 | 2.5 | 2.4 | 1.667 | .5 |
| 1.5 | 2.857 | 2.5 | 2.4 | 1.6 | 0 | -.333 |
| 2.5 | 2.857 | 2.286 | 1.6 | 0 | -1 | -7 |
| 1.701e+38 | 2.857 | 1.143 | 0 | -2.5 | -10.5 | -37.143 |

As Small et al. (2009, 208) explains, the inference that military service affects smoking behavior is somewhat sensitive to an unobserved confounder. For example, a confounder that increased by 20% both the odds of serving in the military, and the odds of having the higher number of pack-years (for the average pair of observations), would not threaten the

inference of a non-zero effect (for $\Delta = \Gamma = 1.2$, $p = .0001$). But, a confounder that increased the odds of serving in the military by 50%, and the odds of having higher pack-years by 150%, would call the inference into doubt (for $\Gamma = 1.5$, $\Delta = 2.5$, $p = .113$); in fact, the minimum Hodges-Lehmann point estimate for such a confounder is 0.

3.10 Saved Results for `arsimsens`

The following saved results are available after `arsimsens`.

`r(hlr)`, the minimum Hodges-Lehmann point estimate, rounded to three decimal places.

`r(hle)`, the minimum Hodges-Lehmann point estimate. (Because `pairsimsens` calculates the two terms in the numerator of (8) with error on the order of 10^{-6} , `r(hlr)`, not `r(hle)`, is reported in the program's output.)

`r(sumvar_imax)`, $\sum_{i=1}^I \sigma_{i \max}^2$.

`r(summu_imax)`, $\sum_{i=1}^I \mu_{i \max}$.

`r(del)`, the value of Δ specified.

`r(gam)`, the value of Γ specified.

`r(tsobs)`, the observed test statistic.

`r(dev)`, the standardized deviate.

`r(pval)`, the maximum probability that the null holds

4. Summary

This paper has presented software that allows for simultaneous sensitivity analysis in Stata. The analysis can be performed after full, 1:k, or pair matching, for a binary treatment dose, and both ordinal/continuous and binary responses. For a hypothesized confounder with user-specified relationship to treatment and to response, the maximum probability that the null holds, and the minimum Hodges-Lehmann point estimate may be obtained. The `.ado` and `.hlp` files for the software presented in this paper are provided in a `.zip` file in the supplementary materials.

Acknowledgments

I thank Jeffrey Budziak, Matthew Hitt, Laszlo Lempert, and the reviewers and editors, for advice, discussion, and suggestions.

Appendix A

Consider the problem of calculating μ_{ik} in (7). (The calculation of σ_{ik}^2 is parallel.) Even noting that the denominators can be factored out, the straightforward method of computing μ_{ik} requires approximately $n_i!(n_i) + n_i! + n_i$ calculations, which is impractical for larger strata. I describe here the computational shortcut `arsimsens` uses, which makes feasible the calculation of μ_{ik} for sets of (essentially) any size.

Below, I suppress the subscript i to reduce clutter, and introduce the following modification in terminology, which will be convenient in subsequent discussion: define \mathbf{k} as $n - k$, so that $\tilde{\mathbf{u}}_{\mathbf{k}}$ is the vector with \mathbf{k} ones followed by $n - \mathbf{k}$ zeros. Also, since the two are the same in the application, write $\mathbf{z}^T \mathbf{r}$ for $\mathbf{z}^T \mathbf{q}(\mathbf{r}, \mathbf{m})$. The basic approach will be to re-write the formula for $\mu_{\mathbf{k}}$ in (7) so that each sum, except one, is expressed as (a multiple of) an elementary symmetric polynomial, which can be more quickly calculated.

Define

$$a = \sum_{\mathbf{r} \in \text{Orb}(\vec{\mathbf{r}}_C)} \sum_{\mathbf{z} \in \text{Orb}(\mathbf{Z})} \mathbf{z}^T \mathbf{r} \exp(\gamma \tilde{\mathbf{u}}_{\mathbf{k}}^T \mathbf{z}) \exp(\delta \tilde{\mathbf{u}}_{\mathbf{k}}^T \mathbf{r}). \quad (9)$$

Then, the formula for $\mu_{\mathbf{k}}$ given in (7) can be written as

$$\mu_{\mathbf{k}} = \frac{1}{\sum_{\mathbf{w} \in \text{Orb}(\vec{\mathbf{r}}_C)} \exp(\delta \tilde{\mathbf{u}}_{\mathbf{k}}^T \mathbf{w})} \frac{1}{\sum_{\mathbf{b} \in \text{Orb}(\mathbf{Z})} \exp(\gamma \tilde{\mathbf{u}}_{\mathbf{k}}^T \mathbf{b})} a. \quad (10)$$

Define $y_1 = \exp(\gamma \vec{\mathbf{r}}_{C1})$, $y_2 = \exp(\gamma \vec{\mathbf{r}}_{C2})$, and so on, for $y_1 \dots y_n$. Let $e_{\mathbf{k}}(y_1, y_2, \dots, y_n)$ be the \mathbf{k}^{th} elementary symmetric polynomial, $\sum_{1 \leq j_1 < j_2 < \dots < j_{\mathbf{k}} \leq n} y_{j_1} \dots y_{j_{\mathbf{k}}}$. Then, the first denominator in (10) can be written as

$$\sum_{\mathbf{w} \in \text{Orb}(\vec{\mathbf{r}}_C)} \exp(\delta \tilde{\mathbf{u}}_{\mathbf{k}}^T \mathbf{w}) = \mathbf{k}!(n - \mathbf{k})! e_{\mathbf{k}}(y_1, \dots, y_n). \quad (11)$$

This is useful since it is fast to compute $e_{\mathbf{k}}(y_1, \dots, y_n)$ for $\mathbf{k} = 0, 1, \dots, n$ recursively. `arsimsens` uses the ‘‘Summation Formula’’ (see Baker and Harwell, 1996, 176-178).⁶ The second denominator can be calculated analogously.

Now, let S_n be the set of all permutations of any n elements, and $S_n \times S_n$ the set of all pairs (π, π') of permutations. Divide $S_n \times S_n$ into classes based on the value of $\rho \equiv \pi^{-1} \pi'$; thus, there are $n!$ classes, with each class D containing $n!$ permutation pairs. Then,

$$a = \sum_D \sum_{\substack{(\pi, \pi') \\ \pi^{-1} \pi' = \rho_D}} \pi'(\mathbf{Z})^T \pi(\vec{\mathbf{r}}_C) \exp(\gamma \tilde{\mathbf{u}}_{\mathbf{k}}^T \pi'(\mathbf{Z})) \exp(\delta \tilde{\mathbf{u}}_{\mathbf{k}}^T \pi(\vec{\mathbf{r}}_C)). \quad (12)$$

Note that within a class, $\pi'(\mathbf{Z})^T \pi(\vec{\mathbf{r}}_C)$ is constant, and that in each class there is one pair (π, π') in which π is the identity permutation (such that $\pi(\vec{\mathbf{r}}_C) = \vec{\mathbf{r}}_C$). Also, $\pi^{-1} \pi' = \rho$ means that $\pi' = \pi \rho$. Thus, (12) can be written

$$a = \sum_D \rho_D(\mathbf{Z})^T \vec{\mathbf{r}}_C \sum_{\substack{(\pi, \pi') \\ \pi^{-1} \pi' = \rho_D}} \exp(\tilde{\mathbf{u}}_{\mathbf{k}} \pi(\gamma \rho(\mathbf{Z}) + \delta \vec{\mathbf{r}}_C)). \quad (13)$$

6. For pseudocode, see Rehman and Ipsen (2011, 101).

Denote by $Z_{\rho(j)}$ the j^{th} coordinate of $\rho(\mathbf{Z})$ and let $y_1 = \exp(\gamma Z_{\rho(1)} + \delta \vec{r}_{C1})$, $y_2 = \exp(\gamma Z_{\rho(2)} + \delta \vec{r}_{C2})$, and so on, for y_1, \dots, y_n . Then,

$$a = \sum_D \rho_D(\mathbf{Z})^T \vec{r}_C k!(n-k)! e(y_1, \dots, y_n). \quad (14)$$

And because, with full matching, there are only n distinct permutations of \mathbf{Z} , computing (14) requires calculating only n summands.

References

- Baker, F. and Harwell, M. “Computing Elementary Symmetric Functions and Their Derivatives: A Didactic.” *Applied Psychological Measurement*, 20(2):169–192 (1996).
- Fisher, R. A. *The Design of Experiments*. Edinburgh: Oliver & Boyd (1935).
- Gangl, M. “rbounds: module to perform Rosenbaum sensitivity analysis for average treatment effects on the treated.” [http://fmwww.bc.edu/RePEc/bocode/r/rbounds.ado](http://fmwww.bc.edu/RePEc/bocode/r/rbounds ado) (2004).
- Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. “Dual and Simultaneous Sensitivity Analysis for Matched Pairs.” *Biometrika*, 85(4):907–920 (1998).
- . “Asymptotic Separability in Sensitivity Analysis.” *Journal of the Royal Statistical Society, Series B*, 62(3):545–555 (2000).
- Hansen, B. B. “Full Matching in an Observational Study of Coaching for the SAT.” *Journal of the American Statistical Association*, 99(467):609–618 (2004).
- Hansen, B. B. and Klopfer, S. O. “Optimal Full Matching and Related Designs via Network Flows.” *Journal of Computational and Graphical Statistics*, 15(3):609–627 (2006).
- Hodges, J. and Lehmann, E. “Rank Methods for Combination of Individual Experiments in the Analysis of Variance.” *Annals of Mathematical Statistics*, 33(2):482–497 (1962).
- Lehmann, E. L. *Nonparametrics: Statistical Methods Based on Ranks*. Oakland, CA: Holden Day (1975).
- Liu, W., Kuramoto, J., and Stuart, E. A. “An Introduction to Sensitivity Analysis for Unobserved Confounding in Nonexperimental Prevention Research.” *Prevention Science*, 14(6):570–580 (2013).
- McNemar, Q. “Note on the Sampling Error of the Differences between Correlated Proportions or Percentages.” *Psychometrika*, 12(2):153–157 (1947).
- Rehman, R. and Ipsen, I. “Computing Characteristic Polynomials from Eigenvalues.” *SIAM Journal on Matrix Analysis and Applications*, 32(1):90–114 (2011).
- Rosenbaum, P. R. “A Characterization of Optimal Designs for Observational Studies.” *Journal of the Royal Statistical Society*, 53(3):597–610 (1991).

—. *Observational Studies*. New York, NY: Springer, 2nd edition (2002).

—. *Design of Observational Studies*. New York, NY: Springer (2010).

Small, D., Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. “Simultaneous Sensitivity Analysis for Observational Studies Using Full Matching or Matching with Multiple Controls.” *Statistics and Its Interface*, 2(2):203–211 (2009).

Subramanian, H. C. and Overby, E. “mbsens: module to compute sensitivity metric for matched sample using McNemar’s test.” <http://fmwww.bc.edu/RePEc/bocode/m/mbsens.ado> (2014).

Wang, L. and Krieger, A. M. “Causal Conclusions are Most Sensitive to Unobserved Binary Covariates.” *Statistics in Medicine*, 25(13):2257–2271 (2006).

Wilcoxon, F. “Individual Comparisons by Ranking Methods.” *Biometrics Bulletin*, 1(6):80–83 (1945).