# Supplement to Multiple Imputation for Nonresponse in Surveys Using Design Weights and Auxiliary Margins

Kewei Xu* Jerome P. Reiter†

## 1 Introduction

This supplement includes tabular results from the simulations presented in the main text. It also provides more details on the MDAM–sys algorithm for two binary variables used in the simulation studies.

## 2 Tabular Results from Section 4 in Main Text

Results from the simulations in the main text in tabular form are presented in Figure 2.1 and Figure 2.2.

## 3 More Details About MDAM–sys Algorithm

To illustrate the MDAM–sys algorithm, we present the computations in Section 3.2.2 of the main text for binary $X_2$ and $X_1$. We use these in the simulations of Section 4 in the main text. First, we compute $\Pr(X_2 = 1|X_1 = 1, U = 0)$ and $\Pr(X_2 = 1|X_1 = 0, U = 0)$ using

*Department of Statistical Science, Box 90251, Duke University, Durham, NC 27708-0251
†Department of Statistical Science, Box 90251, Duke University, Durham, NC 27708-0251

a survey-weighted, ratio estimator like the one in (3.4) of the main text. To ease notation, define

$$\hat{\beta}_{1,obs} = \log\left[\frac{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=1, x^*_{i1}=1, U=0)}{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=0, x^*_{i1}=1, U=0)}\right] - \log\left[\frac{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=1, x^*_{i1}=0, U=0)}{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=0, x^*_{i1}=0, U=0)}\right] \tag{3.1}$$

$$\hat{\beta}_{0,obs} = \log\left[\frac{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=1, x^*_{i1}=0, U=0)}{\sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=0, x^*_{i1}=0, U=0)}\right]. \tag{3.2}$$

Define the unobserved quantities $\hat{\beta}_{1,mis}$ and $\hat{\beta}_{0,mis}$ for the unit nonrespondents analogously, replacing $U_i = 0$ with $U_i = 1$ in (3.1) and (3.2). We set the log odds ratios equal for the unit respondents and nonrespondents, that is, we set $\hat{\beta}_{1,mis} = \hat{\beta}_{1,obs}$. We also set

$$\frac{\sum_{i\in\mathcal{S}} w_i I(x^*_{i1}=1, U_i=1)}{1+\exp(-(\hat{\beta}_{0,mis}+\hat{\beta}_{1,mis}))} + \frac{\sum_{i\in\mathcal{S}} w_i I(x^*_{i1}=0, U_i=1)}{1+\exp(-\hat{\beta}_{0,mis})} = \hat{T}_2 - \sum_{i\in\mathcal{S}} w_i I(x^*_{i2}=1, U_i=0).$$

To determine $\hat{\beta}_{0,mis}$, we numerically solve the system of nonlinear equations using the modified Powell method as implemented in the R package "nleqslv" (Hasselman, 2018). The imputation probabilities are then

$$\begin{aligned}
\tilde{p}_{211} &= \frac{1}{1+\exp[-(\hat{\beta}_{0,mis}+\hat{\beta}_{1,mis})]} \\
\tilde{p}_{210} &= \frac{1}{1+\exp(-\hat{\beta}_{0,mis})}.
\end{aligned}$$

# References

Hasselman, B. (2018). nleqslv: solve systems of nonlinear equations. R package version 3.2.

| | Estimate | | | | CI Coverage | | | Variance | | | Average Estimated Variance | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | Truth | Premiss | MDAM adj | IH | Premiss | MDAM adj | IH | Premiss | MDAM adj | IH | Premiss | MDAM adj | IH |
| $X_1 = 0 \mid X_2 = 0$ | 0.578 | 0.578 | 0.576 | 0.605 | 95.6 | 99.8 | 68.6 | 1.4 | 1.5 | 1.6 | 1.3 | 3.7 | 2.8 |
| $X_1 = 0 \mid X_2 = 1$ | 0.559 | 0.559 | 0.562 | 0.604 | 95.6 | 99.6 | 26.6 | 1.4 | 2.7 | 2 | 1.4 | 5 | 3.2 |
| $X_2 = 0 \mid X_1 = 0$ | 0.518 | 0.517 | 0.515 | 0.539 | 94.6 | 99.8 | 83.4 | 1.5 | 1.3 | 1.7 | 1.3 | 3.5 | 2.9 |
| $X_2 = 0 \mid X_1 = 1$ | 0.498 | 0.498 | 0.501 | 0.538 | 95 | 99 | 50.8 | 1.6 | 3.6 | 2.4 | 1.5 | 6.2 | 3.9 |
| $X_4 = 0 \mid X_3 = 0$ | 0.825 | 0.826 | 0.836 | 0.83 | 95.8 | 99.2 | 99.8 | 6.7 | 7.2 | 7.5 | 6.9 | 13 | 13.7 |
| $X_4 = 0 \mid X_3 = 1$ | 0.766 | 0.765 | 0.771 | 0.76 | 94.8 | 98 | 97.2 | 0.6 | 0.7 | 0.7 | 0.6 | 1.4 | 1.3 |
| $X_3 = 0 \mid X_4 = 0$ | 0.075 | 0.075 | 0.077 | 0.078 | 96.6 | 98.8 | 97.2 | 0.3 | 0.4 | 0.4 | 0.3 | 0.5 | 0.6 |
| $X_3 = 0 \mid X_4 = 1$ | 0.053 | 0.053 | 0.053 | 0.053 | 96.4 | 98.8 | 99 | 0.8 | 0.9 | 0.9 | 0.7 | 1.6 | 1.5 |
| $X_1 = 0, X_2 = 0$ | 0.294 | 0.294 | 0.293 | 0.326 | 95.4 | 99.2 | 21.4 | 0.7 | 0.9 | 0.8 | 0.6 | 1.6 | 1.4 |
| $X_1 = 1, X_2 = 0$ | 0.275 | 0.275 | 0.276 | 0.279 | 94.8 | 100 | 97.2 | 0.6 | 0.4 | 0.8 | 0.6 | 1.8 | 1.3 |
| $X_1 = 0, X_2 = 1$ | 0.215 | 0.215 | 0.216 | 0.213 | 95.4 | 100 | 98.4 | 0.5 | 0.4 | 0.6 | 0.5 | 1.6 | 1 |
| $X_1 = 1, X_2 = 1$ | 0.216 | 0.216 | 0.215 | 0.183 | 94.8 | 99 | 1.6 | 0.5 | 1.3 | 0.6 | 0.5 | 1.8 | 0.9 |
| $X_3 = 0 \mid X_1 = 0, X_2 = 0$ | 0.064 | 0.064 | 0.066 | 0.066 | 94.2 | 98.6 | 98.8 | 0.7 | 0.8 | 0.8 | 0.7 | 1.3 | 1.2 |
| $X_3 = 0 \mid X_1 = 0, X_2 = 1$ | 0.108 | 0.107 | 0.111 | 0.111 | 95.4 | 99.4 | 99.4 | 1.1 | 1.4 | 1.4 | 1.1 | 2.4 | 2.3 |
| $X_3 = 0 \mid X_1 = 1, X_2 = 0$ | 0.049 | 0.049 | 0.044 | 0.044 | 96.6 | 97.8 | 98.6 | 0.7 | 0.7 | 0.7 | 0.7 | 1.2 | 1.2 |
| $X_3 = 0 \mid X_1 = 1, X_2 = 1$ | 0.052 | 0.052 | 0.058 | 0.058 | 95.8 | 98.6 | 97.8 | 0.8 | 1 | 1 | 0.8 | 1.8 | 1.9 |
| $X_4 = 0 \mid X_1 = 0, X_2 = 0$ | 0.638 | 0.637 | 0.646 | 0.648 | 96.8 | 99.6 | 99.8 | 2.1 | 3 | 2.9 | 2.4 | 6.5 | 6.4 |
| $X_4 = 0 \mid X_1 = 0, X_2 = 1$ | 0.601 | 0.599 | 0.619 | 0.619 | 93.2 | 97.2 | 97.4 | 3.1 | 3.8 | 3.6 | 2.7 | 7.6 | 7.5 |
| $X_4 = 0 \mid X_1 = 1, X_2 = 0$ | 0.963 | 0.964 | 0.964 | 0.964 | 96.6 | 99.6 | 99.6 | 0.5 | 0.5 | 0.5 | 0.5 | 1.2 | 1.1 |
| $X_4 = 0 \mid X_1 = 1, X_2 = 1$ | 0.973 | 0.973 | 0.965 | 0.965 | 96 | 97.6 | 98 | 0.4 | 0.5 | 0.5 | 0.4 | 1.3 | 1.4 |
| $T_{X_1}$ | 1671915 | 1672653 | 1671777 | 1570759 | 89.8 | 100 | 41.2 | 19 | 1.9 | 20.1 | 12.6 | 33.4 | 21.2 |
| $T_{X_2}$ | 1468008 | 1467818 | 1467904 | 1346825 | 93 | 100 | 14.4 | 13.2 | 1.4 | 14.3 | 10.1 | 28.7 | 16.4 |
| $T_{X_3}$ | 3166289 | 3166630 | 3160272 | 3158568 | 83.4 | 85.2 | 84.2 | 33.6 | 34.1 | 34.2 | 16.3 | 18.5 | 18.5 |
| $T_{X_4}$ | 783369 | 784835 | 762556 | 799664 | 95.2 | 95.6 | 97.4 | 8.8 | 10.9 | 9.7 | 7.7 | 15 | 14.1 |
| $T_{X_5}$ | 37338600 | 37205563 | 36288521 | 35344643 | 95.6 | 95.2 | 83.4 | 2.8 | 2.9 | 3.1 | 2.6 | 3.9 | 3.5 |
| $T_{X_6}$ | 57865307 | 57876603 | 58852221 | 57404909 | 91.4 | 96.6 | 93.6 | 1.9 | 1.7 | 2.2 | 1.4 | 2.5 | 2.2 |

Figure 2.1: Results from the simulation in the main text for IH and MDAM–adj. Columns headed by "Premiss" include results for the survey-weighted analysis of the Poisson samples before introduction of any missing data. The column headed by "Truth" includes the quantities computed from all $N$ individuals in the constructed population. Variances for the probabilities are multiplied by $10^4$. Variances for each of $(T_{X_1}, T_{X_2}, T_{X_3}, T_{X_4})$ are multiplied by $10^{-8}$. Variances for $T_{X_5}$ and $T_{X_6}$ are multiplied by $10^{-12}$.

| | Estimate | | | | CI Coverage | | | Variance | | | Average Estimated Variance | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | Truth | MDAM adj | MDAM sys | MDAM yr | MDAM adj | MDAM sys | MDAM yr | MDAM adj | MDAM sys | MDAM yr | MDAM adj | MDAM sys | MDAM y |
| $X_1 = 0 \mid X_2 = 0$ | 0.578 | 0.576 | 0.577 | 0.576 | 99.8 | 100 | 99.8 | 1.5 | 1.4 | 1.1 | 3.7 | 3.7 | 3.1 |
| $X_1 = 0 \mid X_2 = 1$ | 0.559 | 0.562 | 0.561 | 0.562 | 99.6 | 99.8 | 100 | 2.7 | 2.7 | 1.3 | 5 | 4.9 | 4.2 |
| $X_2 = 0 \mid X_1 = 0$ | 0.518 | 0.515 | 0.516 | 0.515 | 99.8 | 100 | 100 | 1.3 | 1.4 | 0.9 | 3.5 | 3.5 | 2.9 |
| $X_2 = 0 \mid X_1 = 1$ | 0.498 | 0.501 | 0.5 | 0.502 | 99 | 99.6 | 99.8 | 3.6 | 3.4 | 1.7 | 6.2 | 6 | 4.8 |
| $X_4 = 0 \mid X_3 = 0$ | 0.825 | 0.836 | 0.836 | 0.836 | 99.2 | 99.2 | 99 | 7.2 | 7.2 | 7 | 13 | 13 | 11.9 |
| $X_4 = 0 \mid X_3 = 1$ | 0.766 | 0.771 | 0.772 | 0.771 | 98 | 98.2 | 97.8 | 0.7 | 0.6 | 0.6 | 1.4 | 1.4 | 1.2 |
| $X_3 = 0 \mid X_4 = 0$ | 0.075 | 0.077 | 0.077 | 0.078 | 98.8 | 98.8 | 97.6 | 0.4 | 0.3 | 0.3 | 0.5 | 0.6 | 0.5 |
| $X_3 = 0 \mid X_4 = 1$ | 0.053 | 0.053 | 0.053 | 0.053 | 98.8 | 99 | 98.8 | 0.9 | 0.9 | 0.9 | 1.6 | 1.6 | 1.4 |
| $X_1 = 0, X_2 = 0$ | 0.294 | 0.293 | 0.294 | 0.293 | 99.2 | 99.4 | 100 | 0.9 | 0.9 | 0.3 | 1.6 | 1.6 | 1.3 |
| $X_1 = 1, X_2 = 0$ | 0.275 | 0.276 | 0.275 | 0.276 | 100 | 100 | 100 | 0.4 | 0.4 | 0.3 | 1.8 | 1.8 | 1.5 |
| $X_1 = 0, X_2 = 1$ | 0.215 | 0.216 | 0.215 | 0.216 | 100 | 100 | 100 | 0.4 | 0.4 | 0.3 | 1.6 | 1.6 | 1.3 |
| $X_1 = 1, X_2 = 1$ | 0.216 | 0.215 | 0.216 | 0.215 | 99 | 99.2 | 100 | 1.3 | 1.3 | 0.4 | 1.8 | 1.8 | 1.5 |
| $X_3 = 0 \mid X_1 = 0, X_2 = 0$ | 0.064 | 0.066 | 0.066 | 0.066 | 98.6 | 99 | 98.6 | 0.8 | 0.8 | 0.8 | 1.3 | 1.3 | 1.2 |
| $X_3 = 0 \mid X_1 = 0, X_2 = 1$ | 0.108 | 0.111 | 0.111 | 0.111 | 99.4 | 99.6 | 99 | 1.4 | 1.4 | 1.4 | 2.4 | 2.3 | 2.1 |
| $X_3 = 0 \mid X_1 = 1, X_2 = 0$ | 0.049 | 0.044 | 0.044 | 0.044 | 97.8 | 98.2 | 97 | 0.7 | 0.7 | 0.7 | 1.2 | 1.2 | 1 |
| $X_3 = 0 \mid X_1 = 1, X_2 = 1$ | 0.052 | 0.058 | 0.058 | 0.058 | 98.6 | 98.6 | 96.6 | 1 | 0.9 | 0.9 | 1.8 | 1.8 | 1.5 |
| $X_4 = 0 \mid X_1 = 0, X_2 = 0$ | 0.638 | 0.646 | 0.646 | 0.646 | 99.6 | 99.6 | 99.4 | 3 | 3 | 2.9 | 6.5 | 6.5 | 6.2 |
| $X_4 = 0 \mid X_1 = 0, X_2 = 1$ | 0.601 | 0.619 | 0.619 | 0.619 | 97.2 | 98 | 97.2 | 3.8 | 3.7 | 3.7 | 7.6 | 7.6 | 7 |
| $X_4 = 0 \mid X_1 = 1, X_2 = 0$ | 0.963 | 0.964 | 0.964 | 0.964 | 99.6 | 99.6 | 99.2 | 0.5 | 0.5 | 0.5 | 1.2 | 1.1 | 1 |
| $X_4 = 0 \mid X_1 = 1, X_2 = 1$ | 0.973 | 0.965 | 0.965 | 0.965 | 97.6 | 98.2 | 96.2 | 0.5 | 0.5 | 0.5 | 1.3 | 1.3 | 1.1 |
| $T_{X_1}$ | 1671915 | 1671777 | 1671585 | 1671055 | 100 | 100 | 100 | 1.9 | 1.7 | 1.1 | 33.4 | 33.5 | 24.7 |
| $T_{X_2}$ | 1468008 | 1467904 | 1468056 | 1467455 | 100 | 100 | 100 | 1.4 | 1.4 | 1.2 | 28.7 | 28.7 | 23.1 |
| $T_{X_3}$ | 3166289 | 3160272 | 3160289 | 3160605 | 85.2 | 83.8 | 100 | 34.1 | 34.1 | 3 | 18.5 | 18.5 | 14.5 |
| $T_{X_4}$ | 783369 | 762556 | 762215 | 762672 | 95.6 | 95.8 | 98.4 | 10.9 | 10.7 | 6 | 15 | 15.1 | 13.5 |
| $T_{X_5}$ | 37338600 | 36288521 | 36307042 | 36290757 | 95.2 | 94.8 | 92.8 | 2.9 | 3 | 2.7 | 3.9 | 3.9 | 3.2 |
| $T_{X_6}$ | 57865307 | 58852221 | 58876057 | 58874748 | 96.6 | 96.6 | 96.6 | 1.7 | 1.6 | 1.1 | 2.5 | 2.5 | 2.1 |

Figure 2.2: Results from the simulation in the main text for MDAM–adj, MDAM–sys, and MDAM–yr. The column headed by "Truth" includes the quantities computed from all $N$ individuals in the constructed population. Variances for the probabilities are multiplied by $10^4$. Variances for each of $(T_{X_1}, T_{X_2}, T_{X_3}, T_{X_4})$ are multiplied by $10^{-8}$. Variances for $T_{X_5}$ and $T_{X_6}$ are multiplied by $10^{-12}$.