Tilted sensitivity analysis in matched observational studies

Colin B. Fogarty *

Abstract

We present a new procedure for conducting a sensitivity analysis in matched observational studies. For any candidate test statistic, the approach defines tilted modifications dependent upon the proposed strength of unmeasured confounding. The framework subsumes both (i) existing approaches to sensitivity analysis for sign-score statistics; and (ii) sensitivity analyses using conditional inverse probability weighting, wherein one weights the observed test statistic based upon the worst-case assignment probabilities for a proposed strength of hidden bias. Unlike the prevailing approach to sensitivity analysis after matching, there is a closed form expression for the limiting worst-case distribution when matching with multiple controls. Moreover, the approach admits a closed form for its design sensitivity, a measure used to compare competing test statistics and research designs, for matching with multiple controls, whereas the conventional approach generally only does so for pair matching. The tilted sensitivity analysis improves design sensitivity under a host of generative models. The proposal may also be adaptively combined with the conventional approach to attain a design sensitivity no smaller than the maximum of the individual design sensitivities. Data illustrations indicate that tilting can provide meaningful improvements in the reported robustness of matched observational studies.

*Department of Statistics, University of Michigan, Ann Arbor, MI.

1 Introduction

In an observational study employing matching, treated and control individuals with similar observed covariates are placed into matched sets through the solution of an optimization problem (Rosenbaum, 1991; Hansen, 2004; Zubizarreta, 2012; Pimentel et al., 2015; Brumberg et al., 2024). Despite their similarity with respect to available covariates, individuals in the same matched set may differ in terms of their probability of receiving the treatment due to unobserved factors. A sensitivity analysis in an observational study examines the extent to which adjustment for observables alone may have failed to produce genuine evidence for a treatment effect due to the potential presence of hidden bias.

Rosenbaum (1987) introduced a model for sensitivity analysis suitable for matched observational studies, which bounds the odds ratio of the probabilities of receiving the treatment for any two individuals in the same matched set. While the model for biased treatment assignments of Rosenbaum (1987) is applicable regardless of the structure of the matched sets, methods and supporting theory for conducting a sensitivity analysis under this model vary substantially depending upon whether or not one has a paired observational study (a study with exactly one treated and one control unit in each matched set). For paired observational studies, when testing the sharp null hypothesis of no treatment effect one can both provide bounds on the p-value through the construction of a stochastically dominating random variable. These bounds are valid in finite samples and provide an interpretable characterization of the worst-case pattern of hidden bias: individuals with larger responses are given a higher probability of receiving the treatment. For other forms of matching, similar results are attainable for a class of test statistics referred to as sign-score statistics in Rosenbaum (1988); see §3.1 for more details. While this class includes useful members, most notably the Mantel-Haenszel test statistic when outcomes are binary, it does not contain commonly deployed statistics such as the difference in means with continuous outcomes.

Outside of sign-score statistics, new complications arise. When computing the worst-case p-value under the sharp null there are $\prod_{i=1}^{I}(n_i - 1)$ candidate patterns of hidden bias to consider, where n_i is the number of individuals in the ith matched set and I is the number of matched sets (Rosenbaum and Krieger, 1990). The exponential scaling as a function of the number matched sets renders exact calculation of the worst-case pattern of hidden bias infeasible in practice when matching with multiple controls. A major breakthrough in facilitating sensitivity analysis in matched designs beyond pair matching was attained by Gastwirth et al. (2000), who showed that one may find asymptotically valid bounds on the p-value through the solution of I optimization problems, each with only $n_i - 1$ decision variables. The procedure, known as the *separable algorithm*, is reviewed in §2.2. Unlike in the paired case, the worst-case pattern of unmeasured confounding returned by the separable algorithm is not available in closed form. Furthermore, the worst-case pattern of hidden bias can change as the allowed degree of hidden bias within the sensitivity model changes.

The differences in sensitivity analysis between paired studies and the broader class of matched designs have significant implications for theoretical comparisons of competing test statistics. One particular manifestation is in the calculation of design sensitivity (Rosenbaum, 2004), a metric describing a test statistic's ability to distinguish non-negligible treatment effects from biased treatment assignments. For paired studies, closed form expressions for the design sensitivity are readily available, facilitating analytic comparisons between test statistics and furnishing broader intuition about the drivers of robustness in paired observational studies. In more general matched designs, with few exceptions one must rely upon Monte-Carlo simulation to compare design sensitivities between competing test statistics, hindering the development of the deeper insights that are available in the paired setting.

This paper introduces a new procedure for sensitivity analysis in matched observational studies. For any candidate test statistic, we introduce in §3 a *tilted* modification based upon the degree of hidden biased assumed to exist at any given stage of the sensitivity analysis. The modification is constructed in such a way that even beyond paired designs, the worst-case pattern of unmeasured confounding is both available in closed form and does not vary as the allowed degree of hidden bias increases. Rather than allowing the output of the separable algorithm to vary as the allowed strength of confounding increases, one instead modifies the test statistic itself in such a way that the bound on the worst-case expectation remains constant. The resulting inference is equivalent to that produced by the usual approach in paired designs, but differs from the conventional approach to sensitivity analysis outside of paired studies. Through this, tilting may be viewed as an alternative extension of sensitivity analysis to matching with multiple controls. Despite the matched context, we also highlight connections between the tilted modifications and approaches to sensitivity analysis involving inverse probability weighting (Zhao et al., 2019; Dorn and Guo, 2022). In contrast with the conventional approach, we show in §4 that the tilted sensitivity analysis furnishes test statistics with a closed form for their design sensitivity for any matched structure. The tilted sensitivity analysis provides an improvement in design sensitivity relative to the conventional approach for a wide range of generative models, but does not uniformly dominate the conventional approach. In §4.4, we note that the tilted and conventional approaches can be adaptively combined to attain a design sensitivity no smaller than the maximum of tilted and convtional design sensitivities. In §5, we highlight the benefits conferred by tilting in multiple real-data applications.

2 Background and review

2.1 Sensitivity analysis after matching

There are I matched sets formed on the basis of pretreatment covariates. For notational ease we assume that the ith of I matched sets contains 1 treated individual and $n_i - 1$ controls, but the proposals herein extend immediately to matched structures returned by full matching (Hansen, 2004; Kang et al., 2016); see Rosenbaum (2002b, Problem 4.12) for details of the extension. Let Z_{ij} be the treatment indicator for the jth individual in the ith matched set, taking the value 1 if treated and 0 otherwise. Let y_{1ij} and y_{0ij} be the potential outcomes under treatment and control, let x_{ij} be the vector of observed covariates, and let $0 \leq u_{ij} \leq 1$ be an unobserved covariate for the ijth individual. The observed response for individual ij is $Y_{ij} = Z_{ij}y_{1ij} + (1 - Z_{ij})y_{0ij}$; implicit in this representation is that the Stable Unit-Treatment Value Assumption (SUTVA) holds

(Rubin, 1974). Let $\mathbf{Z} = (Z_{11}, ..., Z_{In_I})^T$ be the vector of treatment assignments across all matched sets, let $\mathbf{Z}_i = (Z_{i1}, ..., Z_{in_i})^T$ be the assignments in matched set i, and let the analogous notation hold for other quantities such as $\mathbf{u}_i = (u_{i1}, ..., u_{in_i})^T$. Let $\mathbf{Y_Z}$ be the lexicographically ordered vector of observed responses under assignment $\mathbf{Z}$. Let $\mathcal{F} = \{y_{0ij}, y_{1ij}, x_{ij}, u_{ij}\}$ be the features of the observed study population, let $\Omega = \{\mathbf{z} : \sum_{j=1}^{n_i} z_{ij} = 1, \;\; i = 1, ..., I\}$ be the set of treatment assignments satisfying the matched design, and let $\mathcal{Z} = \{\mathbf{Z} \in \Omega\}$.

Letting $\pi_{ij} = \mathrm{pr}(Z_{ij} \mid \mathcal{F})$, the sensitivity model introduced in Rosenbaum (1987) proposes the following logit model for the assignment probabilities:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \kappa_i + \log(\Gamma) u_{ij}, \tag{1}$$

where $\Gamma \geq 1$ is a sensitivity parameter controlling the impact of unmeasured confounding on the assignment probabilities. Observe that at $\Gamma = 1$, all individuals in the same matched set have the same assignment probabilities. Conditioning on $\mathcal{Z}$ to return attention to the matched structure at hand, we have

$$\begin{aligned} \varrho_{ij} := \mathrm{pr}(Z_{ij} = 1 \mid \mathcal{F}, \mathcal{Z}) &= \frac{\exp\{\log(\Gamma) u_{ij}\}}{\sum_{\ell=1}^{n_i} \exp\{\log(\Gamma) u_{i\ell}\}}; \\ \mathrm{pr}_{\mathbf{u}}(\mathbf{Z} = \mathbf{z} \mid \mathcal{F}, \mathcal{Z}) &= \frac{\exp\{\log(\Gamma) \mathbf{z}^T \mathbf{u}\}}{\sum_{\mathbf{b} \in \Omega} \exp\{\log(\Gamma) \mathbf{b}^T \mathbf{u}\}}, \end{aligned} \tag{2}$$

such that conditioning on $\mathcal{Z}$ removes dependence on the nuisance parameters κ_i. At $\Gamma = 1$, conditioning on $\mathcal{Z}$ recovers the usual randomization distribution in a block-randomized experiment. $\Gamma > 1$ allows for departures from this idealized experiment.

Consider a test statistic of the form $T = \mathbf{Z}^T \mathbf{q} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} Z_{ij} q_{ij}$, where $\mathbf{q} = q(\mathbf{Y_Z})$ is a function of the observed responses under assignment vector $\mathbf{Z}$, $\mathbf{Y_Z}$; such statistics are called sum statistics, and include the treated-minus-control difference in means, Wilcoxon's rank sum test, and the Mantel-Haenszel test among many (Rosenbaum, 2002b). Assuming that (1) holds at Γ, the right-tail probability for T is

$$\mathrm{pr}(T \geq a \mid \mathcal{F}, \mathcal{Z}) = \frac{\sum_{\mathbf{b} \in \Omega} 1\{\mathbf{b}^T q(\mathbf{Y_b}) \geq a\} \exp\{\log(\Gamma) \mathbf{b}^T \mathbf{u}\}}{\sum_{\mathbf{b} \in \Omega} \exp\{\log(\Gamma) \mathbf{b}^T \mathbf{u}\}}. \tag{3}$$

This distribution is unknown to the practitioner, due to its dependence on observed outcomes $\mathbf{Y_b}$ for assignments $\mathbf{b} \neq \mathbf{Z}$ and on the unobserved covariates $\mathbf{u}$. With a few notable exceptions (Rosenbaum, 2002a; Fogarty, 2023), the literature on sensitivity analysis after matching typically considers tests of sharp null hypotheses under the finite population model, hypotheses whose truth impute the missing values of the potential outcomes for each individual. For ease of exposition, we consider the sharp null hypothesis of no effect at all, $H_F : y_{0ij} = y_{1ij} \quad (i = 1, ..., I, ; j = 1, ..., n_i)$. Under this null, $q(\mathbf{Y_b}) = q(\mathbf{Y_z})$ for any $\mathbf{b}, \mathbf{z} \in \Omega$, as the observed responses $\mathbf{Y_Z}$ provide the hypothetical responses $\mathbf{Y_b}$ for all possible randomizations. At $\Gamma = 1$, the right-tail probability displayed in (3) can be directly computed under H_F, providing the basis for Fisher randomization tests. At $\Gamma > 1$, this tail probability is still unknown due to its dependence on the unmeasured confounders u_{ij}. In a sensitivity analysis, one instead computes upper bounds on the tail probability (3) at any given $\Gamma > 1$. One then finds the largest value of Γ for which H_F can still be rejected. This changepoint Γ, referred to as the *sensitivity value* (Zhao, 2019), provides a measure of the robustness of a study's findings to unmeasured confounding.

For various extensions of the sensitivity model (1) to accommodate features ranging from stochastic potential outcomes to heterogeneous bias across matched sets, see Small et al. (2009); Fogarty and Hasegawa (2019); Chen et al. (2024); Zhang et al. (2024) and Wu and Li (2025).

2.2 The separable algorithm

For a given Γ, asymptotic bounds on (3) under the sharp null hypothesis can be attained using the separable algorithm introduced in Gastwirth et al. (2000). In short, the algorithm finds the vector(s) of unmeasured confounders yielding the worst-case expectation under the sharp null. If multiple vectors provide the same worst-case expectation, the one providing the largest variance is selected. One then conducts inference using a normal random variable with this expectation and variance.

Let $T_i = \sum_{j=1}^{n_i} Z_{ij} q_{ij}$, such that $T = \sum_{i=1}^{I} T_i$. Rearrange the values q_{ij} in each matched set i such that $q_{i1} \geq q_{i2} \geq ... \geq q_{in_i}$. Using results in Rosenbaum and Krieger

(1990), the worst-case expectation for T_i under the sharp null hypothesis, optimized over all possible $\mathbf{u}_i \in [0,1]^{n_i}$, can be restricted to a search over $n_i - 1$ candidates:

$$\mu_{\Gamma i} = \max_{a=1,\dots,n_i-1} \frac{\Gamma \sum_{j=1}^{a} q_{ij} + \sum_{j=a+1}^{n_i} q_{ij}}{\Gamma a + (n_i - a)}. \tag{4}$$

Let $\mathcal{A}_i$ be set of values for a attaining the maximal expectation $\mu_{\Gamma i}$, and define the largest variance for T_i for patterns of hidden bias within $\mathcal{A}_i$ as $\nu_{\Gamma i}^2 := \max_{a \in \mathcal{A}_i} \left\{\Gamma \sum_{j=1}^{a} q_{ij}^2 + \sum_{j=a+1}^{n_i} q_{ij}^2\right\} / \{\Gamma a + (n_i - a_i)\} - (\mu_{\Gamma i})^2$. Under mild regularity conditions, the tail probability for T in (3) is asymptotically upper bounded by the tail probability of a normal random variable with expectation $\sum_{i=1}^{I} \mu_{\Gamma i}$ and variance $\sum_{i=1}^{I} \nu_{\Gamma i}^2$. This asymptotic approximation requires solving I tractable optimization problems, each requiring the enumeration of only $n_i - 1$ candidate solutions.

The following proposition, proven in the web-based supporting materials, provides an alternative representation for (4) as a stratum-specific Z-estimator (a zero of an estimating equation), which will be useful in what follows.

Proposition 1. *Let* $\bar{q}_i = n_i^{-1} \sum_{j=1}^{n_i} q_{ij}$. *Then,* $\mu_{\Gamma i} = \bar{q}_i + \{(\Gamma - 1)/(1+\Gamma)\} M_{\Gamma i}$, *where*

$$M_{\Gamma i} := \texttt{SOLVE}\left\{c : \frac{1}{n_i}\sum_{j=1}^{n_i}\left|q_{ij} - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right)c\right| = c\right\}, \tag{5}$$

where $\texttt{SOLVE}(\cdot)$ *denotes the solution to the displayed equation.*

3 Tilted sensitivity analysis

We now introduce a modified approach to conducting a sensitivity analysis when matching with multiple controls. Suppose that the sharp null hypothesis H_F. For each matched set i, define $T_{\Gamma i}$, referred to as the *tilted* version of T_i henceforth, as

$$T_{\Gamma i} := T_i - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right)|T_i - \bar{q}_i|, \tag{6}$$

which has the equivalent representation

$$T_{\Gamma i} = \left(\frac{2\Gamma}{1+\Gamma}\right)\frac{T_i - \bar{q}_i}{\exp\{\log(\Gamma)\mathbb{1}\{T_i > \bar{q}_i\}\}}. \tag{7}$$

The representation (7) inspires the name for the modification: $f_\Gamma(x) := x/\exp\{\log(\Gamma)\mathbb{1}\{x \geq 0\}\}$ is the 45 degree line through the origin for $x \leq 0$, with slope equal to 1. For $x > 0$, the slope changes from 1 to $1/\Gamma$, tilting towards the horizontal axis.

Let $w_{\Gamma i} > 0$, $i = 1, ..., I$, be nonnegative multipliers which are fixed under the sharp null, and consider as a test statistic $T_\Gamma(w_\Gamma) = \sum_{i=1}^{I} w_{\Gamma i} T_{\Gamma i}$. Rosenbaum (2014) describes the benefits of weighting for improving the power of a sensitivity analysis. As will be demonstrated, different choices of weights will also provide connections between sensitivity analyses using tilted statistics and both (i) using sign-score statistics (Rosenbaum, 1988); and (ii) using inverse probability weighting.

We now describe how to conduct a sensitivity analysis after tilting when (1) is assumed to hold at Γ. For each matched set i, define the variance $\tilde{\nu}^2_{\Gamma i}$ as

$$\tilde{\nu}^2_{\Gamma i} := \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i}\mathbb{1}\{q_{ij} > \bar{q}_i\}}\right)\sum_{j=1}^{n_i}\frac{(q_{ij} - \bar{q}_i)^2}{\exp\{\log(\Gamma)\mathbb{1}\{q_{ij} > \bar{q}_i\}\}}. \tag{8}$$

Further define $c_{\Gamma,\alpha}(w_\Gamma) = \Phi^{-1}(1-\alpha)\sqrt{\sum_{i=1}^{I} w^2_{\Gamma i}\tilde{\nu}^2_{\Gamma i}}$, where $\Phi^{-1}(\cdot)$ is the quantile function for the standard normal distribution. The following Theorem, proven in the web-based supporting material, justifies conducting a sensitivity analysis for each Γ by rejecting the sharp null hypothesis when $T_\Gamma(w_\Gamma) \geq c_{\Gamma,\alpha}(w_\Gamma)$. That is, $c_{\Gamma,\alpha}(w_\Gamma)$ is an asymptotically valid critical value when (1) holds at Γ.

Theorem 1. *Suppose that the sharp null hypothesis holds and that the sensitivity model (1) holds at Γ. Suppose that the stratum sizes are bounded, $n_i \leq \tilde{N}$ for all i, and suppose further that the constants q_{ij} and weights $w_{\Gamma i}$ satisfy*

$$\frac{\max_{i=1,\dots,I} \;\; w^2_{\Gamma i}(q_{i1} - q_{in_i})^2}{\sum_{i=1}^{I} w^2_{\Gamma i}(q_{i1} - q_{in_i})^2} \rightarrow 0$$

as $I \rightarrow \infty$, where q_{ij} are again sorted so that $q_{i1} \geq q_{i2} \geq ... \geq q_{in_i}$. Let $\vartheta_{\Gamma i}$ and $\sigma^2_{\Gamma i}$

denote the true expectation and variance for $T_{\Gamma i}$. Let $\mathcal{B}_I = \{i : \sigma^2_{\Gamma i} > \tilde{\nu}^2_{\Gamma i}\}$, and suppose that $\sqrt{I^{-1}\sum_{i\in\mathcal{B}_I} w^2_{\Gamma i}(\sigma^2_{\Gamma i} - \tilde{\nu}^2_{\Gamma i})} = o\left(I^{-1/2}|\sum_{i=1}^I w_{\Gamma i}\vartheta_{\Gamma i}|\right)$. Then, for all $\epsilon > 0$, there exists an I^* such that $I > I^*$ implies that for any $\alpha < 1/2$,

$$pr\{T_\Gamma(w_\Gamma) \geq c_{\Gamma,\alpha}(w_\Gamma) \mid \mathcal{F}, \mathcal{Z}\} \leq \alpha + \epsilon.$$

Theorem 1 implies that $1 - \Phi\left(T_\Gamma(w_\Gamma)/\sqrt{\sum_{i=1}^I w^2_{\Gamma i}\tilde{\nu}^2_{\Gamma i}}\right)$ provides an asymptotic upper bound on the p-value for testing the sharp null when (1) holds at Γ. The condition relating $\max w^2_{\Gamma i}(q_{i1} - q_{in_i})^2$ to $\sum w^2_{\Gamma i}(q_{i1} - q_{in_i})^2$ provides a Gaussian limiting distribution by Theorem 1 of Hájek et al. (1999, §6.1.2). A condition analogous to $\sqrt{I^{-1}\sum_{i\in\mathcal{B}_I} w^2_{\Gamma i}(\sigma^2_{\Gamma i} - \tilde{\nu}^2_{\Gamma i})} = o\left(I^{-1/2}|\sum_{i=1}^I w_{\Gamma i}\vartheta_{\Gamma i}|\right)$ is needed to justify use of the separable algorithm of Gastwirth et al. (2000) for *any* test statistic, and is not required solely because of tilting. For intuition on what is required, note that if $I^{-1}\sum_{i=1}^I w_{\Gamma i}\vartheta_{\Gamma i}$ and $I^{-1}\sum_{i=1}^I w^2_{\Gamma i}\sigma^2_{\Gamma i}$ have finite nonzero limits, then $\sqrt{I^{-1}\sum_{i\in\mathcal{B}_I} w^2_{\Gamma i}(\sigma^2_{\Gamma i} - \tilde{\nu}^2_{\Gamma i})} = O(1)$ and $I^{-1/2}|\sum_{i=1}^I w_{\Gamma i}\vartheta_{\Gamma i}| = O(I^{1/2})$, hence satisfying the condition. For a pathological example where the condition fails, see the web-based supporting material.

The following lemma, proven in the web-based supporting materials, allows one to find a closed-form solution to output of the separable algorithm described in §2.2 even when matching with multiple controls, which is generally not possible without deploying tilting. Armed with this lemma, the remainder of the proof of Theorem 1 follows in a straightforward way from the proof of Proposition 1 of Gastwirth et al. (2000), and Theorem 1 of Hájek et al. (1999, §6.1.2).

Lemma 1. *Suppose that (1) holds at Γ. Then, $E(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) \leq 0$ for all $i = 1, ..., I$, with $E(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) = 0$ when $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$. When $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$, the variance of $T_{\Gamma i}$ equals $\tilde{\nu}^2_{\Gamma i}$ in (8). If there exists another pattern of unmeasured confounding $\mathbf{u}'_i \neq \mathbf{u}_i$ for which $E(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) = 0$, the variance for $T_{\Gamma i}$ under pattern $\mathbf{u}'_i$ will be no larger than $\tilde{\nu}^2_{\Gamma i}$.*

Remark 1. Unlike for T_i, the worst-case expectation for $T_{\Gamma i}$ when (1) is assumed to hold at Γ may be calculated *without* resorting to the separable algorithm described in §2.2: replacing q_{ij} by $q_{ij} - \bar{q}_i - \{(\Gamma - 1)/(1 + \Gamma)\}|q_{ij} - \bar{q}_i|$ in (4) yields $\mu_{\Gamma i} = 0$. Moreover, in the case where multiple values of a in (4) attain $\mu_{\Gamma i} = 0$, setting $a = \sum_{j=1}^{n_j} \mathbb{1}\{q_{ij} > \bar{q}_i\}$ provides the largest variance. Owing to this, the worst-case deviate returned by the separable algorithm has a closed-form solution when using $T_\Gamma(w_\Gamma)$ as a test statistic.

Remark 2. The pattern of unmeasured confounding yielding the worst-case expectation takes on a particularly intuitive form for tilted test statistics: for any value of Γ, individuals ij above the match-set specific average of q_{ij}, $\bar{q}_i$, are given $u_{ij} = 1$, making them relatively more likely to be treated, while individuals with q_{ij} at or below the stratum-specific mean are given $u_{ij} = 0$, making them relatively less likely to be treated. The resulting pattern of unmeasured confounding maximizes $\mathrm{pr}\,(T_i > \bar{q}_i \mid \mathcal{F}, \mathcal{Z})$, the probability that the observed q_{ij} is above the average value $\bar{q}_i$. Contrast this with the pattern of hidden bias yielding the worst-case expectation for T_i without tilting described in Proposition 1, which in general can change as a function of the value of Γ at which the sensitivity analysis is being conducted. Consequently, the unmeasured confounder returned by the separable algorithm when applied to T_i need not maximize $\mathrm{pr}\,(T_i > \bar{q}_i \mid \mathcal{F}, \mathcal{Z})$.

3.1 Tilting, pair matching, and sign-score statistics

To build intuition for the tilted sensitivity analysis, we first describe connections between sensitivity analyses using the tilted test statistics $T_\Gamma(w_\Gamma)$ and existing results for sensitivity analyses for an important subset of statistics commonly used in matched observational studies. Suppose that for $i = 1, ..., I$, T_i admits the representation

$$T_i = a_{i2} + (a_{i1} - a_{i2}) \sum_{j=1}^{n_i} Z_{ij} \mathbb{1}\{q_{ij} = a_{i1}\}. \tag{9}$$

with $a_{i1} \geq a_{i2}$. Recalling that $\sum_{j=1}^{n_i} Z_{ij} = 1$ for all i, any T_i which takes on at most two distinct values for each matched set i may be expressed in this form. From (9), we see that $\sum_{i=1}^{I} T_i$ is equivalent to an important class of statistics known as sign-score statistics (Rosenbaum, 1988, 2002b, §§4.3-4.4). Statistics admitting the representation (9) include any test statistic in a matched pairs design, along with several important statistics when matching with multiple controls such as the Mantel-Haenszel test statistic with binary outcomes. For such test statistics, one can construct a stochastically dominating random variable B_i when (1) holds at Γ by assigning $u_{ij} = 1$ to observations with $q_{ij} = a_{i1}$, and $u_{ij} = 0$ to observations with $q_{ij} = a_{2i}$; see Proposition 14 of Rosenbaum (2002b, §4.4.1) for details. It follows that $\mu_{\Gamma i}$ in (4), the worst-case expectation that would be returned by the separable algorithm, equals the expectation of the stochastically bounding random variable B_i. The following illustrates a direct relationship between $T_i - \mu_{\Gamma i}$ and $w_{\Gamma i}T_{\Gamma i}$ for a particular choice of the weights, $w_{\Gamma i}^{(ss)}$.

Proposition 2. *Suppose that T_i admits the representation (9). Then,*

$$T_i - \mu_{\Gamma i} = \underbrace{\left(\frac{\Gamma+1}{2}\right)\left(\frac{n_i}{n_i + (\Gamma-1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}\right)}_{=:w_{\Gamma i}^{(SS)}} T_{\Gamma i}. \tag{10}$$

First consider the implications of Proposition 2 for matched pair designs ($n_i = 2$). The representation for $T_{\Gamma i}$ in (6) is commonly encountered when conducting sensitivity analyses with pairs; see, for instance, the numerator of the second display in Equation 6 of Rosenbaum (2007). In paired designs, $|T_i - \bar{q}_i|$ is constant across all randomizations when testing the sharp null, and the worst-case expectation when conducting a sensitivity analysis at Γ is $\mu_{\Gamma i} = \bar{q}_i + \{(\Gamma-1)/(1+\Gamma)\}|T_i - \bar{q}_i|$. For paired designs, $w_{\Gamma i}^{(SS)} = 1$, and $T_{\Gamma i} = T_i - \mu_{\Gamma i}$ is thus the observed value for the test statistic T_i in stratum i minus the worst-case expectation when (1) holds at Γ. Proposition 2 shows that in the case of pair matching, the tilted sensitivity analysis reproduces the standard approach to sensitivity analysis without modification. Beyond paired observational studies, Proposition 2 illustrates a more general connection between the tilted sensitivity analysis and the conventional sensitivity analysis when restricting attention to sum statistics of the form (9). Notably, statistics of the form (9) are the statistics for which one can conduct the conventional sensitivity analysis without using the separable algorithm. With the choice of weights $w_{\Gamma i}^{(ss)}$ in (10), the tilted sensitivity analysis recovers the conventional sensitivity analysis for sign-score statistics.

3.2 Tilting and inverse probability weighting

The equivalent form for $T_{\Gamma i}$ in (7), in concert with Theorem 1, inspire alternative motivation for the tilted statistics through connections with inverse probability weighting (IPW). Suppose first that there was no unmeasured confounding, such that $\Gamma = 1$. Then, under the sharp null H_F and for each matched set i, we would have $E(T_i - \bar{q}_i \mid \mathcal{F}, \mathcal{Z}) = E\left\{\sum_{j=1}^{n_i} Z_{ij}(q_{ij} - \bar{q}_i) \mid \mathcal{F}, \mathcal{Z}\right\} = 0$ as $\mathrm{pr}(Z_{ij} = 1 \mid \mathcal{F}, \mathcal{Z}) = 1/n_i$. Suppose now that (1) holds at $\Gamma > 1$, and that the true conditional assignment probabilities are $\varrho_{ij} = \exp\{\log(\Gamma)u_{ij}\}/\sum_{\ell=1}^{n_i}\exp\{\log(\Gamma)u_{i\ell}\}$. At $\Gamma > 1$, the unmeasured covariates tilt the distribution of $\mathbf{Z}_i = (Z_{i1}, ..., Z_{in_i})^T$ away from uniform, such that $E\left\{\sum_{j=1}^{n_i} Z_{ij}(q_{ij} - \bar{q}_i) \mid \mathcal{F}, \mathcal{Z}\right\}$ need not equal zero. The inverse probability weighted $n_i^{-1}\sum_{j=1}^{n_i} Z_{ij}(q_{ij} - \bar{q}_i)/\varrho_{ij}$ *does* have expectation zero, and at $\Gamma = 1$ it is precisely $\sum_{j=1}^{n_i} Z_{ij}(q_{ij} - \bar{q}_i)$.

When conducting a sensitivity analysis, we do not have access to the true assignment probabilities ϱ_{ij} because u_{ij} is unknown. Consequently, the realization of the random variable $n_i^{-1}\sum_{j=1}^{n_i} Z_{ij}(q_{ij} - \bar{q}_i)/\varrho_{ij}$ is unknown to us. To address this, one may consider instead the *worst-case* weighting subject to (1) holding at Γ. Towards this end, define $\tilde{\varrho}_{ij}$ as

$$\tilde{\varrho}_{ij} = \frac{\exp\{\log(\Gamma)\mathbb{1}\{q_{ij} > \bar{q}_i\}\}}{\sum_{\ell=1}^{n_i}\exp\{\log(\Gamma)1\{q_{i\ell} > \bar{q}_i\}\}} = \frac{\exp\{\log(\Gamma)\mathbb{1}\{q_{ij} > \bar{q}_i\}\}}{n_i + (\Gamma - 1)\sum_{\ell=1}^{n_i}\mathbb{1}\{q_{i\ell} > \bar{q}_i\}}.$$

Using (2), $\tilde{\varrho}_{ij}$ are the conditional assignment probabilities when $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$. Under the sharp null and when (1) holds at Γ, these assignment probabilities maximize the probability that $T_i - \bar{q}_i > 0$. The corresponding IPW statistic using these worst case conditional assignment probabilities is

$$
\begin{aligned}
IPW_{\Gamma i} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{Z_{ij}(q_{ij}-\bar{q}_i)}{\tilde{\varrho}_{ij}} \\
&= \underbrace{\left(\frac{\Gamma+1}{2\Gamma}\right)\left\{\frac{n_i+(\Gamma-1)\sum_{j=1}^{n_i}\mathbb{1}\{q_{ij}>\bar{q}_i\}}{n_i}\right\}}_{=:w_{\Gamma i}^{(IPW)}} T_{\Gamma i}, \qquad (11)
\end{aligned}
$$

where the last line uses (7). Noting that $w_{\Gamma i}^{(IPW)}$ in (11) is a constant under the sharp null, Lemma 1 implies that under the sharp null, when (1) holds at Γ, $E(IPW_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) \leq 0$, with equality when $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$. In other words, weighting by $\tilde{\varrho}_{ij}$ yields a random variable whose expectation is bounded above by 0 when (1) holds at Γ, with equality when $\tilde{\varrho}_{ij}$ are the actual assignment probabilities.

Equation (11) illustrates strong connections between tilted statistics $T_{\Gamma i}$ and the worst-case IPW statistic $IPW_{\Gamma i}$: they differ only in the form of the matched-set weights $w_{\Gamma i}$, with the unweighted tilted statistic ignoring the normalizing constant $n_i + (\Gamma-1)\sum_{j=1}^{n_i}\mathbb{1}\{q_{ij} > \bar{q}_i\}$ for the worst-case assignment probabilities. When T_i is a sign-score statistic, equations (10) and (11) further demonstrates striking yet previously unexplored similarities between sensitivity analyses after matching using sign-score statistics and sensitivity analyses using worst-case inverse probability weighting within each matched set: the methods are both weighted versions of the tilted sensitivity analysis, differing in the choice of weights $w_{\Gamma i}$. See the web-based supporting materials for an illustration of these connections using the Mantel-Haenszel statistic with binary outcomes. Tilting provides a unifying framework for considering these seemingly disparate approaches to conducting a sensitivity analysis after matching.

4 Comparing the tilted and conventional sensitivity analyses through design sensitivity

4.1 Generative model for comparison

While Proposition 2 reveals connections between tilted statistics of the form $T_\Gamma(w_\Gamma) = \sum_{i=1}^I w_{\Gamma i} T_{\Gamma i}$ and sensitivity analyses using sign-score statistics, in general the sensitivity analyses based upon tilted and untilted statistics are fundamentally different. Without restricting to sign-score statistics, the tilted and conventional approaches are no longer equal up to the choice of weights $w_{\Gamma i}$. In particular, the unmeasured confounders yielding the worst-case expectation for $T_{\Gamma i}$ and for T_i can differ. While the tilted sensitivity analysis is more straightforward to conduct because it avoids the separable algorithm, one would be hesitant to apply the tilted sensitivity analysis to the broader class of sum statistics if it demonstrated deficiencies in terms of statistical performance relative to the conventional approach. In this section, we highlight circumstances under which the tilted sensitivity analysis proves advantageous in terms of a metric known as *design sensitivity* (Rosenbaum, 2004).

For theoretical comparisons among competing approaches to sensitivity analysis, is common for ease of analysis to assume (i) a superpopulation generative model for the responses in each matched set; and (ii) that each matched set has 1 treated individual and J controls, such that $n_i = J+1$ for all i (Rosenbaum, 2004, 2013). For $i = 1, .., I$; $j = 1, ..., J+1$, potential outcomes and treatment assignments are generated as

$$y_{1ij} = \alpha_i + \tau + \varepsilon_{1ij}; \;\; y_{0ij} = \alpha_i + \varepsilon_{0ij}, \tag{12}$$

where α_i are matched set fixed effects, $\tau \geq 0$ is the treatment effect, and $(\varepsilon_{0ij}, \varepsilon_{1ij})^T$ are *iid* draws from a mean zero, exchangeable multivariate error distribution $F_\varepsilon(\cdot)$. Treatment assignments are generated satisfying $Z_{ij} \perp\!\!\!\perp (\varepsilon_{0ij}, \varepsilon_{1ij}); \mathrm{pr}(\mathbf{Z} = \mathbf{z} \mid \mathcal{F}, \mathcal{Z}) = |\Omega|^{-1}$. The observed responses remain $Y_{ij} = Z_{ij} y_{0ij} + (1 - Z_{ij}) y_{i1j}$, but are now random due to both randomized treatment assignment and randomness across draws for $(\varepsilon_{0ij}, \varepsilon_{1ij})$. The test statistics under consideration remain of the form $\mathbf{Z}^T\mathbf{q} = \sum_{i=1}^I \sum_{j=1}^{n_i} Z_{ij} q_{ij} =$

$\sum_{i=1}^{I} T_i$, with the vector $q = q(\mathbf{Y_Z})$ also varying across draws from the generative model.

We consider the tilted sensitivity analysis with $w_{\Gamma i} = 1$ for all i, comparing the conventional sensitivity analysis to the unweighted version of the tilted sensitivity analysis; results extend naturally to weighted variants. $T_{\Gamma i}$ and $T_i - \mu_{\Gamma i}$, the observed statistic minus its worst-case expectation under the tilted and conventional approaches when (1) is assumed to hold at Γ, may be expressed using (5) and (6) as

$$\begin{aligned} T_{\Gamma i} &= T_i - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right)|T_i - \bar{q}_i|; \\ T_i - \mu_{\Gamma i} &= T_i - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right) M_{\Gamma i}. \end{aligned}$$

The right-hand sides of the above displays differ only in the random variable by which $\{(\Gamma-1)/(1+\Gamma)\}$ is multiplied: $|T_i - \bar{q}_i|$ for the tilted sensitivity analysis, and $M_{\Gamma i}$ for the conventional approach. As suggested by the notation, the random variable $M_{\Gamma i}$ changes with the value of Γ at which the sensitivity analysis is conducted, while $|T_i - \bar{q}_i|$ does not. Differences in the distributions of $M_{\Gamma i}$ and $|T_i - \bar{q}_i|$ will play a fundamental role in determining the comparative performance of the tilted and conventional sensitivity analyses.

Throughout this section, we make the following high-level assumptions about the limiting behavior of T_i, $T_{\Gamma i}$, and $M_{\Gamma i}$. Primitive conditions guaranteeing the assumed behavior will depend upon the particular function $q(\mathbf{Y_Z})$ used to form a given test statistic.

Assumption 1 (Convergence in probability of first and second sample moments)**.** *For all $\Gamma \geq 1$, $I^{-1}\sum_{i=1}^{I}(T_i - \bar{q}_i)$, $I^{-1}\sum_{i=1}^{I}|T_i - \bar{q}_i|$, $I^{-1}\sum_{i=1}^{I} M_{\Gamma i}$, $I^{-1}\sum_{i=1}^{I}\nu_{\Gamma i}^2$, and $I^{-1}\sum_{i=1}^{I}\tilde{\nu}_{\Gamma i}^2$ converge in probability to constant limiting values.*

Assumption 2 (Asymptotic normality)**.** *For all $\Gamma \geq 1$, $I^{-1}\sum_{i=1}^{I}(T_{\Gamma i} - T'_{\Gamma i}) = o_p(I^{-1/2})$ and $I^{-1}\sum_{i=1}^{I}(T_i - T'_i) = o_p(I^{-1/2})$ where $\{T_{\Gamma i'}\}$ and $\{T'_i\}$ are sequences of independent random random variables. $I^{-1/2}\sum_{i=1}^{I}\{T'_i - E(T'_i)\}/\sqrt{I^{-1}\sum_{i=1}^{I} Var(T'_i)}$ and$I^{-1/2}\sum_{i=1}^{I}\{T'_{\Gamma i} - E(T'_{\Gamma i})\}/\sqrt{I^{-1}\sum_{i=1}^{I} Var(T'_{\Gamma i})}$ converge jointly in distribution to a multivariate normal.*

Moving forwards, let $\theta = \text{plim } I^{-1}\sum_{i=1}^{I}(T_i - \bar{q}_i)$, let $\eta = \text{plim } I^{-1}\sum_{i=1}^{I}|T_i - \bar{q}_i|$, and let $m_\Gamma = \text{plim } I^{-1}\sum_{i=1}^{I} M_{\Gamma i}$, where plim denotes the limit in probability.

4.2 Design sensitivity for the tilted sensitivity analysis

Rosenbaum (2004) introduced design sensitivity to compare the limiting performance of test statistics in a sensitivity analysis. The design sensitivity is derived under the favorable setting for an observational study of a positive treatment effect and no hidden bias. Whether this favorable setting actually holds is unknown to the practitioner, so even under the favorable setting a researcher would like to assess the robustness of their findings to hidden bias through conducting a sensitivity analysis. Under the generative model described in §4.1 with $\tau > 0$ in (12) and under Assumptions 1 and 2, there exists a number $\tilde{\Gamma}$, the design sensitivity, such that a sensitivity analysis under (1) will reject the null of no effect with probability tending to 1 at $\Gamma < \tilde{\Gamma}$, and with probability tending to zero for $\Gamma > \tilde{\Gamma}$. Test statistics conferring larger values for $\tilde{\Gamma}$ are preferred.

The design sensitivity for the conventional sensitivity analysis is

$$\tilde{\Gamma}_{conv} = \texttt{SOLVE}\left\{\tilde{\Gamma} : \theta = \left(\frac{\tilde{\Gamma} - 1}{1 + \tilde{\Gamma}}\right) m_{\tilde{\Gamma}}\right\}. \tag{13}$$

In words, it is the value $\tilde{\Gamma}_{conv}$ such that the limiting value for the observed test statistic equals the limiting value of the worst-case expectation (assuming the sharp null) returned by the separable algorithm, making use of the representation from Proposition 1. Unfortunately, for $J \geq 2$ controls the design sensitivity $\tilde{\Gamma}_{conv}$ does not admit a closed form for the general class of sum statistics precisely because the separable algorithm does not provide a closed form for $M_{\Gamma i}$. In practice, it is instead calculated through Monte-Carlo simulation and numeric root finding. This has hindered theoretical comparisons of different choices of test statistic using design sensitivity as a guide when matching with multiple controls.

For the tilted sensitivity analysis, no such difficulty arises. By using a test statistic with a known upper bound on the worst-case expectation under the null when (1) is assumed to hold at Γ, the tilted sensitivity analysis *does* admit a closed form solution for its design sensitivity, even when matching with multiple controls. The following theorem provides the closed form.

Theorem 2. *Under Assumptions 1 and 2, the design sensitivity for the tilted sensitivity analysis is*

$$\tilde{\Gamma}_{tilt} = \mathit{SOLVE}\left\{\tilde{\Gamma} : \theta = \left(\frac{\tilde{\Gamma}-1}{1+\tilde{\Gamma}}\right)\eta\right\} = \frac{\eta+\theta}{\eta-\theta}. \tag{14}$$

The form of the design sensitivity, available for matching with multiple controls, aligns with existing closed-form solutions for design sensitivity in the conventional approach for the case of matched pairs (Rosenbaum, 2013, 2016). Essential to the proof of Theorem 2 is that the output of the separable algorithm when applied to the tilted sensitivity analysis is available in closed form by Lemma 1. In the web-based supporting materials, we provide an illustration of the additional insight provided by a closed-form solution for the design sensitivity when matching with multiple controls. In particular, we discuss the role that the number of matched controls plays in improving design sensitivity for a class of statistics known as M-statistics (Maritz, 1979; Huber, 1981), previously investigated for the conventional sensitivity analysis in Rosenbaum (2013) using Monte-Carlo simulation.

4.3 Comparing design sensitivities

Comparing (13) and (14) provides the following simple condition for determining whether or not the tilted sensitivity analysis outperforms the conventional approach in terms of design sensitivity.

Proposition 3. *Suppose that Assumptions 1 and 2 hold, and let $\tilde{\Gamma}_{tilt}$ and $\tilde{\Gamma}_{conv}$ be the design sensitivity for the tilted sensitivity analysis and conventional sensitivity analysis respectively. Then, the design sensitivity for the tilted sensitivity analysis is greater than that of the conventional sensitivity analysis if and only if $m_{\tilde{\Gamma}_{tilt}} > \eta$, or equivalently, if and only if $m_{\tilde{\Gamma}_{conv}} > \eta$.*

The lack of a closed form solution for m_Γ at any Γ makes it challenging to precisely characterize (in analytic form) circumstances under which the design sensitivity of the tilted sensitivity analysis will be larger (better) than that of the conventional approach. Nonetheless, expansions based upon Proposition 1 reveal the essential role played by the distribution of $\mathbf{q}_i = (q_{i1}, ..., q_{in_i})^T$ in determining which method should be preferred.

Proposition 4. *Let* $\hat{G}_i(t) = (J+1)^{-1} \sum_{j=1}^{J+1} \mathbb{1}\{q_{ij} - \bar{q}_i \leq t\}$ *be the stratum-specific empirical distribution function of* $\{q_{i1} - \bar{q}_i, ..., q_{i(J+1)} - \bar{q}_i\}$*, and let* $\kappa_\Gamma = (\Gamma - 1)/(1 + \Gamma)$*. For any* Γ *and any* t*, the following bounds on* $I^{-1}\sum_{i=1}^{I} M_{\Gamma i}$ *hold for any* I*:*

$$I^{-1}\sum_{i=1}^{I}(\kappa_\Gamma M_{\Gamma i} - t)\left\{2\hat{G}_i(t) - 1\right\} \tag{15}$$

$$\leq I^{-1}\sum_{i=1}^{I} M_{\Gamma i} - \{I(J+1)\}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J+1}|q_{ij} - \bar{q}_i - t| \tag{16}$$

$$\leq I^{-1}\sum_{i=1}^{I}(\kappa_\Gamma M_{\Gamma i} - t)\left\{2\hat{G}_i(\kappa_\Gamma M_{\Gamma i}) - 1\right\}. \tag{17}$$

To illustrate the intuition conferred by Proposition 4, set $\Gamma = \tilde{\Gamma}_{conv}$ and $t = \theta$ in (15)-(17). A sufficient condition for the tilted approach having larger design sensitivity is that the probability limit of (16) minus that of (15) is greater than $m_{\tilde{\Gamma}_{conv}} - \eta$. Note that $I^{-1}\sum_{i=1}^{I}(\kappa_\Gamma M_{\Gamma i} - \theta)$ converges in probability to 0 when $\Gamma = \tilde{\Gamma}_{conv}$ by (13). By the representation for $M_{\Gamma i}$ in Proposition 1 as a Z-estimator, (15) and (17) become smaller in magnitude, converging to zero, as the number of matched controls J increase by consistency of Z-estimators (Van der Vaart, 2000, Theorem 5.7). This suggests that $\{I(J+1)\}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J+1}|q_{ij} - \bar{q}_i - \theta|$ may be viewed as a first-order approximation of $I^{-1}\sum_{i=1}^{I} M_{\Gamma i}$ at $\Gamma = \tilde{\Gamma}_{conv}$ when matching with multiple controls. Taking $I \to \infty$, the implied approximation for $J > 1$ is

$$\mathrm{sgn}(m_{\tilde{\Gamma}_{conv}} - \eta) \approx \mathrm{sgn}\left(\mathrm{plim}\ \{I(J+1)\}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J+1}|q_{ij} - \bar{q}_i - \theta| - \eta\right), \tag{18}$$

where $\mathrm{sgn}(\cdot)$ is 1 if positive, -1 if negative, 0 otherwise.

In light of Proposition 3, this inspires using the right-hand side of (18) as a heuristic for whether the tilted sensitivity analysis will outperform the conventional approach.

The sign of the displayed difference varies considerably across distributions, often positive with right-skewed generative models in (12) and negative for left-skewed distributions. In the web-based supporting materials, we show for the difference in means statistic that the sign is non-negative whenever the marginal distribution of $\varepsilon_{0ij} - \sum_{\ell \neq j} \varepsilon_{0i\ell}/J$ is symmetric, and joint distribution of $\left(\varepsilon_{0i1} - \sum_{\ell \neq 1} \varepsilon_{0i\ell}/J, ..., \varepsilon_{0in_i} - \sum_{\ell \neq n_i} \varepsilon_{0i\ell}/J\right)^T$ is convex unimodal, with ε_{0ij} defined as in (12). A multivariate distribution is convex unimodal when it has a density f such that, for every $c > 0$, the set $\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq c\}$ is a centrally convex symmetric convex set (Dharmadhikari and Jogdeo, 1976, Definition 3.3). This condition holds for the multivariate normal among many other common multivariate distributions, suggesting that the tilted sensitivity analysis may confer a benefit for a broad class of symmetric generative models.

While heuristic arguments using Proposition 4 can aid intuition, we now use Monte-Carlo simulation to compare the design sensitivities of the tilted and conventional approaches under the generative model described in §4.1. We vary the number of matched controls the distribution F_ε in (12), and the choice of test statistic. We consider matching with $J = 2, 3$ and 5 controls and compare performance with three different test statistics. We use the difference in means statistic; the aligned rank statistic of Hodges and Lehmann (1963) with alignment by stratum means; and the m-statistic $T = \sum_{i=1}^{I} \sum_{j=1}^{J+1} Z_{ij} \sum_{\ell=1}^{J+1} \psi_{hu,2.5}(Y_{ij} - Y_{i\ell})$, where $\psi_{hu,2.5}(x) = \text{sgn}(x) \min\{|x|/\hat{s}, 2.5\}$ and $\hat{s}$ is the sample median of the $IJ(J-1)/2$ values $|Y_{ij} - Y_{i\ell}|$, $j \neq \ell$. This m-statistic is inspired by the loss function of Huber (1981), and is described in detail in Rosenbaum (2007). We consider distributions for the errors in (12) that are Normal, t_3, $\text{Exp}(1) - 1$, and $1 - \text{Exp}(1)$, where $\text{Exp}(\lambda)$ is an exponential distribution with rate parameter λ. The Normal and t_3 distributions are symmetric, while $\text{Exp}(1) - 1$, and $1 - \text{Exp}(1)$ are right-skewed and left-skewed respectively. We set $\tau/\sigma = 1/2$ where τ is defined as in (12) and $\sigma^2 = \text{Var}(\varepsilon_{1ij} - \varepsilon_{0i\ell})$, $j \neq \ell$; this accounts for differences in the variances of the generative models, measuring treatment effects in units of the standard deviations of the potential treated minus control differences for individuals in the same matched set. For each setting, we compute the design sensitivities of the tilted and conventional approaches through $I = 100,000$ Monte-Carlo simulations.

Table 1: Design sensitivities when matching with $J = 2$, $J = 3$ and $J = 5$ controls for the tilted and conventional sensitivity analysis. Three of test statistics are displayed across the columns. Across the rows, there are four error distributions $F_\varepsilon(\cdot)$: Normal, t_3, Exp(1)-1, and 1-Exp(1), where Exp(1) is the exponential distribution with rate parameter 1. The signal to noise ratio is $\tau/\sigma = 1/2$ for all settings.

		Difference in Means			Huber m, trim = 2.5			Aligned Rank		
		$J = 2$	$J = 3$	$J = 5$	$J = 2$	$J = 3$	$J = 5$	$J = 2$	$J = 3$	$J = 5$
Norm.	Conv.	4.08	4.35	4.86	3.88	4.20	4.64	3.69	4.06	4.43
	Tilt.	4.32	4.62	5.10	4.09	4.45	4.84	3.87	4.17	4.50
t_3	Conv.	4.97	5.38	5.89	5.19	5.80	6.54	5.13	5.76	6.70
	Tilt.	5.22	5.70	6.31	5.52	6.25	6.92	5.37	5.95	6.60
E(1)-1	Conv.	4.14	4.22	4.40	4.16	4.54	4.86	4.34	5.07	5.94
	Tilt.	5.37	6.37	8.06	5.45	7.13	9.24	5.62	7.54	11.0
1-E(1)	Conv.	5.09	6.15	8.01	5.23	6.17	7.96	4.96	5.91	7.08
	Tilt.	4.21	4.36	4.58	4.25	4.44	4.67	4.19	4.43	4.56

Table 1 shows the results. We see that as J increases, the design sensitivities of both the tilted and conventional approaches increase; see the web-based supporting information for a discussion of why this holds for tilted m-statistics. We see that tilting, in addition to providing a closed form for the design sensitivity, provides an improvement in design sensitivity over the conventional approach for the symmetric and right-skewed error distributions for all three test statistics and all choices of J except for the aligned rank statistic with $J = 5$ and a t_3 generative model, where the conventional approach slightly outperforms the tilted variant. The conventional approach has superior design sensitivity for the left-skewed generative distribution, 1-Exp(1). To assess the quality of the heuristic (18), we evaluated its Monte-Carlo estimate for each test statistic, generative model, and number of matched controls reflected in Table 1. The heuristic successfully predicted whether the tilted sensitivity analysis would outperform the conventional approach in all settings except for the aligned rank statistic with $J = 5$ and a t_3 generative model.

In the web-based supporting materials, we include an additional simulation study which plots the difference between $m_{\tilde{\Gamma}_{tilt}}$ and η as a function of τ/σ across a range of generative models. This plot illustrates that the comparisons in design sensitivity persist beyond a fixed value of the effect size.

4.4 Adaptive combination

Table 1 reveals that the tilted sensitivity analysis does not uniformly improve upon the conventional approach in terms of design sensitivity; rather, the improvement is tied to the underlying distribution of $q_{ij} - \bar{q}_i$. In the web-based supporting materials, we describe how one need not choose between the tilted and conventional approaches. Rather, the approaches may be combined to attain a design sensitivity no smaller than the maximum of the tilted and conventional design sensitivities by employing the approach for multivariate one-sided testing introduced in Cohen et al. (2020). In short, the method adaptively chooses a weighted combination of the tilted and conventional test statistics, allowing the method to choose whichever approach has better performance on the observed data set. In the web-based supporting material we also present simulation studies comparing the power of a sensitivity analysis using the tilted, conventional, and adaptive approaches in finite samples. These illustrate the minimal loss in power from adaptivity relative to whichever of the tilted and conventional approaches performed better in a given generative model. Furthermore, the simulations show the potential for large gains relative to the worse of the tilted and conventional approaches. See also Heng et al. (2021) for an alternative approach to adaptive combination of test statistics for improved design sensitivity.

5 Data examples

Design sensitivity provides but one of many metrics by which methods for sensitivity analysis may be judged. While the favorable setting of no bias and a treatment effect enables theoretical comparisons of competing methods, whether benefits under this setting reflect benefits in practice is not a foregone conclusion. In real-world examples, there may well be a treatment effect, but it is unlikely that the true value of Γ for which (1) holds is $\Gamma = 1$ (no unmeasured confounding). In this section, we compare the performance of the tilted, conventional, and adaptive approaches in several real data examples. For each matched comparison, we consider a sensitivity analysis for the sharp null hypothesis of no effect at $\alpha = 0.05$. For each method, we report the *sensitivity value*

(Zhao, 2019), the largest value of Γ in (1) for which the sensitivity analysis rejects the null hypothesis. Larger values for the design sensitivity are preferred, as they imply the findings are more robust to hidden bias.

We consider the following data sets, all of which are either available within packages on `CRAN` or are included in tabular form within the cited papers.

(SL) Smoking and lead levels in the blood. $I = 150$. One treated, four controls in each matched set (Rosenbaum, 2013).

(FM) Fish consumption and mercury levels in the blood. $I = 397$. One treated, two controls in each matched set (Rosenbaum, 2014).

(DC) Dropping out of high school and cognitive achievement. $I = 12$. One treated, two controls in each matched set (Gastwirth et al., 2000).

(AB) Alcohol and blood pressure. $I = 206$. One treated, two controls in each matched set (Rosenbaum, 2023a).

(LH) Light drinking and HDL cholesterol levels. $I = 200$. One treated, four controls in each matched set (Rosenbaum, 2023b).

(CM) Coffee and myocardial infarction. $I = 115$. Case-referent study. 27 pairs with one case, one referent. 88 triples with one case, two referents (Jick et al., 1973).

For the data sets with continuous outcomes, we consider four test statistics. As in the simulations in §4.3, we consider the difference in means statistic, the m-statistic $\psi_{hu,2.5}$, and the aligned-rank test of Hodges and Lehmann (1963). In addition, we consider the weighted rank test denoted `u868` in Rosenbaum (2024), highlighted therein for its impressive performance in sensitivity analysis. For the data set on coffee and myocardial infarction, we use the Mantel-Haenszel test statistic.

Table 2 shows the results. We see that for the data sets FM, DC, AB, and CM, the tilted sensitivity analysis outperformed the conventional approach for all choices of test statistic. For LH, the conventional sensitivity analysis outperforms the tilted approach for all test statistics. For SL, the tilted sensitivity analysis outperformed the conventional approach for all test statistics except for the `u868` rank test from Rosenbaum (2024). The adaptive approach generally lags behind the better of the two

	Difference in Means			Huber, trim=2.5			Aligned Rank			Rank Sum, `u868`		
	C	T	A	C	T	A	C	T	A	C	T	A
SL	1.49	1.53	1.52	2.07	2.18	2.15	2.00	2.10	2.07	1.50	1.49	1.49
FM	15.9	20.8	20.4	14.0	19.9	19.4	15.3	21.2	20.6	18.1	37.0	32.1
DC	1.32	1.34	1.32	1.30	1.34	1.31	1.36	1.38	1.35	1.63	1.64	1.57
AB	2.18	2.20	2.18	2.17	2.18	2.16	2.11	2.11	2.10	2.02	2.05	2.02
LH	3.74	3.64	3.69	3.38	3.30	3.34	3.30	3.15	3.25	4.47	4.11	4.86
CM	1.32	1.34	1.33	-	-	-	-	-	-	-	-	-

Table 2: The largest value of Γ such that the null hypothesis is rejected at $\alpha = 0.05$ using the conventional (C), tilted (T), and adaptive (A) approaches with various test statistics. The first five rows correspond to the observational studies with continuous outcomes, while the final row only uses the Mantel-Haenszel test since the data were binary.

approaches, but can perform materially better than the laggard; this is particularly striking in the findings from data set FM, where choosing the conventional sensitivity analysis would have yielded materially lower sensitivity values than the tilted or adaptive approaches. Observe that for `u868` rank sum test in data set AH, the adaptive procedure outperforms both the conventional *and* the tilted sensitivity analysis.

Unlike the other studies, the results for study CM involve a comparison of the tilted and conventional sensitivity analyses when using a sign-score statistic, as described in §3.1. The differences in performance are solely driven by different choices of weights for the Mantel-Haenszel test. For this data set, the sign-score weights yield slightly worse performance than the implict weights of the tilted approach.

6 Discussion

The tilted sensitivity analysis has appealing practical and theoretical properties: it may implemented in closed form without use of the separable algorithm of Gastwirth et al. (2000); it has a closed form expression for its design sensitivity even when matching with multiple controls; it provides improvements in design sensitivity for a range of generative models; and provides improvements in reported robustness to hidden bias in many real-data examples as illustrated in Table 2. Furthermore, one need not choose between the tilted and conventional approaches to sensitivity analysis with matched controls: the adaptive approach outlined in §4.4 provides insurance against a poor choice. One question which remains unaddressed is whether weights inspired by inverse conditional propensity score weighting, defined in (11), provide improvements in terms of design sensitivity and/or sensitivity values. These weights are linear functions of $\hat{G}_i(0) = n_i^{-1} \sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} - \bar{q}_i \leq 0\}$, which suggests that the underlying distribution of $q_{ij} - \bar{q}_i$ will play a role in describing circumstances under which these the weights are beneficial. More broadly, the connections between tilting and inverse probability weighting illustrated in §3.2 necessitate a detailed comparison between tilting and the broader classes of weighting-based approaches to sensitivity analysis under the marginal sensitivity model (Tan, 2006; Zhao et al., 2019; Dorn and Guo, 2022).

While the exposition herein has focused on tests of sharp null hypotheses for the general class of sum statistics, for the difference in means statistic Fogarty (2023) employed tilting in an investigation of tests of the weak null of no effect on average. There it was shown that the untilted difference in means fails to provide a valid sensitivity analysis for the average treatment effect, and that left-skewed generative models can provide large gaps between the nominal and actual Type I error rates; see Table 1, setting (f) of the online supplementary material of Fogarty (2023). It is of interest that left-skewed generative models are precisely the models where the conventional approach outperformed the tilted variant, as demonstrated in Table 1. Part of the deficit in power for the tilted approach in this setting may be attributed to the fact that it also provides valid inference for the average treatment effect under mild conditions.

That the tilted approach provided an improvement in design sensitivity relative to the conventional approach for all of the symmetric generative models considered in §4.3 raises the possibility of a theoretical guarantee along these lines. The following result represents a step in this direction.

Proposition 5. *Suppose that the joint distribution of* $\mathbf{q}_i - \bar{q}_i = (q_{i1} - \bar{q}_i, ..., q_{in_i} - \bar{q}_i)^T$ *is convex unimodal. Then, for any* $\Gamma > 1$, $M_{\Gamma i}$ *stochastically dominates* $n_i^{-1} \sum_{j=1}^{n_i} |q_{ij} - \bar{q}_i|$. *That is, for any* c,

$$pr\left(\frac{1}{n_i}\sum_{j=1}^{n_i} |q_{ij} - \bar{q}_i| \geq c\right) \leq pr(M_{\Gamma i} \geq c).$$

At $\tau = 0$ under the favorable setting described in §4.1, we have that $E\left(|T_i - \bar{q}_i|\right) = E\left(n_i^{-1}\sum_{j=1}^{n_i}|q_{ij} - \bar{q}_i|\right)$, Therefore, a corollary of Proposition 5 is that at $\tau = 0$ in (12), $E|T_i - \bar{q}_i| \leq E(M_{\Gamma i})$ for any $\Gamma > 1$ under convex unimodality. Providing a set of sufficient conditions under which $E|T_i - \bar{q}_i| \leq E(M_{\Gamma i})$ at $\Gamma = \tilde{\Gamma}_{tilt}$ when $\tau > 0$, implying by Proposition 3 that the tilted sensitivity analysis has superior design sensitivity to the conventional approach, remains an open question.

Acknowledgments

This work is supported in part by funds from the National Science Foundation (NSF DMS-2413484).

Appendix

A The impact of matching with multiple controls on design sensitivity after tilting

A closed form solution for the design sensitivity enables analytical comparisons between competing research designs and test statistics. As an illustration, we consider the impact of the number of matched controls on design sensitivity when using m-statistics, studied for the conventional sensitivity analysis by Monte-Carlo simulation in Rosenbaum (2013). M-statistics are statistics asymptotically equivalent to

$$T' = \sum_{i=1}^{I}\sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell}), \tag{19}$$

where ψ is a non-stochastic function assumed odd, monotone non-decreasing and uniformly continuous. Examples of m-statistics include $\psi(x) = x$, which returns $(n_i - 1)$ times the treated-minus-control difference in means in each matched set, along with various statistics inspired by the works of Maritz (1979) and Huber (1981). Observe that $q_{ij} = \sum_{\ell=1}^{n_i}\psi(Y_{ij} - Y_{i\ell})$ for all ij, and that $\bar{q}_i = 0$ for all i by oddness of $\psi(\cdot)$. If $n_i = J + 1$ for all i, then T' is, under the generative model considered in §4.1, the sum of I *iid* random variables. Moreover, for each i, $T_i' := \sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell})$ is the sum of J permutation symmetric random variables.

Let $\theta = E\{J^{-1} \sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell})\} = E\{\psi(y_{0i1} - y_{0i2} + \tau)\}$, and let $\eta_J = E|J^{-1} \sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell})\}| = E|J^{-1} \sum_{j=2}^{J+1} \psi(y_{0i1} - y_{0ij} + \tau)|$. Then, when matching with J controls and using m-statistics, the design sensitivity for the tilted sensitivity analysis is $\tilde{\Gamma}_{tilt,J} = (\eta_J + \theta)/(\eta_J - \theta)$. This closed form helps to clarify the role that the number of matched controls plays in determining the design sensitivity of the tilted sensitivity analysis: so long as η_J decreases with J, the design sensitivity will increase as the number of matched controls increases. For intuition on why this generally holds, let $v^2 = \text{var}\{\psi(y_{0i1} - y_{0i2} + \tau)\}$ and for $J > 1$ let $\rho = \text{cor}\{\psi(y_{0i1} - y_{0i2} + \tau), \psi(y_{0i1} - y_{0i3} + \tau)\}$. Then, in the generative model of §4.1, $\text{var}\left(J^{-1} \sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell})\right) = \rho v^2 + (1-\rho)v^2/J$. As $\rho > 0$ by Theorem 2.1 of Esary et al. (1967), increasing the number of matched controls leaves the expectation of $J^{-1} \sum_{j=1}^{n_i} Z_{ij} \sum_{\ell=1}^{n_i} \psi(Y_{ij} - Y_{i\ell})$ unchanged while reducing its variance. For many distributions, including the normal distribution, the decrease in variance as J increases implies a reduction in η_J as a function of J, and hence an improvement in design sensitivity for the tilted statistic.

B Comparing design sensitivities as a function of effect size

Proposition 3 provides a straightforward means of comparing the design sensitivities of the tilted and conventional approaches by way of Monte-Carlo simulation. It is particularly convenient to leverage Proposition 3 at $\Gamma = \tilde{\Gamma}_{tilt}$, given the closed-form expression for the tilted sensitivity analysis available from Theorem 2 as a function of η and θ. We now compare the tilted and conventional sensitivity analyses under the generative model described in §4.1, varying the number of matched controls and the distribution F_ε. We consider $\psi_{hu,2.5}(x) = \text{sgn}(x)\min\{|x|/\hat{s}, 2.5\}$, where $\hat{s}$ is the sample median of the $IJ(J-1)/2$ values $|Y_{ij} - Y_{i\ell}|$, $j \neq \ell$; the value 2.5 is referred to as the degree of *trimming* of the statistic. This particular choice of ψ in (19) is an m-statistic inspired by the loss function of Huber (1981); see Rosenbaum (2007) for more details. We consider choices for the marginal distributions for $(\varepsilon_{0ij}, \varepsilon_{1ij})$ that are (i) standard normal; (ii) t_3; (iii) standard double exponential; (iv) $\text{Exp}(1) - 1$; and (v) $1 - \text{Exp}(1)$, where $\text{Exp}(\lambda)$ is an exponential distribution with rate λ. All distributions have mean zero. Distributions (i)-(iii) are symmetric about zero, distribution (iv) is right-skewed, and distribution (v) is left-skewed. For each choice of generative distribution, and number of matched controls, we evaluate Monte-Carlo estimates of $m_{\tilde{\Gamma}_{tilt}} - \eta$ for a range of values of $\tau > 0$ based upon $I = 100,000$ matched sets. For $\psi_{hu,2.5}$, the probability limits for η and $m_{\tilde{\Gamma}_{tilt}}$ are the expectations of $|T_i - \bar{q}_i| = |T_i|$ and $M_{\Gamma i}$ evaluated at $\Gamma = \tilde{\Gamma}_{tilt}$ respectively when $\psi_{hu,2.5}$ is replaced by $\tilde{\psi}_{hu,2.5}(x) = \text{sgn}(x)\min\{|x|/s, 2.5\}$, where s is the probability limit for $\hat{s}$.

Figure 1 shows the Monte-Carlo estimates of $m_{\tilde{\Gamma}_{tilt}} - \eta$ for $\psi_{hu,2.5}$ with different choices of numbers of control J and generative distributions as a function of τ/σ, where $\sigma^2 = \text{Var}(\varepsilon_{1ij} - \varepsilon_{0i\ell})$, $j \neq \ell$; this accounts for differences in the variances of the generative models, measuring treatment effects in units of the standard deviations of the potential treated minus control differences for individuals in the same matched set. We see that for the normal, t_3, double exponential, and $\text{Exp}(1) - 1$ distributions, the y coordinates of the plots are above zero for values of τ/σ and all values for the number of controls displayed. By Proposition 3, this implies that the titled sensitivity analysis had a larger design sensitivity than the conventional approach for each of these scenarios and for all values of τ/σ. When the distribution is instead $1 - \text{Exp}(1)$, a left-skewed distribution, we see that the conventional sensitivity analysis outperforms the tilted approach for all test statistics and numbers of controls displayed.

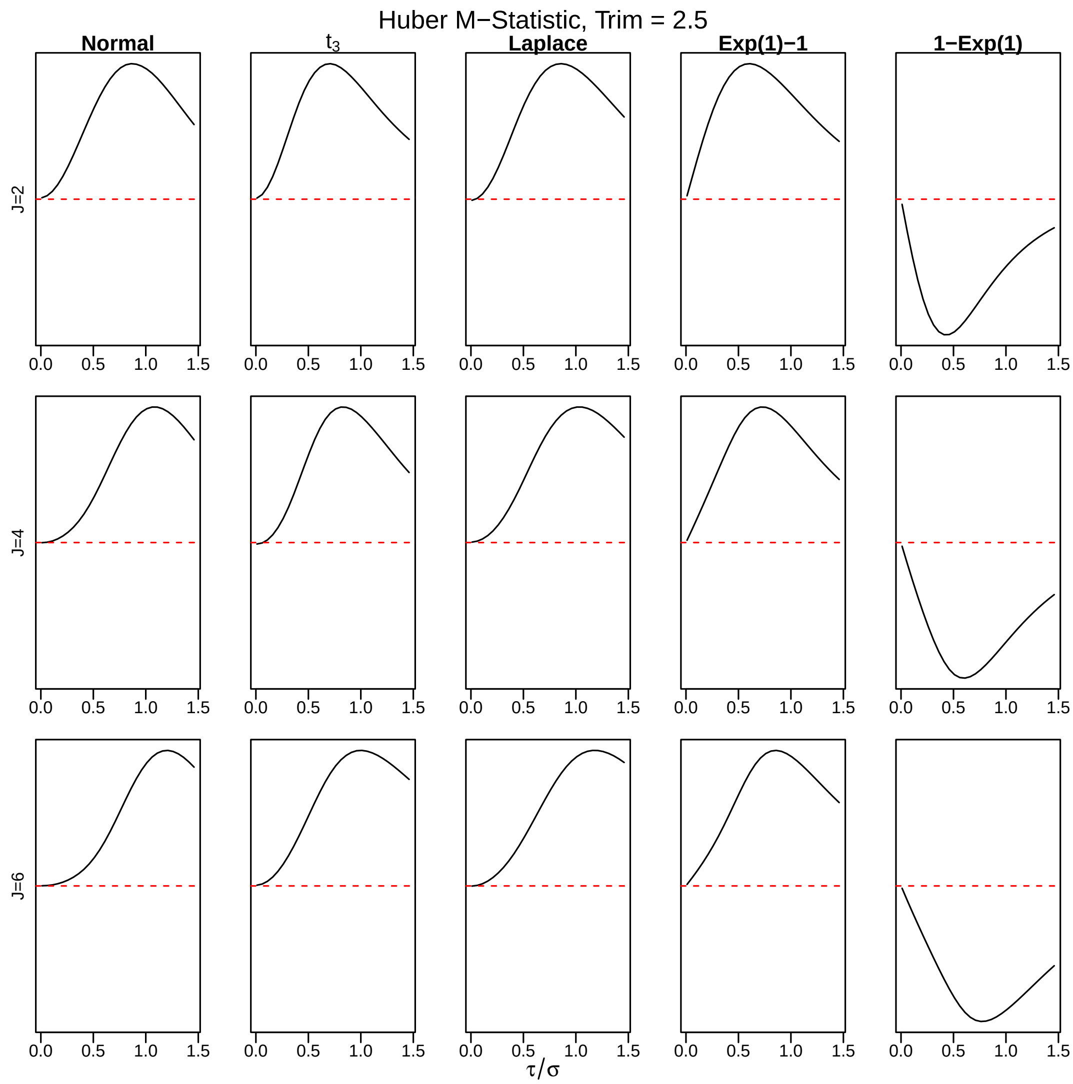

Figure 1: Comparing the design sensitivities of the tilted and conventional approaches for the m-statistic $\psi_{hu,2.5}$. The plots show Monte-Carlo estimates of $m_{\tilde{\Gamma}_{tilt}} - \eta$ as a function of τ/σ, along with a horizontal dotted line at $y = 0$ for a given choice of the marginal distributions for $(\varepsilon_{0ij}, \varepsilon_{1ij})$ (varying across columns) and number of matched controls (varying across rows) in the generative model (1). By Proposition 3, if for any given generative model and number of matched controls J the function lies above the dotted line at a given value of τ/σ, the tilted sensitivity analysis has a better design sensitivity than the conventional approach for that value of τ/σ; otherwise, the conventional approach has the larger design sensitivity.

C Adaptive combination of the tilted and conventional approaches

C.1 Adaptive weighting through a two-person game

We now describe how the approach of Cohen et al. (2020) for multivariate one-sided testing may be used to adaptively combine the tilted and conventional sensitivity analyses. As will be asserted in Proposition 6, the resulting procedure has design sensitivity that is no smaller than the *maximum* of the design sensitivities of the tilted and conventional approaches. For a fixed $\Gamma \geq 1$ let $\mathbf{a}_{ij} = (a_{ij1}, a_{ij2})^T$ with $a_{ij1} = q_{ij}$, $a_{ij2} = w_{\Gamma i}[q_{ij} - \bar{q}_i - \{(\Gamma-1)/(1+\Gamma)\}|q_{ij} - \bar{q}_i|]$. Consider the vector of test statistics $\mathbf{A}_\Gamma = (T, T_\Gamma(w_\Gamma))^T$, which contains the conventional test statistic in its first element and the tilted test statistic with weights w_Γ in its second. For any fixed value for the assignment probabilities $\boldsymbol{\varrho} = (\varrho_{11}, \varrho_{12}, ..., \varrho_{In_I})^T$, the expectation and covariance for $\mathbf{A}_\Gamma$ are, for $k, \ell = 1, 2$,

$$\mu_{\Gamma k}(\varrho) = \sum_{i=1}^{I}\sum_{j=1}^{n_i} a_{ijk}\varrho_{ij}; \;\; \Sigma_{\Gamma k\ell}(\varrho) = \sum_{i=1}^{I}\left\{\sum_{j=1}^{n_i} a_{ijk}a_{ij\ell}\varrho_{ij} - \left(\sum_{j=1}^{n_i} a_{ijk}\varrho_{ij}\right)\left(\sum_{j=1}^{n_i} a_{ij\ell}\varrho_{ij}\right)\right\}.$$

Let $\mathcal{P}_\Gamma$ be the set of possible assignment probabilities when (1) holds at Γ. Then, for any $\Gamma \geq 1$ in (1), define B_Γ as

$$B_\Gamma = \min_{\boldsymbol{\varrho}\in\mathcal{P}_\Gamma} \sup_{\boldsymbol{\lambda}\in\Lambda_+} \max\left[0, \frac{\boldsymbol{\lambda}^T\{\mathbf{A}_\Gamma - \boldsymbol{\mu}_\Gamma(\boldsymbol{\varrho})\}}{\{\boldsymbol{\lambda}^T\Sigma_\Gamma(\boldsymbol{\varrho})\boldsymbol{\lambda}\}^{1/2}}\right]^2, \tag{20}$$

where $\Lambda_+ = \{(\lambda_1, \lambda_2)^T : \lambda_1 \geq 0, \lambda_2 \geq 0, (\lambda_1, \lambda_2)^T \neq (0,0)^T\}$ is the nonnegative orthant excluding the zero vector. Observe that for any fixed λ, (20) would conduct a sensitivity analysis based upon the linear combination $\lambda_1 T + \lambda_2 T_\Gamma(w_\Gamma)$ with a greater-than alternative when (1) holds at Γ. Assuming bivariate asymptotic normality for A_Γ, for $\alpha \leq 0.5$ the worst-case deviate for a fixed $\boldsymbol{\lambda}$ could be compared to the square of the $1-\alpha$ quantile of a standard normal quantile to determine whether or not the sensitivity analysis rejects at Γ. Optimizing $\boldsymbol{\lambda}$ over Λ_+ further allows for adaptively chosen linear combinations of the tilted and conventional sensitivity analysis that may outperform either approach on its own. The critical value used to perform inference, $c_{\Gamma,\Lambda_+}(\alpha)$, must account for the additional optimization over Λ, and may no longer be chosen based on a standard normal quantile. Instead, $c_{\Gamma,\Lambda_+}(\alpha)$ is a quantile from chi-bar-square distribution; see Cohen et al. (2020) for details. The proposed method for combining the conventional and tilted approaches rejects the sharp null when $B_\Gamma \geq c_{\Gamma,\Lambda_+}(\alpha)$. As $c_{\Gamma,\Lambda_+}(\alpha) \geq \{\Phi^{-1}(1-\alpha)\}^2$, adaptively optimizing the test statistic over all nonnegative linear combinations requires a larger critical value. So long as $\mathbf{A}_\Gamma$ has a bivariate normal limiting distribution under the sharp null, the adaptive sensitivity analysis asymptotically controls the Type I error rate at α if (1) holds at Γ. Under the generative model considered in §4.1, the procedure can further provide uniform improvements in design sensitivity, as described in the following proposition.

Proposition 6. *Let $\tilde{\Gamma}_{tilt}$ and $\tilde{\Gamma}_{conv}$ be the design sensitivities for the tilted and conventional sensitivity analyses respectively when using test statistics $T_\Gamma(w_\Gamma)$ and T. Consider the procedure that rejects the sharp null when $B_\Gamma \geq c_{\Gamma,\Lambda_+}(\alpha)$; call its design sensitivity $\tilde{\Gamma}_{adapt}$. Then, under Assumptions 1 and 2, $\tilde{\Gamma}_{adapt} \geq \max\{\tilde{\Gamma}_{tilt}, \tilde{\Gamma}_{conv}\}$.*

Proposition 6 is an immediate consequence of Theorem 2 of Cohen et al. (2020), and the proof is omitted. It states that in terms of design sensitivity, one need not choose *a priori* between the tilted and conventional sensitivity analyses. Rather, the two approaches may be combined with a resulting design sensitivity that is at least as good as the better of the two individual approaches. .

C.2 The power of a sensitivity analysis

Proposition 6 suggests that in terms of design sensitivity alone, the adaptive procedure should be preferred: it is never worse than the tilted and conventional sensitivity analyses, and it can be strictly better than both of them when $n_i > 2$. Recall however that design sensitivity calculations focus solely on the limit as $I \to \infty$ and ignore small-sample considerations. In moderate sample sizes, the differences in the sizes of the critical values used to perform inference in the adaptive approach (based upon the $\bar{\chi}^2$ distribution) versus either the conventional or titled approaches (based upon the standard normal) will impact the power of these methods, such that the adaptive approach may not provide an improvement in finite samples.

To investigate this further, we consider power simulations under the generative model in §4.1 for moderate sample sizes I. In each of 10,000 Monte Carlo simulations, we simulate data sets with $I = 200$ and $J = 3$, so that each matched set has $n_i = 4$, with one treated and three control observations. We use the generative model described in §4.1 to generate potential responses under treatment and control. We consider three marginal distributions for ε_{ijz} in (12): (i) t_3; (ii) $\text{Exp}(1) - 1$; and (iii) $1 - \text{Exp}(1)$, where $\text{Exp}(\lambda)$ denotes an exponential distribution with rate λ. The treatment effect to noise ratio is set at $\tau/\sigma = 1/2$ for all three distributions, with $\sigma = \sqrt{\text{Var}(\varepsilon_{1ij} - \varepsilon_{0i\ell})}$, $j \neq \ell$, as before, and we proceed using the difference in means as the test statistic. From Table 1, we see that the tilted sensitivity analysis has higher design sensitivity than the conventional approach in generative model (i) and (ii) (5.70 versus 5.38 and 6.37 versus 4.22 respectively), while the conventional approach has higher design sensitivity than the tilted sensitivity analysis in generative model (iii) (6.15 versus 4.36). By Proposition 6, the adaptive procedure has a design sensitivity no smaller than the larger of the two individual design sensitivities. Therefore, the design sensitivity for the adaptive procedure will be lower bounded by 5.70, 6.37, and 6.15 in settings (i)-(iii).

For each choice of generative model, we produce Monte-Carlo estimates across our 10,000 simulations for the power of the tilted, conventional, and adaptive sensitivity analyses for Γ ranging from 1 to 6. Figure 2 shows the results. The plots in the top half of the figure are from setting (i), the $t_{df=3}$ generative model, and reveal that while the tilted sensitivity analysis has higher design sensitivity than the conventional approach, it does not provide uniformly higher power than the conventional approach for all Γ in finite samples. The zoomed-in view between $\Gamma = 2.5$ and $\Gamma = 4.0$ in the top-right shows this more clearly. For $\Gamma < 3.3$, the conventional approach improves upon the power of the tilted sensitivity analysis, while for $\Gamma \geq 3.3$ the tilted sensitivity analysis has higher power. The figure illustrates additional benefits of the adaptive procedure: despite using a larger critical value, the adaptive procedure attains power larger than that of both the tilted sensitivity analysis and the conventional sensitivity analysis for

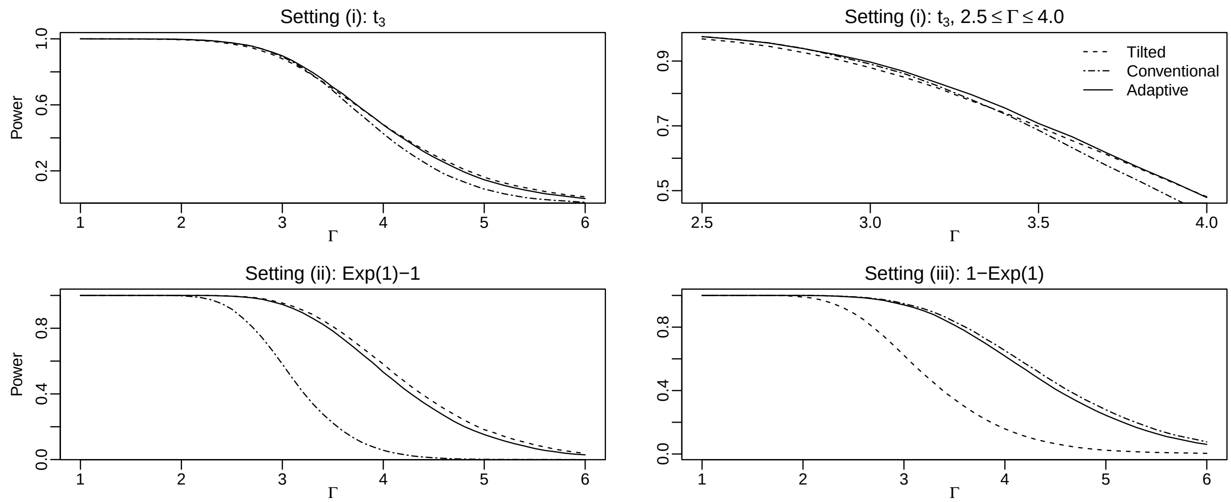

Figure 2: The power of a sensitivity analysis for the tilted (dashed line), conventional (dot-dashed line), and adaptive (solid line) procedures with $I = 200$, $J = 3$, $\tau/\sigma = 1/2$, with the marginal distribution for ε_{ijz} in (12) chosen to be (i) $t_{df=3}$; (ii) $\text{Exp}(1) - 1$, and (ii) $1 - \text{Exp}(1)$, where $\text{Exp}(1)$ is an exponential distribution with rate 1. The plot in the top-right zooms in on the power plots for setting (i) between $\Gamma = 2.5$ and $\Gamma = 4$ to better visualize behavior within that range. Curves are Monte-Carlo estimates calculated using 10,000 data sets.

$\Gamma \in [1.8, 3.9]$. Moreover, for $\Gamma > 3.9$ the adaptive approach hews closely to the power curve of the tilted sensitivity analysis (which attains the largest power in that range). The largest gap in the power between the adaptive and tilted sensitivity analyses in that range is 0.017, reflecting a small price for adaptivity.

The plots from setting (ii), bottom-left, and setting (iii), bottom-right correspond to generative models for which there is a large gap between the design sensitivities of the titled and conventional approaches. The tilted sensitivity analysis has larger design sensitivity in setting (ii), and the conventional approach has larger design sensitivity in setting (iii). The power profiles reflect a similar behavior: in terms of power, the tilted sensitivity analysis dominates the conventional approach in the plot from setting (ii), but is dominated by the conventional approach in the plot from setting (iii). The discrepancies are particularly large for Γ ranging from 3 to 5 in those plots. The power of the conventional approach can lag behind that of the tilted sensitivity analysis by as much as 0.60 in the setting (ii) plot, and the tilted approach's power is smaller than that of the conventional approach by as much as 0.51 in the setting (iii) plot. As anticipated, the adaptive procedure closely tracks the power profile of the superior method in each plot by emphasizing the superior procedure in its weighted combination, providing insurance against choosing an underperforming method. The cost of this insurance is the gap between the power of the adaptive approach and the superior method for each model. The maximal value for this gap is 0.05 in the second plot, and 0.04 in the third.

D Proofs and Examples

D.1 Proof of Proposition 1

As described in §2.2, the worst-case expectation $\mu_{\Gamma i}$ returned by the separable algorithm is

$$\mu_{\Gamma i} = \max_{a=1,\ldots,n_i-1} \frac{\Gamma \sum_{j=1}^{a} q_{ij} + \sum_{j=a+1}^{n_i} q_{ij}}{\Gamma a + (n_i - a)},$$

where $q_{i1} \geq \ldots \geq q_{in_i}$

We claim that $\mu_{\Gamma i} = \bar{q}_i + \{(\Gamma - 1)/(1+\Gamma)\} M_{\Gamma i}$, where

$$M_{\Gamma i} = \texttt{SOLVE}\left\{ c : \frac{1}{n_i} \sum_{j=1}^{n_i} \left| q_{ij} - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right) c \right| = c \right\}.$$

To show this, first note

$$\begin{aligned}
\mu_{\Gamma i} &= \max_{a=1,\ldots,n_i-1} \frac{\Gamma \sum_{j=1}^{a} q_{ij} + \sum_{j=a+1}^{n_i} q_{ij}}{\Gamma a + (n_i - a)} \\
&= \bar{q}_i + \max_{a=1,\ldots,n_i-1} \frac{\Gamma \sum_{j=1}^{a} (q_{ij} - \bar{q}_i) + \sum_{j=a+1}^{n_i} (q_{ij} - \bar{q}_i)}{\Gamma a + (n_i - a)}. \\
&= \bar{q}_i + \max_{a=1,\ldots,n_i-1} \frac{(\Gamma - 1) \sum_{j=1}^{a} (q_{ij} - \bar{q}_i)}{(\Gamma - 1)a - (\Gamma - 1)n_i/2 + (\Gamma + 1)n_i/2}. \\
&= \bar{q}_i + \left(\frac{\Gamma - 1}{1+\Gamma}\right) \max_{a=1,\ldots,n_i-1} \frac{\sum_{j=1}^{a} (q_{ij} - \bar{q}_i)}{n_i/2 + \frac{\Gamma-1}{1+\Gamma}(a - n_i/2)},
\end{aligned}$$

where the third line uses that $\sum_{j=1}^{n_i} (q_{ij} - \bar{q}_i) = 0$, so $\sum_{i=1}^{a} (q_{ij} - \bar{q}_i) = -\sum_{i=a+1}^{n_i} (q_{ij} - \bar{q}_i)$.

We now show that

$$M_{\Gamma i} := \max_{a=1,\ldots,n_i-1} \frac{\sum_{j=1}^{a}(q_{ij}-\bar{q}_i)}{n_i/2+\frac{\Gamma-1}{1+\Gamma}(a-n_i/2)} = \texttt{SOLVE}\left\{c:\ \frac{1}{n_i}\sum_{j=1}^{n_i}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right| = c\right\}.$$

Let $N(a) = \sum_{j=1}^{a}(q_{ij}-\bar{q}_i)$ and $D(a) = n_i/2+\frac{\Gamma-1}{1+\Gamma}(a-n_i/2)$ be the numerator and denominator of the objective function on the left hand side. Then, $M_{\Gamma i} = \max_{a=1,\ldots,n_i-1} N(a)/D(a)$, which is a fractional optimization problem. For any c, consider instead the parametric problem

$$H_{\Gamma i}(c) = \max_{a=1,\ldots,n_i-1} N(a) - cD(a). \tag{21}$$

By Dinkelbach (1967), we have that $H_{\Gamma i}(M_{\Gamma i}) = 0$ and that the zero is unique. That is, the zero of the parametric problem as a function of c is attained when setting c equal to the maximizer of the fractional problem. Moreover, we see that $H_{\Gamma i}(c) > 0$ if and only if $M_{\Gamma i} > c$.

Consider a fixed value for c in $H_{\Gamma i}(c)$:

$$H_{\Gamma i}(c) = \max_{a=1,\ldots,n_i-1}\sum_{j=1}^{a}\left\{q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right\} + \left(\frac{\Gamma-1}{1+\Gamma}\right)(n_i/2)c - (n_i/2)c.$$

The arg max of the above is simply $a^* = \sum_{j=1}^{n_i}\mathbb{1}\left\{q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c > 0\right\}$, resulting in only the non-negative values of $q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c$ being summed. As $\sum_{j=1}^{n_i}\left\{q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right\} = -n_i\left(\frac{\Gamma-1}{1+\Gamma}\right)c$ since $\sum_{j=1}^{n_i}(q_{ij}-\bar{q}_i) = 0$, we have

$$\begin{aligned}
\sum_{j=1}^{a^*}\left\{q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right\} &= \sum_{j=1}^{a^*}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right| \\
&= -n_i\left(\frac{\Gamma-1}{1+\Gamma}\right)c - \sum_{j=a^*+1}^{n_i}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right| \\
&= -(n_i/2)\left(\frac{\Gamma-1}{1+\Gamma}\right)c + (1/2)\sum_{j=1}^{n_i}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right|
\end{aligned}$$

Hence,

$$H_{\Gamma i}(c) = (1/2)\sum_{j=1}^{n_i}\left|q_{ij} - \bar{q}_i - \left(\frac{\Gamma - 1}{1+\Gamma}\right)c\right| - (n_i/2)c \tag{22}$$

As $M_{\Gamma i}$ is the zero of the above, it must satisfy $n_i^{-1}\sum_{j=1}^{n_i}\left|q_{ij} - \bar{q}_i - \left(\frac{\Gamma-1}{1+\Gamma}\right)M_{\Gamma i}\right| = M_{\Gamma i}$ as desired.

D.2 Theorem 1 and the separable algorithm

D.2.1 Proof of Lemma 1

The bound on the expectation was proved as Lemma 1 in Fogarty (2023), but is included here for completeness. For any vector of unmeasured confounders $\mathbf{u}_i = (u_{i1}, ..., u_{in_i})$, if (1) holds at Γ we have

$$E_{\mathbf{u}_i}(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) = \frac{\sum_{j=1}^{n_i}\exp\{\log(\Gamma)u_{ij}\}[q_{ij} - \bar{q}_i - \{(\Gamma-1)/(1+\Gamma)\}\,|q_{ij} - \bar{q}_i|]}{\sum_{j=1}^{n_i}\exp\{\log(\Gamma)u_{ij}\}} \tag{23}$$

Observe that

$$q_{ij} - \bar{q}_i - \left(\frac{\Gamma-1}{1+\Gamma}\right)|q_{ij} - \bar{q}_i| = \begin{cases} 2(q_{ij} - \bar{q}_i)/(1+\Gamma) & q_{ij} > \bar{q}_i \\ 2\Gamma(q_{ij} - \bar{q}_i)/(1+\Gamma) & q_{ij} \leq \bar{q}_i \end{cases}.$$

The numerator in (23) is maximized by setting $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$, resulting in the largest possible multiplier, $\exp\{\log(\Gamma)u_{ij}\} = \Gamma$, for positive values of $q_{ij} - \bar{q}_i$, and the smallest possible value of the multiplier, $\exp\{\log(\Gamma)u_{ij}\} = 1$ to nonpositive values of $q_{ij} - \bar{q}_i$. Under this choice of u_{ij}, the numerator of (23) becomes $2\Gamma/(1+\Gamma)\sum_{j=1}^{n_i}(q_{ij} - \bar{q}_i) = 0$ as $\sum_{j=1}^{n_i}(q_{ij} - \bar{q}_i) = 0$. The confounder $u_{ij} = \mathbb{1}\{q_{ij} \geq \bar{q}_i\}$ thus also maximizes (23), as any other choice u'_{ij} will yield an expectation that is less than or equal to zero.

If there are values of q_{ij} for which $q_{ij} = \bar{q}_i$, $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$ is not the unique maximizer of the expectation. In this case, $\tilde{u}_{ij} = \mathbb{1}\{q_{ij} \geq \bar{q}_i\}$ would also maximize the expectation. In the separable algorithm, if there are multiple values for the unmeasured confounder attaining the maximal expectation for $T_{\Gamma i}$, the one which maximizes the variance is chosen. To complete the proof of the lemma, we must thus show that of the values for the unmeasured confounder attaining the largest expectation, $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$ produces the largest possible variance for $T_{\Gamma i}$.

For any vector u_i for which $E_{\mathbf{u}_i}(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) = 0$, the corresponding variance for $T_{\Gamma i}$ is

$$\mathrm{var}_{u_i}(T_{\Gamma i} \mid \mathcal{F}, \mathcal{Z}) := \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} u_{ij}}\right)\sum_{j=1}^{n_i}\frac{(q_{ij} - \bar{q}_i)^2}{\exp\{\log(\Gamma)\mathbb{1}\{q_{ij} > \bar{q}_i\}\}}.$$

From this, since $\Gamma \geq 1$ we see that the variance is maximized by minimizing the value for $\sum_{j=1}^{n_i} u_{ij}$. This is accomplished by setting $u_{ij} = \mathbb{1}\{q_{ij} > \bar{q}_i\}$, rather than $u_{ij} = \mathbb{1}\{q_{ij} \geq \bar{q}_i\}$.

D.2.2 Proof of Theorem 1

Suppose that the sharp null hypothesis holds and that the sensitivity model (1) holds at Γ. Under the conditions on q_{ij} and $w_{\Gamma i}$ provided in the Theorem's statement, Theorem 1 of Hájek et al. (1999) applies to $T_\Gamma(w_\Gamma)$. Let $\vartheta_{\Gamma i}$ and $\sigma^2_{\Gamma i}$ be the true expectation and variance of $T_{\Gamma i}$. We thus have that for every ϵ there exists an I_1 such that for any k, $I \geq I_1$ implies

$$\mathrm{pr}\left(\frac{T_\Gamma(w_\Gamma) - \sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i}}{\sqrt{\sum_{i=1}^{I} w_{\Gamma i}^2 \sigma_{\Gamma i}^2}} \geq k\right) \leq 1 - \Phi(k) + \epsilon$$

As $\vartheta_{\Gamma i}$ and $\sigma^2_{\Gamma i}$ are unknown, the separable algorithm attempts to upper bound the above tail probability when (1). By Lemma 1, $\vartheta_{\Gamma i} \leq 0$ for all i. Moreover, for matched sets i where $\vartheta_{\Gamma i} = 0$, we have that $\tilde{\nu}^2_{\Gamma i} \geq \sigma^2_{\Gamma i}$, where $\tilde{\nu}^2_{\Gamma i}$ is the candidate variance returned by the separable algorithm.

For any $k > 0$, it suffices to show that there exists an I_2 such that $I > I_2$ implies

$$\mathrm{pr}\left(T_\Gamma(w_\Gamma) - \sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i} - k\sqrt{\sum_{i=1}^{I} w_{\Gamma i}^2 \tilde{\sigma}_{\Gamma i}^2} \geq 0\right) \leq \mathrm{pr}\left(T_\Gamma(w_\Gamma) - k\sqrt{\sum_{i=1}^{I} w_{\Gamma i}^2 \tilde{\nu}_{\Gamma i}^2} \geq 0\right).$$

Equivalently, we show that there exists an I_2 such that $I > I_2$ implies

$$k\left(\sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2 \sigma_{\Gamma i}^2} - \sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2 \tilde{\nu}_{\Gamma i}^2}\right) \leq -I^{-1/2}\sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i}$$

Again by Lemma 1 we have $\vartheta_{\Gamma i} \leq 0$ for all i. Because of this, when $I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\sigma_{\Gamma i}^2 \leq I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2$, the inequality holds. Consider instead the case that $I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\sigma_{\Gamma i}^2 > I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2$. Using concavity of $f(x) = \sqrt{x}$ and recalling $\mathcal{B}_I = \{i : \sigma_{\Gamma i}^2 > \tilde{\nu}_{\Gamma i}^2\}$, we have

$$\begin{aligned}
&k\left(\sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2 \sigma_{\Gamma i}^2} - \sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2 \tilde{\nu}_{\Gamma i}^2}\right) \\
&\leq \frac{k}{2}\left(\frac{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2(\sigma_{\Gamma i}^2 - \tilde{\nu}_{\Gamma i}^2)}{\sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}}\right) \\
&\leq \frac{k}{2}\sqrt{I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2(\sigma_{\Gamma i}^2 - \tilde{\nu}_{\Gamma i}^2)}\sqrt{\frac{I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2\sigma_{\Gamma i}^2 - I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}}.
\end{aligned}$$

Under the assumption of bounded stratum sizes, there exist constants $C_{\Gamma 1}, C_{\Gamma 2}$ such that $\sigma_{\Gamma i}^2 \leq C_{\Gamma 1}\sum_{j=1}^{n_i} q_{ij}^2$ and $\tilde{\nu}_{\Gamma i}^2 \geq C_{\Gamma 2}\sum_{j=1}^{n_i} q_{ij}^2$ for all i. Therefore,

$$\sqrt{\frac{I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2\sigma_{\Gamma i}^2 - I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}} = O(1).$$

We have further assumed that $\sqrt{I^{-1}\sum_{i\in\mathcal{B}_I} w_{\Gamma i}^2(\sigma_{\Gamma i}^2 - \tilde{\nu}_{\Gamma i}^2)} = o(I^{-1/2}\sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i})$, which in turn implies $k\left(\sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\sigma_{\Gamma i}^2} - \sqrt{I^{-1}\sum_{i=1}^{I} w_{\Gamma i}^2\tilde{\nu}_{\Gamma i}^2}\right) = O(1)o(I^{-1/2}\sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i}) = o(I^{-1/2}\sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i})$. As $I^{-1/2}\sum_{i=1}^{I} w_{\Gamma i}\vartheta_{\Gamma i} \leq 0$ by Lemma 1, the proof is complete.

D.2.3 A pathological example where the separable algorithm fails

Let $n_i = 4$ for all $i = 1, ..., I$, and let $q_{i1} = 1, q_{i2} = i^{-2}, q_{i3} = -i^{-2}, q_{i4} = -1$; note $\bar{q}_i = 0$ for all i. The tilted test statistic applies the separable algorithm to $q_{ij\Gamma} = q_{ij} - \{(\Gamma - 1)/(1 + \Gamma)\}|q_{ij}|$ when conducting a sensitivity analysis when (1) holds at Γ, yielding by Lemma 1 $u_{i1} = u_{i2} = 1$, $u_{i3} = u_{i4} = 0$, a worst-case expectation of 0, and a variance

$$\tilde{\nu}^2_{\Gamma i} = \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{2\Gamma + 2}\right)(1 + i^{-4})(1 + \Gamma^{-1})$$

Suppose that in reality $u_{i1} = 1$, $u_{i2} = u_{i3} = u_{i4} = 0$. Then, the true expectation and variance would be

$$\begin{aligned}
\vartheta_{\Gamma i} &= \frac{2\Gamma}{1+\Gamma}\frac{(\Gamma^{-1} - 1)i^{-2}}{\Gamma + 3} \\
\sigma^2_{\Gamma i} &= \frac{2\Gamma + 2}{\Gamma + 3}\tilde{\nu}^2_{\Gamma i} + \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{\Gamma + 3}\right) i^{-4}(\Gamma^{-2} - \Gamma^{-1}) - \vartheta^2_{\Gamma i}
\end{aligned}$$

Evaluating the Riemann zeta function at $s = 2$ and $s = 4$, we have $\sum_{i=1}^{\infty} i^{-2} = \pi^2/6$, and $\sum_{i=1}^{\infty} i^{-4} = \pi^4/90$. Therefore:

$$\begin{aligned}
I^{-1}\sum_{i=1}^{I}\tilde{\nu}^2_{\Gamma i} &\to \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{2\Gamma + 2}\right)(1 + \Gamma^{-1}) \\
I^{-1}\sum_{i=1}^{I}\sigma^2_{\Gamma i} &\to \frac{4\Gamma^2}{(1+\Gamma)^2}\left(\frac{1}{\Gamma + 3}\right)(1 + \Gamma^{-1}) \\
I^{-1/2}\sum_{i=1}^{I}\vartheta_{\Gamma i} &\to 0.
\end{aligned}$$

Noting that the limiting value of $I^{-1}\sum_{i=1}^{I}\tilde{\nu}^2_{\Gamma i}$ is strictly smaller than that of $I^{-1}\sum_{i=1}^{I}\tilde{\nu}^2_{\Gamma i}$ for $\Gamma > 1$ by a factor $(\Gamma + 3)/(2\Gamma + 2)$, this implies that for any $k > 0$, there exists an I_2 such that $I > I_2$ implies

$$k\left(\sqrt{I^{-1}\sum_{i=1}^{I}\sigma^2_{\Gamma i}} - \sqrt{I^{-1}\sum_{i=1}^{I}\tilde{\nu}^2_{\Gamma i}}\right) \geq -I^{-1/2}\sum_{i=1}^{I}\vartheta_{\Gamma i},$$

which implies under asymptotic normality that the separable algorithm provides a Type I error rate larger than α for any $\alpha < 1/2$.

To explain why the separable algorithm fails here, the particular form of q_{ij} considered allows one to find a pattern of hidden bias not used by the separable algorithm for which the expectation is smaller for each i, in an asymptotically negligible way, than the expectation deployed in the separable algorithm, but where the variance is larger in a manner that persists asymptotically. Since the separable algorithm focuses first on finding the largest expectation, it fails to identify this alternative pattern of hidden bias with a larger variance and an asymptotically equivalent expectation at $\sqrt{I}$ scaling.

D.3 Proposition 2, and an illustration with the Mantel-Haenszel statistic

D.3.1 Proof of Proposition 2

For statistics of the form (10), recalling $a_{i1} \geq a_{i2}$ we have

$$\begin{aligned}
q_{ij} &\in \{a_{i1}, a_{i2}\}\\
\bar{q}_i &= a_{i2} + (a_{i1} - a_{i2})\frac{\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} = a_{i1}\}}{n_i}\\
\mathbb{1}\{q_{ij} > \bar{q}_i\} &= \mathbb{1}\{q_{ij} = a_{i1}\}\\
\mu_{\Gamma i} &= a_{i2} + \frac{\Gamma(a_{i1} - a_{i2})\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}
\end{aligned}$$

$T_i - \mu_{\Gamma i}$ can take on two possible values: $a_{i1} - \mu_{\Gamma i}$ or $a_{i2} - \mu_{\Gamma i}$. These may be expressed as

$$a_{i1} - \mu_{\Gamma i} = \frac{(a_{i1} - a_{i2})(n_i - \sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\})}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}} = \left(\frac{n_i}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}\right)(a_{i1} - \bar{q}_i) \tag{24}$$

$$a_{i2} - \mu_{\Gamma i} = \frac{-\Gamma(a_{i1} - a_{i2})\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}} = \left(\frac{n_i}{n_i + (\Gamma - 1)\sum_{j=1}^{n_i} \mathbb{1}\{q_{ij} > \bar{q}_i\}}\right)\Gamma(a_{i2} - \bar{q}_i), \tag{25}$$

The tilted test statistic $T_i - \bar{q}_i - \{(\Gamma-1)/(1+\Gamma)\}|T_i - \bar{q}_i|$ takes on the following two values (note $a_{i1} \geq \bar{q}_i \geq a_{i2}$):

$$a_{i1} - \bar{q}_i - \left(\frac{\Gamma-1}{1+\Gamma}\right)|a_{i1} - \bar{q}_i| = \left(\frac{2}{1+\Gamma}\right)(a_{i1} - \bar{q}_i); \tag{26}$$

$$a_{i2} - \bar{q}_i - \left(\frac{\Gamma-1}{1+\Gamma}\right)|a_{2i} - \bar{q}_i| = \left(\frac{2}{1+\Gamma}\right)\Gamma(a_{2i} - \bar{q}_i) \tag{27}$$

Comparing (24) and (25) to (26) and (27) completes the proof.

D.3.2 Illustration: Mantel-Haenszel test with binary outcomes

To illustrate the connections made within §§3.1-3.2, consider the use of the Mantel-Haenszel test statistic with binary potential outcomes, $y_{zij} \in \{0,1\}$ for $z = 0,1$. The Mantel-Haenszel statistic is $T = \sum_{i=1}^{I}\sum_{j=1}^{n_i} Z_{ij}Y_{ij}$, summing the number of events among the treated units; this corresponds to $q_{ij} = Y_{ij}$, and to $a_{i1} = 1$, $a_{i2} = 0$ in (9). Under the sharp null hypothesis, $Y_{ij} = y_{0ij}$.

For any matched set i, assuming (1) holds at Γ, the tilted statistic $T_{\Gamma i}$ is

$$\begin{aligned} T_{\Gamma i} &= \sum_{j=1}^{n_i} Z_{ij}\left\{Y_{ij} - \bar{Y}_i - \left(\frac{\Gamma-1}{1+\Gamma}\right)|Y_{ij} - \bar{Y}_i|\right\} \\ &= \left(\frac{2\Gamma}{1+\Gamma}\right)\sum_{j=1}^{n_i}\frac{Z_{ij}(Y_{ij} - \bar{Y}_i)}{\exp\{\log(\Gamma)\mathbb{1}\{Y_{ij} > \bar{Y}_i\}\}} \end{aligned}$$

where $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$.

In the conventional sensitivity analysis for the Mantel-Haenszel statistic, the worst-case expectation $\mu_{\Gamma i}$ is calculated by setting $u_{ij} = \mathbb{1}\{Y_{ij} = 1\}$, such that individuals who experienced the event have the higher probability of treatment. The test statistic minus this worst-case expectation, $T_i - \mu_{\Gamma i}$, is

$$\begin{aligned} T_i - \mu_{\Gamma i} &= \sum_{j=1}^{n_i} Z_{ij}Y_{ij} - \frac{\Gamma\sum_{j=1}^{n_i} Y_{ij}}{n_i + (\Gamma-1)\sum_{j=1}^{n_i} Y_{ij}} \\ &= \left(\frac{\Gamma n_i}{n_i + (\Gamma-1)\sum_{j=1}^{n_i}\mathbb{1}\{Y_{ij} > \bar{Y}_i\}}\right)\sum_{j=1}^{n_i}\frac{Z_{ij}(Y_{ij} - \bar{Y}_i)}{\exp\{\log(\Gamma)\mathbb{1}\{Y_{ij} > \bar{Y}_i\}\}}, \end{aligned}$$

with the second line using Proposition 2. Finally, letting $\tilde{\varrho}_{ij}$ be the values for $\mathrm{pr}(Z_{ij} =$

$1 \mid \mathcal{F}, \mathcal{Z})$ when $u_{ij} = \mathbb{1}\{Y_{ij} > \bar{Y}_i\}$, from (11) the corresponding worst-case IPW statistic would instead take the form

$$IPW_{\Gamma i} = \left(\frac{n_i + (\Gamma-1)\sum_{j=1}^{n_i}\mathbb{1}\{Y_{ij} > \bar{Y}_i\}}{n_i}\right)\sum_{j=1}^{n_i}\frac{Z_{ij}(Y_{ij}-\bar{Y}_i)}{\exp\{\log(\Gamma)\mathbb{1}\{Y_{ij} > \bar{Y}_i\}\}}.$$

From this we see that $T_{\Gamma i}$, $T_i - \mu_{\Gamma i}$, and $IPW_{\Gamma i}$ differ only in the weights they apply to the contribution $\sum_{j=1}^{n_i} Z_{ij}(Y_{ij}-\bar{Y}_i)/\exp\{\log(\Gamma)\mathbb{1}\{Y_{ij} > \bar{Y}_i\}$ from each set i. At $\Gamma = 1$ (no unmeasured confounding) the weights are equal across the three approaches, but differ at $\Gamma > 1$ when $n_i > 2$. The weights of the conventional sensitivity analysis and the IPW approach are inversely proportional, and depend upon the normalizing constant for the worst-case assignment probabilities under the sharp null, $n_i + (\Gamma-1)\sum_{j=1}^{n_i}\mathbb{1}\{Y_{ij} > \bar{Y}_i\}$. The tilted statistic does not make use of the normalizing constant.

D.4 Proof of Theorem 2

At level α, the sensitivity analysis using Theorem 1 rejects the null when assuming that (1) holds at Γ when

$$\frac{I^{-1/2}\sum_{i=1}^{I}(T_i - \bar{q}_i) - I^{-1/2}\left(\frac{\Gamma-1}{1+\Gamma}\right)\sum_{i=1}^{I}|T_i - \bar{q}_i|}{\sqrt{I^{-1}\sum_{i=1}^{I}\tilde{\nu}_{\Gamma i}^2}} \geq \Phi(1-\alpha)$$

Under Assumption 1, the denominator converges to a constant limiting value, and $I^{-1}\sum_{i=1}^{I}(T_i - \bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i - \bar{q}_i|$ (the numerator divided by $I^{1/2}$) also converges to a constant limiting value. By Assumption 2, the random variable on the left hand side will follow a Gaussian distribution when $E\left\{I^{-1/2}\sum_{i=1}^{I}(T_i - \bar{q}_i) - I^{-1/2}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i - \bar{q}_i|\right\}$ tends to a constant, in which case the power of the resulting procedure will be some value $\beta \in (0,1)$. The power of the sensitivity analysis converges to 1 when plim $I^{-1}\sum_{i=1}^{I}(T_i - \bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i - \bar{q}_i| = c > 0$ as the numerator tends to ∞ in this case, and converges to 0 when plim $I^{-1}\sum_{i=1}^{I}(T_i - \bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i - \bar{q}_i| = c < 0$ as the numerator tends to $-\infty$. The design sensitivity is the value Γ such that the difference in the probability limits exactly equals 0, i.e. the Γ that solves the equation plim $I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i-\bar{q}_i| = 0$. Solving for Γ, we attain

$$\tilde{\Gamma}_{tilt} = \operatorname*{plim}_{I\to\infty} \frac{I^{-1}\sum_{i=1}^{I}|T_i-\bar{q}_i| + I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i)}{I^{-1}\sum_{i=1}^{I}|T_i-\bar{q}_i| - I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i)} = \frac{\eta+\theta}{\eta-\theta}.$$

D.5 Proof of Proposition 3

The design sensitivities $\tilde{\Gamma}_{tilt}$ and $\tilde{\Gamma}_{conv}$ are the values Γ that solve, respectively

$$\text{plim } I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}|T_i-\bar{q}_i| = 0;$$
$$\text{plim } I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}M_{\Gamma i} = 0.$$

We prove the result for $m_{\tilde{\Gamma}_{tilt}}$; the proof for $m_{\tilde{\Gamma}_{conv}}$ is analogous. The function plim $I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\Gamma-1)/(1+\Gamma)\}\sum_{i=1}^{I}M_{\Gamma i}$ is monotone nonincreasing in Γ. Letting $\tilde{\Gamma}_{tilt}$ be the design sensitivity for the tilted statistic, we see that if plim $I^{-1}\sum_{i=1}^{I}M_{\tilde{\Gamma}_{tilt}i} - I^{-1}\sum_{i=1}^{I}|T_i-\bar{q}_i| > 0$, then plim $I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\tilde{\Gamma}_{tilt}-1)/(1+\tilde{\Gamma}_{tilt})\}\sum_{i=1}^{I}M_{\tilde{\Gamma}_{tilt}i} < 0$. As the function is monotone nondecreasing, this implies that $\tilde{\Gamma}_{conv}$, the zero of this equation, falls below $\tilde{\Gamma}_{tilt}$. Conversely, if plim $I^{-1}\sum_{i=1}^{I}M_{\tilde{\Gamma}_{tilt}i} - I^{-1}\sum_{i=1}^{I}|T_i-\bar{q}_i| < 0$, then plim $I^{-1}\sum_{i=1}^{I}(T_i-\bar{q}_i) - I^{-1}\{(\tilde{\Gamma}_{tilt}-1)/(1+\tilde{\Gamma}_{tilt})\}\sum_{i=1}^{I}M_{\tilde{\Gamma}_{tilt}i} > 0$, implying that $\tilde{\Gamma}_{conv} > \tilde{\Gamma}_{tilt}$.

D.6 Proposition 4

D.6.1 Proof of Proposition 4

Note that $f(t) = n_i^{-1}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-t|$ is convex in t, with a subderivative $f'(t) = n_i^{-1}\sum_{i=1}^{I}(\mathbb{1}\{q_{ij}-\bar{q}_i \le t\} - \mathbb{1}\{q_{ij}-\bar{q}_i > t\}) = 2\hat{G}_i(t) - 1$. The proof follows by using the representation from Proposition 1 that $M_{\Gamma i} = n_i^{-1}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-\kappa_\Gamma M_{\Gamma i}|$ with $\kappa_\Gamma := (\Gamma-1)/(1+\Gamma)$, along with bounds on a convex function using its subgradient:

$$
\begin{aligned}
M_{\Gamma i} &= \frac{1}{n_i}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-\kappa_\Gamma M_{\Gamma i}| \\
&\geq \frac{1}{n_i}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-t| + (\kappa_\Gamma M_{\Gamma i}-t)\{2\hat{G}_i(t)-1\};
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{n_i}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-t| &\geq \frac{1}{n_i}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-\kappa_\Gamma M_{\Gamma i}| + (t-\kappa_\Gamma M_{\Gamma i})\{2\hat{G}_i(\kappa_\Gamma M_{\Gamma i})-1\} \\
&= M_{\Gamma i} + (t-\kappa_\Gamma M_{\Gamma i})\{2\hat{G}_i(\kappa_\Gamma M_{\Gamma i})-1\}.
\end{aligned}
$$

Rearranging the above inequalities yields the result.

D.6.2 Proposition 4 under convex unimodality

In this section, we show for the difference in means statistic that $E\left(n_i^{-1}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i-\theta|\right) \geq E|T_i-\bar{q}_i|$ under a suitable convex unimodality assumption to be described. Without loss of generality, assume that the first individual in each matched set received the treatment. For the difference in means test statistic, we have that $\theta = E(T_i-\bar{q}_i) = \tau$. Let $\boldsymbol{\zeta} \in \mathbb{R}^{n_i}$ with $\zeta_1 = \theta$, $\zeta_j = -\theta/(n_i-1)$ for $j = 2, ..., n_i$, and let $\mathcal{E}_{ij} = Z_{ij}\varepsilon_{1ij}+(1-Z_{ij})\varepsilon_{0ij}$. Then, for the difference in means test, $q_{ij} = \tilde{q}_{ij}+\theta_j$, where $\tilde{q}_{ij} = \mathcal{E}_{ij}-\sum_{\ell\neq j}\mathcal{E}_{i\ell}/(n_i-1)$ are mean zero, permutation symmetric random variables and $n_i^{-1}\sum_{j=1}^{n_i}\tilde{q}_{ij} = 0$.

Suppose that $\tilde{\mathbf{q}}_i = (\tilde{q}_{i1}, ..., \tilde{q}_{in_i})^T$ has symmetric marginal distributions and is convex unimodal (about the origin). A multivariate distribution is convex unimodal when it has a density f such that, for every $c > 0$, the set $\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq c\}$ is a centrally convex symmetric convex set (Dharmadhikari and Jogdeo, 1976, Definition 3.3). Symmetric marginals imply that $\eta = E|\tilde{q}_{ij}+\theta| = E|\tilde{q}_{ij}-\theta|$. Furthermore, Proposition 4.4.1 of Tong (1990) shows that the density of $\tilde{q}_i$ is Schur-concave because of convex unimodality and permutation symmetry. Observe that $\boldsymbol{\zeta}-\theta$ majorizes the constant vector $(-\theta, ..., -\theta)^T \in \mathbb{R}^{n_i}$ and that $g(\mathbf{x}) = -n_i^{-1}\sum_{j=1}^{n_i}|x_j|$ is a Schur-concave function. By the convolution theorem for Schur-concave functions (Marshall et al., 2011, Theorem J.1), we have that $E(-n_i^{-1}\sum_{j=1}^{n_i}|\tilde{q}_{ij}-b_j|)$ is a Schur-concave function of $\mathbf{b}=(b_1,...,b_{n_i})^T$, and hence $E(n_i^{-1}\sum_{j=1}^{n_i}|\tilde{q}_{ij}-b_j|)$ is Schur-convex. Consequently,

$$\begin{aligned} E\left(n_i^{-1}\sum_{j=1}^{n_i}|q_{ij}-\theta|\right) &= E\left(n_i^{-1}\sum_{j=1}^{n_i}|\tilde{q}_{ij}-\theta+\zeta_j|\right) \\ &\geq E\left(n_i^{-1}\sum_{j=1}^{n_i}|\tilde{q}_{ij}-\theta|\right) \\ &= E|\tilde{q}_{ij}-\theta| = E|\tilde{q}_{ij}+\theta| = \eta. \end{aligned}$$

D.7 Proof of Proposition 5

By Dinkelbach (1967), we have for any c that $M_{\Gamma i}\geq c$ if and only if $H_{\Gamma i}(c)\geq 0$, with $H_{\Gamma i}(c)$ defined as in (21). Using the equivalent form (22), we have

$$\mathrm{pr}(M_{\Gamma i}\geq c) = \mathrm{pr}\left\{n_i^{-1}\sum_{j=1}^{n_i}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right|\geq c\right\}.$$

Let $\mathcal{C}_c=\{\mathbf{x}: n_i^{-1}\sum_{j=1}^{n_i}|x_j|\leq c\}$, and observe that for any fixed c this is a symmetric convex set. As $\mathbf{q}_i-\bar{q}_i=(q_{i1}-\bar{q}_i,...,q_{in_i}-\bar{q}_i)^T$ is convex unimodal, by Anderson (1955) we have

$$\begin{aligned} \mathrm{pr}\left\{n_i^{-1}\sum_{j=1}^{n_i}\left|q_{ij}-\bar{q}_i-\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right|\geq c\right\} &= 1-\mathrm{pr}\left\{\mathbf{q}_i-\bar{q}_i\in\mathcal{C}_c+\left(\frac{\Gamma-1}{1+\Gamma}\right)c\right\} \\ &\geq 1-\mathrm{pr}\left\{\mathbf{q}_i-\bar{q}_i\in\mathcal{C}_c\right\} \\ &= \mathrm{pr}\left\{n_i^{-1}\sum_{j=1}^{n_i}|q_{ij}-\bar{q}_i|\geq c\right\} \end{aligned}$$

as desired.

References

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176.

Brumberg, K., Small, D. S., and Rosenbaum, P. R. (2024). Optimal refinement of strata to balance covariates. *Biometrics*, 80(3):ujae061.

Chen, Z., Cho, M. H., and Zhang, B. (2024). Manipulating a continuous instrumental variable in an observational study of premature babies: Algorithm, partial identification bounds, and inference under randomization and biased randomization assumptions. *arXiv preprint arXiv:2404.17734*.

Cohen, P. L., Olson, M. A., and Fogarty, C. B. (2020). Multivariate one-sided testing in matched observational studies as an adversarial game. *Biometrika*, 10(4):809–825.

Dharmadhikari, S. and Jogdeo, K. (1976). Multivariate unimodality. *The Annals of Statistics*, pages 607–613.

Dinkelbach, W. (1967). On nonlinear fractional programming. *Management Science*, 13(7):492–498.

Dorn, J. and Guo, K. (2022). Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, pages 1–13.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474.

Fogarty, C. B. (2023). Testing weak nulls in matched observational studies. *Biometrics*, 79(3):2196–2207.

Fogarty, C. B. and Hasegawa, R. B. (2019). Extended sensitivity analysis for heterogeneous unmeasured confounding with an application to sibling studies of returns to education. *The Annals of Applied Statistics*, 13(2):767–796.

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2000). Asymptotic separability in sensitivity analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(3):545–555.

Hájek, J., Šidák, Z., and Sen, P. K. (1999). *Theory of Rank Tests*. Academic press.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618.

Heng, S., Kang, H., Small, D. S., and Fogarty, C. B. (2021). Increasing power for observational studies of aberrant response: An adaptive approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(3):482–504.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34(2):598–611.

Huber, P. J. (1981). *Robust Statistics*. Springer, New York.

Jick, H., Miettinen, O. S., Neff, R. K., Shapiro, S., Heinonen, O. P., and Slone, D. (1973). Coffee and myocardial infarction. *New England Journal of Medicine*, 289(2):63–67.

Kang, H., Kreuels, B., May, J., and Small, D. S. (2016). Full matching approach to instrumental variables estimation with application to the effect of malaria on stunting. *The Annals of Applied Statistics*, 10(1):335–364.

Maritz, J. (1979). A note on exact robust confidence intervals for location. *Biometrika*, 66(1):163–170.

Marshall, A., Olkin, I., and Arnold, B. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer.

Pimentel, S. D., Kelz, R. R., Silber, J. H., and Rosenbaum, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *Journal of the American Statistical Association*, 110(510):515–527.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74(1):13–26.

Rosenbaum, P. R. (1988). Sensitivity analysis for matching with multiple controls. *Biometrika*, 75(3):577–581.

Rosenbaum, P. R. (1991). A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):597–610.

Rosenbaum, P. R. (2002a). Attributing effects to treatment in matched observational studies. *Journal of the American Statistical Association*, 97(457):183–192.

Rosenbaum, P. R. (2002b). *Observational Studies*. Springer, New York.

Rosenbaum, P. R. (2004). Design sensitivity in observational studies. *Biometrika*, 91(1):153–164.

Rosenbaum, P. R. (2007). Sensitivity analysis for M-estimates, tests, and confidence intervals in matched observational studies. *Biometrics*, 63(2):456–464.

Rosenbaum, P. R. (2013). Impact of multiple matched controls on design sensitivity in observational studies. *Biometrics*, 69(1):118–127.

Rosenbaum, P. R. (2014). Weighted M-statistics with superior design sensitivity in matched observational studies with multiple controls. *Journal of the American Statistical Association*, 109(507):1145–1158.

Rosenbaum, P. R. (2016). Using Scheffé projections for multiple outcomes in an observational study of smoking and periodontal disease. *The Annals of Applied Statistics*, 10(3):1447–1471.

Rosenbaum, P. R. (2023a). A second evidence factor for a second control group. *Biometrics*, 79(4):3968–3980.

Rosenbaum, P. R. (2023b). Sensitivity analyses informed by tests for bias in observational studies. *Biometrics*, 79(1):475–487.

Rosenbaum, P. R. (2024). Bahadur efficiency of observational block designs. *Journal of the American Statistical Association*, 119(547):1871–1881.

Rosenbaum, P. R. and Krieger, A. M. (1990). Sensitivity of two-sample permutation inferences in observational studies. *Journal of the American Statistical Association*, 85(410):493–498.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Small, D., Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (2009). Simultaneous sensitivity analysis for observational studies using full matching or matching with multiple controls. *Statistics and its Interface*, 2(2):203–211.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637.

Tong, Y. L. (1990). *The multivariate normal distribution*. Springer Science & Business Media.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press, Cambridge.

Wu, D. and Li, X. (2025). Sensitivity analysis for quantiles of hidden biases in matched observational studies. *Journal of the American Statistical Association*, pages to appear, DOI 10.1080/01621459.2024.2441527.

Zhang, J., Small, D. S., and Heng, S. (2024). Sensitivity analysis for matched observational studies with continuous exposures and binary outcomes. *Biometrika*, 111(4):1349–1368.

Zhao, Q. (2019). On sensitivity value of pair-matched observational studies. *Journal of the American Statistical Association*, 114(526):713–722.

Zhao, Q., Small, D. S., and Bhattacharya, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(4):735–761.

Zubizarreta, J. R. (2012). Using mixed integer programming for matching in an observational study of kidney failure after surgery. *Journal of the American Statistical Association*, 107(500):1360–1371.