Estimating Causal Effects Using Coarsened Treatments as Instruments*

John Henderson†

Assistant Professor
Dept. of Political Science
Yale University

November 2, 2015

Abstract

Researchers often estimate causal effects in experimental or observational studies after coarsening continuous measures of treatments. In the statistical matching context, in particular, non-discrete interventions are frequently discretized to facilitate pair-stratification using traditional matching approaches for binary treatments. A well-known issue in studying coarsened interventions is that any coarsening induces measurement error that attenuates estimates, while inflating estimator standard errors. While this bias is known, there is yet no standard correction for it. This research note illustrates the *error-in-variables* structure underlying the use of discrete transformations of non-discrete (or dose) interventions. It also recommends the use of the standard IV estimator to recover an unbiased estimate of the uncoarsened treatment effect. Particular attention will be given to the problem of matching with a continuous intervention, which motivates simulations.

Key words: experiments, matching, continuous interventions, discrete data, instrumental variables, measurement error.

*All errors are the author's responsibility.

†john.henderson@yale.edu, http://jahenderson.com/, Institution for Social and Policy Studies, Yale University, 77 Prospect Street, New Haven, CT 06520-8209

Motivation

Researchers routinely discretize continuous interventions to facilitate experimental or observational analysis. This transformation is inherent in experiments where researchers decide what marginal doses will be assigned to subjects as dichotomous treatment or control values. In observational studies, continuous features of subjects are frequently captured by discrete categories to facilitate measurement, for example, to elicit honest responses to questions people might be reticent to answer accurately (e.g., their income). Moreover, dichotomization is often used by researchers to simplify the interpretation of their findings (e.g., Gelman and Park 2008), or to allow non-parametric estimation of treatment effects through matching with continuous interventions (e.g., Henderson 2015, Lu et al. 2011).

In spite of widespread use, the shortcomings to discretizing variables have been well-documented (Cox 1957, Feldt 1961, Morgan and Elashoff 1986). Coarsening continuous interventions reduces estimator precision and statistical power of hypotheses tests, biases the interpretation of marginal effects, and often relies on arbitrary cutoffs that can further bias results (Harrell 2015). Researchers are advised to avoid dichotomizing variables. Yet, in many instances discretization is unavoidable, and can be inferentially desirable. For example, scholars may be able to recruit only enough subjects to compare the average outcomes across two or three levels of treatment, but in the ideal would study marginal effects over a much broader range. Also, a recent development in observational research has been devising statistical matching approaches to study continuous interventions (Lu et al. 2011, 2001). Across these matching methods, continuous treatments are necessarily discretized prior to analysis. This discretization fundamentally alters the quantity being estimated, and raises concerns about the efficiency of such matching estimators. Yet, this choice may be necessary to obtain balance on covariates after matching.

In this research note, I extend the *error-in-variables* instrument variables (IV) framework to the problem of discretizing continuous interventions. In particular, I show that

a Wald IV estimator corrects both the bias and imprecision associated with coarsening continuous treatments. The key insight is that discretizing an intervention induces measurement error that can be rescaled back to the marginal dose effect by accounting for the shift in doses across the discrete treatment categories. Importantly, this result reduces some of the concern with discretizing interventions, and especially the arbitrary reductions in information that typically bias estimates. Throughout, I frame this finding in the context of matching using a continuous intervention, and present simulation results to illustrate the method.

Discretizing Doses as an *Error-in-Variables* Problem

The main concern with coarsening interventions is that doing so introduces error, which can attenuate the apparent association between the uncoarsened measurements. Correcting for this measurement error is well-understood in the context of the *error-in-variables* framework, motivating the original development of instrumental variables (IV). Here I extend and clarify this framework in the context of measurement error induced by coarsening a treatment measure prior to estimating its effects on y. Though this extension is natural, interestingly it has yet to be explored by scholars.

Define a coarsening to be some function $z' = c(z, \rho)$. A coarsening is any (weakly) monotonic transformation that utilizes threshold points ρ to rescale a measure to be 'less continuous.' Formally, coarseness is defined as, for continuous z, $\lim_{z \to \rho + \epsilon} c(z) - c(\rho + \epsilon) < 0$, as $\epsilon \leftarrow 0$, such that there is always a discrete jump in $c(z)$ as z crosses ρ. The coarsening operator is straightforward. All $z \in (\rho_0, \rho_1) \equiv z' = \rho_0 + k$. In words, all values of z between some low (ρ_0) and high (ρ_1) threshold, take on the same value $z' = \rho_0 + k$, where k is some rescaling constant applied to all coarsened units.

Weak monotonicity is clearly retained through coarsening so long as $\rho_1 > \rho_0$, k is constant, and τ is wider than the interval defined by (ρ_0, ρ_1). This latter condition precludes a special case where $z'_i = \rho_0$ for all i, so that z' is at least a dichotomous

variable. This ensures $E[zz'] \neq 0$, and thus that z' contains some information about z. From this, we can observe that informativeness of z' is a function of the distribution of ρ. Consider a vector of ρ thresholds that contains many points (i.e., $|\rho|$ is large). The informativeness of z' increases as $(\tau_1 - \tau_0)/|\rho| \to 0$, where τ_1 and τ_0 define the upper and lower bound of the treatment domain τ. As $|\rho|$ increases on a fixed domain τ, the distances between any two thresholds must shrink to zero (such that ρ approaches continuity). Consequently, $|\rho| \to \infty \implies Pr\{z \in (\rho_0, \rho_1)\} \to 0$. With a near-continuous vector of thresholds ρ, the probability that any two values of z will be coarsened to be identical is zero, so that $z = c(z)$. Information can be summarized by a R^2 statistic, which is maximized when $R^2(z, z') = 1$. It can be shown that the least informative case is when z' is dichotomous. Further, any dichotomous coarsening loses information as $var(z')$ is minimized. It is this loss in information that is the main concern when studying variables measured with error.

Unbiasedness of the Wald Estimator

Under random assignment of z, the coarsened instrument z' is independent of potential *dose-response* outcomes $y_i = Y_i(z)$, denoted here as a real value function of any assignment of z. Following the above conditions, any sufficiently informative coarsening can be used in an IV framework to recover the (linear) effect of z on y. This linear effect is estimated through two-stage-least-squares (2SLS):

$$\begin{aligned} \beta_{IV} &= E[y|z] \\ &= \frac{E[y|z']}{E[z|z']} \\ &= \frac{Cov(y, z')}{Cov(z, z')}. \end{aligned} \tag{1}$$

This estimator requires that the coarsened variable be associated with y only through its association on z. This is the exclusion restriction that no other pathway exists between z'

and y besides that through z. Since any coarsening only depends on ρ thresholds and z, this restriction is assured, unless ρ depends on y. In standard empirical approaches, such dependency should not emerge. The stable unit treatment value (SUTVA) assumption also holds, since again the value for z_i' can only depend on z_i and ρ. As discussed above, monotonicity also holds assuming strictly monotonic thresholds in ρ. Finally, for estimates of Equation 1 to be consistent, the denominator $Cov(z, z')$ must be sufficiently greater than zero. There is a well-known issue in IV analyses with weak instruments, in which low covariance in the denominator can inflate estimates. Researchers should aim to construct a coarsened instrument with enough information about treatment so that the expected bias, $E[\beta - \hat{\beta}_V]$, is small.

A similar finding holds when units are randomized conditional on covariates X, as assumed in observational studies. The conditional form of Equation 1 is

$$\begin{aligned} \beta_{IV|X} &= E[y|z, X] \\ &= \frac{E[y|z', X]}{E[z|z', X]} \\ &= \frac{Cov(y, z'|X)}{Cov(z, z'|X)}, \end{aligned} \tag{2}$$

which can be estimated using 2SLS. If discretizing ρ are independent of X, being derived from some exogenous process, then the above conditional expectation yields an unbiased estimate of $\beta_{IV|X}$. This is clear since $z'|X = c(z|X, \rho|X) \in \{\rho_0|X - k|X, \rho_1|X - k|X\}$, so that $z'|X \in \{\rho_0 - k, \rho_1 - k\}$, whenever $\rho_0, \rho_1, k \perp\!\!\!\perp X$. Yet in some cases, (e.g., non-bipartite matching on continuous treatments), the particular thresholds that discretize treatments may depend on variation in X (Henderson 2015, Lu et al. 2011). It is necessary then to show that conditioning on X is sufficient to account for this dependency during estimation.

The argument requires that no excluded factors U influence both $\rho(z, X)$ and y after controlling for X. This can be illustrated for the case of matching on X. For a matched

pair, s_{ij}, in which $X_i = X_j$, treatment is dichotomized by construction, so that there is a *high* dose z_{is}^+ and a *low* dose z_{js}^- in each s stratification. Within s there is at least one common threshold $\rho_s \in (z_{js}^-, z_{is}^+)$ that separates z_{is}^+ and z_{js}^- doses. (In fact, a range of thresholds lies between z_{is}^+ and z_{js}^-.) It is sufficient to show that if either if $U \perp\!\!\!\perp y|X$ or $U \perp\!\!\!\perp z|X$, then the estimator in Equation 2 is unbiased. Consider any U that influences y but not z, when units are matched on X. Matching with continuous interventions typically pairs units based on their multivariate distances on some function

$$md(X, Z) \;=\; \left\{ \frac{(X_i - X_j)^2}{(z_i - z_j)^2} \right\}^{1/2}. \tag{3}$$

Clearly distances do not depend on U. Thus, both the matched pairs and ρ thresholds remain unchanged for any distribution of U orthogonal to treatment.[1] As a result, after matching, the expected marginal dose effect can be estimated as

$$\begin{aligned} \beta_{IV|X} &= E[y|z, X] \\ &= \frac{E[y|z', X]}{E[z|z', X]} \\ &= \frac{E[y|z_{is}^+, X] - E[y|z_{js}^-, X]}{E[z|z_{is}^+, X] - E[z|z_{js}^-, X]}. \end{aligned} \tag{4}$$

Next, if U influences z but not y, this can impact the matches using the above distance function, though will not confound causal estimates. For units $i \neq j \neq l$, take some binary U where $X_i = X_j = X_l$, but $U_i = U_l = 1$ and $U_j = 0$. If U influences treatment by ν (e.g., $z = f(X) + \nu U$), then

$$\begin{aligned} E[z + \nu U|X] &\equiv E[z_i + \nu|X] = E[z_l + \nu|X] = E[z_j|X] \\ &\equiv E[z_i|X] = E[z_l|X] = E[z_j|X] - \nu. \end{aligned}$$

[1] The more general result is $z \perp\!\!\!\perp U \implies E[Y_i(z|U, z, X)] = E[Y_i(z|X)] = E[y|z, X]$, while $y \perp\!\!\!\perp U \implies E[Y_i(z)|z, X, U] = E[Y_i(z|U)|z, X] = E[Y_i(z + \nu U)|z, X]$. Redefining $\tilde{z} = z + \nu U$ then $E[Y_i(\tilde{z})|z, X]/E[z(\tilde{z})|z, X] \implies E[Y_i(\tilde{z}|z)|X]/E[z(\tilde{z}|z)|X] = E[Y_i(\tilde{z})|X]/E[z(\tilde{z})|X] = E[y|\tilde{z}, X]/E[z|\tilde{z}, X]$.

Consider two set of matches, one where all i are matched to j (*imbalanced*), and another where all i are matched to l (*balanced*). Under *balanced* in all s_{il}, there will be some $\rho_{il} \in (z_{ls}^-, z_{is}^+)$, that under *imbalanced* will be $\rho_{ij} \in (z_{ls}^- - \nu, z_{is}^+)$. Since $U \perp\!\!\!\perp y|X$ this implies $Y_i(z_i - \nu|X) = Y_i(z_i|X) - Y_i(\nu|X)$. Thus $E[y|z_{is}^+, X] - E[y|z_{ls}^-, X] \equiv E[y|z_{is}^+, X] - E[y|z_{js}^-, X] + E[y|\nu]$. Recall in the denominator, $E[z|z_{is}^+, X] - E[z|z_{ls}^-, X] \equiv E[z|z_{is}^+, X] - E[z|z_{js}^-, X] + E[\nu]$, so that

$$\begin{aligned}
\beta_{IV|X} &= \frac{E[y|z_{is}^+, X] - E[y|z_{ls}^-, X]}{E[z|z_{is}^+, X] - E[z|z_{ls}^-, X]} \\
&= \frac{E[y|z_{is}^+, X] - E[y|z_{js}^-, X] + E[y|\nu]}{E[z|z_{is}^+, X] - E[z|z_{js}^-, X] + E[\nu]} \\
&= \frac{E[y|z_{is}^+, X] - E[y|z_{js}^-, X] + \Delta y/\Delta\nu}{E[z|z_{is}^+, X] - E[z|z_{js}^-, X] + \Delta z/\Delta\nu}.
\end{aligned}$$

Clearly this estimator is unbiased since it is the marginal dose effect given change in treatment $dz = z + \nu U$. However, any U influencing z, but orthogonal to the potential outcomes, *can* affect the mean difference estimator in the numerator. We see that $E[y|z_{is}^+, X] - E[y|z_{js}^-, X] \neq E[y|z_{is}^+, X] - E[y|z_{ls}^-, X]$, which differ by $E[y|\nu]$, a term that wholly depends on the association between treatment and an unobservable. Without correcting for this shift, matching on continuous treatments will yield biased inferences, and this bias can depend on unobserved factors that should be theoretically ignorable.

In sum, the resulting linear estimator of the marginal dose effect is independent of any unobserved factor orthogonal to the potential outcomes or to treatment, conditional on covariates. This finding relies on a rescaling in the denominator that accounts for the shift in y emerging due to 'imbalance' on U that can influence how units are matched. Though seemingly obvious, this is an important finding since it ensures that matching on X is sufficient to recover $\beta_{IV|X}$ through *any* discretization operation (independent of y) under the conditional exchangeability assumption.

Estimating the Matching Variance of $\beta_{IV|X}$

Statistical inference for $\hat{\beta}_{IV|X}$ proceeds in the standard way. Under the classical IV assumptions, the Wald estimator in Equation 2 is asymptotically normally distributed, so that $\sqrt{N}(\beta_{IV|X} - \beta) \xrightarrow{d} N(0, Avar(\hat{\beta}_{IV|X}))$, where $Avar(\hat{\beta}_{IV|X})$ is the asymptotic variance of $\hat{\beta}_{IV|X}$. This finding yields a standard t-test for hypothesis testing, with

$$Avar(\hat{\beta}_{IV|X}) \quad = \quad \frac{\hat{\sigma}^2}{R^2_{z,z'}}(\mathbb{X}'\mathbb{X})^{-1} \tag{5}$$

for regression variance $\hat{\sigma}^2 = \hat{u}'\hat{u}/N$, and the first-stage $R^2_{z,z'}$ statistic. Here $\mathbb{X}$ is the block matrix $\left[\vec{1}_{n\times 1} \;\; X_{n\times k} \;\; z_{n\times 1}\right]$. The loss in precision associated with the IV estimator with a discretized instrument is $1/R^2_{z,z'} = \sum_i (z - \bar{z})^2 / \sum_i (\hat{z} - \bar{z})^2$, or the amount of information loss from z to z'. This test variance $Avar(\hat{\beta}_{IV|X})$ is computed analytically from the above. A straightforward (sampling-based) approach for matched data is to first stratify the sample into pairs, and then estimate the variance over the stratifications after matching. Importantly, this approach will approximate the additional inefficiency associated with matched estimators (Abadie and Imbens 2006).[2]

Simulation Results

In this section, I present a series of simulations to illustrate the bias correction of the Wald estimator, with and without matching. I generate data using six models of y and z, drawing $k = 1000$ times from each model space.[3] In Model 1, z is fully randomized, while in Models 2 - 6, treatment is confounded with X through increasing complexity.

[2] In exploring the asymptotic properties of matching, Abadie and Imbens (2006) find the estimators generally are not root-N consistent. The authors develop an estimator of asymptotic variance that provides proper coverage accounting for the additional inefficiency in matching. In practice, this variance estimator is similar to the recommended approach above. Alternatively, Rosenbaum (2002) recommends a permutation inference approach that provides correct coverage under the sharp null hypothesis. The loss is that these permutation standard errors are more complex and costly to compute. See the Online Appendix for more details.

[3] See the Online Appendix for details on the sampling space of the models.

Throughout, z is dichotomized by ρ to produce z', where ρ is a random variable independent of (Models 1 - 2) or conditional on X (Models 3 - 6). For each k draw, I sample $N = \{100, 500, 1000\}$ units from each model. I then estimate the following IV system of equations *before matching* using 2SLS:

$$\begin{aligned} z &= \alpha + \eta z' + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \\ y &= \alpha + \beta_0 z + \psi_1 X_1 + \psi_2 X_2 + \psi_3 X_3 + \psi_4 X_4 + \nu. \end{aligned}$$

After matching, I run the reduced form model:

$$\begin{aligned} z &= \alpha + \eta z' + \epsilon \\ y &= \alpha + \beta_0 z + \nu. \end{aligned}$$

Here I use nearest-neighbor matching on multivariate Mahalanobis distances to estimate an average treatment effect (ATE) for the conditional linear dose response.

Table 1: Mean Square Error Simulation Results

Parameters	Estimation	M1	M2	M3	M4	M5	M6
N = 100							
$0 < \mu_{z'} < 1$	2SLS	0.062	37.157	10.213	4.937	1.130	1.910
	Matching	0.166	0.207	0.213	0.094	0.154	0.060
$.3 < \mu_{z'} < .7$	2SLS	0.028	4.092	0.932	0.312	0.350	0.026
	Matching	0.057	0.153	0.175	0.061	0.145	0.050
N = 500							
$0 < \mu_{z'} < 1$	2SLS	0.014	0.151	0.107	1.406	0.075	0.075
	Matching	0.032	0.133	0.118	0.041	0.099	0.037
$.3 < \mu_{z'} < .7$	2SLS	0.007	0.061	0.036	0.013	0.015	0.133
	Matching	0.011	0.108	0.078	0.035	0.059	0.018
N = 1000							
$0 < \mu_{z'} < 1$	2SLS	0.008	0.249	112.470	2.365	0.023	0.019
	Matching	0.013	0.100	0.103	0.038	0.071	0.022
$.3 < \mu_{z'} < .7$	2SLS	0.004	0.031	0.017	0.003	0.011	0.009
	Matching	0.006	0.072	0.078	0.027	0.053	0.018

Results for the simulations are presented in Table 1. The table shows the Mean Squared Error (MSE) for 2SLS and Matching estimates drawn from the simulated data processes at various specifications. Notably, across the simulations, the Wald estimator generally provides the appropriate correction for the measurement error introduced through coarsening. However, this does depend on the amount of information in z', and to a lesser extent the size of the sample N. When the mean of z' is allowed to range close to 0 or 1, the unmatched 2SLS estimates tend to inflate for certain model specifications, and especially under simulation Models 2, 3 and 4. In contrast, perhaps unexpectedly, the Matching ATE estimator actually appears to be relatively robust to the amount of information in the coarsening instrument. Indeed, the MSE for Matching never reaches above 0.213, and generally produces unbiased estimates of the true underlying effect β_0 regardless of the information or model complexity. The is likely due to matching in 'both directions,' that is from controls to treated *and* treated to controls. Doing so appears to provide additional estimator stability when either the effect of treatment for the treated or for the controls is weakly informed by z'. Ensuring sufficient information in z', by stratifying simulations so that $.3 < \mu_z < .7$, produces much better results across both methods. Naturally, increasing the sample size also produces lower MSE.

Conclusion

In this note, I extend the *error-in-variables* framework to the problem of discretizing continuous interventions, and provide a correction for attenuation error associated with this discretization. Quantitative scholars frequently coarsen non-discrete measures to facilitate empirical analysis. In doing so, researchers alter the interpretation of their findings by restricting consideration to subsets of treatment, generally biasing estimates of the overall marginal dose effect. I show here that researchers can both discretize treatments *and* use these findings to assess broader marginal effect by rescaling their estimates appropriately.

This rescaling result can reduce many of the concerns associated with discretization, and especially that the use of arbitrary reductions in information pose unnecessary restrictions on identifying and interpreting causal estimates. Assuming thresholds are independent of treatment or outcomes, otherwise arbitrary cutoff are not a concern, so long as the discretized instrument is sufficiently informative. Simulation findings go further to show that even with limited information, matching can produce stable and unbiased estimates, though unmatched estimators are sensitive. Counter to common wisdom, this note should give scholars greater confidence in dichotomizing treatments to recover unbiased causal estimates of marginal dose effects, particularly when using statistical matching estimators.

A Online Appendix

A.1 Permutation Inference for Matching IV Standard Errors

Permutation inference can be used to identify confidence intervals and standard errors for matching IV estimates. The approach utilizes Hodges-Lehmann (HL) point estimates, which has the property of yielding confidence intervals proportional to the first-stage association between the dichotomized instrument z' and the continuous treatment z. Consequently, if the instrument is weak, HL standard errors will be very large, indicating confidence bounds of near-infinite length.

For a binary intervention, z', define $d_i^T = \mathbf{z}_i(z'_i = 1)$ and $d_i^C = \mathbf{z}_i(z'_i = 0)$ to be the change in the continuous intervention over binary levels. Similarly, define $y_i^T = \mathbf{y}_i(z'_i = 1, \mathbf{z}_i)$ and $y_i^C = \mathbf{y}_i(z'_i = 0, \mathbf{z}_i)$ for the outcome. The exogeneity assumption states $z' \perp \mathbf{y}_i, \mathbf{z}_i$, which is assured if thresholds for z' are independent of (fully randomized) treatment and the outcome. We can identify two unbiased causal quantities, $\tau^y = E\{\mathbf{y}(1, \mathbf{z}) - \mathbf{y}(0, \mathbf{z})\}$ and $\tau^z = E\{\mathbf{z}(1) - \mathbf{z}(0)\}$, the ratio of which is the classical IV estimator. A causal quantity may be recovered by conditioning on X, so that $\{z' \perp \mathbf{y}_i, \mathbf{z}_i\}|X$ holds. (This conditionality is always assumed, but left implicit in the notation.)

From these primitives, Rosenbaum (2002) develops a non-parametric permutation inference approach to identify correct intervals for instruments after matching. Assume an additive model of treatment effects: $y_i^T - y_i^C = \beta(z_i^T - z_i^C) = y_i^T - \beta z_i^T = y_i^C - \beta z_i^C = \mathbb{Y} - \beta\mathbb{Z}$. Note that $\mathbb{Y} - \beta\mathbb{Z}$ is constant with respect to z', allowing for a sharp permutation test. Assuming (conditional) exogeneity, the instrumental effect β is identified as:

$$\begin{aligned} t &= SOLVE\{\hat{t} = t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)\} \\ &= \left(\inf\{\hat{t} > t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)\} + \sup\{\hat{t} < t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)\}\right)/2, \end{aligned} \tag{6}$$

where $t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)$ is a rank test statistic (e.g., Wilcoxon sign rank). For matched estimators, β is solved for $t = m(\frac{n}{2} + 1)/4$, where m is number of treated, n is number

of units, and t is a target rank statistic under the null of no differences in the ranked outcomes for treated and control units. (Unmatched estimators use the solution at $t = m(n+1)/2$.)

The HL estimator is usually solved numerically. If $m = 556$, and $n = 1294$, then the HL estimate is found by solving Equation (6) setting and $\tilde{t} = 90072$, using the analytic permutation solutions uncovered in Rosenbaum (2002). If the data is matched, then $t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)$ test statistic is Wilcoxon sign rank

$$t(z', \mathbf{y} - \hat{\beta}\mathbf{z}) \;=\; \sum_{s}^{S} \texttt{rank}\left\{\left|(\mathbf{y}_{s1} - \hat{\beta}\mathbf{z}_{s1}) - (\mathbf{y}_{s2} - \hat{\beta}\mathbf{z}_{s2})\right|\right\} \sum_{i}^{2} z'_{si} c_{si}$$

where `rank` is a ranking function, and

$$c_{s1} = \begin{cases} 1 & \text{if } (\mathbf{y}_{s1} - \hat{\beta}\mathbf{z}_{s1}) > (\mathbf{y}_{s2} - \hat{\beta}\mathbf{z}_{s2}) \\ 0 & \text{otherwise,} \end{cases} \quad c_{s2} = \begin{cases} 1 & \text{if } (\mathbf{y}_{s1} - \hat{\beta}\mathbf{z}_{s1}) < (\mathbf{y}_{s2} - \hat{\beta}\mathbf{z}_{s2}) \\ 0 & \text{otherwise.} \end{cases}$$

A solution is obtained by incrementing (decrementing) a proposed $\hat{\beta}$ over some range in a positive (negative) direction, until the conditions in Equation (6) are satisfied.

The algorithm to estimate $\hat{\beta}_{HL}$ is as follows:

1. Set $\hat{\beta}_0 = 0$; set target $\tilde{t} = m(\frac{n}{2} + 1)/4$
2. Calculate $t_0 = t(z', \mathbf{y}_i - \hat{\beta}_0\mathbf{z}_i)$ and $t_\epsilon = t(z', \mathbf{y}_i - (\hat{\beta}_\epsilon)\mathbf{z}_i)$, where $\hat{\beta}_\epsilon = \hat{\beta}_0 + \epsilon$
3. Stop if $(t_0 + t_\epsilon)/2 \approx \tilde{t}$ and $t_0 < \tilde{t}$ and $t_\epsilon > \tilde{t}$
4. Otherwise increment, setting new $\hat{\beta}_0 = \hat{\beta}_\epsilon$
5. Repeat until Step 3. is satisfied, setting $\hat{\beta}_{HL} = \hat{\beta}_\epsilon$

Confidence bounds for a 95% interval can be estimated using the same approach through permutation inference. A permutation inference considers a distribution $\mathbb{T}$ of test statistics $t(z', \mathbf{y}_i - \hat{\beta}\mathbf{z}_i)$, produced from all (or a large sample of) permutations of

possible intervention assignments. This distribution can be inverted to identify correct confidence bounds for some β. To do so for $H_0 : \beta = \beta_0$, estimate the above Hodges-Lehmann (HL) point statistic. Next, randomly permute treatment assignments in z', and collect the resulting distribution $\mathbb{T}(\beta_0) = t(z', \mathbf{y}_i - \beta_0 \mathbf{z}_i)$. Then set t^α to be the $(1-\alpha)$ quantile of $\mathbb{T}$, and again solve for t^α. This produces the confidence interval under the testing null β_0. These intervals can be used to find standard error estimates, and p-values for the HL estimator.

A.2 Simulation Data Generating Models

Model 1:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho, \sigma_\rho^2) \end{aligned}$$

Model 2:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho, \sigma_\rho^2) \end{aligned}$$

Model 3:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4, \sigma_\rho^2) \end{aligned}$$

Model 4:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3^2 + \gamma_4 X_4^2, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3^2 + \gamma_4 X_4^2, \sigma_\rho^2) \end{aligned}$$

Model 5:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_3 X_4, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_3 X_4, \sigma_\rho^2) \end{aligned}$$

Model 6:

$$\begin{aligned} y &\sim N(\alpha_y + \beta_0 z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma_y^2) \\ z &\sim N(\alpha_z + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3^2 + \gamma_4 X_4^2 + \gamma_5 X_3^2 X_4^2, \sigma_z^2) \\ \rho &\sim N(\alpha_\rho + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3^2 + \gamma_4 X_4^2 + \gamma_5 X_3^2 X_4^2, \sigma_\rho^2) \end{aligned}$$

$$\begin{aligned} X_k &\sim N(\alpha_{X_k}, \sigma_{X_k}^2) \\ \alpha, \beta, \gamma, \gamma &\sim N(0, 1) \\ \sigma &\sim \text{Inv-Gamma}(1.6, 1) \end{aligned}$$

References

Abadie, Alberto, and Guido Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74: 235–267.

Cox, David R. 1957. "Note on Grouping." *Journal of the American Statistical Association* 52 (3): 543–547.

Feldt, Leonard S. 1961. "The Use of Extreme Groups to Test for the Presence of a Relationship." *Psychometrika* 26 (1): 307–316.

Gelman, Andrew, and David K. Park. 2008. "Splitting a Predictor at the Upper Quarter or Third and the Lower Quarter or Third." *The American Statistician* 62 (4).

Harrell, Frank. 2015. "Problems Caused by Categorizing Continuous Variables." Available online at http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/CatContinuous.

Henderson, John A. 2015. "A Genetic Matching Approach to Estimating Treatment Effects Using Non-Binary Interventions." Working Paper.

Lu, Bo, Elaine Zutto, Robert Hornik, and Paul R. Rosenbaum. 2001. "Matching With Doses in an Observational Study of a Media Campaign Against Drug Abuse." *Journal of the American Statistical Association* 96 (456): 1245–1253.

Lu, Bo, Robert Greevy, Xinyi Xu, and Cole Beck. 2011. "Optimal Nonbipartite Matching and Its Statistical Applications." *The American Statistician* 65 (1): 21–30.

Morgan, Timothy M., and Robert M. Elashoff. 1986. "Effect of Categorizing a Continuous Covariate on the Comparison of Survival Time." *Journal of the American Statistical Association* 81 (4): 917–921.

Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer-Verlag.