**Statistica Sinica Preprint No: SS-2016-0396.R2**

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Fully Efficient Joint Fractional Imputation for Incomplete Bivariate Ordinal Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>SS-2016-0396.R2</td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td>10.5705/ss.202016.0396</td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
<td>Changbao Wu</td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
FULLY EFFICIENT JOINT FRACTIONAL IMPUTATION FOR INCOMPLETE BIVARIATE ORDINAL RESPONSES

Xichen She and Changbao Wu

University of Waterloo

Abstract: We propose a fully efficient joint fractional imputation method for handling bivariate ordinal responses with missing observations. We show that the method is ideally suited for bivariate ordinal responses to create a single imputed data file and provides valid and efficient inferences for the joint and marginal probabilities, association measures as well as regression analysis. Asymptotic properties of estimators based on the joint fractionally imputed data set are developed and their superiority over existing methods, including available-case analysis, propensity score adjustment and sequential regression multiple imputation methods, is demonstrated through both theoretical results and simulation studies. The proposed joint fractional imputation strategy employs modelling procedures that could be used for the sequential regression multiple imputation method but creates a single imputed data set which can be easily analyzed using existing softwares with minor modifications. Variance estimation and test of independence are also discussed under the proposed joint fractional imputation method.
1. Introduction

Ordinal responses are categorical variables with an ordered scale and are routinely collected and analyzed by researchers from many scientific fields. For example, ordinal variables are commonly used in medical studies to measure the severity of injuries (i.e., minor, mild, severe or life-threatening), the stage of progression of a disease, the effect of a treatment, and many others. Bivariate ordinal responses are also commonly observed, such as conditions on two related parts of the body or measures of two contrasting treatments.

There are two major problems in statistical analysis of bivariate ordinal responses: (i) contingency table analysis; and (ii) regression analysis. The contingency table analysis focuses mainly on the joint and the marginal distribution of the two variables and, more importantly, the interrelation between responses, whereas the regression modelling explores the dependence of both responses on covariates while simultaneously taking into consideration the correlation between the two response variables. Statistical methods developed for bivariate nominal responses are also applicable to ordinal contingency tables. Agresti [2010, 2013] contain an excellent review
on related techniques. There have been methods developed specifically for ordinal responses to better handle the ordering nature of the variables. Kendall (1945), Goodman and Kruskal (1954) and Somers (1962) proposed different association measures to summarize correlation between ordinal responses. Alternatively, several association models were built to characterize the dependence, see, for instance, Haberman (1974) and Goodman (1979, 1985). Regression analysis with ordinal responses did not attract much attention until the emergence of generalized linear models (McCullagh and Nelder (1989)). The generalized estimating equation (GEE) method, initially proposed by Liang and Zeger (1986) as a tool for longitudinal and clustered data, can be applied for regression analysis with ordinal responses. See, for example, Lumley (1996), Parsons et al. (2006) and Touloumis et al. (2013). Transitional models which include other responses as predictors are another alternative approach to incorporating correlation between responses. For more detailed discussions on modeling techniques for ordinal responses, see Supplementary Material.

If one or both ordinal responses contain missing observations, none of the existing analysis tools is directly applicable. There have been enormous developments in recent years on theory and application of methods for handling missing data. Multiple imputation (MI), formally proposed by Rubin
XICHEN SHE AND CHANGBAO WU

(1987), has gained tremendous popularity among users of incomplete data. However, most studies focus on cases with continuous or nominal responses and little attention has been given to ordinal responses. The sequential regression multiple imputation (SRMI) method, proposed by Raghunathan et al. (2001) and also known as multiple imputation with chained equations (MICE) (van Buuren and Groothuis-Oudshoorn, 2011), is a flexible and practical procedure for generating multiple imputed data sets, and the method is technically applicable to ordinal responses. In Supplementary Material, we elaborate on key steps to implement the method for bivariate ordinal responses with missing values. One of the major drawbacks of SRMI, however, is the lack of theoretical justifications. The popularity of SRMI in practical applications rests largely on empirical studies rather than theoretical arguments (White et al., 2011).

Multiple imputation requires the creation of multiple data files and separate storage and analysis of those files by the users. From an operational point of view and for large survey agencies, it is more appealing to have a single imputed data file, especially if the file is to be released for public use with multiple users (Brick and Kalton, 1996). Single imputation, however, is criticized for its lack of efficiency due to the potential variation, known as the imputation variance, induced by random imputation.
procedures. Fractional imputation (FI), originally proposed by Kalton and Kish (1984) and later further studied by Kim and Fuller (2004) and Kim (2011), has surged as an attractive alternative to multiple imputation for reducing the imputation variance. It replaces each missing observation by a cluster of plausible values with each imputed value also receiving a fractional weight. Observed components are duplicated for fractionally imputed units, resulting in a single enlarged data file. With appropriate fractional weights, standard analyses can be applied directly to the imputed data file with minor modifications to incorporate the weights and lead to valid and efficient inferences. See Yang and Kim (2016) for an insightful review of recent developments in the FI literature.

In this paper, we propose a fully efficient joint fractional imputation (JFI) procedure for handling incomplete bivariate ordinal responses by creating a single imputed data file, which can be released for public use. The proposed method is fully efficient in the sense that it does not incur any additional variation from the imputation and leads to valid inferences for the joint and marginal probabilities, association measures as well as regression analysis. Tests of independence can also be carried out based on the association estimators. We justify the validity of our proposed procedure by revealing its deep link to the EM algorithm (Dempster et al. (1977)).
The rest of the paper is organized as follows. Section 2 introduces basic settings and notation and inferential problems with bivariate ordinal responses. In Section 3, we present our proposed method and establish major theoretical results. Results from simulation studies with comparisons to existing methods are reported in Section 4. Some concluding remarks are given in Section 5.

2. Basic Settings and Notation

Suppose that the sample data set is given by \( D = \{ (y_i, \delta_i, x_i), \ i = 1, \ldots, n \} \), where \( y_i = (y_{i1}, y_{i2}) \) are two ordinal responses on \( R \)-level and \( J \)-level scales, respectively, and both are partially observed. Let \( \delta_i = (\delta_{i1}, \delta_{i2}) \) be the corresponding response indicators: \( \delta_{it} = 1 \) if \( y_{it} \) is observed and \( \delta_{it} = 0 \) otherwise, \( t = 1, 2 \). The vector \( x_i \) consists of fully observed auxiliary variables in the data file. We assume that the data set is an independent sample of size \( n \) from \( (y, \delta, x) \). Furthermore, units in the sample can be partitioned into four groups, depending on the missing pattern of the responses:

\[
\mathcal{R} = \{ i : \delta_{i1} = 1, \delta_{i2} = 1 \}, \quad \mathcal{P}_1 = \{ i : \delta_{i1} = 1, \delta_{i2} = 0 \}, \\
\mathcal{P}_2 = \{ i : \delta_{i1} = 0, \delta_{i2} = 1 \}, \quad \mathcal{M} = \{ i : \delta_{i1} = 0, \delta_{i2} = 0 \}.
\]

We consider scenarios where the responses are missing-at-random (MAR) as
FRACTIONAL IMPUTATION FOR ORDINAL RESPONSES

termed by Little and Rubin (2002) such that $(\delta_1, \delta_2 \perp y_{mis}) \mid (y_{obs}, x)$, where $y_{mis}$ and $y_{obs}$ are respectively the missing and the observed component(s) of $y$. It implies that $P(\delta_1 = 1, \delta_2 = 0 \mid y, x) = P(\delta_1 = 1, \delta_2 = 0 \mid y_1, x)$, $P(\delta_1 = 0, \delta_2 = 1 \mid y, x) = P(\delta_1 = 0, \delta_2 = 1 \mid y_2, x)$ and $P(\delta_1 = 0, \delta_2 = 0 \mid y, x) = P(\delta_1 = 0, \delta_2 = 0 \mid x)$. Note that above MAR assumption is less restrictive than the monotone missingness often used for longitudinal data and is sufficient for the justification of our proposed procedure to be presented in Section 3.

In the absence of missing values, observations for bivariate ordinal responses can be cross-classified into an $R \times J$ table of cell counts based on the response values. For a fixed sample size $n$, the cell counts of the contingency table follow a multinomial distribution. We denote the probability of the bivariate ordinal responses falling into the cell in the $r$th row and $j$th column by $\pi_{rj} = P(y_1 = r, y_2 = j)$, $r = 1, \ldots, R, j = 1, \ldots, J$.

Let $\pi = (\pi_{11}, \ldots, \pi_{1J}, \ldots, \pi_{R1}, \ldots, \pi_{RJ})'$ be the vector of all cell probabilities. We have $\sum_{r=1}^{R} \sum_{j=1}^{J} \pi_{rj} = 1$. The marginal distributions of the responses are of basic interest and are denoted by $\pi_1 = (\pi_{1+}, \ldots, \pi_{R+})'$ and $\pi_2 = (\pi_{+1}, \ldots, \pi_{+J})'$, where $\pi_{r+} = \sum_{j=1}^{J} \pi_{rj}$ and $\pi_{+j} = \sum_{r=1}^{R} \pi_{rj}$. The dependence between the two ordinal responses, however, is often the main focus for analysis of bivariate data. In such cases measures of asso-
cation are of primary concern. A simple example is the conditional distribution of \( y_1 \) given \( y_2 \) at level \( j \): \( \pi_{1j} = (\pi_{11j}, \ldots, \pi_{1Rj})', j = 1, \ldots, J \), where \( \pi_{rj} = P(y_1 = r | y_2 = j) = \pi_{rj}/\pi_{+j} \). A more popular example is a set of different types of ordinal odds ratios, including the local, the cumulative and the global odds ratios. See Supplementary Material for detailed definitions.

It is sometimes more appealing to characterize the association between two ordinal variables by a single summary index rather than a set of odds ratios. Several such measures have been proposed based on the probabilities of concordance and discordance. Two ordinal observations \((y_{i1}, y_{i2})\) and \((y_{m1}, y_{m2})\) are concordant if the subject ranking higher on \( y_1 \) also ranks higher on \( y_2 \); while they are discordant if the one ranking higher on \( y_1 \) ranks lower on \( y_2 \). Goodman and Kruskal (1954) proposed to use the parameter \( \gamma \) defined as

\[
\gamma = \frac{(\Pi_c - \Pi_d)}{(\Pi_c + \Pi_d)}, \tag{2.1}
\]

where \( \Pi_c = 2 \sum_{r<k} \sum_{j<l} \pi_{rj} \pi_{kl} \) and \( \Pi_d = 2 \sum_{r<k} \sum_{j>l} \pi_{rj} \pi_{kl} \), corresponding to the probabilities of concordance and discordance for two randomly selected observations. The value of \( \gamma \) ranges from \(-1\) to \(1\). When \( |\gamma| = 1 \), there is a monotone relationship between \( y_1 \) and \( y_2 \), but not necessarily strictly monotone. For example, \( \gamma = 1 \) indicates that if \( y_{i1} < y_{m1} \) then \( y_{i2} \leq y_{m2} \). When \( y_1 \) and \( y_2 \) are independent, we have \( \gamma = 0 \), but the re-
verse statement is not true. Other examples of association measures include Kendall’s Tau-b \cite{Kendall1945} and Somers’ d \cite{Somers1962}, both having the same numerator $\prod_c - \prod_d$. The plug-in estimator of $\prod_c - \prod_d$ is given by $C - D$, where $C = 2 \sum_{r<k} \sum_{j<l} \hat{\pi}_{rj} \hat{\pi}_{kl}$ and $D = 2 \sum_{r<k} \sum_{j>l} \hat{\pi}_{rj} \hat{\pi}_{kl}$. Simon \cite{Simon1978} showed that any estimated measures based on $C - D$ are equivalent in terms of efficacy for testing independence. The Wald-type test statistic for independence is given by

$$z = \frac{(C - D)}{\hat{\sigma}_{C-D}}, \quad (2.2)$$

where $\hat{\sigma}_{C-D}$ can be the nonnull standard error of $C - D$ or the null standard error using the relations $\pi_{rj} = \pi_{r+} \pi_{+j}$ under independence. Agresti \cite{Agresti2010} recommended to use the latter one and claimed that the test statistic with null standard error converges to normal distribution faster under the null hypothesis. Since ordinal responses are a special type of categorical data, the Pearson $\chi^2$ test is also applicable. However, the latter is designed for a general alternative and may not have good power for testing a trend, which is of primary interest for ordinal responses. On the contrary, the $z$ statistic given in (2.2) is very natural for alternative hypotheses such as $\Pi_c > \Pi_d$ or $\Pi_c < \Pi_d$, corresponding to a positive and negative trend.

When one or both ordinal responses contain missing values, the naive “available-case analysis” (ACA) approach by deleting observations with
missing values is usually invalid unless the missing rate is very low or the
data are missing completely at random (MCAR) (Little and Rubin (2002)).
Two existing approaches for handling missing values in this case are the
propensity score adjustment (PSA) method and the sequential regression
multiple imputation (SRMI) method. Details of these two methods are
given in Supplementary Material and their performances compared to our
proposed method are presented in Section 4.

3. Fully Efficient Joint Fractional Imputation

In this section we present our proposed joint fractional imputation ap-
proach to bivariate ordinal responses with missing values. We combine the
modelling strategies from the SRMI method with the specific feature of or-
dinal variables to create a single fractionally imputed data set which is well
suited for both marginal and joint analyses. The efficiency of the approach
is demonstrated through a maximum likelihood interpretation of the pro-
cedure, asymptotic properties of the estimators and results of simulation
studies.

3.1. Joint fractional imputation

The imputation models we use are inspired by the transitional modelling
briefly mentioned in Section 1 and the sequential regression modelling used
by the SRMI method. We impose a marginal regression model on one of
the responses and a transitional regression model on the other with the first
response as a predictor. To be more specific, we consider the following two
models:

Marginal: \[ g_1(\eta_{r1}) = \alpha_{r1} - \beta'_1 x, \]

Transitional: \[ g_2(\eta_{j2}) = \alpha_{j2} - \beta'_2 x - \sum_{r=2}^{R} \nu_r I(y_1 = r), \] (3.1)

where \( \eta_{r1} = P(y_1 \leq r \mid x) \) and \( \eta_{j2} = P(y_2 \leq j \mid y_1, x) \) are the cumulative
probabilities given the covariates, and \( g_1 \) and \( g_2 \) are link functions. Let
\( \theta_1 = (\alpha_{11}, ..., \alpha_{R1}, \beta'_1)' \) be the parameters in the marginal model and \( \theta_2 = (\alpha_{12}, ..., \alpha_{J2}, \beta'_2, \nu_2, ..., \nu_R)' \) be the parameters in the transitional model. Both
models in (3.1) belong to the cumulative link model family. Popular choices
for the link functions include the \text{logit}, \text{probit} and \text{c-log-log} functions. There
exists an interesting latent variable interpretation for models with different
links. See She (2017) for further details. Our proposed method can be
easily adapted to more complex parametric forms of (3.1), for example, one
with nonlinear systematic components, and other ordinal regression models
based on continuation ratios or adjacent-categories.

A practical question regarding (3.1) is on which response variable to be
used for the marginal model. The decision could be based on results from
two preliminary model fittings for each response variable using available-
case analysis and choose the better fitted model. Another important factor to consider is the observed sample sizes for the four groups of units discussed in Section 2. Modelling the response with a larger proportion of observed values provides a more accurate starting point. Let $y_1$ be the response variable chosen for the marginal model.

The joint, marginal and conditional probabilities of $(y_1, y_2)$ given $x$ are fully determined by (3.1). To see this, we first note that $P(y_1 = r \mid x; \theta_1)$ and $P(y_2 = j \mid x, y_1 = r; \theta_2)$ are directly available from (3.1) and hence

$$P(y_1 = r, y_2 = j \mid x; \theta_1, \theta_2) = P(y_1 = r \mid x; \theta_1)P(y_2 = j \mid x, y_1 = r; \theta_2).$$

(3.2)

It follows that

$$P(y_2 = j \mid x; \theta_1, \theta_2) = \sum_{r=1}^{R} P(y_1 = r, y_2 = j \mid x; \theta_1, \theta_2),$$

(3.3)

which further leads to

$$P(y_1 = r \mid x, y_2 = j; \theta_1, \theta_2) = \frac{P(y_1 = r, y_2 = j \mid x; \theta_1, \theta_2)}{P(y_2 = j \mid x; \theta_1, \theta_2)}. $$

(3.4)

The two models specified by (3.1) and the relations described in (3.2) - (3.4) are used for our proposed joint fractional imputation method. The single imputed data set is created in two stages as follows.

Stage One: Create imputed values for the bivariate ordinal responses
We impute each missing value by using all possible outcomes while keeping observed values unchanged. For fully observed units in $\mathcal{R}$, the corresponding observations remain the same. Imputed values for the missing responses are created based on the missing patterns.

1. For units in $\mathcal{P}_1$ with only $y_2$ missing, we replicate each observation $J$ times and impute the missing $y_2$ with values $1, 2, \ldots, J$.

2. For units in $\mathcal{P}_2$ with only $y_1$ missing, we replicate each observation $R$ times and impute the missing $y_1$ with values $1, 2, \ldots, R$.

3. For units in $\mathcal{M}$ with both $y_1$ and $y_2$ missing, each observation is replicated $RJ$ times with the missing responses $(y_1, y_2)$ replaced by all possible combinations $(r, j)$, $r = 1, 2, \ldots, R$ and $j = 1, 2, \ldots, J$.

Table 1 shows the structure of the imputed data set for a toy example with $n = 4$ observations, one for each of the four groups $\mathcal{R}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{M}$. The bivariate ordinal response variables each has two levels ($R = J = 2$) and there are three auxiliary variables. The imputed data set is an enlarged data file with the same number of variables as the initial sample and a total number of $n^* = n_r + Jn_{p1} + Rn_{p2} + RJn_m$ observations, where $n_r$, $n_{p1}$, $n_{p2}$ and $n_m$ are the size of group $\mathcal{R}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{M}$, respectively. For the simple example shown in Table 1 we have
Table 1: A Simple Example of Fractionally Imputed Data Set with $n = 4$
and $R = J = 2$.

<table>
<thead>
<tr>
<th>i</th>
<th>$y_{i1}$</th>
<th>$y_{i2}$</th>
<th>$\delta_{i1}$</th>
<th>$\delta_{i2}$</th>
<th>$x_{i1}$</th>
<th>$x_{i2}$</th>
<th>$x_{i3}$</th>
<th>$m$</th>
<th>$w_{m}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_{11}$</td>
<td>$y_{12}$</td>
<td>1</td>
<td>1</td>
<td>$x_{11}$</td>
<td>$x_{12}$</td>
<td>$x_{13}$</td>
<td>1</td>
<td>$w_{1}^*$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{21}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>$x_{23}$</td>
<td>2</td>
<td>$w_{2}^*$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{21}$</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>$x_{23}$</td>
<td>3</td>
<td>$w_{3}^*$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$y_{32}$</td>
<td>0</td>
<td>1</td>
<td>$x_{31}$</td>
<td>$x_{32}$</td>
<td>$x_{33}$</td>
<td>4</td>
<td>$w_{4}^*$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>$y_{32}$</td>
<td>0</td>
<td>1</td>
<td>$x_{31}$</td>
<td>$x_{32}$</td>
<td>$x_{33}$</td>
<td>5</td>
<td>$w_{5}^*$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$x_{41}$</td>
<td>$x_{42}$</td>
<td>$x_{43}$</td>
<td>6</td>
<td>$w_{6}^*$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>$x_{41}$</td>
<td>$x_{42}$</td>
<td>$x_{43}$</td>
<td>7</td>
<td>$w_{7}^*$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$x_{41}$</td>
<td>$x_{42}$</td>
<td>$x_{43}$</td>
<td>8</td>
<td>$w_{8}^*$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>$x_{41}$</td>
<td>$x_{42}$</td>
<td>$x_{43}$</td>
<td>9</td>
<td>$w_{9}^*$</td>
</tr>
</tbody>
</table>

$n_r = n_{p1} = n_{p2} = n_m = 1$, $J = R = 2$ and $n^* = 9$. We re-index the imputed data set with subscript $m$ and the fractionally imputed data can be represented by $D^* = \{ (y_m^*, \delta_m^*, x_m^*, w_m^*), m = 1, \ldots, n^* \}$, where values of $y_m^* = (y_{m1}^*, y_{m2}^*)$ are either observed or imputed, indicated by $\delta_m^* = (\delta_{m1}^*, \delta_{m2}^*)$. The imputed data file has an added column for the fractional weights $w_m^*$. It is a crucial part of the data file production and details are given below in “Stage Two”. For public-use data files, the columns for
Stage Two: Calculate fractional weights

Each observation in the imputed data set is accompanied by a fractional weight \( w^*_m \), which can be calculated iteratively by the following procedures.

1. Choose initial values \( \theta_1^{(0)}, \theta_2^{(0)} \) for the parameters in the models (3.1).

2. Define the general weight function as

\[
W(y, \delta, x; \theta_1, \theta_2) = \delta_1 \delta_2 + \delta_1 (1 - \delta_2) P(y_2 = y_2 | x, y_1 = y_1; \theta_2) \\
+ (1 - \delta_1) \delta_2 P(y_1 = y_1 | x, y_2 = y_2; \theta_1, \theta_2) \\
+ (1 - \delta_1)(1 - \delta_2) P(y_1 = y_1, y_2 = y_2 | x; \theta_1, \theta_2). \tag{3.5}
\]

Note that by the relations described in (3.2) - (3.4), the weight function is fully determined by the models in (3.1).

3. Calculate the initial fractional weights

\[
w^{(0)}_m = W(y^*_m, \delta^*_m, x^*_m; \theta_1^{(0)}, \theta_2^{(0)}), \quad m = 1, \ldots, n^*. \tag{3.6}
\]

4. Fit the two models in (3.1) using the imputed data set \( D^* \) with the weights \( w^{(0)}_m \) for the first iteration or the weights \( w^{(1)}_m \) from Step (5) for subsequent iterations and obtain updated estimates \( \theta_1^{(1)} \) and \( \theta_2^{(1)} \).
(5) Update the fractional weights as
\[ w_m^{(1)} = W(y_m^*, \delta_m^*, x_m^*, \theta_1^{(1)}, \theta_2^{(1)}), \quad m = 1, \ldots, n^*. \]

(6) Repeat Steps (4) and (5) until the fractional weights converge. Denote the final converged weights by \( w^* = (w_1^*, \ldots, w_n^*). \)

From the general weight function defined in Step (2), it can be seen that fully observed units from \( R \) receive weights 1. The imputed observations for units from the other three groups receive different fractional weights depending on which group the corresponding original unit belongs to.

The initial values \( \theta_1^{(0)}, \theta_2^{(0)} \) in Step (1) could be the estimates obtained by the available-case analysis method for the models in (3.1). More specifically, we can fit the marginal model with data from \( R \) and \( P_1 \), and fit the transitional model with data from \( R \) alone and use the resulting estimates as \( \theta_1^{(0)}, \theta_2^{(0)} \). A practical issue is that when the size of group \( R \) is too small, the transitional model may not be numerically identifiable. Should that be the case, we take initial values of \( \nu_r \) in the transitional model as 0 and estimate the remaining parameters in \( \theta_2 \) with data from \( R \) and \( P_2 \). Further details on using weights for Step (4) are given in Section 3.2. Issues with convergence for the final fractional weights are addressed in Section 3.3.

3.2. Analysis with fractionally imputed data set
With fractionally imputed data sets, estimation methods for complete data can be applied with a simple modification to incorporate the fractional weights. For example, the cell probabilities $\pi_{rj}$ can be estimated by

$$
\hat{\pi}_{rj}^{fi} = \frac{\sum_{m=1}^{n^*} w_m^* I(y_{m1}^* = r, y_{m2}^* = j)}{\sum_{m=1}^{n^*} w_m^*},
$$

(3.7)

where the superscript “$fi$” denotes “fractional imputation”. It is apparent from the procedures described in Section 3.1 that $\sum_{m=1}^{n^*} w_m^* = n$. The general expression in (3.7) is more convenient for complex survey data, not discussed in this paper, where the fractional weights are absorbed into the survey weights. The marginal probabilities $\pi_{r+}$ of $y_1$ can be similarly estimated by

$$
\hat{\pi}_{r+}^{fi} = \frac{\sum_{m=1}^{n^*} w_m^* I(y_{m1}^* = r)}{\sum_{m=1}^{n^*} w_m^*},
$$

(3.8)

The association parameter $\gamma$ can be estimated by $
\hat{\gamma}^{fi} = \frac{(C^{fi} - D^{fi})}{(C^{fi} + D^{fi})},
$
where $C^{fi} = 2 \sum_{r<k} \sum_{j<l} \hat{\pi}_{rj}^{fi} \hat{\pi}_{kl}^{fi}$ and $D^{fi} = 2 \sum_{r<k} \sum_{j>l} \hat{\pi}_{rj}^{fi} \hat{\pi}_{kl}^{fi}$. In general, for parameters defined as $g(\pi)$ where $g(\cdot)$ is a differentiable function, we can use the simple plug-in estimator $g(\hat{\pi}^{fi})$, where $\hat{\pi}^{fi} = (\hat{\pi}_{11}^{fi}, \ldots, \hat{\pi}_{1J}^{fi}, \ldots, \hat{\pi}_{R1}^{fi}, \ldots, \hat{\pi}_{RJ}^{fi})$ with elements given in (3.7).

Fitting regression models such as (3.1) with fractionally imputed data sets and the incorporation of the fractional weights can be carried out in similar ways as in (3.7) and (3.8) by solving weighted estimating equa-
tions. Further details can be found in She (2017). Variance estimation for fractionally imputed estimators will be discussed in Section 3.5.

3.3. Maximum likelihood interpretation

We now demonstrate that the weights from the proposed joint fractional imputation procedure do converge to a set of stable values. We show this by starting from the likelihood approach to estimating parameters in (3.1). In this section, the probability mass function of a discrete random variable is denoted by \( f(\cdot) \). The likelihood function of the observed data is given by

\[
L_{\text{obs}} = \prod_{i=1}^{n} \int f(\delta_i | x_i, y_i) f(y_i | x_i; \theta_1, \theta_2) d\mu(y_{i,\text{mis}}),
\]

where \( y_{i,\text{mis}} \) is the missing part of the bivariate responses. Under the MAR assumption, response indicators and the missing responses are conditionally independent given the observed responses and covariates, i.e., \( f(\delta | x, y) = f(\delta | x, y_{\text{obs}}) \), which does not involve \( y_{\text{mis}} \) and hence can be taken to the outside of the integral. We can re-write \( L_{\text{obs}} \) into two parts as

\[
L_{\text{obs}} = \prod_{i=1}^{n} f(\delta_i | x_i, y_{i,\text{obs}}) \prod_{i=1}^{n} \int f(y_{i1}, y_{i2} | x_i; \theta_1, \theta_2) d\mu(y_{i,\text{mis}}),
\]

with only the second part involving parameters \( \theta_1 \) and \( \theta_2 \). Noting that \( y_{1}, y_{2} \) are discrete variables, the integrals can be written as summations over all possible values. By considering the four groups of sampled units
separately, we can re-write $L_{\text{obs}}$ as

$$L_{\text{obs}} \propto \prod_{i \in R} f(y_{i1}, y_{i2} \mid x_i; \theta_1, \theta_2) \times \prod_{i \in P_1} \left[ \sum_{y_{i2} = 1}^R f(y_{i1}, y_{i2} \mid x_i; \theta_1, \theta_2) \right]$$

$$\times \prod_{i \in P_2} \left[ \sum_{y_{i1} = 1}^R f(y_{i1}, y_{i2} \mid x_i; \theta_1, \theta_2) \right],$$

(3.9)

where $f(y_{i1}, y_{i2} \mid x; \theta_1, \theta_2) = f(y_{i1} \mid x; \theta_1)f(y_{i2} \mid y_{i1}, x; \theta_2)$, which can be obtained from (3.1). The term involving group $\mathcal{M}$ vanishes because the double summation of the joint probability mass function equals 1. By taking derivatives of $l_{\text{obs}} = \log L_{\text{obs}}$ with respect to $\theta_1$ and $\theta_2$ and setting them equal to zeros, we obtain the set of score functions as

$$0 = \sum_{i \in R, P_1} S_1(y_{i1}, x_i; \theta_1) + \sum_{i \in P_2, y_{i1} = 1}^R S_1(y_{i1}, x_i; \theta_1) f(y_{i1} \mid y_{i2}, x_i; \theta_1, \theta_2),$$

$$0 = \sum_{i \in R} S_2(y_{i2}, y_{i1}, x_i; \theta_2) + \sum_{i \in P_2, y_{i1} = 1}^R S_2(y_{i2}, y_{i1}, x_i; \theta_2) f(y_{i1} \mid y_{i2}, x_i; \theta_1, \theta_2),$$

(3.10)

where $S_1(y_{i1}, x; \theta_1) = \partial \log f(y_{i1} \mid x; \theta_1) / \partial \theta_1$ and $S_2(y_{i2}, y_{i1}, x; \theta_2) = \partial \log f(y_{i2} \mid y_{i1}, x; \theta_2) / \partial \theta_2$ are the score functions of $\theta_1$ and $\theta_2$ when the marginal model and the transitional model are fitted separately with complete data and

$$f(y_{i1} \mid y_{i2}, x; \theta_1, \theta_2) = \frac{f(y_{i1} \mid x; \theta_1)f(y_{i2} \mid y_{i1}, x; \theta_2)}{\sum_{y_{i1} = 1}^R f(y_{i1} \mid x; \theta_1)f(y_{i2} \mid y_{i1}, x; \theta_2)}$$

(3.11)

is the derived conditional probability mass function of $y_{i1}$ given $y_{i2}$ and $x$. 

is the derived conditional probability mass function of $y_{i1}$ given $y_{i2}$ and $x$. 

It is difficult to solve the score equations \((3.10)\) directly. An alternative approach is to apply the EM algorithm (Dempster et al. (1977)) to find the maximum likelihood estimators of \(\theta_1, \theta_2\). Let \(\theta = (\theta_1', \theta_2')'\) be all the parameters and \(\theta^{(t)} = (\theta_1^{(t)}', \theta_2^{(t)}')'\) be the values after the \(t\)th iteration.

**E-step:** Calculate \(Q(\theta | \theta^{(t)}) = E \left\{ \sum_{i=1}^{n} \log f(y_i | x_i; \theta) | y_{obs}, \delta, x; \theta^{(t)} \right\} \), where \(y_{obs}\) denotes the observed part of \(y\). Following the same partition used for \(L_{obs}\), we can re-write \(Q(\theta | \theta^{(t)})\) into four terms:

\[
Q(\theta | \theta^{(t)}) = \sum_{i \in R} \log f(y_{i1}, y_{i2} | x_i; \theta) + \sum_{i \in P_1} \sum_{y_2=1}^{J} [\log f(y_{i1}, y_2 | x_i; \theta)] f(y_2 | y_{i1}, x_i; \theta_2^{(t)})
+ \sum_{i \in P_2} \sum_{y_1=1}^{R} [\log f(y_1, y_{i2} | x_i; \theta)] f(y_1 | y_{i2}, x_i; \theta_1^{(t)}, \theta_2^{(t)})
+ \sum_{i \in M} \sum_{y_1=1}^{R} \sum_{y_2=1}^{J} [\log f(y | x_i; \theta)] f(y | x_i; \theta^{(t)}).
\tag{3.12}
\]

**M-step:** Obtain \(\theta^{(t+1)}\) which maximize \(Q(\theta | \theta^{(t)})\) with respect to \(\theta\). Note that \(\theta_1\) and \(\theta_2\) in \(Q(\theta | \theta^{(t)})\) are separable. This leads to simpler forms of
score functions. For example, for $\theta_1$, the maximum point solves equations

$$\begin{align*}
0 &= \sum_{i \in R} S_1(y_{i1}, x_i; \theta_1) + \sum_{i \in P_1} \sum_{y_2 = 1}^J f(y_2 | y_{i1}, x_i; \theta_2^{(t)}) S_1(y_{i1}, x_i; \theta_1) \\
& \quad + \sum_{i \in P_2} \sum_{y_1 = 1}^R f(y_1 | y_{i2}, x_i; \theta_1^{(t)}, \theta_2^{(t)}) S_1(y_{i1}, x_i; \theta_1) \\
& \quad + \sum_{i \in M} \sum_{y_1 = 1}^R \sum_{y_2 = 1}^J f(y | x_i; \theta^{(t)}) S_1(y_{i1}, x_i; \theta_1). \tag{3.13}
\end{align*}$$

It is important for our following arguments to note that (3.13) are the same as the score equations obtained by fitting the marginal model with the imputed data set weighted by $w^*(t) = (w_1^*(t), \ldots, w_n^*(t))$, where $w_m^{(t)} = W(y_m^*, \delta_m^*, x_m^*; \theta_1^{(t)}, \theta_2^{(t)})$. Same results can also be shown for $\theta_2$. In other words, our proposed joint fractional imputation procedures have the same spirit as the EM algorithm.

The convergence properties of the EM algorithm were studied by Wu (1983). In our case, $Q(\theta | \theta^{(t)})$ is continuous with respect to $\theta$ and $\theta^{(t)}$, and hence the EM sequence $\{\theta_1^{(t)}, \theta_2^{(t)}\}$ converges to a stationary point $(\hat{\theta}_1, \hat{\theta}_2)$ which is the solution to the score equations (3.10). We summarize the above discussions with the following theorem.

**Theorem 3.1.** The fractional weights $\{w^{*(t)}\}$ defined in the proposed joint fractional imputation procedures converge to a stable set of values denoted...
by $w^*$ as $t \to \infty$, and the $m$th element of $w^*$ is given by

$$w^*_m = W(y^*_m, \delta^*_m, x^*_m; \hat{\theta}_1, \hat{\theta}_2),$$

where $(\hat{\theta}_1, \hat{\theta}_2)$ is the solution to the score equations (3.10).

From (3.9), data from group $M$ can be omitted for estimating $\theta_1, \theta_2$, which makes the fourth term in (3.12) unnecessary. It implies that our proposed JFI procedures can be simplified by excluding imputed units of group $M$ in iterations of Steps (4) and (5) and only updating the fractional weights for these units with the final estimates $\hat{\theta}_1, \hat{\theta}_2$.

3.4. Asymptotic properties of fractionally imputed estimators

We begin with the estimator $\hat{\pi}^{fi} = (\hat{\pi}_{r_11}^{fi}, ..., \hat{\pi}_{r_1J}^{fi}, ..., \hat{\pi}_{R_11}^{fi}, ..., \hat{\pi}_{R_J}^{fi})'$ of the vector $\pi$ of joint cell probabilities, where $\pi_{r_j}^{fi}$ is given in (3.7). Note that $\hat{\pi}_{r_j}^{fi}$ is a weighted sum of indicator functions of “non-independent” observations in the imputed data file. To investigate the asymptotic behaviour of $\hat{\pi}_{r_j}^{fi}$, it is essential to write it in the form of the original sample.

For the joint fractionally imputed data set, each observation in the original sample with one or both missing responses corresponds to a “bundle of observations” in the imputed file. For example, the $i_0$th observation $(y_{i_01}, *, 1, 0, x_{i_0})$ from group $P_1$ with $y_{i_02}$ missing corresponds to the bundle $\{(y_{i_01}, 1, 1, 0, x_{i_0}), \ldots, (y_{i_01}, J, 1, 0, x_{i_0})\}$. Suppose that this bundle of $J$
imputed data points are listed from the \(m_0\)th to \((m_0 + J - 1)\)th observations in the imputed data file \(D^*\). By the definition of \(w_m^*\) in Theorem 3.1, it is easy to see that \(\sum_{m=m_0}^{m_0+J-1} w_m^* = 1\). Since the \(J\) imputed values for \(y_{i_02}\) are deterministically filled as \(1, \ldots, J\), we further have

\[
\sum_{m=m_0}^{m_0+J-1} w_m^* I(y_{i_1}^* = r, y_{i_2}^* = j) = \sum_{m=m_0}^{m_0+J-1} w_m^* I(y_{i_01} = r, m - m_0 + 1 = j),
\]

and at most one term on the right hand side is non-zero, which is \(w_{m_0+j-1}^* I(y_{i_01} = r) = W((y_{i_01}, j), (1, 0), \hat{\theta}_1, \hat{\theta}_2) I(y_{i_01} = r)\). Similar arguments can be made for observations from other groups. Define the estimating function for \(\pi_{r_1}\) as

\[
U_{r_1}(y, \delta, x; \pi_{r_1}, \hat{\theta}_1, \hat{\theta}_2) = \delta_1 \delta_2 I(y_1 = r, y_2 = j) + \delta_1 (1 - \delta_2) W((y_1, j), (1, 0), \theta_1, \theta_2) I(y_1 = r) + (1 - \delta_1) \delta_2 W((r, y_2), (0, 1), \theta_1, \theta_2) I(y_2 = j) + (1 - \delta_1)(1 - \delta_2) W((r, j), (0, 0), \theta_1, \theta_2) - \pi_{r_1}. \tag{3.14}
\]

It can be seen that \(\hat{\pi}_{r_1}^{f_{i_1}}\) given in (3.7) is the same as the solution to the estimating equation

\[
0 = \sum_{i=1}^{n} U_{r_1}(y_i, \delta_i, x_i; \pi_{r_1}, \hat{\theta}_1, \hat{\theta}_2), \tag{3.15}
\]

which depends on preliminary estimators of \(\theta_1\) and \(\theta_2\). This two-step estimator \(\hat{\pi}_{r_1}\) can be more conveniently handled as a component of solutions.
to an extended system of estimating equations. Let
\[
S_{\text{obs}}^{(1)}(y, \delta, x; \theta_1, \theta_2) = E[S_1(y_1, x; \theta_1) \mid y_{\text{obs}}, \delta, x; \theta_1, \theta_2],
\]
\[
S_{\text{obs}}^{(2)}(y, \delta, x; \theta_1, \theta_2) = E[S_2(y_2, y_1, x; \theta_2) \mid y_{\text{obs}}, \delta, x; \theta_1, \theta_2].
\] (3.16)

The estimators \((\hat{\theta}_1, \hat{\theta}_2)\) are initially defined as the solution to the score equations \[3.10\] and can be re-written as the solution to
\[
0 = \sum_{i=1}^{n} S_{\text{obs}}^{(1)}(y_i, \delta_i, x_i; \theta_1, \theta_2), \quad 0 = \sum_{i=1}^{n} S_{\text{obs}}^{(2)}(y_i, \delta_i, x_i; \theta_1, \theta_2). \] (3.17)

Let \(U(\pi, \theta_1, \theta_2) = (U_{11}, ..., U_{1J}, ..., U_{R1}, ..., U_{RJ})', \ S_{\text{obs}}(\theta_1, \theta_2) = (S_{\text{obs}}^{(1)}', S_{\text{obs}}^{(2)}')', \ S(\theta_1, \theta_2) = (S_1', S_2')', \) where \(U_{rj}, S_{\text{obs}}^{(1)}, S_{\text{obs}}^{(2)}, S_1\) and \(S_2\) are short forms of functions defined in \[3.14\], \[3.10\] and \[3.16\]. The following theorem summarizes the asymptotic properties of \(\hat{\pi}_{fi}\). Proofs are outlined in the Supplementary Material.

**Theorem 3.2.** Let \(\pi_0, \theta_{10}\) and \(\theta_{20}\) be the true values of \(\pi, \theta_1\) and \(\theta_2\).

Under the regularity conditions specified in Supplementary Material, \(\hat{\pi}_{fi}\) with elements given by \[3.7\] is a consistent estimator of \(\pi\). Furthermore,
\[
n^{1/2}(\hat{\pi}_{fi} - \pi_0) \sim N\left(0, \text{Var}\left[U(\pi_0, \theta_{10}, \theta_{20}) + \kappa I_{\text{obs}}^{-1}S_{\text{obs}}(\theta_{10}, \theta_{20})\right]\right),
\]
where “\(\sim\)” represents “is asymptotically distributed as”,
\[
I_{\text{obs}} = \left(E[-\partial S_{\text{obs}}(\theta_1, \theta_2)/\partial \theta_1], E[-\partial S_{\text{obs}}(\theta_1, \theta_2)/\partial \theta_2]\right),
\]
evaluated at the true values of the parameters, \( \kappa = (\kappa'_{11}, ..., \kappa'_{1J}, ..., \kappa'_{R1}, ..., \kappa'_{RJ})' \), and

\[
\kappa_{rj} = E\left\{ I(y_1 = r, y_2 = j) \left[ S((r, j), x; \theta_{10}, \theta_{20}) - S_{obs}((r, j), \delta, x; \theta_{10}, \theta_{20}) \right] \right\}.
\]

**Corollary 3.2.1.** Let \( g(\pi) \) be a differentiable function of \( \pi \), either scalar or vector valued. Denote the asymptotic variance of \( n^{1/2}(\hat{\pi}_{fi} - \pi_0) \) given in Theorem 3.2 as \( \Sigma_{fi} \). Then \( g(\hat{\pi}_{fi}) \) is a consistent estimator of \( g(\pi) \) and

\[
n^{1/2}\left[ g(\hat{\pi}_{fi}) - g(\pi_0) \right] \sim N(0, \Gamma \Sigma_{fi} \Gamma'),
\]

where \( \Gamma = \partial g(\pi)/\partial \pi \) and is evaluated at \( \pi_0 \).

The corollary follows directly from the Continuous Mapping Theorem and the Delta method. The marginal probabilities, various types of odds ratios and association measures are all special cases with different \( g(\cdot) \). For example, the marginal probabilities of \( y_1 \) can be written as \( \pi_1 = C\pi \), where \( C = \text{diag}(1', \ldots, 1') \) is a \( R \times (RJ) \) block diagonal matrix and \( 1' = (1, \ldots, 1)' \) with length \( J \). It follows that \( \Gamma = C \) in this case.

### 3.5. Variance estimation

We now briefly discuss issues with variance estimation. The linearization method uses the expressions of asymptotic variances given in Corollary 3.2.1 and replaces unknown population quantities by estimates using the
imputed data set. For example, the quantity $\kappa_{rj}$ defined in Theorem 3.2 can be estimated by $\hat{\kappa}_{rj}$, which is computed as

$$\frac{1}{n} \sum_{m=1}^{n^*} I(y_{m1} = r, y_{m2} = j) \left[ S((r, j), \mathbf{x}_{m}^*; \hat{\theta}_1, \hat{\theta}_2) - S_{obs}((r, j), \delta_{m}^*, \mathbf{x}_{m}^*; \hat{\theta}_1, \hat{\theta}_2) \right].$$

The linearization method, however, requires detailed derivations of the asymptotic variance, which can be cumbersome for parameters with a complex structure such as $\gamma$. More importantly, the linearization method relies on full access to the information used in the imputation procedure, including the response indicators $\delta_i$ and all of the covariates $\mathbf{x}_i$. For public-use data files, some information is suppressed and not available to the data users, in which cases, the linearization method is not applicable.

Resampling methods such as the jackknife (Rao and Shao (1992)) and the bootstrap (Efron (1994)) are an attractive alternative approach for variance estimation with imputed estimators. Let $\mathbf{b}_i = (y_i, \delta_i, \mathbf{x}_i)$ denote the $i$th observation in the original data file $\mathcal{D}$. The bootstrap variance estimator of $g(\hat{\pi}^{fi})$ can be computed through the following steps:

1. Draw a simple random sample of size $n$ from the original sample $\mathcal{D}$ with replacement; denote the bootstrap sample as $\mathcal{B}_1 = \{\tilde{\mathbf{b}}_i^{(1)}, i = 1, \ldots, n\}$.

2. Apply the joint fractional imputation procedure to the bootstrap sample $\mathcal{B}_1$; let $\hat{\mathbf{\theta}}^{(1)} = (\hat{\theta}_1^{(1)Y}, \hat{\theta}_2^{(1)Y})'$ and $\hat{\pi}^{(1)}$ be the resulting estimate of $\theta = \ldots$
(\theta_1', \theta_2')' and \pi; compute g(\hat{\pi}^{(1)})

3) Repeat Steps (1) and (2) a large number B times; let \{g(\hat{\pi}^{(1)}), \ldots, g(\hat{\pi}^{(B)})\} be the resulting estimates from the repeated bootstrap samples. The bootstrap variance estimator of g(\hat{\pi}^{fi}) is computed as

$$\text{var}[g(\hat{\pi}^{fi})] = \frac{1}{B} \sum_{k=1}^{B} [g(\hat{\pi}^{(k)}) - B^{-1} \sum_{k=1}^{B} g(\hat{\pi}^{(k)})]^2.$$ 

The validity of the bootstrap variance estimator is discussed in the Supplementary Material. The resampling methods are often preferred for creating public-use files, where the fractional weights based on the bootstrap samples are attached as additional columns of replication weights to the data file and variance estimation is done by repeatedly applying the standard analysis with these replication weights.

A practical issue with the resampling methods, especially for the bootstrap approach, is that when the sample size is small, the algorithm may not converge numerically for some bootstrap samples. In our simulation studies discussed in Section 4 with sample size n = 200 and n = 500, the occurrence rate of such “singular” cases is negligible. But for smaller sample sizes, this problem needs to be properly dealt with. From the arguments given in the Supplementary Material, most of the variation of the bootstrap estimator \hat{\theta}^{(k)} = (\hat{\theta}_1^{(k)'}, \hat{\theta}_2^{(k)'}')' can be captured by the first-order Taylor expansion.
around $\hat{\theta}$. Therefore, a possible workaround is to use the one-step Newton method discussed in Yang and Kim (2016), where for every bootstrap sample, $\hat{\theta}^{(k)}$ is calculated by one-step iteration from $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')'$:

$$
\hat{\theta}^{(k)} = \hat{\theta} - \left\{ \sum_{m=1}^{n^{(k)*}} \hat{H}(y_m^*, \delta_m^*, x_m^*; \hat{\theta}_1, \hat{\theta}_2) \right\}^{-1} \left\{ \sum_{m=1}^{n^{(k)*}} H(y_m^*, \delta_m^*, x_m^*; \hat{\theta}_1, \hat{\theta}_2) \right\},
$$

where $n^{(k)*}$ is the size of the imputed data file created in Stage One based on the $k$th bootstrap sample,

$$
H(y, \delta, x; \theta_1, \theta_2) = W(y, \delta, x; \theta_1, \theta_2)S(y, x; \theta_1, \theta_2)
$$

and $\hat{H}(\theta_1, \theta_2) = \partial H(\theta_1, \theta_2)/\partial (\theta_1', \theta_2')$. We then obtain the fractional weights and estimator $\hat{\pi}^{(k)}$ based on $\hat{\theta}^{(k)}$.

4. Simulation Studies

We report results from simulation studies on the finite sample performance of the proposed estimators under the joint fractional imputation, with comparisons to existing methods. We consider bivariate ordinal responses $(y_1, y_2)$, each with three categories, and two covariates: a continuous variable $x_1$ generated from $\text{Exp}(1)$ and a discrete variable $x_2$ following $\text{Bernoulli}(0.5)$. The responses $(y_1, y_2)$ follow the marginal and the transitional models given in (3.1). In order to apply the PSA method, we simulate the response indicators in a way that the propensity scores follow a baseline-category logit model.
The parameters in the propensity score models are carefully chosen such that the proportions of units in the four groups $\mathcal{R}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{M}$ are controlled to have desirable patterns to mimic two real world scenarios. The first scenario has the majority of the sample fully observed, with proportions being $(0.5, 0.2, 0.2, 0.1)$ for the four groups. For the second scenario, only one of the two responses is observed for the majority of sampled units, with the proportions being $(0.2, 0.3, 0.4, 0.1)$. The simulation studies consist of three parts: (i) Point estimators; (ii) Variance estimators; and (iii) Tests of independence.

Table 2 presents results from the first part of the simulation on Absolute Relative Bias (ARB, in %) and Mean Squared Error (MSE, multiplied by $10^4$) of different estimators of the first element $\pi_{+1}$ of the marginal probabilities of $y_2$ under the two response patterns (RP, indicated by 5221 and 2341) and two sample sizes $n = 200$ and $n = 500$. The complete sample estimator without any missing values is denoted by COMP and is listed as the gold-standard reference; the estimator from available-case analysis is denoted as ACA; the propensity score adjusted estimator is indicated by PSA; the SRMI method with 5 imputed data sets is denoted by SRMI5. Our proposed joint fractional imputation estimator is denoted by JFI. Simulation results for the association measure $\gamma$ are summarized in Table 3.
Table 2: ARB (in %) and MSE (×10^4) for Estimating \( \pi_{+1} \)

<table>
<thead>
<tr>
<th>RP</th>
<th>n</th>
<th>COMP</th>
<th>ACA</th>
<th>PSA</th>
<th>SRMI5</th>
<th>JFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>5221</td>
<td>200</td>
<td>ARB</td>
<td>0.2</td>
<td>7.1</td>
<td>0.04</td>
<td>1.0</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>MSE</td>
<td>(8.9)</td>
<td>(14.7)</td>
<td>(12.1)</td>
<td>(11.8)</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>MSE</td>
<td>(3.6)</td>
<td>(8.3)</td>
<td>(5.2)</td>
<td>(5.0)</td>
</tr>
<tr>
<td>2341</td>
<td>200</td>
<td>ARB</td>
<td>—</td>
<td>8.8</td>
<td>0.008</td>
<td>1.5</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>MSE</td>
<td>—</td>
<td>(20.6)</td>
<td>(12.9)</td>
<td>(12.5)</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>MSE</td>
<td>—</td>
<td>(10.5)</td>
<td>(5.1)</td>
<td>(5.1)</td>
</tr>
</tbody>
</table>

The simulation results show clearly that the ACA estimator is not consistent for either the marginal probability \( \pi_{+1} \) or the association measure \( \gamma \). The other three methods PSA, SRMI5 and JFI provide comparably results for estimating marginal probabilities with negligible biases. However, for the estimation of \( \gamma \), the PSA estimator is far less efficient than the two imputation-based estimators. The SRMI estimator is close to the proposed JFI estimator under the first response pattern but has unreasonably large biases under the second scenario where there are only 20% of the
Table 3: ARB (in %) and MSE ($\times 10^4$) for Estimating $\gamma$

<table>
<thead>
<tr>
<th>RP</th>
<th>$n$</th>
<th>COMP</th>
<th>ACA</th>
<th>PSA</th>
<th>SRMI5</th>
<th>JFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>5221</td>
<td>200</td>
<td>ARB</td>
<td>0.5</td>
<td>8.7</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>(7.3)</td>
<td>(15.0)</td>
<td>(25.4)</td>
<td>(13.2)</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>0.04</td>
<td>9.1</td>
<td>0.4</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>(2.8)</td>
<td>(6.9)</td>
<td>(13.1)</td>
<td>(5.3)</td>
<td>(5.0)</td>
</tr>
<tr>
<td>2341</td>
<td>200</td>
<td>ARB</td>
<td>—</td>
<td>10.9</td>
<td>5.0</td>
<td>10.6</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>—</td>
<td>(42.3)</td>
<td>(87.8)</td>
<td>(27.9)</td>
<td>(25.1)</td>
</tr>
<tr>
<td>500</td>
<td>ARB</td>
<td>—</td>
<td>12.5</td>
<td>3.9</td>
<td>7.8</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>—</td>
<td>(17.2)</td>
<td>(56.8)</td>
<td>(11.3)</td>
<td>(9.5)</td>
</tr>
</tbody>
</table>

sampled units having both responses observed. Our proposed JFI estimator performs well for all cases and is uniformly better than the alternative methods considered in the simulation.

The second part of the simulation is on variance estimation. For the SRMI method, the variance estimator uses Rubin’s combining rule; for the JFI method, two versions of variance estimators are considered: the linearization method (JFIL) and the bootstrap method (JFIB). Table 4 reports the Absolute Relative Bias (ARB, in %) of the variances estimators.
for the two parameters $\pi_{+1}$ and $\gamma$. For estimating $\pi_{+1}$, all variance estimators have acceptable ARB. For estimating $\gamma$, the variance estimator of the SRMI estimator has large negative biases, which implies that the variance estimator based on Rubin’s combining rule underestimates the true variance. Both the linearization and the bootstrap variance estimators for the JFI method are consistent.

Table 4: ARB (in %) of Variance Estimators for $\pi_{+1}$ and $\gamma$

<table>
<thead>
<tr>
<th></th>
<th>$\pi_{+1}$</th>
<th>SRMI5</th>
<th>JFIL</th>
<th>JFIB</th>
<th>$\gamma$</th>
<th>SRMI5</th>
<th>JFIL</th>
<th>JFIB</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5221</td>
<td>200</td>
<td>3.0</td>
<td>5.7</td>
<td>3.7</td>
<td>(−)16.9</td>
<td>5.1</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>8.0</td>
<td>1.3</td>
<td>2.2</td>
<td>(−)16.0</td>
<td>4.1</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2341</td>
<td>200</td>
<td>4.4</td>
<td>4.0</td>
<td>3.7</td>
<td>(−)24.2</td>
<td>1.4</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>7.5</td>
<td>2.1</td>
<td>1.3</td>
<td>(−)26.4</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The third part of the simulation is on test of independence between the two ordinal responses. We use the Wald-type test statistic given in (2.2) based on a particular pair of point and variance estimators with significant level at 0.05. By tuning the parameters in (3.1), we simulate the power of tests for a series of cases where the true value of the association measure
\( \gamma \) increases from 0 to 1, departing gradually from the null hypothesis of independence.

The power of a test is computed as the simulated rejection probability under the given scenario. Plots of the power function for missing pattern 2341 are shown in Figures 1 and 2, corresponding to sample sizes at \( n = 200 \) and 500. Each plot shows the power functions of three different tests: \texttt{JFI\_non}, \texttt{JFI\_nul} and \texttt{SRMI}. The first test uses the regular linearization variance estimator without considering the null hypothesis; the second test uses the linearization variance estimator under the null hypothesis (i.e., \( \pi_{rj} = \pi_r + \pi_j \)); the third test uses the regular point and variance estimators for the \texttt{SRMI} method. Test results using bootstrap variance estimators are very similar to the ones using linearization variance estimators and not reported here to save space. The horizontal line in each figure represents the nominal value 0.05 for the level of the test. Plots for missing pattern 5221 are presented in Supplementary Material.

There are three major observations from the power functions: (i) The test based on the \texttt{SRMI} method has type I errors bigger than the nominal value 0.05, and it becomes more pronounced when the sample size is small or the proportion of units in \( \mathcal{R} \) is small. (ii) The type I errors for the two \texttt{JFI}-based tests are very close to the nominal value and both tests have
similar power. (iii) The response patterns have significant impact on the power of the tests, with the pattern 5221 producing more powerful tests than the pattern 2341. The first observation is in line with the results on underestimation of variance for the SRMI method. The second observation shows that there is no significant advantage of using the variance estimator under the null hypothesis. The last observation is in agreement with common sense since data with the pattern 5221 provide more information on the association between the two response variables than the other pattern.

5. Concluding Remarks

Statistical analysis with missing data faces two distinct scenarios. It could be a data set of small or moderate size collected for specific scientific purposes and the analysis is carried out by specific researchers who have full access to the data set and are equipped with a profound knowledge of statistics. It has become increasingly common, however, that data sets are collected by a large research team or a statistical agency and contain information on many variables. The researchers handling missing data only serve as data suppliers who create one or several complete data sets with missing values properly treated. The processed data sets are supposed to be released to or can be accessed by multiple users with possibly restricted access for different research objectives. Imputation for missing values is
widely accepted for creating public-use data files to provide a consistent platform for multiple users.

Our proposed joint fractional imputation method for bivariate ordinal responses possesses several attractive features. It is fully capable of dealing with the first scenario. The procedure produces a single imputed data set that leads to valid and efficient inferences for commonly encountered analysis problems. Our discussions in this paper on validity and efficiency of analysis with the fractionally imputed data set focus on estimation of joint and marginal probabilities and association measures and on test of independence. Regression analysis is only discussed as part of the model building process for the imputation procedure. It is shown in [She (2017)] that the fractionally imputed data set also leads to valid regression analysis involving one or both ordinal responses if the set of regressors for the analysis model is the same or a subset of the covariates used in the imputation model (3.1). The proposed procedure accompanied by the resampling methods described in Section 3.5 is ideally suited for creating public-use data files in the second scenario, particularly for large complex survey data. The factional weights become part of the survey weights and variance estimation is done through the use of additional columns of replication weights. The proposed procedure still provides valid inference even when the data users
only have partial access to the available information.

**Supplementary Material**

The Supplementary Material contains elaborate discussions on modeling techniques for complete ordinal responses and a detailed review on existing methods for handling missing ordinal observations. Regularity conditions, proof of Theorem 3.2 and justification of the bootstrap variance estimator, as well as two additional plots for the power functions of the tests, are also presented.

**Acknowledgements**

This research is supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada and a Collaborative Research Team grant from the Canadian Statistical Sciences Institute (CANSSI).

**References**


*Biometrika, 73*, 13–22.


Figure 1: Power Function with \( n = 200 \) and Pattern 2341

Figure 2: Power Function with \( n = 500 \) and Pattern 2341