

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2011

Multiple Auxiliary Variables in Nonresponse Adjustment

Frauke Kreuter

University of Maryland - Baltimore, fkreuter@survey.umd.edu

Kristen Olson

University of Nebraska-Lincoln, kolson5@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/sociologyfacpub>

 Part of the [Sociology Commons](#)

Kreuter, Frauke and Olson, Kristen, "Multiple Auxiliary Variables in Nonresponse Adjustment" (2011).
Sociology Department, Faculty Publications. 151.

<https://digitalcommons.unl.edu/sociologyfacpub/151>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Multiple Auxiliary Variables in Nonresponse Adjustment

Frauke Kreuter,

University of Maryland, College Park, USA, and
Institute for Employment Research, Nuremberg, Germany;
Corresponding author — 1218 LeFrak Hall, College Park, MD, USA;
email fkreuter@survey.umd.edu

Kristen Olson

University of Nebraska-Lincoln, USA

Abstract

Prior work has shown that effective survey nonresponse adjustment variables should be highly correlated with both the propensity to respond to a survey and the survey variables of interest. In practice, propensity models are often used for nonresponse adjustment with multiple auxiliary variables as predictors. These auxiliary variables may be positively or negatively associated with survey participation, they may be correlated with each other, and can have positive or negative relationships with the survey variables. Yet the consequences for nonresponse adjustment of these conditions are not known to survey practitioners. Simulations are used here to examine the effects of multiple auxiliary variables with opposite relationships with survey participation and the survey variables. The results show that bias and mean square error of adjusted respondent means are substantially different when the predictors have relationships of the same directions compared to when they have opposite directions with either propensity or the survey variables. Implications for nonresponse adjustment and responsive designs will be discussed.

Keywords: nonresponse bias adjustment, response propensity models, mean square error, survey participation

Introduction

It is common practice to apply postsurvey adjustment methods to survey data that suffer from nonresponse. Weighting is one such postsurvey adjustment method. With nonresponse weighting (whether it is a weighting class adjustment method, Kalton and Flores-Cervantes 2003; Little 1986; or a response propensity weighting method, Rosenbaum and Rubin 1983), survey respondents are assigned a nonzero weight to compensate for their differential probability of participation given their selection into the sample. These weighting adjustments are aimed at reducing nonresponse bias in the final survey estimates.

Although the goal of most nonresponse weighting adjustments is to decrease potential bias in survey estimates due to nonresponse, it has long been known that survey analysts risk increasing the variability of the survey estimates by the use of weights. Kish (1965) showed that ineffective nonresponse weights (that is, those that do not reduce nonresponse bias) increase the variance of adjusted respondent means by the squared coefficient of variation of the weights ($1 + cv^2$). An inflation in the variance, or standard errors, of survey estimates increases the width of confidence intervals and reduces the ability of an analyst to detect significant results.

Adjustment variables that are successful at reducing potential nonresponse bias have two properties: they are predictive of the sampled person's probability of responding to a survey request and of the survey variables of interest (Groves 2006; Kalton and Flores-Cervantes 2003; Kalton and Maligalig 1991; Little 1986; Little and Vartivarian 2003, 2005). Little and Vartivarian (2005) demonstrated the need for both criteria – predicting both the probability of response and the survey variables – to be met in order to reduce nonresponse bias without increasing the variance of the estimate. If the association between the auxiliary variable (used for weighting adjustment) and the survey variable of interest is low, while the association between that same auxiliary variable and the response indicator is high, then the weighted mean will have increased variance without decreasing the nonresponse bias. Their work (Little and Vartivarian 2005) examined the correlation of one adjustment variable with the response indicator and a survey outcome variable. In practice, researchers find themselves in a much more complicated situation:

- First, few auxiliary variables are available for both respondents and nonrespondents in most surveys.
- Second, threatened by the decrease in response rates, survey re-

searchers have extensively researched and developed covariates of survey participation and moved toward data collection systems that collect such covariates.¹ Thus, most available auxiliary variables might be stronger predictors of response while their relationship to key survey variables is unknown.

- Third, any given auxiliary variable is likely to differ in the strength of its association with key survey outcome variables (Kreuter et al. 2010).
- Fourth, no single variable is likely to be highly correlated with both the response indicator and all key survey outcome variables (Groves, Wagner, and Peytcheva 2007; Kreuter, Lemay, and Casas-Cordero 2007; Peytchev and Olson 2007; Yan and Raghunathan 2007). Therefore, researchers have to make choices on where to concentrate their efforts when expanding their collection of auxiliary variables and on how to combine multiple auxiliary variables in their postsurvey adjustments.

When examining nonresponse adjustment of respondent means across a variety of surveys, Kreuter et al. (2010) noticed a fifth challenge. Strong predictors of both response and the survey outcome variables—that is, those variables meeting the criteria set out by Little and Vartivarian (2005)—led to large shifts in adjusted survey estimates relative to their unadjusted counterparts when considered individually but had little to no effect when used jointly.

For example, in the American National Election Study, adding a commonly used strong predictor of response (voiced negative statements about the survey) to the already existing nonresponse weighting variables sometimes led to substantial changes in the adjusted estimate and at other times to no changes. The changes in the adjusted estimate were not clearly related to the bivariate association between the auxiliary variable and the survey variables, but to the fact that the associations of these auxiliary variables with the response indicator and the survey outcome variable were in opposite directions (Peytchev and Olson 2007). Similarly, in the Wisconsin Divorce Study, Olson (2007) found two strong predictors of both propensity and the survey variable of interest (age and the number of children) in an examination of nonresponse adjustment for the mean length of marriage. Although both variables individually met the criteria for effective adjustment variables, when both auxiliary variables were included in a propensity model, there was no change in the adjusted estimate. Here, too, the two variables were in opposite directions, with

number of children being positively correlated to both participation as well as length of marriage, whereas age was negatively related to response and positively associated with the length of marriage.

Inspired by these findings, we examine here the effects of multiple auxiliary variables on bias and variance of the adjusted respondent means. Of particular interest in this research are the consequences on bias and variance of adjusted respondent means when the associations of these auxiliary variables with the response indicator and the survey outcome variable are in the same or opposite directions. With this work we hope to stimulate a discussion among survey methodologists and survey practitioners and give guidance for the field. We will use a simple graphical model and a simulation with one response indicator, one survey outcome variable, and two potential adjustment variables. Except for the binary response indicator, all variables are assumed to be continuous and normally distributed. We also discuss whether the effects observed with continuous adjustment variables hold when the auxiliary variables are categorical.

Graphical Illustration

The unexpected findings of Kreuter et al. (2010) and Olson (2007) suggest that different auxiliary variables may have counteracting effects on either response propensity or the survey variables. Thus, nonresponse bias in survey estimates may not necessarily occur when there is more than one influence of auxiliary variables on either survey participation or the survey variables.

Under a stochastic model for survey nonresponse (Groves 2006; Lessler and Kalsbeek 1992), nonresponse bias of an unadjusted respondent mean can be expressed as:

$$\text{Bias}(\bar{Y}_R) \sim \frac{\text{Cov}(P, Y)}{\bar{P}} = \text{Corr}(P, Y) * SD(Y) * CV(P) \quad (1)$$

where P represents the response propensity and Y the survey variable of interest, $SD(Y)$ the standard deviation of Y , and $CV(P)$ the coefficient of variation of P . The average response propensity P translates to the survey response rate. As can be seen in equation (1), stronger associations between Y and P will lead to increased nonresponse bias of the unadjusted respondent mean. The question is how competing influences affect $\text{Corr}(P, Y)$, the correlation between response propensity and the survey variables of interest. In particular, when will two competing influences on response propensity and the survey vari-

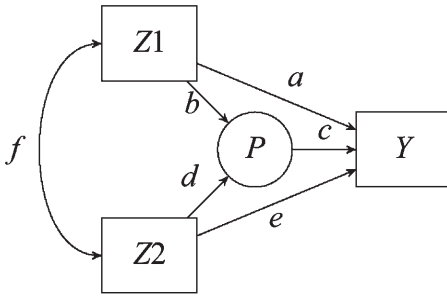


Figure 1. Path diagram displaying the relationship between auxiliary variables Z1 and Z2, response propensity P, and survey variable of interest Y.

ables increase $Corr(P, Y)$ and when will two competing influences decrease $Corr(P, Y)$?

We start with a simple path diagram in Figure 1. Here, two auxiliary variables, Z1 and Z2, are measured for both respondents and nonrespondents. Let us assume both auxiliary variables influence the sampled person’s probability of participating in the survey P as well as the observed survey outcome variable Y. We call $\hat{P}_{1,2}$ the estimate of the response propensity using both auxiliary variables as predictors, \hat{P}_1 the estimate of response propensity using only Z1, and likewise \hat{P}_2 the estimate using only Z2.

Each lower case letter in Figure 1 represents a standardized partial regression coefficient, sometimes called path coefficient (Loehlin 1998), except f, which represents the correlation between the two auxiliary Z variables.²

Using path rules (Loehlin 1998), we can express the correlation between response propensity and the survey outcome variable when there are two auxiliary variables as:

$$Corr(\hat{P}_{1,2}, Y) = c + ba + de + bfe + dfa \tag{2}$$

We asked earlier how the estimated correlation $Corr(P, Y)$ will change as a function of the specification of the response propensity model. To address this, imagine that only one auxiliary variable, Z1, that influences both response propensity and the survey variable of interest, is used in the adjustment model. In this situation, then the paths between Z2, P, and Y disappear, and $Corr(\hat{P}_{1,2}, Y)$ (as represented by equation (2)) reduces to:

$$Corr(\hat{P}_1, Y) = c + ba \tag{3}$$

Of interest here is under which circumstances will $Corr(\hat{P}_1, Y)$ be larger than $Corr(\hat{P}_{1,2}, Y)$ when adding the second variable to the response propensity model:

$$c + ba > c + ba + de + bfe + dfa \quad (4)$$

That is: When does the full model specification, including the second auxiliary variable, reduce the correlation between response propensity and the outcome variable? Subtracting $c + ba$ from both sides of the equation, we see that this will be true when

$$0 > de + bfe + dfa \quad (5)$$

For simplicity, let us first assume $Z1$ and $Z2$ are uncorrelated, that is $f = 0$. When $f = 0$, then the bfe and dfa terms in Equation (5) are zero, and the inequality reduces to $0 > de$. This simplified inequality will hold when the second auxiliary variable, $Z2$, is either negatively related to the probability of responding to the survey (represented by d) or $Z2$ is negatively associated with Y (represented by e). When one of these relationships is negative and the other one is positive, then the correlation between P and Y will be reduced by the introduction of $Z2$. In other words, when the direction of the association between $Z2$ and Y is opposite that of $Z2$ and P , then $Corr(\hat{P}_{1,2}, Y)$ will be reduced compared to $Corr(\hat{P}_1, Y)$. Reverse coding either one of the auxiliary variables or Y will not change this result (see Appendix for more detail).

However, if $f \neq 0$ and $Z2$ has no relationship with Y , then $e = 0$ and Equation (5) becomes $0 > dfa$. This means that if one of the three correlations represented by d , f , and a is negative, then the addition of $Z2$ reduces $Corr(\hat{P}_{1,2}, Y)$ compared to $Corr(\hat{P}_1, Y)$. If two of the three are negative and one is positive, then the product of these three terms is positive and the correlation increases. A similar situation holds when $d = 0$. If f , d , and e are not equal to zero, the situation is more complicated. We will address the various situations in our simulation.

A relevant question is when these counteracting effects might occur in a survey context. Take, for example, a survey on income and wealth, where information on age and education of the sample members is part of the sampling frame. For simplicity, we assume linear relationships between the variables. Wealth tends to be positively associated with age and education (Kominski and Sutterlin 1993). That is, on average, older people and more educated people are more wealthy. At the same time, older people tend to be less likely to cooperate with a survey request while people with higher levels of education tend to be more likely to cooperate with survey requests (Abraham, Maitland, and Bianchi 2006).³ Thus, for one variable (education), values associated with greater wealth are also associated with higher levels of cooperation, while for the other variable (age), values asso-

ciated with greater wealth are associated with lower levels of cooperation. For a commonly observed variable such as income or wealth, these counteracting effects may lead to reduced nonresponse bias due to noncooperation on mean income. (We return to the distinction between contact and cooperation in the discussion.)

We examine in detail how varying the strength and direction of the association changes nonresponse bias and variance properties of unadjusted and adjusted survey means. Since any real survey data is subject to multiple influences outside our control and potentially unmeasured, we use simulation methods for this research.

Simulation Set-up

The first step in this simulation is to create populations of size $N = 1,000,000$ with one outcome variable Y and two auxiliary variables Z_1 and Z_2 . The auxiliary variables are both multivariate normally distributed with a mean of zero and a standard deviation of one ($MN(0, 1)$), with varying correlation strength between these two variables. In the terms of Figure 1, $f = 0$, $f = 0.2$, or $f = -0.2$. The outcome variable is a function of Z_1 and Z_2 as well as a standard normally distributed error term u :

$$y_i = 10 + \beta_1 * z_{1i} + \beta_2 * z_{2i} + u_i \quad (6)$$

To permit easy comparison across simulations, the average value of each of the survey variables for the full sample, \bar{y} , is set to 10 for all conditions. The β s in equation (6) correspond to unstandardized versions of a and e in the path diagram.⁴ For the different populations we vary the β s to create weak, moderate, and strong associations with the survey outcome variables. It does not matter which of the two coefficients is negative for our analysis of opposite effects, thus we assign negative values for only one of the two beta coefficients. This creates a total of 54 simulation cells (i.e., populations), with all possible combinations of $\beta_1 : 0.1, 2, 4$; $\beta_2 : 0.1, 2, 4, -0.1, -2, -4$; and $f : 0, 0.2, -0.2$.

From each of these 54 populations we randomly select samples of size $n = 2,500$ and repeat the sampling $m = 500$ times for each population. For each sampled person in each of the 500 samples, we generate a response probability, p_i , where p_i is a function of z_{1i} and z_{2i} (see Equation (7)). Similarly to creating the outcome variable, we generate response probabilities by systematically varying coefficients for the Z s. That is, we allow γ s in Equation (6) to equal either 0.1, 1, or 3.

These values correspond to odds ratios of 1.1, 2.7, and 20 for a one-unit change in either z variable, respectively.

$$p_i = \frac{e^{1 + \gamma_1 * z_{1i} + \gamma_2 * z_{2i}}}{1 + e^{1 + \gamma_1 * z_{1i} + \gamma_2 * z_{2i}}} \quad (7)$$

We then create a binary response indicator r , which is drawn from a binomial distribution with probability p_i ($r \sim B(1, p_i)$) (essentially, a Bernoulli draw where the probability of participating is different for each person). The γ s in this function correspond roughly to b and d in the aforementioned path diagram.⁵

As a result, we get a set of respondents of size n_R from each sample of $n = 2,500$ and a set of nonrespondents of size $n - n_R$. For the responding sample $r_i = 1$ we create weights to adjust for nonresponse bias. To create such adjustment weights, we use a logistic regression model to predict the response indicator r as a function of either z_1 alone, z_2 alone, or both z_1 and z_2 . That is, we estimated:

$$\begin{aligned} \text{logit}(\text{Pr}(r = 1)) &= \hat{\gamma}_{01} + \hat{\gamma}_{11} * z_{1i} \\ \text{logit}(\text{Pr}(r = 1)) &= \hat{\gamma}_{02} + \hat{\gamma}_{21} * z_{2i} \\ \text{logit}(\text{Pr}(r = 1)) &= \hat{\gamma}_{03} + \hat{\gamma}_{12} * z_{1i} + \hat{\gamma}_{22} * z_{2i} \end{aligned} \quad (8)$$

The inverse of the estimated response propensities from each of these logistic models formed the adjustment weights. We estimate the resulting response rate (n_R/n) (averaged over the $m = 500$ samples drawn out of each of the populations), the unadjusted respondent mean \bar{y}_R , and three different adjusted respondent means. Our estimates of interest from these analyses are then the bias (Equation (9)), and the root mean square error (RMSE, Equation (10)) of the respondent means using none or one of the three nonresponse adjustments. In essence, these results examine the effects for nonresponse adjustment of one omitted predictor, one extra predictor, and two important predictors of propensity and the survey variables.

$$\text{Bias}(\bar{y}_R) = 1/m \sum (\bar{y} - \bar{y}_R) \quad (9)$$

$$\begin{aligned} \text{RMSE}(\bar{y}_R) &= \sqrt{[\text{Bias}(\bar{y}_R)]^2 + \text{Var}(\bar{y}_R)} \\ &= \sqrt{1/m \sum (\bar{y} - \bar{y}_R)^2} \end{aligned} \quad (10)$$

γ_1	γ_2		
	.1	1	3
.1	73.0	69.6	61.3
1	69.6	67.3	60.7
3	61.3	60.7	58.4

Figure 2. Average response rates for different combinations of γ_1 and γ_2 .

The bias and RMSE of the unadjusted respondent mean is of interest as the “baseline” value for most analysts. If the RMSE for the adjusted means is greater than those for the unadjusted respondent mean, then the analyst is better off (in RMSE terms) by not using the nonresponse adjustments.⁶

To examine the sensitivity of our results, we altered our simulations in two ways. First, we created dichotomous adjustment variables by splitting Z_1 and Z_2 at their medians. These categorical variables are closer to adjustment variables seen in common surveys. Second, to examine the function of sample size, we also selected samples of size $n = 10,000$ from our populations.

Simulation Results

As shown in Figure 2, the average response rates varied across the cells formed by the different specifications of γ s. If both auxiliary variables z_1 and z_2 strongly influence survey participation ($\gamma_1 = 3$ and $\gamma_2 = 3$), we see the lowest average response rate with 58 percent. However, if both auxiliary variables z_1 and z_2 only weakly influence survey participation ($\gamma_1 = .1$ and $\gamma_2 = .1$), an average response rate of 73 percent is achieved.

The influence of two auxiliary variables that vary in strength and direction of the relationship on both response and the survey outcome has consequences for both bias and variance (expressed here as root mean square error) of unadjusted and adjusted respondent means. Figures 3, 4, and 5 present results from the simulation. Within each panel (representing a “survey administration”), we present results for nine survey outcome variables Y , created using varying strengths of the β coefficients without correlation between the Z variables. Each panel varies in the combination of γ values. (The results are consistent when the negative coefficient is in the propensity model rather than the Y model; thus, we show only the negative coefficients for β s.)

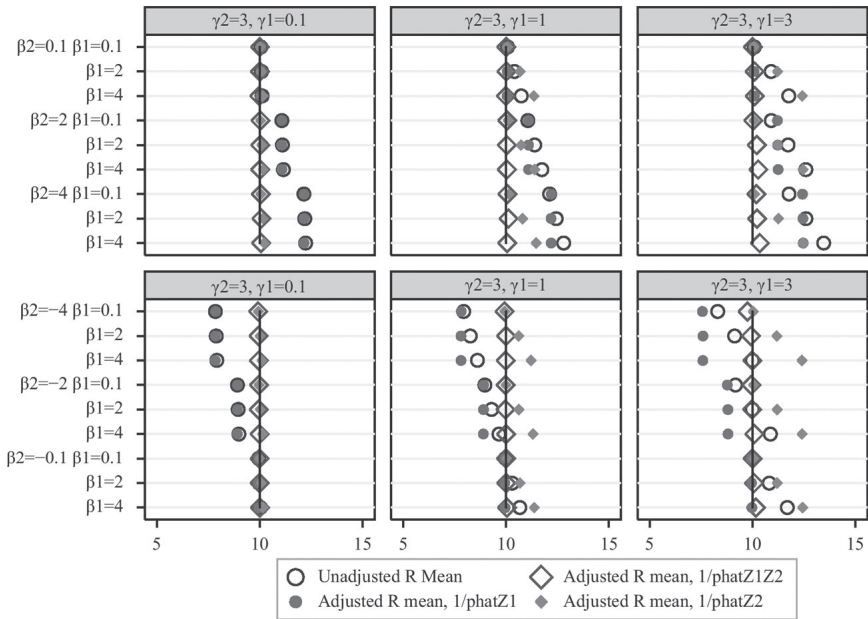


Figure 3. Estimated means for unadjusted and adjusted respondent sample, $n = 2,500, f = 0$.

One way to think about these panels is that each of them represents a survey with a given response rate and distribution of response propensities, and each line in the panel represents different survey variables, each varying in their association with the determinants of nonresponse in that survey. The four values within each line represent the estimate without nonresponse adjustment, adjusted using only either z_1 and z_2 and using both z variables in the adjustment. We will look in detail at the figures for bias and RMSE separately.

Effect of Adjustment on Bias

In Figure 3 the estimates for the unadjusted respondent means are displayed as hollow circles and serve as a reference value to which we compare the effectiveness of the nonresponse adjustment. Since the overall population mean for all simulations is 10, deviation from the vertical line indicates bias in these respondent means. For simplicity, we show results when one auxiliary variable (z_1) is highly predic-

tive of response ($\gamma_1 = 3$) and the other auxiliary variable (z_2) varies in its association with response.⁷

In panel 1 of Figure 3, z_1 is only very weakly predictive of response ($\gamma_1 = 0.1$); as we move across the top panels to the right, the effect of z_1 on the probability to respond increases ($\gamma_1 = 1$ and $\gamma_1 = 3$).

In the first line in panel 1, both of the z variables are only weakly predictive of the survey outcome variable Y , with $\beta_1 = 0.1$ and $\beta_2 = 0.1$. As a result, all of the respondent estimates are unbiased, whether unadjusted or adjusted means, regardless of the strength of γ_1 . As we move down the lines in panel 1, the association of z_1 and z_2 with the survey outcome variable increases.

We see that the unadjusted respondent mean is now upwardly biased by nonresponse. The value of the unadjusted respondent mean in lines 4 through 6 is the same regardless of the strength of the association between z_1 and Y . This makes sense as z_2 is a strong predictor of propensity, but z_1 is not. In this panel, the adjustment model that accounts only for z_1 (dark grey dot) does little to shift the mean toward the population value; however, the adjustment models that include either the strong predictor of response z_2 (grey diamond) or both z variables (hollow diamond) essentially remove all of the bias of the estimate. This is as expected as these models appropriately account for the missing data mechanism, that is, the probability of being a respondent is largely determined by z_2 .

As we move across the top panels, we see a different story. As the association between z_1 and the probability of responding to the survey increases, the additional auxiliary variable has a correspondingly increasing effect on the nonresponse bias of the unadjusted respondent mean. With two moderate to strong influences on nonresponse bias, adjusting for both z_1 and z_2 (hollow diamond) always removes nonresponse bias. The adjustment models that include only one predictor (e.g., only z_1) eliminate nonresponse bias when the other variable's (e.g., z_2) influence on the survey variables is weak ($\beta_1 = 0.1$) but actually increase the bias of the adjusted mean when the influence of the other auxiliary variable on the survey variables is strong ($\beta_1 = 2$ or $\beta_1 = 4$).

Moving down in Figure 3, we now see situations in which one of the auxiliary variables (z_2) is negatively associated with the survey outcome variable. The negative correlation of z_2 and y does not change the effect size compared to the panel above with all positive correlations, but the bias is now negative. As we increase the effect of z_1 on the response propensity, an interesting pattern emerges: depending on the strength of each of the associations, the effects can

cancel each other out, and the unadjusted mean will be unbiased even though both variables are strong predictors of both response propensity and the survey variables. This happens in rows one, five, and nine of the last panel in Figure 3. This is likely what Olson (2007) found in the Wisconsin Divorce Study.

As in the case when both variables are positively associated with the survey variables, including both predictors in an adjustment model always yields the least biased adjusted estimate. Unlike previously, however, when either z variable is a moderate to strong predictor of the Y variable and a moderate to strong predictor of propensity, then including only one z variable in the adjustment model substantially increases the bias of the estimate. That is, accounting for only one of the competing influences on propensity will actually damage (i.e., increase the bias) survey estimates.

We also looked at the same conditions but with a correlation between z_1 and z_2 and of 0.2 or -0.2 . Surprisingly little differed in terms of bias reduction when the auxiliary variables were correlated. Some slight differences are visible with respect to RMSE.⁸ Future research could examine stronger correlations between the two auxiliary variables.

Impact of the Adjustment on Variance

Figures 4 and 5 show the relative root mean square error (Equation (10)) of the nonresponse adjusted respondent mean from the same simulation conditions as discussed previously. Each symbol represents the ratio of the average RMSE for a given set of adjustment variables divided by the RMSE of the unadjusted respondent mean. A value of 1 (the dashed vertical reference line) indicates no change in the RMSE after applying nonresponse adjustments, values below 1 indicate decrease in estimated RMSE relative to the unadjusted respondent mean (desirable), while values above 1 indicate increased RMSE estimates relative to the unadjusted respondent mean (undesirable).

Starting in panel 1 of Figure 4, the case where z_1 is a weak predictor ($\gamma_1 = 0.1$), we see two distinct patterns. First, when z_2 is a weak predictor of Y ($\beta_2 = 0.1$), then the change in RMSE is determined by the strength of the association between z_1 and Y and whether z_2 is included in the adjustment model. When z_1 is the only variable included in the adjustment model, then the RMSE is reduced (relative RMSE < 1) as β_1 increases (the hollow circles move to the left as we go down the lines in panel 1). When z_2 – the strong predictor of participation – is included in the adjustment model, then the relative RMSE of the

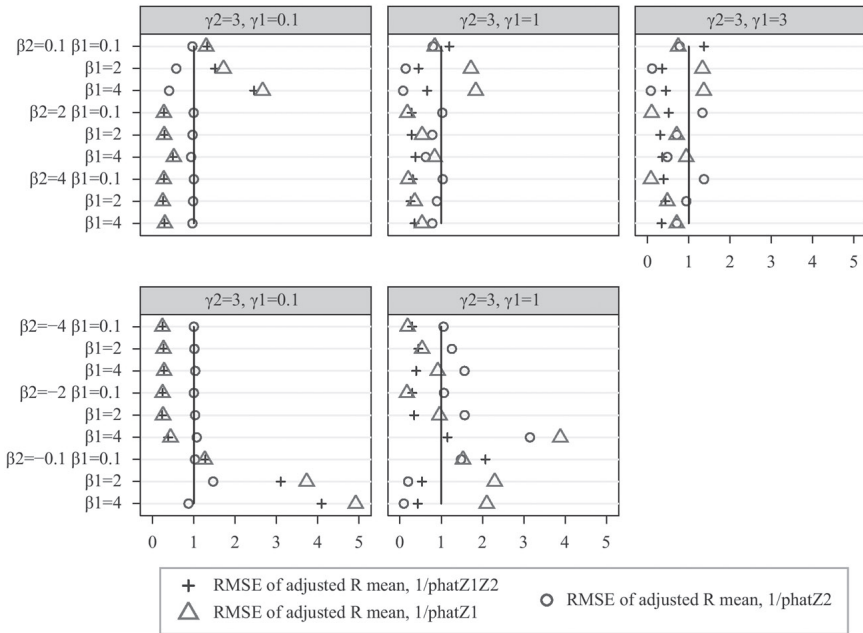


Figure 4. Relative ratio of root mean square error for the adjusted respondent mean to the unadjusted respondent mean, $n = 2,500, f = 0$.

adjustment mean increases, even if z_1 is also included in the model. That is, in a setting where there was no nonresponse bias in the unadjusted respondent mean (as was show in panel 1 of Figure 4), including z_2 (the strong predictor of response) in the adjustment model increases the RMSE when it is not a moderate or strong predictor of the survey variables. This is the instance identified by Kish (1965) for variance inflation due to weighting.

What happens when both z_1 and z_2 are positively associated with survey participation and are of at least moderate strength? Here, the story is not nearly as clear cut. When the auxiliary variables are each only weakly associated with the survey variables of interest, there is an increase in RMSE when both variables are included in the adjustment model, but a slight decrease in RMSE when only one auxiliary variable is included in the adjustment model.

When one auxiliary variable is moderately or strongly associated with Y , but the other is only weakly associated with Y , then the inclusion of only the strong Y predictor in an adjustment model reduces RMSE relative to the unadjusted respondent mean, and the in-

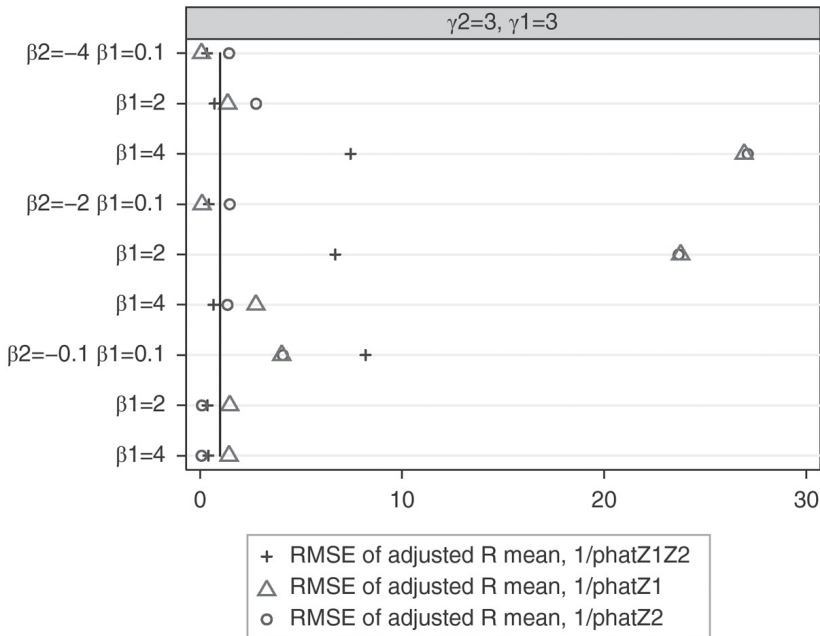


Figure 5. Relative ratio of root mean square error for the adjusted respondent mean to the unadjusted respondent mean, $n = 2,500, f = 0$.

clusion of only the weak Y predictor increases RMSE. Including both variables in the adjustment model reduces RMSE relative to the unadjusted respondent mean, but the best scenario (in RMSE terms) is to include only the strong Y predictor. That is, in this situation, maximizing propensity model fit reduces bias but increases variance of the adjusted mean. Finally, when both auxiliary variables are moderate to strong positive predictors of propensity and the survey variables, then inclusion of both of the auxiliary variables in the adjustment model has the largest reduction on RMSE. In this situation, adjustment models that include either z_1 or z_2 also reduce RMSE relative to the unadjusted respondent mean, with the magnitude of the reduction varying as expected with strength of the association of each auxiliary variable with the Y variables.

Moving down to the bottom panel of Figure 4, we see that when one of the auxiliary variables has a negative association with the survey variables, the increase in RMSE can be quite dramatic. As before, first consider the case where there is only one strong predictor of propensity (panel 4) and that predictor is included in the adjustment model (the plus signs and triangles), but it is not a strong predictor of the survey

variables (lines 7 through 9). In this situation, the RMSE is at least four times greater than the unadjusted respondent mean, substantially more than the same scenario when both are positive predictors (around 1.5 to 2.5 times). When the strong propensity predictor is also a strong predictor of the survey variables and it is included in the adjustment model, then the RMSE is approximately the same size as that of the unadjusted mean. Unlike when all of the associations between the auxiliary variables, Y , and propensity are positive, including only the strong predictor of the survey variables in the adjustment model does not lead to a reduction in RMSE. In this panel, the other possible scenarios for the relationship with Y mimic those of the “all positive” case.

With increasing strength of the association between z_2 and the response propensity, the effect on RMSE increases substantially. When z_1 and z_2 are only weakly associated with Y (line 7 of panel 5 and Figure 5), then the RMSE increases by 2 to 10 times over the unadjusted respondent mean when both variables are included in the adjustment model; the increase in RMSE is only 1.5 to 4 times when only one auxiliary variable is included in the adjustment model. This magnitude is substantially greater than that of the “all positive” case. If either z_1 or z_2 is moderately or strongly associated with Y while the other is weakly associated with Y , then the inclusion of only the strong predictor in the adjustment model does better (larger relative reduction in RMSE) than including both variables, as we saw when all the predictors are positive.

The biggest difference in RMSE arises when z_1 and z_2 are moderate or strong predictors of both propensity and Y . In the “all positive” case, any adjustments—whether including only one variable or both auxiliary variables—lead to decreases in the RMSE relative to the unadjusted respondent mean. When z_1 and z_2 have associations of equal strength with survey participation (Figure 5) and have associations of equal but opposite strength with the survey variables (e.g., $\beta_1 = -2$ and $\beta_1 = 2$), then any adjustments substantially increase the RMSE relative to the unadjusted mean, even when the missing data mechanism is appropriately modeled by including both z_1 and z_2 . When the association of z_1 and z_2 with Y are not of equal strength, then adjustments that include only one auxiliary variable increase RMSE slightly relative to the unadjusted respondent mean.

We replicated all of the aforementioned simulations using categorical z_1 and z_2 . As can be seen in the Appendix, in each instance, the pattern of the findings for the bias of the adjusted respondent means is identical. The primary difference is in the effectiveness of the adjustment. With continuous adjustment variables, inclusion of both z_1 and z_2 leads to unbiased adjusted respondent means. With categorical ad-

justment variables, the adjusted respondent means are still slightly biased, even when both variables are included in the adjustment model. Interestingly, the RMSE for the categorical adjustments when the adjustment model is misspecified is substantially smaller than that for the continuous adjustments. This is due to the reduced variability of the weights with categorical variables (Kish 1965; Little and Vartivarian 2005).

Discussion

Finding strong predictors of both response propensity and the survey variables of interest substantially affects both bias and variance properties of adjusted respondent means, as shown by Little and Vartivarian (2005). However, the importance of the direction of the association between the auxiliary variables, response propensity, and the survey variables of interest has not previously been explored. The implications for bias and mean square error of adjusted respondent means are substantially different when the predictors have relationships of the same directions compared to when they have opposite directions with either propensity or the survey variables.

In practice, survey researchers often look at the difference between weighted and unweighted estimates to hypothesize about reduction of bias from the use of weighting adjustments. Gauging nonresponse adjustments to be effective by monitoring the difference in the adjusted respondent mean from the unadjusted respondent mean may not necessarily be informative. Auxiliary variables that have strong but opposite relationships with the survey variables may reduce or eliminate nonresponse bias of the unadjusted respondent mean. In this instance, the adjusted respondent mean may not shift greatly from the unadjusted mean, even though the bivariate associations between the auxiliary variables and the survey variables are strong. On the other hand, when all of the auxiliary variables are positively associated with both propensity and the survey variables, then lack of movement likely reflects no bivariate association between the auxiliary variables and the survey variables. Survey practitioners monitor the variability of the weights to examine the potential effects on variance of the estimates, often using the $1 + cv^2$ shortcut for the variance increase (Kish 1965; Little and Vartivarian 2005). Weight developers examine functions of variance because estimates of bias are not available. In our simulation we are able to look at both. The implications of including all available predictors for RMSE, however, vary substantially over survey variables. As long as these predictors are positively and moderately associated

with both propensity and the survey variables, then RMSE will tend to decrease by including them in the adjustment model. This is not true when a predictor is weakly but positively associated with the survey variables of interest. In this instance, including extraneous predictors in an adjustment model increases sampling variability, thereby increasing confidence interval width and decreasing the power to detect significant differences. If the relationship between the auxiliary variables and the survey variables are of opposite signs, then excluding important predictors has much more serious implications for both bias and variance of survey estimates.

A survey practitioner monitoring the correlation between an estimated response propensity and the survey variables may be surprised to find that adding predictors to an estimated propensity model weakens that association, even though the bivariate relationships show that the auxiliary variable is a strong predictor of both the survey variables and propensity. We have shown here that this will happen when the relationship between the auxiliary variables and either propensity or the survey variables is of opposite sign from the previously included predictors. An example that would occur in many surveys is the relationship between age, education, and income. Older people are commonly found to be less likely to cooperate with a survey request (a negative association between age and response propensity), but tend to have higher income (a positive association between age and the survey variable). Higher education levels tend to positively predict both survey participation and income. If there were no other influence on survey participation (admittedly a strong assumption), the effect of these two competing influences could result in no bias in the unadjusted respondent mean due to noncooperation nonresponse.

Although we focused here on overall response, the same auxiliary variable can have opposite associations with noncontact and noncooperation nonresponse. For example, although age tends to be negatively associated with cooperation, it is often positively associated with contactability. That is, older persons are easier to contact than younger persons, but they tend to be less cooperative once contacted (Groves and Couper 1998). If the strength of the associations are of equal magnitude for both noncontact and noncooperation nonresponse, the net result could be no nonresponse bias on the mean respondent age or other strongly correlated variables. The conditional nature of contact and cooperation and the concomitant effect on nonresponse bias should be examined through future simulations and observational research.

In addition, we focused here on relationships that were composed solely of main effects. Many real-world examples exist in which the

effect of one auxiliary variable is moderated by the effect of a second auxiliary variable. For example, many survey practitioners use “tree” modeling procedures to create nonresponse adjustment weights. In these procedures, including CHAID and CART methods, all auxiliary variables are categorized. Each category is subsequently subdivided to form subgroups that maximize the fit of the propensity model. Although the results reported here do not directly address these tree methods, they suggest that building adjustment models on important survey variables, rather than on survey participation, would minimize both bias and RMSE of the adjusted survey estimates. This would hold true even when the predictors themselves do not strongly discriminate on response propensity.

We also focused here on two adjustment variables that vary in their strength with propensity and the important survey variables. The effects varied substantially when one was omitted, added but not important, or both added and both important. The number of adjustment variables used in large-scale surveys varies from 1 or 2 (e.g., the National Election Studies) to more than 10 (e.g., the Medical Expenditure Panel Survey; see Wun and Ezzati-Rice 2007). Selecting among multiple available adjustment variables is not directly addressed here.

The results presented here have implications for responsive design decisions (Groves and Heeringa 2006). In a responsive design, decisions are made about changes in survey recruitment protocols during a field period to deliberately change error properties of the final respondent data set. Nonresponse error is the most commonly tackled error source in a responsive design, in which variability of response rates over important subgroups are monitored and changes in key statistics are tracked over the field data collection. This type of monitoring focuses primarily on differences in response propensity across these subgroups; characteristics that predict differential response rates during the data collection process have also been used in nonresponse adjustment (see e.g., the National Survey of Family Growth Cycle 6; Lepkowski et al. 2006). The aforementioned results suggest that auxiliary variables over which to monitor subgroup response rates should be selected specifically for both the strength and the relationship between the direction of their relationship with P and the direction of their relationship with Y . In a study design with continuous interviewing, such as Cycle 7 of the National Survey of Family Growth (Groves, Mosher, and Kirgis 2009), information gained about the correlates of survey participation and the survey variables from one replicate of the study can be used to tailor the design for later survey sample replicates.

Survey analysts often voice frustration over being encouraged to use survey weights in their analyses and find that they make little to no difference in point estimates. Our results offer insights into when the weights will or will not change point estimates and variance estimates of descriptive statistics. These results also indicate that a reduction in mean square error of survey estimates can be obtained with adjustment variables that are related to survey variables, even if they are not related to propensity and thus do not change the point estimates. With these results and the weighting models for a particular survey, the analyst will be able to determine whether their results do not change because of a lack of relationship between the predictors of nonresponse and their key variables of interest (and thus leading to an increase in variability of the estimates) or because of competing effects (and thus leading to a decrease in variance). Future research could examine the effects on analytic statistics such as regression coefficients.

Clearly more empirical study is needed to understand whether auxiliary variables that have long been shown to be predictors of propensity are associated with the survey variables (e.g., information on the survey process obtained from contact forms or call records). It may be the case that the long research focus on developing propensity predictors actually yielded auxiliary variables that are only weakly correlated with the important survey variables in most surveys (Kreuter and Kohler 2009). That is, in many real-life surveys, there may be unobserved correlates of both survey participation and the important survey variables. To the extent that this is true, extended efforts must be devoted to development of auxiliary predictors that predict both survey participation and the survey variables of interest in the same way.

Acknowledgments – We thank Jill Dever, Ulrich Kohler, Michael Lemay, Austin Nichols, Elizabeth Stuart, and Richard Valliant for stimulating discussions about path diagrams and simulations. We also thank Mick Couper, Stephanie Eckman, Carolina Casas-Cordero, and Juliana Werneburg for comments on previous drafts, and three anonymous reviewers for *Sociological Methods & Research* for suggestions that improved the article.

Notes

1. The U.S. Census Bureau, for example, now uses an automated system for collecting contact histories for their computer-assisted personal interviews (Bates, Dahlhamer, and Singer 2006).

2. Although the path from P to Y appears that P influences the values of Y , the path diagram is drawn to reflect that P influences the observation of Y , rather than the value of Y . Under a case of nonignorable nonresponse, Y would influence P . We assume that not to be the case here.
3. The relationship between age and survey participation may be more complicated than specified in this example. Older individuals tend to be more likely to be contacted, but less likely to cooperate in many surveys, than younger individuals (Groves and Couper 1998). However, this relationship varies across surveys.
4. When Z_1 and Z_2 are uncorrelated, standardized betas are equivalent to a and e in the path diagram.
5. Since the γ s are linear on the logistic scale, the γ s are not simply unstandardized versions of b and d .
6. An alternative evaluation measure is confidence interval coverage. If the adjustment is unnecessary (that is, the unadjusted mean is unbiased) and the variance of the estimate is increased, then adequate confidence interval coverage will be observed with each of the adjustment methods. Yet the width of the confidence interval will be unnecessarily increased. Thus, we use root mean square error (RMSE) as our evaluation criterion rather than confidence interval coverage.
7. The results are identical when we vary the strength of the other auxiliary variable.
8. The graphs are provided in the Appendix.

References

- Abraham, K. G., A. Maitland, and S. M. Bianchi. 2006. "Nonresponse in the American Time Use Survey – Who Is Missing From the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70(5):676-703.
- Bates, N., J. Dahlhamer, and E. Singer. 2006. "Privacy Concerns, Too Busy, or Just Not Interested? Exploring the Relationship Between Doorstep Concerns and Nonresponse Bias." Presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, NE.
- Groves, R. M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70(5):646-75.
- Groves, R., J. Wagner, and E. Peytcheva. 2007. "Use of Interviewer Judgments About Attributes of Selected Respondents in Post-survey Adjustment for Unit Nonresponse: An Illustration With the National Survey of Family Growth." *Proceedings of the Survey Research Methods Section, ASA*. Accessed February 28, 2011, at <http://www.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000782.pdf>
- Groves, R., and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley.
- Groves, R., and S. Heeringa. 2006. "Responsive Design for Household Surveys:

- Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society A* 169:439-57.
- Groves, R., W. Mosher, J. Lepkowski, and N. Kirgis. 2009. "Planning and Development of the Continuous National Survey of Family Growth." *Vital Health Statistic Series 1*(48):1-64.
- Kalton, G., and I. Flores-Cervantes. 2003. "Weighting Methods." *Journal of Official Statistics* 19:81-97.
- Kalton, G., and D. Maligalig. 1991. "A Comparison of Methods of Weighting Adjustment for Nonresponse." Pp. 409-28 in *Proceedings of the 1991 Annual Research Conference*. Washington, DC: U.S. Bureau of the Census.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley.
- Kominski, R., and T. Sutterlin. 1993. *Education: The Ticket to Higher Earnings*. Technical Report Statistical Brief 93-7. Washington, DC: U.S. Department of Commerce, Bureau of the Census; http://www.census.gov/apspd/www/statbrief/sb93_7.pdf
- Kreuter, F., and U. Kohler. 2009. "Analyzing Contact Sequences in Call Record Data. Potential and Limitations of Sequence Indicators for Nonresponse Adjustments in the European Social Survey." *Journal of Official Statistics* 25:203-26.
- Kreuter, F., M. Lemay, and C. Casas-Cordero. 2007. "Using Proxy Measures of Survey Outcomes in Post-survey Adjustments: Examples From the European Social Survey (ESS)." *Proceedings of the Survey Research Methods Section, ASA*. Accessed February 28, 2011, at <http://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000527.pdf>
- Kreuter, F., K. Olson, J. Wagner, T. Yan, T. Ezzati-Rice, C. Casas-Cordero, M. Lemay, A. Peytchev, R. Groves, and T. Raghunathan. 2010. "Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples From Multiple Surveys." *Journal of the Royal Statistical Society Series A* 173(2):389-407.
- Lepkowski, J., W. Mosher, K. Davis, R. Groves, J. van Hoewyk, and J. Willem. 2006. "National Survey of Family Growth, Cycle 6: Sample Design, Weighting, Imputation, and Variance Estimation." *Vital Health Statistic Series 2*(142):1-82.
- Lessler, J. T., and W. D. Kalsbeek. 1992. *Nonsampling Errors in Surveys*. New York: John Wiley.
- Little, R. 1986. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review* 54:139-57.
- Little, R. J., and S. Vartivarian. 2003. "On Weighting the Rates in Non-response Weights." *Statistics in Medicine* 22(9):1589-599.
- Little, R., and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31:161-68.
- Loehlin, J. 1998. *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Olson, K. 2007. "An Investigation of the Nonresponse-measurement Error Nexus." PhD dissertation, University of Michigan.

- Peytchev, A., and K. Olson. 2007. "Using Interviewer Observations to Improve Nonresponse Adjustments: NES 2004." *Proceedings of the Survey Research Methods Section, ASA*. Accessed February 28, 2011, at <http://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000695.pdf>
- Rosenbaum, P., and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.
- Wun, L.-M., and T. Ezzati-Rice. 2007. "Assessment of the Impact of Health Variables on Nonresponse Adjustment in the Medical Expenditure Panel Survey (MEPS)." *Proceedings of the Survey Research Methods Section, ASA*. Accessed February 28, 2011, at <http://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000336.pdf>
- Yan, T., and T. Raghunathan. 2007. "Using Proxy Measures of the Survey Variables in Post-survey Adjustments in a Transportation Survey." *Proceedings of the Survey Research Methods Section, ASA*. Accessed February 28, 2011, at <http://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000692.pdf>

* * *

The Authors

Frauke Kreuter is associate professor in the Joint Program in Survey Methodology at the University of Maryland and a research director at the Institute for Employment Research, Nuremberg, Germany.

Kristen Olson is an assistant professor in the Department of Sociology and the Program in Survey Methodology at the University of Nebraska-Lincoln.

Appendix

Recoding issue

Assume the following system of equations:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\text{logit}(\text{Pr}(r = 1)) = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$$

Now, assume that there is a negative relationship between X_1 and Y

$$Y = \beta_0 - \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\text{logit}(\text{Pr}(r = 1)) = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$$

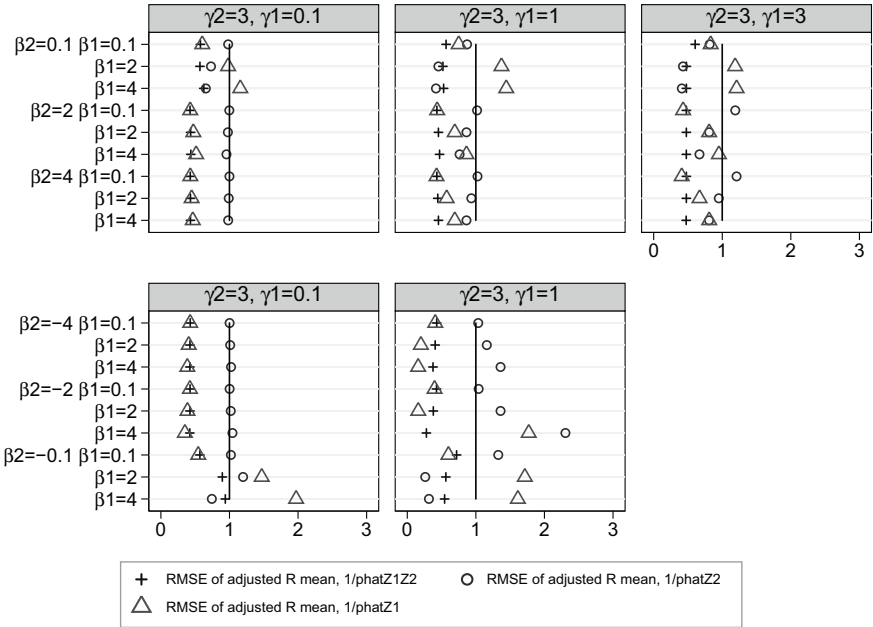
The analyst recodes X_1 so that the relationship between X_1 and Y is now positive, where the * indicates a recoded X_1 .

$$Y = \beta_0 - \beta_1(-X_1^*) + \beta_2 X_2 + \varepsilon = \beta_0 + \beta_1 X_1^* + \beta_2 X_2 + \varepsilon$$

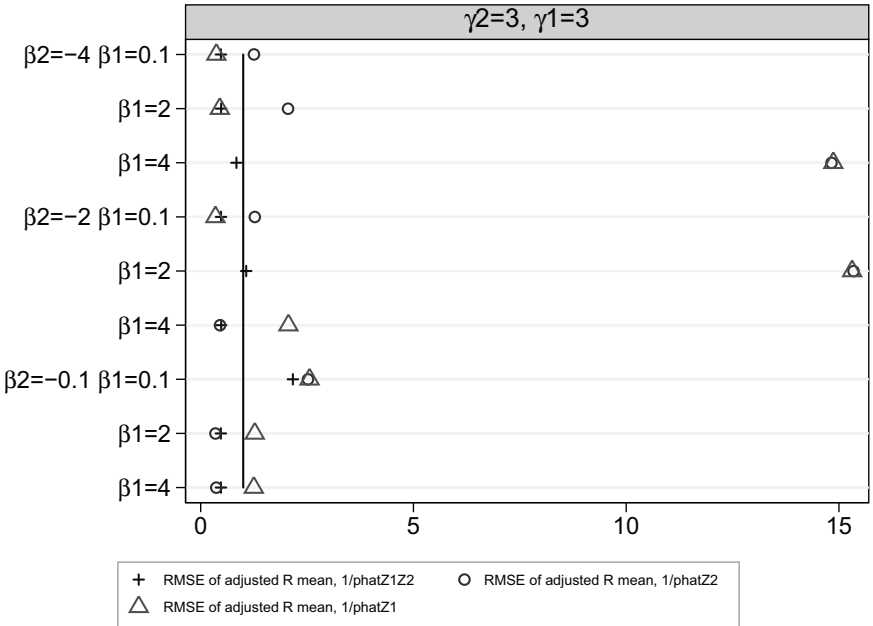
$$\text{logit}(\text{Pr}(r = 1)) = \gamma_0 - \gamma_1 X_1^* + \gamma_2 X_2$$

Note that although this reversed the relationship between X_1 and Y , the relationship between X_1 and $\text{Pr}(r = 1)$ also reversed. That is, recoding for one of the structural relationships also affects the other structural relationships.

Appendix Figure 2



Appendix Figure 3



Appendix Figure 4

