Supplementary Materials for "A Regression-with-Residuals Method for Estimating Controlled Direct Effects"

A: Equivalence between RWR and Sequential G-Estimation Under No Intermediate Interactions

To see the equivalence between RWR and sequential g-estimation, let us consider model (1) in the main text and write the "naive" least squares regression of it as

$$Y = \hat{\beta}_0 + \hat{\beta}_1^T X + \hat{\beta}_2 A + \hat{\beta}_3^T Z + M(\hat{\gamma}_0 + \hat{\gamma}_1^T X + \hat{\gamma}_2 A) + Y_\perp, \tag{1}$$

where $Y_\perp$ denotes the residual. Suppose X is a column vector of p pretreatment confounders and Z is a column vector of q intermediate confounders. For each of the components in Z, it has a least squares fit on X and A. These least squares fits can be combined in matrix form:

$$Z = \hat{\lambda}_0 + \hat{\Lambda}_1 X + \hat{\lambda}_2 A + Z_\perp, \tag{2}$$

where $\hat{\lambda}_0$ and $\hat{\lambda}_2$ are $q \times 1$ vectors, $\hat{\Lambda}_1$ is a $q \times p$ matrix, and $Z_\perp$ is a $q \times 1$ vector of residuals. Substituting equation (2) into equation (1), we have

$$Y = (\hat{\beta}_0 + \hat{\beta}_3^T \hat{\lambda}_0) + (\hat{\beta}_1^T + \hat{\beta}_3^T \hat{\Lambda}_1) X + (\hat{\beta}_2 + \hat{\beta}_3^T \hat{\lambda}_2) A + \hat{\beta}_3^T Z_\perp + M(\hat{\gamma}_0 + \hat{\gamma}_1^T X + \hat{\gamma}_2 A) + Y_\perp. \tag{3}$$

Since $Y_\perp$ is the least squares residual for regression (1), it is orthogonal to the span of $\{1, X, A, Z, M, MX, MA\}$. Because $Z_\perp$ is a linear combination of X, A, and Z, $\{1, X, A, Z_\perp, M, MX, MA\}$ and $\{1, X, A, Z, M, MX, MA\}$ span the same space. Thus equation (3) represents the least squares fit of Y on $\{1, X, A, Z_\perp, M, MX, MA\}$, meaning that the RWR estimator of the CDE is

$$\widehat{\text{CDE}}_{\text{RWR}}(a, a', m) = (\hat{\beta}_2 + \hat{\beta}_3^T \hat{\lambda}_2 + \hat{\gamma}_2 m)(a - a').$$

From equation (3), we also know that the de-mediated outcome can be written as

$$Y_d = (\hat{\beta}_0 + \hat{\beta}_3^T \hat{\lambda}_0) + (\hat{\beta}_1^T + \hat{\beta}_3^T \hat{\Lambda}_1)X + (\hat{\beta}_2 + \hat{\beta}_3^T \hat{\lambda}_2)A + \hat{\beta}_3^T Z_\perp + Y_\perp. \tag{4}$$

Since $Z_\perp$ and $Y_\perp$ are both orthogonal to the span of $\{1, X, A\}$ (from the properties of least squares residuals), $\hat{\beta}_3^T Z_\perp + Y_\perp$ is also orthogonal to the span of $\{1, X, A\}$. Thus equation (4) represents the least squares fit of Y_d on X and A, meaning that the sequential g-estimator of the CDE is

$$\widehat{\text{CDE}}_{\text{SG}}(a, a', m) = (\hat{\kappa}_2 + \hat{\gamma}_2 m)(a - a') = (\hat{\beta}_2 + \hat{\beta}_3^T \hat{\lambda}_2 + \hat{\gamma}_2 m)(a - a').$$

Obviously, the sequential g-estimator is the same as the RWR estimator.

B: Consistency of RWR in the Presence of Intermediate Interactions

First, we explain an implicit modeling assumption that underlies both the sequential g-estimator and the RWR estimator described in the main text. Consider the following SNMM, which is analogous to the observed data regression in equation (1) except that it is a model for the potential outcomes:

$$\mathbb{E}[Y(a, m)|X, A = a, Z] = \beta_0 + \beta_1^T X + \beta_2 a + \beta_3^T Z + m(\gamma_0 + \gamma_1^T X + \gamma_2 a). \tag{5}$$

With the sequential g-estimator, the least squares regression in step 3 implies the linearity of $\mathbb{E}[Y(a, 0)|X, A = a]$ in X and a:

$$\mathbb{E}[Y(a, 0)|X, A = a] = \kappa_0 + \kappa_1^T X + \kappa_2 a. \tag{6}$$

Setting $m = 0$ in model (5), we have

$$\mathbb{E}[Y(a, 0)|X, A = a, Z] = \beta_0 + \beta_1^T X + \beta_2 a + \beta_3^T Z. \tag{7}$$

Taking the expectation of equation (7) over Z, conditional on X, yields

$$\mathbb{E}[Y(a,0)|X, A=a] = \beta_0 + \beta_1^T X + \beta_2 a + \beta_3^T \mathbb{E}[Z|X, A=a]. \tag{8}$$

Comparing equations (6) and (8), we see that $\beta_3^T \mathbb{E}[Z|X, A=a]$ must be linear in X and a. Since β_3 represents model parameters that can vary freely in $\mathbb{R}^q$, the linearity of $\beta_3^T \mathbb{E}[Z|X, A=a]$ implies that each component of $\mathbb{E}[Z|X, A=a]$ must be linear in X and a. Conversely, when each component of $\mathbb{E}[Z|X, A=a]$ is linear in X and a, model 5 implies equation (6). Thus, the sequential g-estimator implicitly assumes each component of $\mathbb{E}[Z|X, A=a]$ is linear in X and a. This assumption is more explicit in the RWR estimator, which requires the user to fit a linear model for each of the intermediate confounders. Thus, both the sequential g-estimator and the RWR estimator are based on the linearity of $\mathbb{E}[Z|X, A=a]$,

$$\mathbb{E}[Z|X, A=a] = \lambda_0 + \Lambda_1 X + \lambda_2 a, \tag{9}$$

although this model can be specified more flexibly in practice by, for example, including higher-order or interaction terms involving X and a. Similar to equation (2) in Appendix A, λ_0 and λ_2 are both $q \times 1$ vectors and Λ_1 is a $q \times p$ matrix.

Next, to see the consistency of the RWR estimator in the presence of intermediate interactions, consider the following SNMM:

$$\mathbb{E}[Y(a,m)|X, A=a, Z] = \beta_0 + \beta_1^T X + \beta_2 a + \beta_3^T Z + m(\gamma_0 + \gamma_1^T X + \gamma_2 a + \gamma_3^T Z). \tag{10}$$

Given equation (9), the CDE can be expressed as

$$\begin{aligned}
\mathbb{E}[Y(a,m) - Y(a',m)] &= \mathbb{E}_X \mathbb{E}[Y(a,m)|X, A=a] - \mathbb{E}_X \mathbb{E}[Y(a',m)|X, A=a'] \quad \text{(because } Y(a,m) \perp\!\!\!\perp A|X, \forall a, m\text{)} \\
&= \mathbb{E}_X \mathbb{E}_{Z|X,A=a} \mathbb{E}[Y(a,m)|X, A=a, Z] - \mathbb{E}_X \mathbb{E}_{Z|X,A=a'} \mathbb{E}[Y(a',m)|X, A=a', Z] \\
&= \beta_2(a-a') + \gamma_2 m(a-a') + \beta_3^T \cdot \mathbb{E}_X[\mathbb{E}[Z|X, A=a] - \mathbb{E}[Z|X, A=a']] \\
&\quad + \gamma_3^T m \cdot \mathbb{E}_X[\mathbb{E}[Z|X, A=a] - \mathbb{E}[Z|X, A=a']] \\
&= \beta_2(a-a') + \gamma_2 m(a-a') + \beta_3^T \lambda_2 (a-a') + \gamma_3^T \lambda_2 m(a-a') \\
&= [(\beta_2 + \beta_3^T \lambda_2) + (\gamma_2 + \gamma_3^T \lambda_2) m](a-a')
\end{aligned}$$

It is easy to show that the RWR estimator based on model (**??**) is equal to

$$\widehat{\mathrm{CDE}}_{\mathrm{RWR}}(a, a', m) = [(\hat{\beta}_2 + \hat{\beta}_3^T \hat{\lambda}_2) + (\hat{\gamma}_2 + \hat{\gamma}_3^T \hat{\lambda}_2)m](a - a').$$

Thus, when linear models for both $\mathbb{E}[Y(a, m)|X, A = a, Z]$ and $\mathbb{E}[Z|X, A = a]$ are correctly specified and the potential outcomes are sequentially ignorable, all coefficient estimates are consistent. It follows that $\widehat{\mathrm{CDE}}_{\mathrm{RWR}}(a, a', m)$ is also consistent.

C: R Code for RWR

In this appendix, we illustrate the implementation of RWR in R for estimating the CDE of media framing on support for immigration. Replication data can be found at Teppei Yamamoto's Dataverse: https://hdl.handle.net/1902.1/19036

```
library(dplyr)
# load data
load("PA-ImaiYamamoto.RData")
# function for demeaning
demean <- function(x) x - mean(x, na.rm = TRUE)
# function for residualizing intermediate confounders
residualize <- function(formula, df) residuals(lm(formula, df))
# data preprocessing
Brader2 <- Brader %>%
  select(immigr, emo, p_harm, tone_eth, ppage, ppeducat, ppgender, ppincimp) %>% na.omit() %>%
  mutate(immigr = 4 - immigr,
         hs = (ppeducat == "high school"),
         sc = (ppeducat == "some college"),
         ba = (ppeducat == "bachelor's degree or higher"),
         female = (ppgender == "female")) %>%
  mutate_at(vars(emo, p_harm, ppage, female, hs, sc, ba, ppincimp), demean) %>%
  mutate(., p_harm_res = residualize(p_harm ~ ppage + female + hs + sc + ba + ppincimp + tone_eth, .))
# total effect model
total_mod <-  lm(immigr ~ ppage + female + hs + sc + ba + ppincimp + tone_eth,
                  data = Brader2)
# rwr without intermediate interactions
rwr1_mod <-  lm(immigr ~  ppage + female + hs + sc + ba + ppincimp + tone_eth + p_harm_res +
                emo * ( ppage + female + hs + sc + ba + ppincimp + tone_eth),
                data = Brader2)
```

```r
# rwr with intermediate interactions
rwr2_mod <-  lm(immigr ~  ppage + female + hs + sc + ba + ppincimp + tone_eth + p_harm_res +
                emo * (ppage + female + hs + sc + ba + ppincimp + tone_eth + p_harm_res),
                data = Brader2)
# bootstrap
nboots <- 500
rwr1_hold <- matrix(NA, nrow = length(coef(rwr1_mod)), ncol = nboots)
rwr2_hold <- matrix(NA, nrow = length(coef(rwr2_mod)), ncol = nboots)
for (i in 1:nboots) {
     star <- sample(1:nrow(Brader2), replace = TRUE)
     Brader2_star <- Brader2[star, ]
     Brader2_star <- Brader2_star %>% tbl_df() %>%
     mutate(., p_harm_res = residualize(p_harm ~ ppage + female + hs + sc + ba + ppincimp + tone_eth, .))
     rwr1_star <-  lm(immigr ~  ppage + female + hs + sc + ba + ppincimp + tone_eth + p_harm_res +
                   data = Brader2_star)
     rwr2_star <-  lm(immigr ~  ppage + female + hs + sc + ba + ppincimp + tone_eth + p_harm_res +
                   data = Brader2_star)
     rwr1_hold[, i] <- coef(rwr1_star)
     rwr2_hold[, i] <- coef(rwr2_star)
}
rownames(rwr1_hold) <- names(coef(rwr1_mod))
rownames(rwr2_hold) <- names(coef(rwr2_mod))

out_coefs <- c("(Intercept)", "tone_eth", "emo", "tone_eth:emo", "p_harm_res", "p_harm_res:emo")
rwr1_est <- coef(rwr1_mod)[out_coefs]
rwr2_est <- coef(rwr2_mod)[out_coefs]
rwr1_se <- apply(rwr1_hold, 1, sd)[out_coefs]
rwr2_se <- apply(rwr2_hold, 1, sd)[out_coefs]
```