| Statistica Sinica Preprint No: SS-2016-0396.R2 | |
|---|---|
| **Title** | Fully Efficient Joint Fractional Imputation for Incomplete Bivariate Ordinal Responses |
| **Manuscript ID** | SS-2016-0396.R2 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202016.0396 |
| **Complete List of Authors** | Xichen She and Changbao Wu |
| **Corresponding Author** | Changbao Wu |
| **E-mail** | cbwu@uwaterloo.ca |

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396

Statistica Sinica **27** (2017), 000-000
doi:http://dx.doi.org/10.5705/ss.20??.???

# 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 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 tests of independence are also discussed under the proposed joint fractional imputation method.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



*Key words and phrases:* Association measure, fractional imputation, contingency table, marginal probabilities, sequential regression multiple imputation.

## 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: contingency table analysis and regression analysis. Contingency table analysis focuses mainly on the joint and the marginal distribution of the two variables and, more importantly, the interrelation between responses; 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 of related techniques.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 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 a considerable development recent years of the theory and application of methods for handling missing data. Multiple imputation (MI), formally proposed by Rubin (1987), has gained tremendous popularity among users of in-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



complete 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 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



later studied by Kim and Fuller (2004) and Kim (2011), is an attractive alternative to multiple imputation for reducing imputation variance. It replaces each missing observation by a cluster of plausible values with each imputed value 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 that 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, and 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



responses. In Section 3, we present our proposed method and establish its 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 $\mathcal{D} = \big\{(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i),\ \ i = 1, \ldots, n\big\}$, where $\boldsymbol{y}_i = (y_{i1}, y_{i2})$ are ordinal responses on $R$-level and $J$-level scales, respectively, and that both are partially observed. Let $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2})$ be the corresponding response indicators: $\delta_{it} = 1$ if $y_{it}$ is observed and $\delta_{it} = 0$ otherwise, $t = 1, 2$. The vector $\boldsymbol{x}_i$ consists of fully observed auxiliary variables in the data file. We assume that the data set is an independent sample of size $n$ from $(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x})$. Units in the sample can be partitioned into four groups, depending on the missing pattern of the responses:

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

We consider scenarios where the responses are missing-at-random (MAR) as termed by Little and Rubin (2002) such that $(\boldsymbol{\delta} \perp \boldsymbol{y}_{mis}) \,|\, (\boldsymbol{y}_{obs}, \boldsymbol{x})$, where $\boldsymbol{y}_{mis}$ and $\boldsymbol{y}_{obs}$ are respectively the missing and the observed component(s) of $\boldsymbol{y}$. This implies that $P(\delta_1 = 1, \delta_2 = 0 \,|\, \boldsymbol{y}, \boldsymbol{x}) = P(\delta_1 = 1, \delta_2 = 0 \,|\, y_1, \boldsymbol{x})$,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



$P(\delta_1 = 0, \delta_2 = 1 \,|\, \boldsymbol{y}, \boldsymbol{x}) = P(\delta_1 = 0, \delta_2 = 1 \,|\, y_2, \boldsymbol{x})$ and $P(\delta_1 = 0, \delta_2 = 0 \,|\, \boldsymbol{y}, \boldsymbol{x}) = P(\delta_1 = 0, \delta_2 = 0 \,|\, \boldsymbol{x})$. The MAR assumption is less restrictive than the monotone missingness often used for longitudinal data and is sufficient for the justification of our proposed procedure 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 $\boldsymbol{\pi} = (\pi_{11}, \ldots, \pi_{1J}, \ldots, \pi_{R1}, \ldots, \pi_{RJ})'$ 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 $\boldsymbol{\pi}_1 = (\pi_{1+}, \ldots, \pi_{R+})'$ and $\boldsymbol{\pi}_2 = (\pi_{+1}, \ldots, \pi_{+J})'$, where $\pi_{r+} = \sum_{j=1}^{J} \pi_{rj}$ and $\pi_{+j} = \sum_{r=1}^{R} \pi_{rj}$. As dependence between the two ordinal responses is often the main focus for analysis of bivariate data, measures of association are of primary concern. A simple example is the conditional distribution of $y_1$ given $y_2$ at level $j$: $\boldsymbol{\pi}_{1|j} = (\pi_{1|j}, \ldots, \pi_{R|j})'$, $j = 1, \ldots, J$, where $\pi_{r|j} = P(y_1 = r \,|\, y_2 = j) = \pi_{rj}/\pi_{+j}$. Another popular example is a set of different types of ordinal odds ratios, including the local, the cumulative and the global odds ratios.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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

$$\gamma = \Big(\prod\nolimits_c - \prod\nolimits_d\Big)\Big/\Big(\prod\nolimits_c + \prod\nolimits_d\Big) \ , \tag{2.1}$$

where $\prod_c = 2\sum_{r<k}\sum_{j<l}\pi_{rj}\pi_{kl}$ and $\prod_d = 2\sum_{r<k}\sum_{j>l}\pi_{rj}\pi_{kl}$ , 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 $y_{i2} \le y_{m2}$. When $y_1$ and $y_2$ are independent, we have $\gamma = 0$, but the reverse statement is not true. Other examples of association measures include Kendall's *Tau-b* (Kendall (1945)) and Somers' $d$ (Somers (1962)), 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 $D = 2\sum_{r<k}\sum_{j>l}\hat{\pi}_{rj}\hat{\pi}_{kl}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



Simon (1978) 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

$$z = \big(C - D\big)/\hat{\sigma}_{\mathrm{C-D}}\,, \tag{2.2}$$

where $\hat{\sigma}_{\mathrm{C-D}}$ 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 (2010) recommended use of the latter and claimed that the test statistic with *null standard error* converges to the normal distribution faster under the null hypothesis. The Pearson $\chi^2$ test is also applicable, but it is designed for a general alternative and may not have good power for testing a trend, which is of primary interest for ordinal responses. The $z$ statistic given in (2.2) is very natural for alternative hypotheses such as $\prod_c > \prod_d$ or $\prod_c < \prod_d$ , 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 propensity score adjustment (PSA) and sequential regression multiple imputation (SRMI). Details of these two methods are given in Supplementary Material

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 approach to bivariate ordinal responses with missing values. We combine the modelling strategies from the SRMI method with the specific feature of ordinal 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 procedure, 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 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



models

$$\begin{aligned} \text{Marginal:} \quad & g_1(\eta_{r1}) \;=\; \alpha_{r1} - \boldsymbol{\beta}_1'\boldsymbol{x}, \\ \text{Transitional:} \quad & g_2(\eta_{j2}) \;=\; \alpha_{j2} - \boldsymbol{\beta}_2'\boldsymbol{x} - \sum_{r=2}^{R} \nu_r \boldsymbol{I}(y_1 = r), \end{aligned} \tag{3.1}$$

where $\eta_{r1} = P(y_1 \le r \,|\, \boldsymbol{x})$ and $\eta_{j2} = P(y_2 \le j \,|\, y_1, \boldsymbol{x})$ are the cumulative probabilities given the covariates, and $g_1$ and $g_2$ are link functions. Let $\boldsymbol{\theta}_1 = (\alpha_{11}, ..., \alpha_{R1}, \boldsymbol{\beta}_1')'$ be the parameters in the marginal model and $\boldsymbol{\theta}_2 = (\alpha_{12}, ..., \alpha_{J2}, \boldsymbol{\beta}_2', \nu_2, ..., \nu_R)'$ 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 *logit*, *probit* and *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, choosing 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 $\boldsymbol{x}$ are fully determined by (3.1). Here $P(y_1 = r \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1)$ and $P(y_2 = j \,|\, \boldsymbol{x}, y_1 = r; \boldsymbol{\theta}_2)$ are directly available from (3.1), and hence

$$P(y_1 = r, y_2 = j \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = P(y_1 = r \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1) P(y_2 = j \,|\, \boldsymbol{x}, y_1 = r; \boldsymbol{\theta}_2)\,. \tag{3.2}$$

It follows that

$$P(y_2 = j \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{r=1}^{R} P(y_1 = r, y_2 = j \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\,, \tag{3.3}$$

which further leads to

$$P(y_1 = r \,|\, \boldsymbol{x}, y_2 = j; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{P(y_1 = r, y_2 = j \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{P(y_2 = j \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}. \tag{3.4}$$

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.

*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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



(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 sizes of groups $\mathcal{R}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{M}$, respectively. For the simple example shown in Table 1 we have $n_r = n_{p1} = n_{p2} = n_m = 1$, $J = R = 2$ and $n^* = 9$. We re-index the imputed data set with subscript $m$ and the fractionally imputed data can be represented by $\mathcal{D}^* = \left\{(\boldsymbol{y}^*_m, \boldsymbol{\delta}^*_m, \boldsymbol{x}^*_m, w^*_m),\ m = 1, \ldots, n^*\right\}$, where values of $\boldsymbol{y}^*_m = (y^*_{m1}, y^*_{m2})$ are either observed or imputed, indicated by $\boldsymbol{\delta}^*_m = (\delta^*_{m1}, \delta^*_{m2})$. The imputed

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



Table 1: A Simple Example of Fractionally Imputed Data Set with $n = 4$ and $R = J = 2$.

| i | $y_{i1}$ | $y_{i2}$ | $\delta_{i1}$ | $\delta_{i2}$ | $x_{i1}$ | $x_{i2}$ | $x_{i3}$ | $m$ | $w_m^*$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | 1 | 1 | $x_{11}$ | $x_{12}$ | $x_{13}$ | 1 | $w_1^*$ |
| 2 | $y_{21}$ | 1 | 1 | 0 | $x_{21}$ | $x_{22}$ | $x_{23}$ | 2 | $w_2^*$ |
| 2 | $y_{21}$ | 2 | 1 | 0 | $x_{21}$ | $x_{22}$ | $x_{23}$ | 3 | $w_3^*$ |
| 3 | 1 | $y_{32}$ | 0 | 1 | $x_{31}$ | $x_{32}$ | $x_{33}$ | 4 | $w_4^*$ |
| 3 | 2 | $y_{32}$ | 0 | 1 | $x_{31}$ | $x_{32}$ | $x_{33}$ | 5 | $w_5^*$ |
| 4 | 1 | 1 | 0 | 0 | $x_{41}$ | $x_{42}$ | $x_{43}$ | 6 | $w_6^*$ |
| 4 | 1 | 2 | 0 | 0 | $x_{41}$ | $x_{42}$ | $x_{43}$ | 7 | $w_7^*$ |
| 4 | 2 | 1 | 0 | 0 | $x_{41}$ | $x_{42}$ | $x_{43}$ | 8 | $w_8^*$ |
| 4 | 2 | 2 | 0 | 0 | $x_{41}$ | $x_{42}$ | $x_{43}$ | 9 | $w_9^*$ |

data file has an added column for the fractional weights $w_m^*$. This 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 $\delta_{m1}^*$ and $\delta_{m2}^*$ and those for components of $\boldsymbol{x}$ that are of sensitive nature might be removed before the release of the file for confidentiality considerations.

*Stage Two: Calculate fractional weights*

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



Each observation in the imputed data set is accompanied by a fractional weight $w_m^*$ that can be calculated iteratively as follows.

(1) Choose initial values $\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}$ for the parameters in the models (3.1).

(2) Define the general weight function as

$$\begin{aligned} W(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \; & \delta_1\delta_2 + \delta_1(1-\delta_2)\, P(y_2 = y_2 \,|\, \boldsymbol{x}, y_1 = y_1; \boldsymbol{\theta}_2) \\ & + (1-\delta_1)\delta_2\, P(y_1 = y_1 \,|\, \boldsymbol{x}, y_2 = y_2; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ & + (1-\delta_1)(1-\delta_2)\, P(y_1 = y_1, y_2 = y_2 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \quad (3.5) \end{aligned}$$

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

$$w_m^{*(0)} = W(\boldsymbol{y}_m^*, \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}), \quad m = 1, \ldots, n^*. \tag{3.6}$$

(4) Fit the two models in (3.1) using the imputed data set $\mathcal{D}^*$ with the weights $w_m^{*(0)}$ for the first iteration or the weights $w_m^{*(1)}$ from Step (5) for subsequent iterations and obtain updated estimates $\boldsymbol{\theta}_1^{(1)}$ and $\boldsymbol{\theta}_2^{(1)}$.

(5) Update the fractional weights as

$$w_m^{*(1)} = W(\boldsymbol{y}_m^*, \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}), \quad m = 1, \ldots, n^*.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



(6) Repeat Steps (4) and (5) until the fractional weights converge. Denote the final converged weights by $\boldsymbol{w}^* = (w_1^*, \ldots, w_{n^*}^*)$.

From the general weight function defined in Step (2), it can be seen that fully observed units from $\mathcal{R}$ receive weight 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 $\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}$ in Step (1) can be the estimates obtained by the available-case analysis method for the models in (3.1): we can fit the marginal model with data from $\mathcal{R}$ and $\mathcal{P}_1$, and fit the transitional model with data from $\mathcal{R}$ alone and use the resulting estimates as $\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}$. A practical issue is that, when the size of group $\mathcal{R}$ 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 $\boldsymbol{\theta}_2$ with data from $\mathcal{R}$ and $\mathcal{P}_2$. 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



weights. For example, the cell probabilities $\pi_{rj}$ can be estimated by

$$\hat{\pi}_{rj}^{fi} = \sum_{m=1}^{n^*} w_m^* \boldsymbol{I}(y_{m1}^* = r, y_{m2}^* = j) / \sum_{m=1}^{n^*} w_m^* \,, \tag{3.7}$$

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 marginal probabilities $\pi_{r+}$ of $y_1$ can be similarly estimated by

$$\hat{\pi}_{r+}^{fi} = \sum_{m=1}^{n^*} w_m^* \boldsymbol{I}(y_{m1}^* = r) / \sum_{m=1}^{n^*} w_m^* \,. \tag{3.8}$$

The association parameter $\gamma$ can be estimated by $\hat{\gamma}^{fi} = \big(C^{fi} - D^{fi}\big)/\big(C^{fi} + D^{fi}\big)$, 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 $\boldsymbol{g}(\boldsymbol{\pi})$ where $\boldsymbol{g}(\cdot)$ is a differentiable function, we can use the simple plug-in estimator $\boldsymbol{g}(\hat{\boldsymbol{\pi}}^{fi})$, where $\hat{\boldsymbol{\pi}}^{fi} = (\hat{\pi}_{11}^{fi}, \ldots, \hat{\pi}_{1J}^{fi}, \ldots, \hat{\pi}_{R1}^{fi}, \ldots, \hat{\pi}_{RJ}^{fi})$ 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 equations. 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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

$$L_{obs} = \prod_{i=1}^{n} \int f(\boldsymbol{\delta}_i \,|\, \boldsymbol{x}_i, \boldsymbol{y}_i) f(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\mu(\boldsymbol{y}_{i,mis}) \,,$$

where $\boldsymbol{y}_{i,mis}$ is the missing part of the bivariate responses. Under the MAR assumption, $f(\boldsymbol{\delta} \,|\, \boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{\delta} \,|\, \boldsymbol{x}, \boldsymbol{y}_{obs})$, which does not involve $\boldsymbol{y}_{mis}$ and hence can be taken to the outside of the integral. We can re-write $L_{obs}$ as

$$L_{obs} = \prod_{i=1}^{n} f(\boldsymbol{\delta}_i \,|\, \boldsymbol{x}_i, \boldsymbol{y}_{i,obs}) \prod_{i=1}^{n} \int f(y_{i1}, y_{i2} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\mu(y_{i,mis}) \,,$$

with only the second part involving parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. As $y_1, y_2$ 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 $L_{obs}$ as

$$\begin{aligned} L_{obs} \;\propto\; & \prod_{i \in \mathcal{R}} f(y_{i1}, y_{i2} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \times \prod_{i \in \mathcal{P}_1} \Big[ \sum_{y_2=1}^{J} f(y_{i1}, y_2 \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \Big] \\ & \times \prod_{i \in \mathcal{P}_2} \Big[ \sum_{y_1=1}^{R} f(y_1, y_{i2} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \Big] \,, \end{aligned} \tag{3.9}$$

where $f(y_1, y_2 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f(y_1 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1) f(y_2 \,|\, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)$, 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



derivatives of $l_{obs} = \log L_{obs}$ with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ and setting them equal to zero, we obtain the set of score functions as

$$\begin{aligned}\mathbf{0} &= \sum_{i\in\mathcal{R},\mathcal{P}_1} \boldsymbol{S}_1(y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_1) + \sum_{i\in\mathcal{P}_2}\sum_{y_1=1}^{R} \boldsymbol{S}_1(y_1, \boldsymbol{x}_i; \boldsymbol{\theta}_1) f(y_1 \,|\, y_{i2}, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\ \mathbf{0} &= \sum_{i\in\mathcal{R}} \boldsymbol{S}_2(y_{i2}, y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_2) + \sum_{i\in\mathcal{P}_2}\sum_{y_1=1}^{R} \boldsymbol{S}_2(y_{i2}, y_1, \boldsymbol{x}_i; \boldsymbol{\theta}_2) f(y_1 \,|\, y_{i2}, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2),\end{aligned} \tag{3.10}$$

where $\boldsymbol{S}_1(y_1, \boldsymbol{x}; \boldsymbol{\theta}_1) = \partial \log f(y_1 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1)/\partial\boldsymbol{\theta}_1$ and $\boldsymbol{S}_2(y_2, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2) = \partial \log f(y_2 \,|\, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2$ are the score functions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ when the marginal model and the transitional model are fitted separately with complete data, and

$$f(y_1 \,|\, y_2, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{f(y_1 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1) f(y_2 \,|\, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)}{\sum_{y_1=1}^{R} f(y_1 \,|\, \boldsymbol{x}; \boldsymbol{\theta}_1) f(y_2 \,|\, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)} \tag{3.11}$$

is the derived conditional probability mass function of $y_1$ given $y_2$ and $\boldsymbol{x}$.

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 $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ be all the parameters and $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\theta}_1^{(t)\prime}, \boldsymbol{\theta}_2^{(t)\prime})'$ be the values after the $t$th iteration.

**E-step**: Calculate $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)}) = E\Big\{\sum_{i=1}^{n} \log f(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}) \,|\, \boldsymbol{y}_{obs}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}^{(t)}\Big\}$, where $\boldsymbol{y}_{obs}$ denotes the observed part of $\boldsymbol{y}$. Following the same partition as

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



used for $L_{obs}$, we can re-write $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})$ as:

$$\begin{aligned} Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)}) = & \sum_{i \in \mathcal{R}} \log f(y_{i1}, y_{i2} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}) \\ & + \sum_{i \in \mathcal{P}_1} \sum_{y_2=1}^{J} \big[\log f(y_{i1}, y_2 \,|\, \boldsymbol{x}_i; \boldsymbol{\theta})\big] f(y_2 \,|\, y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_2^{(t)}) \\ & + \sum_{i \in \mathcal{P}_2} \sum_{y_1=1}^{R} \big[\log f(y_1, y_{i2} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta})\big] f(y_1 \,|\, y_{i2}, \boldsymbol{x}_i; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \\ & + \sum_{i \in \mathcal{M}} \sum_{y_1=1}^{R} \sum_{y_2=1}^{J} \big[\log f(\boldsymbol{y} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta})\big] f(\boldsymbol{y} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}^{(t)}) \,. \end{aligned} \tag{3.12}$$

**M-step**: Obtain $\boldsymbol{\theta}^{(t+1)}$ to maximize $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$. Here $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})$ are separable. This leads to simpler forms of score functions. For example, for $\boldsymbol{\theta}_1$, the maximum point satisfies

$$\begin{aligned} \mathbf{0} \;=\; & \sum_{i \in \mathcal{R}} \boldsymbol{S}_1(y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_1) + \sum_{i \in \mathcal{P}_1} \sum_{y_2=1}^{J} f(y_2 \,|\, y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_2^{(t)}) \boldsymbol{S}_1(y_{i1}, \boldsymbol{x}_i; \boldsymbol{\theta}_1) \\ & + \sum_{i \in \mathcal{P}_2} \sum_{y_1=1}^{R} f(y_1 \,|\, y_{i2}, \boldsymbol{x}_i; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}) \boldsymbol{S}_1(y_1, \boldsymbol{x}_i; \boldsymbol{\theta}_1) \\ & + \sum_{i \in \mathcal{M}} \sum_{y_1=1}^{R} \sum_{y_2=1}^{J} f(\boldsymbol{y} \,|\, \boldsymbol{x}_i; \boldsymbol{\theta}^{(t)}) \boldsymbol{S}_1(y_1, \boldsymbol{x}_i; \boldsymbol{\theta}_1) \,. \end{aligned} \tag{3.13}$$

For our following arguments, (3.13) are the same as the score equations obtained by fitting the marginal model with the imputed data set weighted by $\boldsymbol{w}^{*(t)} = (w_1^{*(t)}, ..., w_{n^*}^{*(t)})$, where $w_m^{*(t)} = W(\boldsymbol{y}_m^*, \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)})$. The same results can be shown for $\boldsymbol{\theta}_2$. Thus our proposed joint fractional imputation procedures have the same spirit as the EM algorithm.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



The convergence properties of the EM algorithm were studied by Wu (1983). In our case, $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ is continuous with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{(t)}$, and hence the EM sequence $\left\{\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}\right\}$ converges to a stationary point $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ that is the solution to the score equations (3.10).

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

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

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

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

### 3.4. Asymptotic properties of fractionally imputed estimators

We begin with the estimator $\hat{\boldsymbol{\pi}}^{fi} = (\hat{\pi}_{11}^{fi}, ..., \hat{\pi}_{1J}^{fi}, ..., \hat{\pi}_{R1}^{fi}, ..., \hat{\pi}_{RJ}^{fi})'$ of the vector $\boldsymbol{\pi}$ of joint cell probabilities, where $\hat{\pi}_{rj}^{fi}$ is given in (3.7). Note that $\hat{\pi}_{rj}^{fi}$ is a weighted sum of indicator functions of "non-independent" observations in the imputed data file. To investigate the asymptotic behaviour of $\hat{\pi}_{rj}^{fi}$, it

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 $(y_{i_01}, *, 1, 0, \boldsymbol{x}_{i_0})$ from group $\mathcal{P}_1$ with $y_{i_02}$ missing corresponds to the bundle $\big\{(y_{i_01}, 1, 1, 0, \boldsymbol{x}_{i_0}),\ \ldots,\ (y_{i_01}, J, 1, 0, \boldsymbol{x}_{i_0})\big\}$. 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 $\mathcal{D}^*$. 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

$$\sum_{m=m_0}^{m_0+J-1} w_m^* \boldsymbol{I}(y_{m1}^* = r, y_{m2}^* = j) = \sum_{m=m_0}^{m_0+J-1} w_m^* \boldsymbol{I}(y_{i_01} = r, m - m_0 + 1 = j),$$

and at most one term on the right hand side is non-zero, $w_{m_0+j-1}^*\ \boldsymbol{I}(y_{i_01} = r) = W((y_{i_01}, j), (1, 0), \boldsymbol{x}_{i_0}; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)\boldsymbol{I}(y_{i_01} = r)$. Similar arguments can be made for observations from other groups. Define the estimating function

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



for $\pi_{rj}$ as

$$\begin{aligned} U_{rj}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \pi_{rj}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \;=\; & \delta_1\delta_2\, \boldsymbol{I}(y_1 = r, y_2 = j) \\ & + \delta_1(1-\delta_2)\, W((y_1, j), (1,0), \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\boldsymbol{I}(y_1 = r) \\ & + (1-\delta_1)\delta_2\, W((r, y_2), (0,1), \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\boldsymbol{I}(y_2 = j) \\ & + (1-\delta_1)(1-\delta_2)\, W((r, j), (0,0), \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - \pi_{rj}\,. \end{aligned} \tag{3.14}$$

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

$$0 = \sum_{i=1}^{n} U_{rj}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \pi_{rj}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2), \tag{3.15}$$

which depends on preliminary estimators of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. This two-step estimator $\hat{\pi}_{rj}$ can be more conveniently handled as a component of solutions to an extended system of estimating equations. Let

$$\begin{aligned} \boldsymbol{S}_{obs}^{(1)}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= E\big[\boldsymbol{S}_1(y_1, \boldsymbol{x}; \boldsymbol{\theta}_1) \,|\, \boldsymbol{y}_{obs}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\big], \\ \boldsymbol{S}_{obs}^{(2)}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= E\big[\boldsymbol{S}_2(y_2, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2) \,|\, \boldsymbol{y}_{obs}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\big]\,. \end{aligned} \tag{3.16}$$

The estimators $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ are initially defined as the solution to the score equations (3.10) and can be re-written as the solution to

$$\mathbf{0} \;=\; \sum_{i=1}^{n} \boldsymbol{S}_{obs}^{(1)}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \quad \mathbf{0} \;=\; \sum_{i=1}^{n} \boldsymbol{S}_{obs}^{(2)}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \tag{3.17}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



Let $\boldsymbol{U}(\boldsymbol{\pi}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (U_{11}, ..., U_{1J}, ..., U_{R1}, ..., U_{RJ})'$, $\boldsymbol{S}_{obs}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\boldsymbol{S}_{obs}^{(1)'}, \boldsymbol{S}_{obs}^{(2)'})'$ and $\boldsymbol{S}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\boldsymbol{S}_1', \boldsymbol{S}_2')'$, where $U_{rj}$, $\boldsymbol{S}_{obs}^{(1)}$, $\boldsymbol{S}_{obs}^{(2)}$, $\boldsymbol{S}_1$, and $\boldsymbol{S}_2$ are short forms of functions defined in (3.14), (3.10), and (3.16). The following theorem summarizes the asymptotic properties of $\hat{\boldsymbol{\pi}}^{fi}$. Proofs are outlined in the Supplementary Material.

**Theorem 3.2.** *Let $\boldsymbol{\pi}_0$, $\boldsymbol{\theta}_{10}$ and $\boldsymbol{\theta}_{20}$ be the true values of $\boldsymbol{\pi}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$. Under the regularity conditions specified in Supplementary Material, $\hat{\boldsymbol{\pi}}^{fi}$ with elements given by (3.7) is a consistent estimator of $\boldsymbol{\pi}$. Furthermore,*

$$n^{1/2}(\hat{\boldsymbol{\pi}}^{fi} - \boldsymbol{\pi}_0) \sim \boldsymbol{N}\Big(\mathbf{0}, Var\big[\boldsymbol{U}(\boldsymbol{\pi}_0, \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) + \boldsymbol{\kappa}\boldsymbol{I}_{obs}^{-1}\boldsymbol{S}_{obs}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})\big]\Big),$$

*where "$\sim$" represents "is asymptotically distributed as",*

$$\boldsymbol{I}_{obs} = \big(E\big[-\partial \boldsymbol{S}_{obs}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_1\big], E\big[-\partial \boldsymbol{S}_{obs}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2\big]\big),$$

*evaluated at the true values of the parameters, $\boldsymbol{\kappa} = (\kappa_{11}', ..., \kappa_{1J}', ..., \kappa_{R1}', ..., \kappa_{RJ}')'$, and*

$$\kappa_{rj} = E\Big\{\boldsymbol{I}(y_1 = r, y_2 = j)\big[\boldsymbol{S}((r,j), \boldsymbol{x}; \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) - \boldsymbol{S}_{obs}((r,j), \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})\big]'\Big\}.$$

**Corollary 3.2.1.** *Let $\boldsymbol{g}(\boldsymbol{\pi})$ be a differentiable function of $\boldsymbol{\pi}$, either scalar or vector valued. If the asymptotic variance of $n^{1/2}(\hat{\boldsymbol{\pi}}^{fi} - \boldsymbol{\pi}_0)$ given in Theorem 3.2 is $\boldsymbol{\Sigma}^{fi}$, then $\boldsymbol{g}(\hat{\boldsymbol{\pi}}^{fi})$ is a consistent estimator of $\boldsymbol{g}(\boldsymbol{\pi})$ and*

$$n^{1/2}\Big[\boldsymbol{g}(\hat{\boldsymbol{\pi}}^{fi}) - \boldsymbol{g}(\boldsymbol{\pi}_0)\Big] \sim \boldsymbol{N}\big(\mathbf{0},\ \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{fi}\boldsymbol{\Gamma}'\big),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



*where* $\boldsymbol{\Gamma} = \partial \boldsymbol{g}(\boldsymbol{\pi})/\partial \boldsymbol{\pi}$ *and is evaluated at* $\boldsymbol{\pi}_0$.

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 $\boldsymbol{g}(\cdot)$. For example, the marginal probabilities of $y_1$ can be written as $\boldsymbol{\pi}_1 = \boldsymbol{C}\boldsymbol{\pi}$, where $\boldsymbol{C} = \text{diag}(\mathbf{1}', \ldots, \mathbf{1}')$ is a $R \times (RJ)$ block diagonal matrix and $\mathbf{1} = (1, \ldots, 1)'$ with length $J$. It follows that $\boldsymbol{\Gamma} = \boldsymbol{C}$ 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

$$\frac{1}{n}\sum_{m=1}^{n^*} \boldsymbol{I}(y_{m1}^* = r, y_{m2}^* = j)\Big[\boldsymbol{S}((r,j), \boldsymbol{x}_m^*; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - \boldsymbol{S}_{obs}((r,j), \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)\Big].$$

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 $\boldsymbol{\delta}_i$ and all of the covariates $\boldsymbol{x}_i$. For public-use

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 $\boldsymbol{b}_i = (\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i)$ denote the $i$th observation in the original data file $\mathcal{D}$. The bootstrap variance estimator of $\boldsymbol{g}(\hat{\boldsymbol{\pi}}^{fi})$ 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 $\mathcal{B}_1 = \{\tilde{\boldsymbol{b}}_i^{(1)},\, i = 1, \ldots, n\}$.

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

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

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

The validity of the bootstrap variance estimator is discussed in the Supplementary Material. The resampling methods are often preferred for

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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. 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 $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\theta}}_1^{(k)'}, \hat{\boldsymbol{\theta}}_2^{(k)'})'$ can be captured by first-order Taylor expansion around $\hat{\boldsymbol{\theta}}$. Therefore, a possible workaround is to use the one-step Newton method discussed in Yang and Kim (2016), where for every bootstrap sample, $\hat{\boldsymbol{\theta}}^{(k)}$ is calculated by one-step iteration from $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2')'$:

$$\hat{\boldsymbol{\theta}}^{(k)} = \hat{\boldsymbol{\theta}} - \left\{ \sum_{m=1}^{n^{(k)*}} \mathring{\boldsymbol{H}}(\boldsymbol{y}_m^*, \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \right\}^{-1} \left\{ \sum_{m=1}^{n^{(k)*}} \boldsymbol{H}(\boldsymbol{y}_m^*, \boldsymbol{\delta}_m^*, \boldsymbol{x}_m^*; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \right\},$$

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

$$\boldsymbol{H}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = W(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \boldsymbol{S}(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



and $\mathring{\boldsymbol{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \partial \boldsymbol{H}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$. We then obtain the fractional weights and estimator $\hat{\boldsymbol{\pi}}^{(k)}$ based on $\hat{\boldsymbol{\theta}}^{(k)}$.

## 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 considered bivariate ordinal responses $(y_1, y_2)$, each with three categories, and two covariates: a continuous variable $x_1$ generated from Exp (1) and a discrete variable $x_2$ following Bernoulli (0.5). The responses $(y_1, y_2)$ followed the marginal and the transitional models given in (3.1). To apply the PSA method, we simulated the response indicators in a way that the propensity scores followed a baseline-category logit model.

The parameters in the propensity score models were carefully chosen such that the proportions of units in the four groups $\mathcal{R}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{M}$ were controlled to have desirable patterns to mimic two real-world scenarios. The first scenario had the majority of the sample fully observed, with proportions $(0.5, 0.2, 0.2, 0.1)$ for the four groups. For the second scenario, only one of the two responses was observed for the majority of sampled units, with the proportions $(0.2, 0.3, 0.4, 0.1)$. The simulation studies consisted of three parts: point estimators, variance estimators, and tests of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



independence.

Table 2: ARB (in %) and MSE ($\times 10^4$) for Estimating $\pi_{+1}$

| RP | $n$ | | COMP | ACA | PSA | SRMI5 | JFI |
|---|---|---|---|---|---|---|---|
| 5221 | 200 | ARB | 0.2 | 7.1 | 0.04 | 1.0 | 0.03 |
| | | MSE | (8.9) | (14.7) | (12.1) | (11.8) | (11.6) |
| | 500 | ARB | 0.33 | 7.7 | 0.3 | 0.2 | 0.3 |
| | | MSE | (3.6) | (8.3) | (5.2) | (5.0) | (4.9) |
| 2341 | 200 | ARB | — | 8.8 | 0.008 | 1.5 | 0.1 |
| | | MSE | — | (20.6) | (12.9) | (12.5) | (12.2) |
| | 500 | ARB | — | 8.5 | 0.2 | 0.2 | 0.3 |
| | | MSE | — | (10.5) | (5.1) | (5.1) | (4.9) |

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 3: ARB (in %) and MSE ($\times 10^4$) for Estimating $\gamma$

| RP | $n$ | | COMP | ACA | PSA | SRMI5 | JFI |
|---|---|---|---|---|---|---|---|
| 5221 | 200 | ARB | 0.5 | 8.7 | 0.9 | 1.5 | 0.04 |
| | | MSE | (7.3) | (15.0) | (25.4) | (13.2) | (12.7) |
| | 500 | ARB | 0.04 | 9.1 | 0.4 | 1.1 | 0.1 |
| | | MSE | (2.8) | (6.9) | (13.1) | (5.3) | (5.0) |
| 2341 | 200 | ARB | — | 10.9 | 5.0 | 10.6 | 0.5 |
| | | MSE | — | (42.3) | (87.8) | (27.9) | (25.1) |
| | 500 | ARB | — | 12.5 | 3.9 | 7.8 | 0.03 |
| | | MSE | — | (17.2) | (56.8) | (11.3) | (9.5) |

The simulation results show clearly that the ACA estimator is not consistent for either the marginal probability $\pi_{+1}$ or the association measure $\gamma$, while PSA, SRMI5, and JFI provide comparable results for estimating marginal probabilities with negligible biases. For the estimation of $\gamma$, the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 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 was on variance estimation. For the SRMI method, the variance estimator used Rubin's combining rule; for the JFI method, two versions of variance estimators were 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 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 suggests 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.

The third part of the simulation was on tests of independence between the two ordinal responses. We used the Wald-type test statistic given in

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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

| | | $\pi_{+1}$ | | | $\gamma$ | | |
|---|---|---|---|---|---|---|---|
| RP | $n$ | SRMI5 | JFIL | JFIB | SRMI5 | JFIL | JFIB |
| 5221 | 200 | 3.0 | 5.7 | 3.7 | (−)16.9 | 5.1 | 3.3 |
| | 500 | 8.0 | 1.3 | 2.2 | (−)16.0 | 4.1 | 3.6 |
| 2341 | 200 | 4.4 | 4.0 | 3.7 | (−)24.2 | 1.4 | 4.5 |
| | 500 | 7.5 | 2.1 | 1.3 | (−)26.4 | 2.3 | 2.3 |

(2.2) based on a particular pair of point and variance estimators with significance level of 0.05. By tuning the parameters in (3.1), we simulated the power of tests for a series of cases in which the true value of the association measure $\gamma$ increased from 0 to 1, departing gradually from the null hypothesis of independence.

The power of a test was 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 tests: `JFI_non`, `JFI_nul` and SRMI. The first test uses the regular linearization variance estimator without considering the null hypothesis; the second test

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



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 SRMI method. Test results using bootstrap variance estimators are very similar to the ones using linearization variance estimators and are not reported here. 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.

Here are three observations from the power functions. The test based on the 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. The type I errors for the two JFI-based tests are very close to the nominal value and both tests have similar power. 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



## 5. Concluding Remarks

Statistical analysis with missing data faces two scenarios. It could be a data set of small or moderate size collected for specific scientific purposes with the analysis carried out by specific researchers who have full access to the data set and are equipped with a solid knowledge of statistics. It is 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 on validity and efficiency of analysis with the fractionally imputed data set have focused on estimation of joint and

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



marginal probabilities and association measures and on test of independence. Regression analysis was 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 discussions on modeling techniques for complete ordinal responses and a detailed review on existing methods for handling missing ordinal observations. Regularity conditions, the 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



## 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

Agresti, A. (2010). *Analysis of Ordinal Categorical Data (2nd edition)*. Wiley, New York.

Agresti, A. (2013). *Categorical Data Analysis (3rd edition)*. Wiley, Hoboken, New Jersey.

Brick, J. M. and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research* **5**, 215–238.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* **39**, 1–38.

Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American' Statistical Association* **89**, 463–475.

Goodman, L. A. (1979). Simple models for analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association* **74** 537–552.

Goodman, L. A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics* **13**, 10–69.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732–764.

Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics* **36**, 589–600.

Kalton, G. and Kish, L. (1984). Some efficient random imputation methods. *Communications in Statistics-Theory and Methods* **13**, 1919–1939.

Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika* **33**, 239–251.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98**, 119–132.

Kim, J. K. and Fuller, W. A. (2004). Fractional hot deck imputation. *Biometrika* **91**, 559–578.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Little, R. J. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (2nd edition)*. John Wiley & Sons, Hoboken, New Jersey.

Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics* **52**, 354–361.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (2nd edition)*. Chapman & Hall, London.

Parsons, N. R., Edmondson, R. N. and Gilmour, S. G. (2006). A generalized estimating equation

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society: Series C* **55**, 507–524.

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J. and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* **27**, 85–96.

Rao, J. N. and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–822.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Statistics, Wiley.

She, X. (2017). *Fractional Imputation for Ordinal and Mixed-type Responses with Missing Observations*, doctoral dissertation, Department of Statistics and Actuarial Science, University of Waterloo, Canada.

Simon, G. (1978). Efficacies of measures of association for ordinal contingency tables. *Journal of the American Statistical Association* **69**, 971–976.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review* **27**, 799–811.

Touloumis, A., Agresti, A. and Kateri, M. (2013). GEE for multinomial responses using a local odds ratios parameterization. *Biometrics* **69**, 633–640.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



equations in R. *Journal of Statistical Software* **45**, 1–67.

White, I. R., Royston, P. and Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* **30**, 377–399.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95–103.

Yang, S. and Kim, J. K. (2016). Fractional imputation in survey sampling: A comparative review. *Statistical Science* **31**, 415–432.

Department of Statistics and Actuarial Science, University of Waterloo,

Waterloo, ON, N2L 3G1, Canada

E-mail: xshe@uwaterloo.ca; cbwu@uwaterloo.ca

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0396



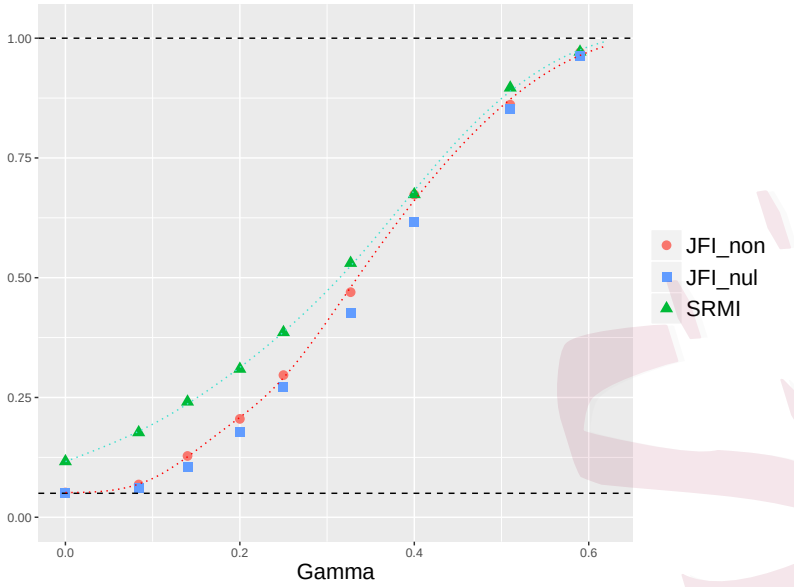


Figure 1: Power Function with $n = 200$ and Pattern 2341

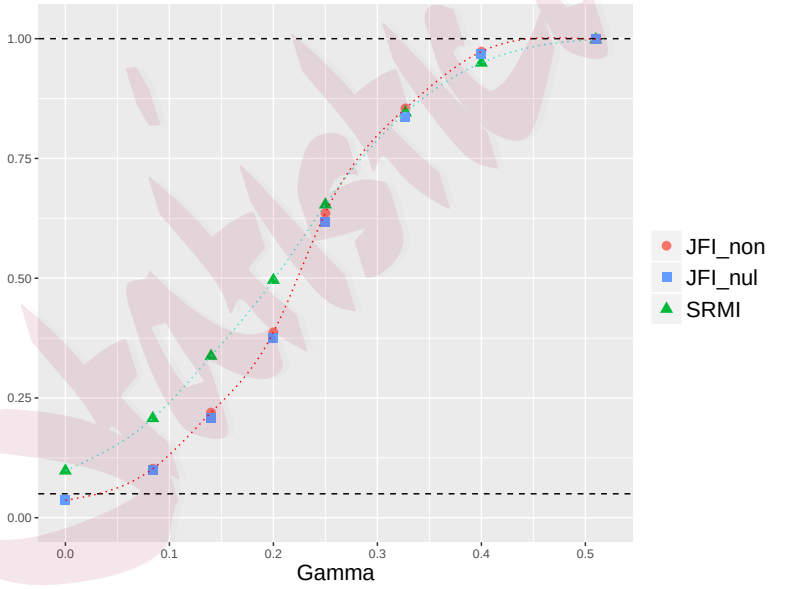


Figure 2: Power Function with $n = 500$ and Pattern 2341