Nonresponse Adjustment of Survey Estimates Based on
Auxiliary Variables Subject to Error

Brady T. West
University of Michigan, Ann Arbor, MI, USA

Roderick J.A. Little
University of Michigan, Ann Arbor, MI, USA

Summary. Auxiliary variables associated with both key survey variables and response propensity are important for post-survey nonresponse adjustments, but rare. Interviewer observations on sample units and linked auxiliary variables from commercially available household databases are promising candidates, but these variables are prone to error. The assumption of missing at random (MAR) that underlies standard weighting or imputation adjustments is thus violated when missingness depends on the true values of these variables, leading to biased survey estimates. This article applies pattern-mixture model estimators to this problem, analyzing data from a survey in Germany (PASS) that links commercial data to a national sample.

Keywords: Auxiliary Variables; Measurement Error; Non-ignorable Missing Data; Nonresponse Adjustment of Survey Estimates; Pattern-Mixture Models; PASS Survey
1. Introduction

We consider nonresponse adjustment of survey estimates based on an auxiliary variable fully observed for a sample of $n$ units from some population. Effective auxiliary variables for nonresponse adjustment should be highly predictive of both key survey variables and the response propensity (Beaumont, 2005; Bethlehem, 2002; Groves, 2006; Lessler and Kalsbeek, 1992; Little and Vartivarian, 2005). In an effort to collect data on auxiliary variables with these properties, some survey programs have requested that interviewers record judgments about selected features of all sample units (Kreuter et al., 2010; West, 2012), but these interviewer observations can be prone to measurement error (Campanelli et al., 1997; Groves et al., 2007; McCulloch et al., 2010; Pickering et al., 2003; Tipping and Sinibaldi, 2010; West, 2012). Some survey programs have also considered linking proxies of key survey variables available in commercial databases to sampling frames, but these variables may also be prone to error (DiSogra et al., 2010). Using these error-prone auxiliary variables in nonresponse adjustments can be problematic. Weighting class or regression nonresponse adjustments based on error-prone auxiliary variables result in bias when missingness depends on the true underlying value (Lessler and Kalsbeek, 1992, p. 190; West, 2012). This article proposes methods for correcting for this bias, and applies them to survey data collected from a national sample in Germany.

We consider data as in Figure 1, where $X_1$ is an auxiliary variable measured with error for all $n$ sampled individuals, $X_2$ is the underlying true value of $X_1$, recorded for each of $r$ survey respondents, and $X_3$ is a survey variable of substantive interest, also measured for the $r$ respondents only. The objective is to make inferences about means of
the variables $X_2$ and $X_3$, using the auxiliary variable $X_1$ to adjust for nonresponse. The auxiliary variable $X_1$ may also represent a proxy variable related to key survey variables and response propensity that combines information on multiple auxiliary covariates, possibly through principal components analysis or linear predictors (e.g., Andridge and Little, 2009, 2011).

![Diagram](image)

**Figure 1:** Missing data pattern under study.

Given the necessary resources, surveys can link error-prone auxiliary proxy variables from varying sources (e.g., interviewer observations, commercially available household databases) to full samples, introducing the scenario illustrated in Figure 1. In this article, we focus on the German Labor Market and Social Security (PASS) survey, a panel study that collects annual labor market, household income, and unemployment benefit receipt data from a nationally representative sample of 12,000 households from the German population. PASS survey managers link auxiliary socio-economic variables from a commercial data source to the PASS sampling frame to assist with stratified
sampling and estimation tasks. In this article, we use these linked variables to apply alternative nonresponse adjustments to respondent data from the first wave of the PASS survey (2006). We contrast the performance of more popular adjustments assuming ignorable, missing at random (MAR) mechanisms with a proposed adjustment method for the case when missingness depends on the true values of the auxiliary proxy that are only measured for survey respondents.

Our proposed method, presented in Section 2, is based on a pattern-mixture model (PMM; Little, 1994; Little and Rubin, 2002, Section 15.5). PMMs stratify the sample cases based on patterns of missing data and formulate distinct models for the variables within each stratum. Unidentified parameters are identified by exploiting parameter restrictions based on assumptions about the missing-data mechanism. Little (1994) derived maximum likelihood (ML) and Bayesian estimators of means and covariances for incomplete data assuming a bivariate normal PMM, under ignorable and non-ignorable mechanisms. Little and Wang (1996) extended this work to multivariate incomplete data with fully observed covariates. More recently, Shardell et al. (2010) applied PMMs to the analysis of normal outcome data provided by proxy respondents in surveys, which may be subject to measurement error, and Baskin et al. (2011) used proxy pattern-mixture analysis, or PPMA (Andridge and Little, 2011), to estimate non-response bias in means of health expenditure variables in the Medical Expenditure Panel Survey (MEPS). In the present application, we develop a trivariate normal PMM suitable for the survey context described by Figure 1.

Previous methods of nonresponse adjustment with error-prone auxiliary variables have assumed that the missing data are MAR, meaning that missingness depends only on
the fully observed auxiliary variables (Rubin, 1976). We develop PMM estimators for the
case where missingness (or a failure to respond to the survey) is assumed to depend on
the true auxiliary variable $X_2$, but not the auxiliary proxy variable $X_1$, after conditioning
on $X_2$. Simulations comparing the PMM estimators with more common estimators are
described in Section 3. In Section 4, we generalize our proposed method to the case of
additional auxiliary variables measured without error. Section 5 presents applications of
our methods to the PASS survey data, and compares our PMM estimates with weighting
class and sequential regression imputation (Raghunathan et al., 2001) estimates that
assume MAR mechanisms. Section 6 summarizes our work and discusses further
extensions. R code implementing the proposed estimators is available upon request from
the authors (email: bwest@umich.edu).

2. Pattern-Mixture Model: Estimation and Inference

2.1. Pattern-Mixture Model (PMM) Estimates

For sample unit $i$, let $m_i$ be a missing data indicator, equal to 0 if a unit responds
to the survey and 1 otherwise. Unit nonrespondents have missing values for $X_2$ and $X_3$.
For the missing data pattern $m_i = m$, we assume

\[
\begin{pmatrix}
  x_{1i} \\
  x_{12i} \\
  x_{13i}
\end{pmatrix}
\sim N_3\left(
\begin{pmatrix}
  \mu_1^{(m)} \\
  \mu_2^{(m)} \\
  \mu_3^{(m)}
\end{pmatrix},
\begin{pmatrix}
  \sigma_{11}^{(m)} & \sigma_{12}^{(m)} & \sigma_{13}^{(m)} \\
  \sigma_{12}^{(m)} & \sigma_{22}^{(m)} & \sigma_{23}^{(m)} \\
  \sigma_{13}^{(m)} & \sigma_{23}^{(m)} & \sigma_{33}^{(m)}
\end{pmatrix}\right)
\equiv N_3\left(\mu^{(m)}, \Sigma^{(m)}\right),
\]  

(1)
a trivariate normal distribution with nine parameters. The marginal distribution of $m_i$
is $m_i \sim Bernoulli(\pi_i)$. There are $2 \times 9 + 1 = 19$ model parameters in total across both
patterns.
The following 12 parameters are clearly identified from the observed data in Figure 1: \( \theta_{id} = (\pi_1, \mu^{(0)}_1, \sigma^{(0)}_{11}, \mu^{(1)}_1, \sigma^{(1)}_{11}, \mu^{(0)}_2, \sigma^{(0)}_{12}, \mu^{(0)}_3, \sigma^{(0)}_{13}, \sigma^{(0)}_{22}, \sigma^{(0)}_{23}, \sigma^{(0)}_{33}). \)

The following 7 parameters are not identified: \( \theta_{nu} = (\mu^{(1)}_2, \mu^{(1)}_3, \sigma^{(1)}_{12}, \sigma^{(1)}_{13}, \sigma^{(1)}_{22}, \sigma^{(1)}_{23}, \sigma^{(1)}_{33}). \)

Let \( \beta^{(m)}_{jk} \) denote the slope coefficient for variable \( k \) in the linear regression of variable \( j \) on variable \( k \) for pattern \( m \), and let \( \beta^{(m)}_{j0k} \) denote the intercept coefficient in this regression.

Also, let \( \sigma^{(m)}_{ji-k} \) denote the residual variance in the regression of variable \( j \) on variable \( k \) for pattern \( m \), and let \( \sigma^{(m)}_{jl-k} \) denote the residual covariance of variable \( j \) and variable \( l \) given variable \( k \) for pattern \( m \). The assumption that missingness of \( X_2 \) and \( X_3 \) depends on \( X_2 \) (the “true” values of the auxiliary variable \( X_1 \), measured in the survey) implies that the distribution of \( X_1 \) and \( X_3 \) given \( X_2 \) is the same for complete and incomplete cases, yielding seven parameter restrictions:

\[
\begin{align*}
\beta^{(0)}_{10-2} &= \beta^{(1)}_{10-2}; \\
\beta^{(0)}_{12-2} &= \beta^{(1)}_{12-2}; \\
\beta^{(0)}_{30-2} &= \beta^{(1)}_{30-2}; \\
\beta^{(0)}_{32-2} &= \beta^{(1)}_{32-2};
\end{align*}
\]

\[
\sigma^{(1)}_{11-2} = \sigma^{(0)}_{11-2}; \\
\sigma^{(1)}_{33-2} = \sigma^{(0)}_{33-2}; \\
\sigma^{(1)}_{13-2} = \sigma^{(0)}_{13-2}.
\]

With seven restrictions and seven unidentified parameters, the model is just-identified, and ML estimates are straightforward extensions of those given in Little (1994). Specifically, we transform \( \theta_{id} \) to the alternative parameterization

\[
\phi_{id} = (\pi_1, \mu^{(0)}_1, \mu^{(1)}_1, \sigma^{(1)}_{11}, \beta_{10-2}, \beta_{12-2}, \beta_{30-2}, \beta_{32-2}, \sigma_{11-2}, \sigma_{13-2}, \sigma_{33-2}),
\]

where the parameter restrictions imply that the last seven parameters are the same for complete and incomplete cases. Define the corresponding sample quantities

\[
\hat{\pi}_i = (n - r) / n , \text{ or the sample proportion of nonrespondents; } \hat{\mu}^{(m)}_i \text{ and } \hat{\sigma}^{(m)}_{i1} , \text{ or the sample mean and variance of } X_1 \text{ for pattern } m \text{ (the variances have denominators } r \text{ and } n - r \text{ respectively, that is, are not corrected for degrees of freedom); and}
\]
\((\hat{\beta}_{102}, \hat{\beta}_{122}, \hat{\beta}_{302}, \hat{\beta}_{322}, \hat{\sigma}_{112}, \hat{\sigma}_{132}, \hat{\sigma}_{332})\), or the least squares estimates of the parameters of the regression of \(X_1\) and \(X_3\) on \(X_2\), for the complete cases \((m = 0)\). These sample quantities are ML estimates of the components of \(\phi_{id}\) provided that \(\hat{\sigma}_{11}^{(i)} > \hat{\sigma}_{11}^{(i)}\), since \(\hat{\sigma}_{11}^{(i)}\) and \(\hat{\sigma}_{11}^{(i)}\) estimate parameters that are subject to the constraint \(\sigma_{11}^{(i)} > \sigma_{11}^{(i)}\); otherwise \(\hat{\sigma}_{11}^{(i)}\) is set to equal \(\hat{\sigma}_{11}^{(i)}\). ML estimates of the components of \(\theta_{id}\) are also the corresponding least squares estimates.

We obtain ML estimates of the remaining non-identified parameters \(\theta_{nid}\) by expressing them as functions of \(\phi_{id}\), and substituting the ML estimates \(\hat{\phi}_{id}\). For example, for \(\mu_{2}^{(i)}\) we have:

\[
\mu_{2}^{(i)} = \beta_{102} + \beta_{122} \mu_{2}^{(i)} \Rightarrow \mu_{2}^{(i)} = \frac{\mu_{2}^{(i)} - \beta_{102}}{\beta_{122}}
\]

\[
\Rightarrow \hat{\mu}_{2}^{(i)} = \frac{\hat{\mu}_{2}^{(i)} - \hat{\beta}_{102}}{\hat{\beta}_{122}} = \hat{\mu}_{2}^{(0)} + \frac{\hat{\mu}_{1}^{(i)} - \hat{\mu}_{1}^{(0)}}{\hat{\beta}_{122}},
\]

where \(\hat{\mu}_{2}^{(0)}\) is the sample mean of \(X_2\) for the complete cases. ML estimates of the other six parameters in \(\theta_{nid}\) are defined in a similar manner, as follows:

\[
\hat{\mu}_{3}^{(i)} = \hat{\mu}_{3}^{(0)} + \hat{\beta}_{322} \frac{\hat{\mu}_{1}^{(i)} - \hat{\mu}_{1}^{(0)}}{\hat{\beta}_{122}}
\]

\[
\hat{\sigma}_{12}^{(i)} = \hat{\sigma}_{12}^{(0)} + \frac{\hat{\sigma}_{11}^{(i)} - \hat{\sigma}_{11}^{(0)}}{\hat{\beta}_{122}}
\]

\[
\hat{\sigma}_{13}^{(i)} = \hat{\sigma}_{13}^{(0)} + \hat{\beta}_{322} \frac{\hat{\sigma}_{11}^{(i)} - \hat{\sigma}_{11}^{(0)}}{\hat{\beta}_{122}}
\]

\[
\hat{\sigma}_{22}^{(i)} = \hat{\sigma}_{22}^{(0)} + \frac{\hat{\sigma}_{11}^{(i)} - \hat{\sigma}_{11}^{(0)}}{\hat{\beta}_{122}}
\]
\[ \hat{\sigma}_{23}^{(1)} = \hat{\sigma}_{23}^{(0)} + \hat{\beta}_{322} \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}}{\hat{\beta}_{122}^2} \]  

(7)

\[ \hat{\sigma}_{33}^{(1)} = \hat{\sigma}_{33}^{(0)} + \hat{\beta}_{322}^2 \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}}{\hat{\beta}_{122}^2} \]  

(8)

The ML estimates of the parameters of the marginal distribution of \( X \) are obtained by combining the parameter estimates of \( \theta_{ni} \) and \( \theta_{ni} \). For example, the ML estimate of the mean \( \mu_2 \) of \( X_2 \) is then (by simple algebra):

\[ \hat{\mu}_2 = \hat{\mu}_2^{(0)} + \hat{\pi}_1 \frac{\hat{\mu}_1^{(1)} - \hat{\mu}_1^{(0)}}{\hat{\beta}_{122}}, \]  

(9)

as in Little (1994). These ML estimators are unstable if the estimated regression coefficient \( \hat{\beta}_{122} \) is close to zero, as when \( X_1 \) has substantial measurement error and is consequently weakly correlated with the true variable \( X_2 \). Thus, the method requires a proxy variable that has a reasonably strong correlation with the true variable.

2.2. Bayesian Inference

Large-sample standard errors for the ML estimates derived above can be based on linearized variance estimators (e.g., Little, 1994). Confidence intervals based on ML estimates and these variance estimates have been shown in simulation studies to yield below nominal coverage, particularly when the sample size is small and the auxiliary variable is weakly associated with the outcome variable (Andridge and Little, 2011, p. 166). Better confidence interval coverage is obtained by a Bayesian approach, assuming noninformative prior distributions and simulating draws from the posterior distribution of
the parameters. We extend the Bayesian methods in Little (1994) to our trivariate normal model. We assume noninformative priors for the 12 identified parameters:

\[ \pi_1 \sim \text{Beta}(0.5, 0.5) \]
\[ p(\mu_1^{(0)}, \Sigma^{(0)}) \propto \left| \Sigma^{(0)} \right|^{-1} \]
\[ p(\mu_i^{(1)}, \sigma_{11}^{(1)}) \propto 1 / \sigma_{11}^{(1)} \]

Draws \( \phi_{id} \) from the posterior distribution of the identified parameters \( \phi_{id} \) are obtained as follows (we assume \( r > 3 \) and \( n - r > 1 \)):

1) \( \pi_{11}^{(d)} \sim \text{Beta}(n - r + 0.5, r + 0.5) \);

2) \( \sigma_{11}^{(0,d)} = r \hat{\sigma}_{11}^{(0)} / u_1^{(d)}, u_1^{(d)} \sim \chi_{r-1}^2 \);

3) \( \mu_i^{(0,d)} = \hat{\mu}_i^{(0)} + z_1^{(d)} \sigma_{11}^{(0,d)} / r, z_1^{(d)} \sim N(0,1) \);

4) \( \sigma_{11}^{(1,d)} = (n-r) \hat{\sigma}_{11}^{(1)} / u_2^{(d)}, u_2^{(d)} \sim \chi_{n-r-1}^2 \);

5) \( \mu_i^{(1,d)} = \hat{\mu}_i^{(1)} + z_2^{(d)} \sigma_{11}^{(1,d)} / (n-r), z_2^{(d)} \sim N(0,1) \);

6) \( \left( \begin{array}{c} \sigma_{11}^{(d)} \\ \sigma_{12}^{(d)} \\ \sigma_{21}^{(d)} \\ \sigma_{32}^{(d)} \\ \sigma_{23}^{(d)} \\ \sigma_{33}^{(d)} \end{array} \right) \sim \text{Inv-Wishart} \left( \left( \begin{array}{c} \hat{\sigma}_{11}^{12} \\ \hat{\sigma}_{12}^{13} \\ \hat{\sigma}_{22}^{13} \\ \hat{\sigma}_{23}^{13} \\ \hat{\sigma}_{32}^{13} \\ \hat{\sigma}_{33}^{13} \end{array} \right), r-2 \right) \);

7) \( \beta_{22}^{(d)} \sim N(\hat{\beta}_{22}^{(0)}, \sigma_{11}^{(0,d)} / (r \hat{\sigma}_{22}^{(0)})); \beta_{02}^{(d)} \sim N(\hat{\mu}_1^{(0)} - \hat{\mu}_2^{(d)}, \hat{\mu}_2^{(d)}, \sigma_{11}^{(d)}/r) \); and

8) \( \beta_{32}^{(d)} \sim N(\hat{\beta}_{32}^{(0)}, \sigma_{11}^{(0,d)} / (r \hat{\sigma}_{32}^{(0)})); \beta_{03}^{(d)} \sim N(\hat{\mu}_3^{(0)} - \hat{\beta}_{32}^{(d)}, \hat{\mu}_3^{(d)}, \sigma_{33}^{(d)}/r) \),

where Inv-Wishart \((S, d)\) denotes the inverse Wishart distribution with \( d \) degrees of freedom and scale matrix \( S \) (see Gelman et al., 2004, Appendix A).

To satisfy the constraint that \( \sigma_{11}^{(1)} > \sigma_{11}^{(2)} \), the draws in 4) and 6) must be such that \( \sigma_{11}^{(1,d)} > \sigma_{11}^{(2,d)} \) (Little, 1994). Draw samples that fail this condition are discarded and repeated. The drawn values from the sequence above then replace the ML estimates in Equations (2) to (9) to generate draws from the posterior distributions of the other parameters. Inferences
are based on a large sample (say, 1,000) of these draws. In particular, the mean of the
draws simulates the posterior mean, and the 2.5% and 97.5% percentiles of the simulated
draws simulate a 95% credible interval for the mean.

2.3. Multiple Imputation

A useful alternative inferential method is multiple imputation (MI; Little
and Rubin, 2002; Andridge and Little, 2011). Parameters of the model are drawn from their
posterior predictive distributions, as above. The missing values of \(X_2\) and \(X_3\) are then
drawn from their conditional distributions given these draws, namely

\[
\begin{align*}
    x_{2i}^{(d)} &\sim N(\beta_{20d}^{(1)} + \beta_{21d}^{(1)} x_{i1}, \sigma_{22d}^{(1)}) \quad \text{and} \\
    x_{3i}^{(d)} &\sim N(\beta_{30d}^{(1)} + \beta_{31d}^{(1)} x_{i1} + \beta_{32d}^{(1)} x_{2i}, \sigma_{33d}^{(1)}),
\end{align*}
\]

(10)

(11)

where the superscript \((d)\) denotes the \(d\)-th set of draws, and the parameters are drawn as
appropriate functions of the draws in Section 2.2. For example,

\[
\begin{align*}
    \beta_{21}^{(1)} &= \frac{\sigma_{11}^{(1)}}{\sigma_{12}^{(1)}} - \sigma_{11d}^{(1)} - \sigma_{11d}^{(1)} \\
    \beta_{21}^{(1)} &= \left( \frac{\sigma_{11}^{(1)} - \sigma_{11d}^{(1)}}{\beta_{12d}^{(1)} \sigma_{11d}^{(1)}} \right)
\end{align*}
\]

This procedure is repeated \(B\) times to create \(B\) complete data sets, which can then
be analyzed using MI combining rules (Rubin, 1987). The within-imputation components
of variance can readily incorporate complex sample design features like sample weights,
which otherwise need to be incorporated by modifying the basic PMM. We also note that
this method does not require draws \(\{\pi_1^{(d)}\}\), since the imputations are exclusively within
pattern \(m = 1\), and the MI analysis of the filled-in data sets does not need to condition on
pattern. This feature is useful when we develop extensions to include other auxiliary
variables in the imputation model (Section 4).
3. Simulation Studies

3.1. Methods Compared

We describe two sets of simulations to compare empirically the performance of the PMM estimates (using Bayesian methods for inference) with other common methods of compensating for unit nonresponse in surveys. Five approaches to estimation and inference for the means of the variables $X_2$ and $X_3$ were compared:

1) PMM estimates and 95% credible intervals for the means based on the Bayesian approach described in Section 2.2 (denoted by PMM).

2) PMM estimates based on the multiple imputation approach described in Section 2.3 (denoted by PMM-MI), with missing values of $X_2$ and $X_3$ are imputed multiple (5) times.

3) Standard multiple imputation (MI), assuming normal data and an ignorable missing data mechanism. Missing values of $X_2$ and $X_3$ are imputed multiple (5) times using an iterative conditional sequential regression imputation approach, as implemented in the \texttt{mi} package of R (Su et al., 2009). Multiple imputation combining rules described by Little and Rubin (2002) are used for estimates and standard errors of the two means, with degrees of freedom for the $t$ distribution computed using the methods for large samples in Barnard and Rubin (1999).

4) A “global” weighting (GW) approach. The complete cases are weighted by the inverses of the individual response propensities, estimated from a logistic regression of the response indicator $(1 - m_i)$ on $X_1$, and weighted estimates of
the means are computed. Taylor series linearization was used to compute estimates of the standard errors of these estimated means, and corresponding 95% confidence intervals for the means.

5) Complete-case (CC) analysis, where analysis is based only on cases with no missing values, with no adjustment of any form for nonresponse, and standard methods for simple random samples are used to compute estimates of means, standard errors, and 95% confidence intervals.

3.2. Simulated Data

We first simulate data from the PMM of Section 2, meaning that the PMM approaches are expected to out-perform the other approaches. Samples are generated from the following PMM:

\[
\begin{pmatrix}
  x_{11} \\
  x_{12} \\
  x_{13} \\
\end{pmatrix}
| m_i = m \sim N_3
\begin{pmatrix}
  \mu_{1}^{(m)} \\
  \mu_{2}^{(m)} \\
  \mu_{3}^{(m)} \\
\end{pmatrix},
\begin{pmatrix}
  1 & \rho & 0.25 \\
  \rho & 1 & 0.5 \\
  0.25 & 0.5 & 1 \\
\end{pmatrix}
\]

for \( m = 0,1; \)

\[ m_i \sim Bernoulli(\pi), \]

where \( \rho = 0.9 \) for low measurement error and \( \rho = 0.6 \) for high measurement error.

When \( \rho = 0.9 \), \((\mu_{1}^{(0)}, \mu_{2}^{(0)}, \mu_{3}^{(0)}) = (1.1,1,9.5)\) and \((\mu_{1}^{(1)}, \mu_{2}^{(1)}, \mu_{3}^{(1)}) = (2,2,10)\), and when \( \rho = 0.6 \), \((\mu_{1}^{(0)}, \mu_{2}^{(0)}, \mu_{3}^{(0)}) = (1.4,1,10.5)\) and \((\mu_{1}^{(1)}, \mu_{2}^{(1)}, \mu_{3}^{(1)}) = (2,2,11)\). The target marginal means of \( X_2 \) and \( X_3 \) are \( \mu_2 = \pi_1 \mu_2^{(1)} + (1 - \pi_1) \mu_2^{(0)} \) and \( \mu_3 = \pi_1 \mu_3^{(1)} + (1 - \pi_1) \mu_3^{(0)} \).

Under this model, nonrespondents have higher means than respondents for the two variables of interest \((X_2 \text{ and } X_3)\), and missingness is a function of values on \( X_2 \). The parameter values are chosen to satisfy the seven parameter restrictions described in
Section 2.1. The parameter $\pi_1$ determining the proportion of missing cases is set to 0.75 or 0.25 (corresponding to high or low unit nonresponse). We generate 1,000 samples of size $n = 1,000$ from this PMM for each value of $\pi_1$ and $\rho$.

The second set of simulations created nonresponse with a nonignorable selection model. Samples were generated from the trivariate normal model

$$
\begin{pmatrix}
    x_{i1} \\
    x_{i2} \\
    x_{i3}
\end{pmatrix} \sim N_3 \left( \begin{pmatrix} 1 \\ 1 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0.25 \\ \rho & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix} \right),
$$

where the parameter $\rho$ was set to 0.9 for low measurement error and 0.6 for high measurement error. The $X_1$ variable has a weaker association with $X_3$ than the “true” auxiliary variable $X_2$, to reflect attenuation of the relationships due to measurement error in $X_1$ (Fuller, 1987). Missing values of $X_2$ and $X_3$ were created using the model

$$
P(m_i = 0 \mid x_{i2}, \alpha, \lambda) = \frac{\exp(\alpha + \lambda x_{i2})}{1 + \exp(\alpha + \lambda x_{i2})},
$$

where $\alpha$ (with possible values 0 and -1) determines the expected response rate, and $\lambda$ (with possible values 2, 1, and 0) determines the dependence of response on the “true” auxiliary variable $X_2$, allowing for analyses of sensitivity to assumptions about the non-ignorable missing data mechanism. For each sample case, a random UNIFORM(0,1) deviate was drawn, and the values of $X_2$ and $X_3$ were retained if this draw was less than or equal to $P(m_i = 0 \mid x_{i2}, \alpha, \beta)$, and deleted otherwise.

For each simulation, we computed the empirical relative bias (%), empirical root mean squared error (RMSE), 95% confidence / credible interval (CI) coverage, and mean 95% CI width for the estimators of the two means defined by the five approaches above, based on 1,000 samples simulated under the alternative missing data mechanisms.
3.3. Results of Simulation Studies

Tables 1 and 2 present simulation results for each of the five estimation methods (PMM, PMM-MI, MI, GW, and CC) under the normal pattern-mixture and selection models specified in Section 3.2. Simulations were performed using R.

**Empirical Bias and RMSE.** When the data are simulated according to a PMM, the PMM and PMM-MI estimators have the smallest empirical bias and RMSE when missingness depends on the true value, $X_2$, as expected (Table 1). Notably, the PMM-MI estimator vastly out-performs the MI estimator, which assumes an ignorable (MAR) mechanism, when the missing data mechanism is nonignorable. The results in Table 1 and Table 2 also show that the empirical bias and RMSE of the MI estimator both increase as a function of measurement error in the auxiliary proxy $X_1$, regardless of the missing data mechanism, and become larger than that of the GW estimator under a PMM with decreased response rates (Table 1). This is also expected, given the bias in regression coefficients engendered by measurement error in the covariates (Fuller, 1987).

The PMM and PMM-MI estimators also perform well (in terms of empirical bias and RMSE) when the data are simulated from a selection model (Table 2). Under the normal selection model and an MCAR mechanism (Table 2), the PMM and PMM-MI estimators have slightly higher empirical RMSEs under high measurement error, reflecting some loss of efficiency from estimating the nonignorable model parameters. Under both missing data mechanisms (Tables 1 and 2), the GW and MI estimators have less empirical bias than the CC estimators when the missing data mechanism is non-ignorable, but are still biased, with a bias that increases as dependence of missingness on
Table 1: Selected simulation results under the pattern-mixture model.

<table>
<thead>
<tr>
<th>ρ</th>
<th>π₁</th>
<th>Method</th>
<th>$\hat{\mu}_2$ Rel. Bias</th>
<th>$\hat{\mu}_2$ RMSE</th>
<th>$\hat{\mu}_2$ 95% CI Cover.</th>
<th>$\hat{\mu}_3$ Rel. Bias</th>
<th>$\hat{\mu}_3$ RMSE</th>
<th>$\hat{\mu}_3$ 95% CI Cover.</th>
<th>$\hat{\mu}_3$ 95% CI Mean Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.75</td>
<td>PMM</td>
<td>3</td>
<td>51</td>
<td>948</td>
<td>194</td>
<td>-2</td>
<td>81</td>
<td>915</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>10</td>
<td>51</td>
<td>968</td>
<td>240</td>
<td>-1</td>
<td>134</td>
<td>952</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI</td>
<td>-1635</td>
<td>290</td>
<td>1</td>
<td>227</td>
<td>-470</td>
<td>471</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GW</td>
<td>-837</td>
<td>170</td>
<td>625</td>
<td>406</td>
<td>-212</td>
<td>225</td>
<td>281</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>-4282</td>
<td>752</td>
<td>0</td>
<td>249</td>
<td>-382</td>
<td>382</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>0.25</td>
<td>PMM</td>
<td>-1</td>
<td>36</td>
<td>951</td>
<td>140</td>
<td>-1</td>
<td>40</td>
<td>929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>7</td>
<td>37</td>
<td>961</td>
<td>165</td>
<td>2</td>
<td>47</td>
<td>984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI</td>
<td>-885</td>
<td>59</td>
<td>714</td>
<td>137</td>
<td>-72</td>
<td>80</td>
<td>556</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GW</td>
<td>-884</td>
<td>61</td>
<td>846</td>
<td>176</td>
<td>-72</td>
<td>79</td>
<td>547</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>-1999</td>
<td>253</td>
<td>0</td>
<td>143</td>
<td>-130</td>
<td>131</td>
<td>86</td>
</tr>
<tr>
<td>0.6</td>
<td>0.75</td>
<td>PMM</td>
<td>-2</td>
<td>104</td>
<td>946</td>
<td>415</td>
<td>-1</td>
<td>90</td>
<td>943</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>59</td>
<td>112</td>
<td>901</td>
<td>437</td>
<td>7</td>
<td>93</td>
<td>889</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI</td>
<td>-2972</td>
<td>524</td>
<td>0</td>
<td>288</td>
<td>-281</td>
<td>314</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GW</td>
<td>-2746</td>
<td>485</td>
<td>1</td>
<td>293</td>
<td>-241</td>
<td>272</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>-4277</td>
<td>751</td>
<td>0</td>
<td>248</td>
<td>-344</td>
<td>380</td>
<td>0</td>
</tr>
<tr>
<td>0.6</td>
<td>0.25</td>
<td>PMM</td>
<td>16</td>
<td>43</td>
<td>959</td>
<td>174</td>
<td>1</td>
<td>39</td>
<td>959</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>7</td>
<td>45</td>
<td>964</td>
<td>189</td>
<td>1</td>
<td>41</td>
<td>948</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI</td>
<td>-1274</td>
<td>163</td>
<td>10</td>
<td>143</td>
<td>-82</td>
<td>95</td>
<td>377</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GW</td>
<td>-1274</td>
<td>163</td>
<td>12</td>
<td>150</td>
<td>-82</td>
<td>94</td>
<td>332</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>-1997</td>
<td>252</td>
<td>0</td>
<td>143</td>
<td>-118</td>
<td>130</td>
<td>72</td>
</tr>
</tbody>
</table>

NOTES: $\rho = \text{corr}(X_1, X_2)$, and defines amount of measurement error in $X_1$; $\pi_1$ defines the proportion of population units with values arising from the model for pattern $m_i = 1$ (non-respondents); PMM = pattern-mixture model estimates based on Bayesian inference approach (Section 2.2); PMM-MI = pattern-mixture model estimates based on the multiple imputation approach (Section 2.3); MI = multiple imputation estimates after regression prediction (assuming a MAR mechanism) and application of Rubin’s combining rules; GW = global weighting estimates; CC = complete case estimates; CI = confidence / credible (for PMM) interval. Rel. Bias = Relative Bias (%) x 100. RMSE = Empirical RMSE x 1000. 95% CI Cover. = Number of intervals covering the true mean out of 1000. 95% CI Mean Width = Mean CI width x 1000.

$X_2$ and measurement error in $X_1$ increases. None of the estimators for the mean of the $X_3$ variable are badly biased in this setting, reflecting the fact that missingness depends on $X_2$. However, higher proportions of nonrespondents in the case of the PMM tend to increase the empirical bias and RMSE of the estimators for the mean of $X_3$ (Table 1), unlike in the case of the normal selection model (Table 2). The PMM and PMM-MI
estimators both appear robust to the model generating the missing data and the amount of measurement error in the auxiliary variable. The pattern of results evident in Table 2 also

Table 2: Selected simulation results under the normal selection model, with $\alpha = 0$ in the response propensity model.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\lambda$</th>
<th>Mean RR (%)</th>
<th>Method</th>
<th>$\hat{\mu}_2$ Rel. Bias</th>
<th>$\hat{\mu}_2$ RMSE</th>
<th>$\hat{\mu}_2$ 95% CI Mean Width</th>
<th>$\hat{\mu}_3$ Rel. Bias</th>
<th>$\hat{\mu}_3$ RMSE</th>
<th>$\hat{\mu}_3$ 95% CI Mean Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>2</td>
<td>78</td>
<td>PMM</td>
<td>2</td>
<td>32</td>
<td>957</td>
<td>130</td>
<td>-1</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>17</td>
<td>34</td>
<td>965</td>
<td>144</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>711</td>
<td>77</td>
<td>379</td>
<td>123</td>
<td>101</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>752</td>
<td>86</td>
<td>574</td>
<td>183</td>
<td>109</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>2911</td>
<td>293</td>
<td>0</td>
<td>122</td>
<td>144</td>
<td>148</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>70</td>
<td>PMM</td>
<td>-18</td>
<td>34</td>
<td>956</td>
<td>132</td>
<td>-1</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>-5</td>
<td>35</td>
<td>959</td>
<td>158</td>
<td>-9</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>541</td>
<td>63</td>
<td>612</td>
<td>128</td>
<td>80</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>540</td>
<td>64</td>
<td>774</td>
<td>165</td>
<td>79</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>2538</td>
<td>256</td>
<td>0</td>
<td>138</td>
<td>127</td>
<td>132</td>
</tr>
<tr>
<td>0.9</td>
<td>0</td>
<td>50</td>
<td>PMM</td>
<td>-24</td>
<td>34</td>
<td>953</td>
<td>138</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>1</td>
<td>37</td>
<td>973</td>
<td>180</td>
<td>6</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>-24</td>
<td>34</td>
<td>951</td>
<td>139</td>
<td>-1</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>-23</td>
<td>33</td>
<td>994</td>
<td>175</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>-22</td>
<td>43</td>
<td>958</td>
<td>176</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>78</td>
<td>PMM</td>
<td>-25</td>
<td>43</td>
<td>951</td>
<td>167</td>
<td>-1</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>6</td>
<td>45</td>
<td>940</td>
<td>181</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>2043</td>
<td>207</td>
<td>0</td>
<td>124</td>
<td>113</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>2064</td>
<td>209</td>
<td>0</td>
<td>127</td>
<td>114</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>2902</td>
<td>292</td>
<td>0</td>
<td>122</td>
<td>146</td>
<td>150</td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>70</td>
<td>PMM</td>
<td>-22</td>
<td>47</td>
<td>940</td>
<td>190</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>1737</td>
<td>177</td>
<td>3</td>
<td>139</td>
<td>94</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>1738</td>
<td>177</td>
<td>2</td>
<td>143</td>
<td>94</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>2568</td>
<td>259</td>
<td>0</td>
<td>138</td>
<td>128</td>
<td>134</td>
</tr>
<tr>
<td>0.6</td>
<td>0</td>
<td>50</td>
<td>PMM</td>
<td>-21</td>
<td>53</td>
<td>945</td>
<td>209</td>
<td>-2</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PMM-MI</td>
<td>13</td>
<td>53</td>
<td>933</td>
<td>209</td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MI</td>
<td>-11</td>
<td>42</td>
<td>943</td>
<td>175</td>
<td>-1</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GW</td>
<td>-15</td>
<td>40</td>
<td>969</td>
<td>176</td>
<td>-1</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>-14</td>
<td>45</td>
<td>954</td>
<td>176</td>
<td>-1</td>
<td>43</td>
</tr>
</tbody>
</table>

NOTES: $\rho = \text{corr}(X_1, X_2)$, and defines amount of measurement error in $X_1$; $\alpha = 0$; $\lambda$ determines dependence of missingness on $X_2$; Mean RR = average response rate across 1000 simulations; PMM = pattern-mixture model estimates based on Bayesian inference approach (Section 2.2); PMM-MI = pattern-mixture model estimates based on the multiple imputation approach (Section 2.3); MI = multiple imputation estimates after regression prediction (assuming a MAR mechanism) and application of Rubin’s combining rules; GW = global weighting estimates; CC = complete case estimates; CI = confidence / credible (for PMM) interval. Rel. Bias = Relative Bias (%) x 100. RMSE = Empirical RMSE x 1000. 95% CI Cover. = Number of intervals covering the true mean out of 1000. 95% CI Mean Width = Mean CI width x 1000.
holds under lower response rates, with $\alpha = -1$ in the normal selection model.

Confidence / Credible Interval Coverage and Width. Under both missing data models, the coverage of 95% confidence intervals based on the MI, GW, and CC estimators is far below nominal when missingness depends on $X_2$, and decreases with increased dependence of missingness on $X_2$ and more measurement error in the auxiliary variable. In contrast, 95% credible intervals based on the PMM and PMM-MI estimators have close to nominal frequentist coverage in nearly all cases. Interestingly, for higher levels of measurement error (under both missing data models), the mean widths of the Bayesian credible intervals based on the PMM estimators and the 95% confidence intervals based on the PMM-MI estimators tend to be higher than that for the other three estimators. This finding reflects the fact that increased measurement error in the auxiliary variable increases the uncertainty in the predictive distribution of the missing values. The PMM-MI approach also tends to produce wider confidence intervals than the other approaches. This finding reflects efficiency losses due to the small number of multiple imputations (5) relative to the information loss from the missing data, and the efficiency can be increased by increasing this number of imputations.

Similar patterns of results were found for the case where $\alpha = -1$ in the normal selection model (introducing lower response rates). In the cases of non-ignorable missing data mechanisms, the lower response rates simply served to increase the bias and RMSE of the MI, GW, and CC estimators while reducing their coverage. The PMM and PMM-MI estimators still performed quite well in the presence of lower response rates, but were once again found to have higher mean confidence interval width in the case of higher measurement error.
4. Including Other Fully Observed Auxiliary Variables

We may wish to include other auxiliary variables as predictors in models for imputing missing values. Suppose that in addition to the data in Figure 1 there is a set of $k$ such fully-recorded auxiliary variables $C$, including a vector of 1s for the intercept, and that missingness of $X_2$ and $X_3$ is assumed to depend on both $X_2$ and $C$. Since the auxiliary variables $C$ are fixed in the model, interactions and nonlinear terms involving the auxiliary variables can be included.

For the missing data pattern $m_i = m$, we assume the following generalization of the model described in Section 2. Conditional on values $c_i$ of the auxiliary variables $C$,

$$
\begin{pmatrix}
X_{i1} \\
X_{i2} \\
X_{i3}
\end{pmatrix}
\sim N_3\left(\begin{pmatrix}
\beta_{1c,c_i}^{(m)}c_i \\
\beta_{2c,c_i}^{(m)}c_i \\
\beta_{3c,c_i}^{(m)}c_i
\end{pmatrix},
\begin{pmatrix}
\sigma_{11c}^{(m)} & \sigma_{12c}^{(m)} & \sigma_{13c}^{(m)} \\
\sigma_{12c}^{(m)} & \sigma_{22c}^{(m)} & \sigma_{23c}^{(m)} \\
\sigma_{13c}^{(m)} & \sigma_{23c}^{(m)} & \sigma_{33c}^{(m)}
\end{pmatrix}\right) 
\equiv N_3\left(\begin{pmatrix}
\beta_{1c,c_i}^{(m)}c_i \\
\beta_{2c,c_i}^{(m)}c_i \\
\beta_{3c,c_i}^{(m)}c_i
\end{pmatrix},
\Sigma_{X_{123c}}^{(m)}\right),
$$

a trivariate normal distribution with $3k + 6$ parameters. In (12), $\beta_{ic,c_i}^{(m)}$ denotes the regression coefficients for the set of auxiliary variables $C$ in the linear regression of variable $i$ on $C$ for pattern $r$, and $\sigma_{ij,c_i}^{(m)}$ denote the residual covariance (variance if $i = j$) of variables $i$ and $j$, given $C$, for pattern $m$. In addition, the marginal distribution of $m_i$ given $c_i$ is

$$m_i \mid c_i, \gamma \sim \text{Bernoulli}\left(\pi_1(c_i, \gamma)\right),$$

where $\pi_1$ is the probability of missingness, and $\gamma$ is a vector of $k$ regression parameters in a logistic regression of the missingness indicator $m_i$ on the auxiliary variables $C$. The following parameters are identified from the observed data:

$$\theta_{id} = (\gamma, \beta_{1c,c_i}^{(0)}, \beta_{2c,c_i}^{(0)}, \beta_{3c,c_i}^{(0)}, \sigma_{11c}^{(0)}, \sigma_{12c}^{(0)}, \sigma_{13c}^{(0)}, \sigma_{22c}^{(0)}, \sigma_{23c}^{(0)}, \sigma_{33c}^{(0)}, \beta_{1c,c_i}^{(1)}, \sigma_{11c}^{(1)}).$$
The following $2k + 5$ parameters are not identified:

$$\theta_{\text{nid}} = (\beta_{2c}^{(1)}, \rho_{2c}^{(1)}, \sigma_{1c}^{(1)}, \sigma_{13c}^{(1)}, \sigma_{23c}^{(1)}, \sigma_{22c}^{(1)}, \sigma_{33c}^{(1)})$$

The assumption that missingness of $X_2$ and $X_3$ depends on $X_2$ and $C$ implies that the distribution of $X_1$ and $X_3$ given $X_2$ and $C$ is the same for complete and incomplete cases, yielding $2k + 5$ parameter restrictions. Hence the model is just-identified (as described earlier).

ML estimates of the identified parameters $\theta_{\text{id}}$ are computed as before, with the regression coefficients on $C$ computed by applying OLS regression to the two patterns. The non-identified parameters $\theta_{\text{nid}}$ are similar functions of the identified parameters given earlier, except that the expressions condition on the auxiliary variables $C$. Define the following sample estimates:

- $\hat{\gamma}$ = ML estimate of $\gamma$ from logistic regression of $M$ on $C$;
- $\hat{\beta}_{1c}^{(m)}$ = OLS regression coefficients of $X_1$ on $C$, missing-data pattern $m$;
- $\hat{\sigma}_{1c}^{(m)}$ = Residual variance of $X_1$ given $C$, missing-data pattern $m$;
- $\hat{\beta}_{jc}^{(0)}$ = OLS regression coefficient of $X_j$ on $C$, complete cases, $j = 2,3$;
- $\hat{\beta}_{j2c}$ = Coefficient of $X_2$ from OLS regression of $X_j$ on $C$ and $X_2$, complete cases, $j = 1,3$;
- $\hat{\sigma}_{jk}^{(0)}$ = Covariance of $X_j, X_k$ given $C$, complete cases.

The ML estimates are then computed as follows, given the notation above (where $C$ includes the column of 1s used for the intercept terms in the models):

$$\hat{\beta}_{2c}^{(1)} = \frac{\hat{\beta}_{2c}^{(0)} + \hat{\beta}_{1c}^{(1)} - \hat{\beta}_{1c}^{(0)}}{\hat{\beta}_{12c}}; \quad \hat{\sigma}_{12c}^{(1)} = \hat{\sigma}_{12c}^{(0)} + \frac{\hat{\sigma}_{11c}^{(0)} - \hat{\sigma}_{11c}^{(0)}}{\hat{\beta}_{12c}}; \quad \hat{\sigma}_{22c}^{(1)} = \hat{\sigma}_{22c}^{(0)} + \frac{\hat{\sigma}_{11c}^{(0)} - \hat{\sigma}_{11c}^{(0)}}{\hat{\beta}_{12c}};$$

$$\hat{\beta}_{3c}^{(1)} = \hat{\beta}_{3c}^{(0)} + \hat{\beta}_{32c}^{(1)} - \hat{\beta}_{32c}^{(0)} \frac{\hat{\beta}_{12c}}{\hat{\beta}_{12c}}; \quad \hat{\sigma}_{13c}^{(1)} = \hat{\sigma}_{13c}^{(0)} + \hat{\beta}_{32c}^{(0)} \frac{\hat{\sigma}_{11c}^{(0)} - \hat{\sigma}_{11c}^{(0)}}{\hat{\beta}_{12c}};$$
For Bayesian inference, assuming noninformative priors for the identified parameters, a sequence of draws from the posterior distribution of the identified parameters in this case can be computed by adding covariates \( C \) to the expressions described earlier, and these draws then replace the ML estimates in the above expressions to simulate draws from the posterior distribution of the other parameters. The following sequence of draws is repeated many times to simulate the posterior distributions and make inferences as before:

1) \( \gamma^{(d)} \sim p(\gamma \mid \text{data}) \), the posterior distribution of \( \gamma \);

2) \( \sigma^{(0)(d)}_{11c} = r \hat{\sigma}^{(0)}_{11c} / u_1^{(d)} \), \( u_1^{(d)} \sim \chi^2_{n-r-k} \);

3) \( \beta^{(0)(d)}_{ic} \sim N(\hat{\beta}^{(0)}_{ic}, S^{(0)}_{cc} \sigma^{(0)(d)}_{11c}) \), where \( S^{(0)}_{cc} \) is the sum of squares and cross-products matrix of the covariates \( C \), for \( m = 0 \);

4) \( \sigma^{(1)(d)}_{11c} = (n-r)\hat{\sigma}^{(1)}_{11c} / u_2^{(d)} \), \( u_2^{(d)} \sim \chi^2_{n-r-k} \);

5) \( \beta^{(1)(d)}_{ic} \sim N(\hat{\beta}^{(1)}_{ic}, S^{(1)}_{cc} \sigma^{(1)(d)}_{11c}) \), where \( S^{(1)}_{cc} \) is the sum of squares and cross-products matrix of the covariates \( C \), for \( m = 1 \);

6) \( \begin{pmatrix} \hat{\sigma}^{(d)}_{112c} \\ \hat{\sigma}^{(d)}_{132c} \\ \hat{\sigma}^{(d)}_{332c} \\ \hat{\sigma}^{(d)}_{333c} \end{pmatrix} \sim \text{Inv-Wishart} \left( \begin{pmatrix} \hat{\sigma}^{(1)}_{112c} \\ \hat{\sigma}^{(1)}_{132c} \\ \hat{\sigma}^{(1)}_{332c} \\ \hat{\sigma}^{(1)}_{333c} \end{pmatrix}, r-k \right) \)

7) \( \beta^{(d)}_{122c} \sim N(\hat{\beta}^{(d)}_{122c}, \sigma^{(d)}_{122c} / (r \hat{\sigma}^{(d)}_{22c})) \); \( \beta^{(d)}_{102c} \sim N(\hat{\mu}^{(d)}_1 - \hat{\beta}^{(d)}_{122c} \hat{\mu}^{(d)}_2, \sigma^{(d)}_{112c} / r) \)

8) \( \beta^{(d)}_{322c} \sim N(\hat{\beta}^{(d)}_{322c}, \sigma^{(d)}_{322c} / (r \hat{\sigma}^{(d)}_{22c})) \); \( \beta^{(d)}_{302c} \sim N(\hat{\mu}^{(d)}_3 - \hat{\beta}^{(d)}_{322c} \hat{\mu}^{(d)}_2, \sigma^{(d)}_{332c} / r) \)

If the objective of the analysis is inference about marginal means of \( X_2 \) or \( X_3 \) (as opposed to the regression parameters or variance-covariance parameters), we can apply the MI approach described in Section 2.3 to make inferences that essentially “integrate
out” values of the auxiliary variables \( C \). We first draw parameters for pattern \( m = 1 \) of the PMM defined in (12) from their posterior distributions (without needing the draws \( \gamma^{(d)} \)), given that our focus is on the pattern \( m = 1 \), and then impute missing values for \( X_2 \) and then \( X_3 \) by taking random draws from their conditional distributions defined by the drawn parameters (as shown in Section 2.3):

\[
x_{2i}^{(d)} \sim N \left( \beta_{2c4c}^{(1)} x_{ci} + \beta_{214c}^{(1)} x_{1i} + \beta_{224c}^{(1)} x_{2i}, \sigma_{224c}^{(1)} \right) \quad \text{and} \quad \tag{13}
\]

\[
x_{3i}^{(d)} \sim N \left( \beta_{3c12c}^{(1)} x_{ci} + \beta_{3112c}^{(1)} x_{1i} + \beta_{3212c}^{(1)} x_{2i} + \beta_{3312c}^{(1)} x_{3i}, \sigma_{3312c}^{(1)} \right). \quad \tag{14}
\]

The SWEEP operator facilitates computation of the parameters in these conditional distributions given the draws for pattern \( m = 1 \) of the PMM; for example, we have \( \beta_{2c4c}^{(1)} = \beta_{2c4c}^{(1)} - \beta_{2c4c}^{(1)} / s_{11c}^{(1)} \). This process is repeated \( B \) times to create \( B \) complete data sets. The means of \( X_2 \) and \( X_3 \) and their standard errors are then computed for each data set using standard complete-case methods (potentially incorporating complex sampling features), and MI combining rules are applied for making inferences.

5. Application: The Labor Market and Social Security (PASS) Survey

We applied our methods to data from the German Labor Market and Social Security (PASS) survey, a panel study that collects annual labor market, household income, and unemployment benefit receipt data from a nationally representative sample of 12,000 households from the German population (Trappmann et al., 2011). According to the PASS survey web site (http://fdz.iab.de/en/FDZ_Individual_Data/PASS.aspx), “PASS is a new central source for analyses of the labour market and poverty situation in Germany as well as the situation of recipients of benefits in accordance with the German Social Code Book II.” German households known to have received unemployment
benefits are sampled at a higher rate than other households, so sampling weights are needed to make representative inferences about the German population. To assist with both stratified sampling and estimation, the PASS survey purchases auxiliary variables describing area-level features for sampled households from the German consumer marketing organization Microm. These variables are then linked to the sampled households at the address level, and linking rates are consistently higher than 95%. See Trappmann et al. (2011) for additional details.

For this application, we identified continuous variables from the Microm database (available for all sample units) and the PASS survey (Wave 1 respondents in 2006) for analysis. Specifically, 48,250 sampled households had information available on a continuous auxiliary variable measuring the average purchasing power (in Euros) of households in the same city block. This variable followed an approximately normal distribution, and was considered as an error-prone auxiliary proxy ($X_1$) of reported monthly household income. Monthly household income and area (in square meters) of the housing unit were both measured for 11,969 respondents to the PASS survey in Wave 1 (a 24.8% unweighted response rate). We also extracted the base sampling weights, stratum identifiers, and sampling error cluster codes for the Wave 1 respondents, given the stratified multistage design employed for the survey.

Monthly household income (log-transformed) was considered as the $X_2$ variable, and unit nonresponse (on $X_2$ and $X_3$) was assumed to be a linear function of this variable. This assumption was supported by strongly significant ($p < 0.001$) associations of both average household purchasing power and the base sampling weight with a response indicator in a logistic regression model fitted to the full sample. For every 10,000 euro
increase in the average purchasing power of households in a given city block, the expected odds of an individual household responding were reduced by about 15% (estimated odds ratio = 0.853, 95% CI = 0.822, 0.885), and larger values on the base sampling weight (generally indicating households not receiving unemployment benefits) were also associated with reduced odds of responding. Area of the housing unit (also log-transformed) was considered as the $X_3$ variable. The correlation between the auxiliary measure of average purchasing power and the reported household income (log-transformed) was 0.223, suggesting substantial error in the auxiliary proxy (the lowest correlation considered in the simulation studies above was 0.6). The correlation of average purchasing power with log-transformed housing unit area was 0.137, while the correlation of housing unit area and household income was 0.642.

5.1 Analysis with One Error-Prone Auxiliary Variable

In the first analysis, we applied the CC, GW, MI, and PMM-MI methods to estimate population means for monthly household income (in Euros) and housing unit area (in meters squared). The GW and MI estimators assumed an ignorable missing data mechanism, where missingness was a function of the auxiliary variable measuring average purchasing power of the households. The PMM-MI estimator assumed a non-ignorable missing data mechanism, where missingness was a function of the household income variable measured in the survey. Each of these four methods also accounted for the complex design features of the Wave 1 PASS sample (weighting for unequal probability of inclusion, stratification, and cluster sampling); see Heeringa et al. (2010) for more details on these types of design-based procedures.
When applying the CC approach for the respondents only, weighted estimates of the means for log-transformed monthly household income and log-transformed housing unit area were computed using the Wave 1 base sampling weight, and TSL was applied (incorporating the stratum and cluster codes and the weighted cluster totals) for variance estimation. When applying the GW approach, the base weights were adjusted by the inverse of the predicted response propensity from a logistic regression model predicting the response indicator with the proxy income variable, and the base weights were ignored when estimating the logistic model (per Little and Vartivarian, 2003). The MI approach was implemented using the \texttt{mi()} function in R to perform multiple sequential regression imputations (as in the simulation studies), and complex sample design features were accounted for in the analysis of each imputed data set using the \texttt{survey} package in R (Lumley, 2010). Finally, we applied the PMM-MI approach described in Section 2.3 for the possible non-ignorable missing data mechanism, given that the standard PMM approach outlined in Section 2.2 does not recognize complex sampling features.

Estimates of population means for household income and housing unit area computed using the four methods were exponentiated to return them to their original scales. Table 3 presents results from applying these four different approaches.

**Table 3:** Estimates of mean reported household income and mean housing unit area (in square meters), based on four different nonresponse adjustment methods*.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Method</th>
<th>Estimated Mean</th>
<th>95% CI</th>
<th>CI Width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>1,814.88</td>
<td>(1,772.99, 1,857.77)</td>
<td>84.78</td>
</tr>
<tr>
<td></td>
<td>GW</td>
<td>1,838.57</td>
<td>(1,795.62, 1,882.54)</td>
<td>86.92</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>1,448.57</td>
<td>(1,412.88, 1,485.15)</td>
<td>72.27</td>
</tr>
<tr>
<td></td>
<td>PMM-MI</td>
<td>1,797.24</td>
<td>(1,744.70, 1,851.36)</td>
<td>106.66</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>89.21</td>
<td>(87.47, 90.99)</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>GW</td>
<td>89.65</td>
<td>(87.91, 91.42)</td>
<td>3.51</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>78.30</td>
<td>(76.92, 79.69)</td>
<td>2.77</td>
</tr>
</tbody>
</table>

*Results may vary due to the use of randomization in the imputation process.
Table 3 shows that inferences based on the CC, GW, and PMM-MI approaches would be similar. We would make different inferences depending on whether the MI approach (assuming an ignorable model) or the PMM-MI approach (assuming a non-ignorable model) is used in this analysis. In the PASS survey, nonrespondents tended to have higher income and significantly higher base sampling weights as a result (given the informative sampling). Given the weak relationship of the error-prone proxy variable with household income observed for the respondents, the imputed values for nonrespondents under the ignorable model all tended to be closer to the mean for the responding cases, which had lower income in general. When the base weights were applied to each imputed data set, these negatively biased predictions were inflated, and this resulted in the substantially different inferences for the means that are evident in Table 3. The PMM-MI approach incorporates the apparent dependence of missingness on income, and is not as heavily affected as a result. However, given the weak relationship of the auxiliary proxy with income (possibly due to error in the proxy), we see the same inefficiency in the PMM-MI estimates that was noted in the simulations.

This analysis demonstrates the sensitivity of multiple imputation inferences based on error-prone auxiliary proxies to assumptions about the missing data mechanism. Given knowledge of the oversampling of low-income households in the PASS survey and the substantial differences in distributions of the base sampling weights between respondents and nonrespondents, use of an error-prone auxiliary proxy under assumptions of an ignorable missing data mechanism may result in bias. In practice, inferences based on the
PMM-MI and MI approaches should be compared to assess the sensitivity of inferences to the assumed missing data model. Better adjustments would include additional auxiliary variables measured with less error and (ideally) having stronger relationships with the key survey variables and response propensity, and we consider such adjustments in the next section.

5.2. Analysis with Multiple Auxiliary Variables

We now compare inferences based on the four approaches that account for the complex sample design features and include multiple auxiliary variables in the adjustments. We consider the informative (and error-free) base sampling weight as an additional auxiliary variable, alongside the auxiliary proxy of household income. The variable containing the base sampling weights was included in the logistic regression model used to compute predicted response propensities for the GW approach, and also included in the imputation models for the MI and PMM-MI approaches. This means that there are $k = 2$ additional auxiliary variables in the vector $C$ from Section 4: a column of 1s for the intercept, and the base sampling weights. The CC analysis results do not change in this case, given that the CC method is not affected by the choice of auxiliary variables for the nonresponse adjustment. Table 4 presents results from including the base sampling weights in the various nonresponse adjustments.

Table 4: Estimates of mean reported household income and mean housing unit area (in square meters), based on four different nonresponse adjustment methods that included the base sampling weight as an additional auxiliary variable*.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Method</th>
<th>Estimated Mean</th>
<th>95% CI</th>
<th>CI Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported</td>
<td>CC</td>
<td>1,814.88</td>
<td>(1,772.99, 1,857.77)</td>
<td>84.78</td>
</tr>
<tr>
<td>Monthly HH</td>
<td>GW</td>
<td>1,860.02</td>
<td>(1,815.87, 1,905.24)</td>
<td>89.37</td>
</tr>
</tbody>
</table>
The results in Table 4 suggest that the CC, GW, and MI estimates are all biased low when these improved adjustments are considered. Inferences based on the PMM-MI method would be significantly different than inferences based on the other three approaches, and suggest that the mean income in the German population is much higher than would be suggested by the approaches assuming ignorable missing data mechanisms. Notably, the GW and MI estimates are very similar to the CC estimates, which suggests that adjustments based on the error-prone auxiliary variable and the base sampling weights are not removing the bias that is arising from what may be a non-ignorable missing data mechanism. Finally, we once again see the same inefficiency in the PMM-MI estimates that was noted in the simulations when the auxiliary proxy is measured with fairly substantial error. As was noted in the simulations, the relative reductions in bias from using the PMM-MI approach may result in estimates with lower RMSE overall despite the decrease in efficiency.

6. Discussion

We have proposed PMM estimators for survey nonresponse, where a fully observed continuous auxiliary variable is measured with error on each of $n$ sample units, true values of the auxiliary variable (along with other continuous survey variables of
interest) are measured on survey respondents, and missingness depends on the true values of the auxiliary variable. Simulation studies suggest that under these conditions, the PMM estimators have reduced empirical bias, reduced empirical RMSE, and 95% credible sets with confidence coverage closer to nominal levels, compared with standard imputation and weighting approaches that assume ignorable (or MAR) missing data models. We also found the PMM estimators to be robust to the model generating the missing data, as these estimators performed equally well when missing data were generated under a normal selection model.

We applied the proposed PMM estimators to descriptive analyses of real data from a large area probability sample survey in Germany (the PASS survey). The applications demonstrated the ability of the proposed PMM-MI estimator to accommodate complex sample design features when a non-ignorable missing data mechanism is suspected and auxiliary variables available for the imputation models may be prone to error. The applications also showed the importance of comparing multiple imputation inferences based on ignorable and non-ignorable models when auxiliary variables are error-prone, and examining the sensitivity of the inferences to assumptions about the missing data mechanism. When incorporating an additional auxiliary variable that was free from error and related to both the survey variables of interest and response propensity (the base sampling weights) in the nonresponse adjustments, the PMM-MI estimator yielded inferences that were substantially different from the methods assuming an ignorable missing data mechanism.

In general, the forms of the proposed PMM estimators indicate situations where one can expect the most bias reduction: 1) missingness is substantially related to the
underlying true value; 2) the auxiliary proxy has substantial measurement error, making the MAR adjustment inadequate; and 3) the missing data rate is high. As shown in the simulation studies, if the measurement error in the auxiliary proxy is large enough that the correlation between the proxy and the true variable is low, then bias reduction will come at the expense of increased variance.

There are many possible extensions of this work. This work only considered a single normally-distributed auxiliary variable measured with error, and extensions to two or more such error-prone variables or non-normal variables would be useful. For instance, some face-to-face surveys request that interviewers record binary (yes/no) judgments about features of sampled households, such as whether young children are present, and these types of judgments can be prone to error (West, 2012). Extensions of the proposed methods to accommodate errors in these types of error-prone binary auxiliary variables are needed. Further extensions might also include development of PMM estimators for additional binary variables measured in the survey, given the importance of binary outcomes in survey research, and work is currently ongoing in this area (Andridge and Little, 2009). We also assumed that there was no measurement error in the survey variables measured for respondents, and the impact of error in these variables on the methods discussed in this study also deserves future research attention.

Finally, applying the proposed PMM methods to real survey data requires that the methods be implemented in statistical software packages. R functions enabling applications of the PMM estimators proposed in this article to real survey data are available upon request from the authors (email: bwest@umich.edu). Data producers could use the proposed methods (and R functions) to impute missing values on key
survey variables if non-ignorable missing data mechanisms are suspected, and then release multiple imputed data sets to the public. Secondary analysts could then apply standard complete case methods when analyzing each data set and make inferences based on straightforward MI combining rules.

References


