Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Semiparametric Fractional Imputation Using Gaussian Mixture Models for Handling Multivariate Missing Data

Hejian Sang, Jae Kwang Kim & Danhyang Lee

To cite this article: Hejian Sang, Jae Kwang Kim & Danhyang Lee (2020): Semiparametric Fractional Imputation Using Gaussian Mixture Models for Handling Multivariate Missing Data, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1796358

To link to this article: https://doi.org/10.1080/01621459.2020.1796358

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 00, NO. 0, 1–10: Theory and Methods
https://doi.org/10.1080/01621459.2020.1796358

Taylor & Francis
Taylor & Francis Group

Semiparametric Fractional Imputation Using Gaussian Mixture Models for Handling Multivariate Missing Data

Hejian Sang[a], Jae Kwang Kim[b], and Danhyang Lee[c]

[a]Google Inc., Mountain View, CA; [b]Department of Statistics, Iowa State University, Ames, IA; [c]Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, AL

ABSTRACT

Item nonresponse is frequently encountered in practice. Ignoring missing data can lose efficiency and lead to misleading inference. Fractional imputation is a frequentist approach of imputation for handling missing data. However, the parametric fractional imputation may be subject to bias under model misspecification. In this article, we propose a novel semiparametric fractional imputation (SFI) method using Gaussian mixture models. The proposed method is computationally efficient and leads to robust estimation. The proposed method is further extended to incorporate the categorical auxiliary information. The asymptotic model consistency and $\sqrt{n}$-consistency of the SFI estimator are also established. Some simulation studies are presented to check the finite sample performance of the proposed method. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received September 2018
Accepted July 2020

KEYWORDS
Item nonresponse; Robust estimation; Variance estimation

1. Introduction

Missing data are frequently encountered in survey sampling, clinical trials, and many other areas. Imputation can be used to handle item nonresponse and several imputation methods have been developed in the literature. Motivated from a Bayesian perspective, Rubin (1996) proposed multiple imputation to create multiple complete datasets. Alternatively, under the frequentist framework, fractional imputation (Kim and Fuller 2004; Kim 2011) makes one complete data with multiple imputed values and their corresponding fractional weights. Little and Rubin (2002) and Kim and Shao (2013) provided comprehensive overviews of the methods for handling missing data.

For multivariate missing data with arbitrary missing patterns, valid imputation methods should preserve the correlation structure in the imputed data. Judkins et al. (2007) proposed an iterative hot deck imputation procedure that is closely related to the data augmentation algorithm of Tanner and Wong (1987) but they did not provide variance estimation. Other imputation procedures for multivariate missing data include the multiple imputation approaches of Raghunathan et al. (2001) and Murray and Reiter (2016), and parametric fractional imputation (PFI) of Kim (2011). The approaches of Judkins et al. (2007) and Raghunathan et al. (2001) are based on conditionally specified models and the imputation from conditional model is generically subject to the model compatibility problem (Chen 2010; Liu et al. 2013; Bartlett et al. 2015). Conditional models for the different missing patterns calculated directly from the observed patterns may not be compatible with each other. The PFI uses the joint distribution to create imputed values and does not suffer from model compatibility problems.

Note that parametric imputation requires correct model specification. Nonparametric imputation methods, such as kernel regression imputation (Cheng 1994; Wang and Chen 2009), are robust but may be subject to the curse of dimensionality. Hence, it is important, often critical, to develop a unified, robust and efficient imputation method that can be used for general purpose estimation. The proposed semiparametric method fills in this important gap by considering a more flexible model for imputation.

In this article, to achieve robustness against model misspecification, we develop an imputation procedure based on Gaussian mixture models (GMMs). GMM is a very flexible model that can be used to handle outliers, heterogeneity, and skewness. McLachlan and Peel (2004) and Bacharoglou (2010) argued that any continuous distribution can be approximated by a finite Gaussian mixture distribution. The proposed method using GMM makes a nice compromise between efficiency and robustness. It is semiparametric in the sense that the number of mixture components is chosen automatically from the data. The computation for parameter estimation in our proposed method is based on EM algorithm and its implementation is relatively simple and efficient.

We note that Elliott and Stettler (2007) and Kim et al. (2014) developed multiple imputation using mixture models. Multiple imputation using Rubin's formula, however, requires congeniality and self-efficiency (Meng 1994), which is not necessarily satisfied for general-purpose estimation (Yang and Kim 2016). Instead of multiple imputation, we use fractional imputation

CONTACT Jae Kwang Kim jkim@iastate.edu Department of Statistics, Iowa State University, Ames, IA 50011.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

which does not require the congeniality and self-efficiency assumptions for general-purpose estimation. Di Zio, Guarnera, and Luzi (2007) also considered using GMM to impute missing data. They developed single imputation using either the conditional mean imputation or random draw. Any theoretical properties of their imputed estimator such as variance estimation is not discussed in Di Zio, Guarnera, and Luzi (2007). We provide a completely theoretical justification for consistency of the proposed imputation estimator. The variance estimation and the model selection for the number of mixture component are also carefully discussed and demonstrated in numerical studies. The proposed method is further extended to handle mixed type data including categorical variable in Section 5. By allowing the proportion vector of mixture component to depend on categorical auxiliary variable, the proposed fractional imputation using GMMs can incorporate the observed categorical variables and provide a very flexible tool for imputation.

The article is structured as follows. The setup of the problem is introduced and a short review of fractional imputation are presented in Section 2. In Section 3, the proposed semiparametric method and its algorithm for implementation are introduced. Some asymptotic results are presented in Section 4. In Section 5, the proposed method is further extended to handle mixed type data. Some numerical studies and a real data application are presented to show the performance of the proposed method in Sections 6 and 7, respectively. In Section 8, some concluding remarks are made. The technical derivations and proof are presented in the Appendix.

2. Basic Setup

Consider a p-dimensional random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)'$. Suppose that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are n independent and identically distributed realizations of the random vector $\mathbf{Y}$. Assume that we are interested in estimating parameter θ which is defined through $E\{U(\theta; \mathbf{Y})\} = 0$, where $U(\cdot; \mathbf{Y})$ is the estimating function of θ. With no missingness, a consistent estimator of θ can be obtained by the solution to

$$\frac{1}{n}\sum_{i=1}^{n} U(\theta; \mathbf{y}_i) = 0. \tag{1}$$

To avoid unnecessary details, we assume that the solution to (1) exists uniquely almost everywhere.

However, due to missingness, the estimating equation in (1) cannot be applied directly. To formulate the multivariate missingness problem, we define the response indicator vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_p)'$ as

$$\delta_j = \begin{cases} 1 \text{ if } Y_j \text{ is observed,} \\ 0 \text{ otherwise,} \end{cases}$$

where $j = 1, 2, \ldots, p$. We assume that the response mechanism is missing at random in the sense of Rubin (1976). We decompose $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$, where $\mathbf{Y}_{\text{obs}}$ and $\mathbf{Y}_{\text{mis}}$ represent the observed and missing parts of $\mathbf{Y}$, respectively. Thus, the missing-at-random assumption is described as

$$P(\boldsymbol{\delta} \in B \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}) = P(\boldsymbol{\delta} \in B \mid \mathbf{Y}_{\text{obs}}), \tag{2}$$

for any measurable set $B \in \sigma(\boldsymbol{\delta})$, the sigma-field generated by $\boldsymbol{\delta}$.

Under the missing-at-random assumption, a consistent estimator of θ can be obtained by solving the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n} E\left\{U(\theta; \mathbf{Y}_i) \mid \mathbf{y}_{i,\text{obs}}\right\} = 0, \tag{3}$$

where it is understood that $E\left\{U(\theta; \mathbf{Y}_i) \mid \mathbf{y}_{i,\text{obs}}\right\} = U(\theta; \mathbf{y}_i)$ if $\mathbf{y}_{i,\text{obs}} = \mathbf{y}_i$. To compute the conditional expectation in (3), the PFI method of Kim (2011) can be developed. To apply the PFI, we can assume that the random vector $\mathbf{Y}$ follows a parametric model in that $F_0(\mathbf{y}) \in \left\{F_\zeta(\mathbf{y}) : \zeta \in \Omega\right\}$. In the PFI, M imputed values for $\mathbf{Y}_{i,\text{mis}}$, say $\{\mathbf{y}_{i,\text{mis}}^{*(1)}, \mathbf{y}_{i,\text{mis}}^{*(2)}, \ldots, \mathbf{y}_{i,\text{mis}}^{*(M)}\}$, are generated from a proposal distribution with the same support of $F_0(\mathbf{y})$ and are assigned with fractional weights, say $\left\{w_{i1}^*, w_{i2}^*, \ldots, w_{iM}^*\right\}$, so that a consistent estimator of θ can be obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^* U(\theta; \mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}}^{*(j)}) = 0,$$

where the fractional weights are constructed to satisfy

$$\sum_{j=1}^{M} w_{ij}^* U(\theta; \mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}}^{*(j)}) \cong E\left\{U(\theta; \mathbf{Y}_i) \mid \mathbf{y}_{i,\text{obs}}\right\}$$

as closely as possible, with $\sum_{j=1}^{M} w_{ij}^* = 1$. In Kim (2011), the fractional weights are computed using the idea of importance sampling.

However, the PFI is based on the assumption of $F_0(\mathbf{y}) \in \left\{F_\zeta(\mathbf{y}) : \zeta \in \Omega\right\}$ and it is not always easy to find the joint distribution family $\left\{F_\zeta(\mathbf{y}) : \zeta \in \Omega\right\}$ correctly. If the joint distribution family $\left\{F_\zeta(\mathbf{y}) : \zeta \in \Omega\right\}$ is misspecified, the PFI can lead to biased inference. All aforementioned concerns motivate us to consider a more robust fractional imputation method using GMMs, which cover a wider class of parametric models.

3. Proposed method

To formulate the proposal, we first assume that the sample is decomposed into G mutually exclusive and exhaustive groups. We then define the group indicator vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_G)'$, where $Z_g = 1$ if the sample unit belongs to the gth group, and $Z_g = 0$ otherwise. We assume that the conditional distribution of $\mathbf{Y}$ given $Z_g = 1$ follows a multivariate normal distribution with parameter $\left\{\mu_g, \Sigma_g\right\}$ and Σ_g can be structured in the sense that $\Sigma_g = \Sigma(\phi_g)$ for some ϕ_g. For example, if $Y_1, \ldots, Y_p$ are repeated measurement over p time period, one may impose the first-order autoregressive model to get

$$\Sigma_g = \sigma_g^2\left\{(1-\rho_g)\mathbf{I}_p + \rho_g \mathbf{J}_p\right\}$$

for some σ_g^2 and $\rho_g \in (-1, 1)$, where $\mathbf{I}_p$ and $\mathbf{J}_p$ are the $p \times p$ identity matrix and matrix of ones, respectively. In this case, the variance-covariance matrix Σ_g is determined by $\phi_g = (\sigma_g^2, \rho_g)$. For the unstructured case, ϕ_g is equal to Σ_g.

Under this setup, the marginal distribution of $\mathbf{Y}$ follows a GMM with density

$$f(\mathbf{y};\alpha,\zeta)=\sum_{g=1}^{G}\alpha_g f(\mathbf{y}\mid z_g=1;\zeta_g), \tag{4}$$

where G is the number of mixture component, $\alpha_g = P(Z_g = 1) \in (0,1)$ is the mixture proportion satisfying $\sum_{g=1}^{G}\alpha_g = 1$, and $f(\cdot \mid z_g = 1;\zeta_g)$ is the density function of normal with parameter $\zeta_g = (\mu_g, \Sigma(\phi_g))$. We assume that the GMM in (4) satisfies the strong first-order identifiability assumption (Chen 1995; Liu and Shao 2003; Chen and Khalili 2008), where the first-order derivatives of $f(\mathbf{y};\alpha,\zeta)$ respect to all parameters are linearly independent.

To handle item nonresponse, we first estimate the parameters from the marginal distribution of the observed data and then apply the fractional imputation from the estimated prediction model. Under model (4) and MAR, the marginal distribution of the observed data is also GMM in the sense that

$$f_{\text{obs}}(\mathbf{y}_{\text{obs}};\alpha,\zeta)=\sum_{g=1}^{G}\alpha_g f(\mathbf{y}_{\text{obs}}\mid z_g=1;\zeta_g),$$

where $f(\mathbf{y}_{\text{obs}} \mid z_g = 1;\zeta_g) = \int f(\mathbf{y} \mid z_g = 1;\zeta_g)d\mathbf{y}_{\text{mis}}$ is also Gaussian. Thus, the EM algorithm maximizing the observed log-likelihood $l_{\text{obs}}(\alpha,\zeta) = \sum_{i=1}^{n}\log f_{\text{obs}}(\mathbf{y}_{i,\text{obs}};\alpha,\zeta)$ can be described as follows:

E-step: Using the current parameter values $(\alpha^{(t)},\zeta^{(t)})$, compute

$$\begin{aligned}p_{ig}^{(t)} &= P\left(z_{ig}=1\mid \mathbf{y}_{i,\text{obs}};\alpha^{(t)},\zeta^{(t)}\right)\\ &= \frac{f(\mathbf{y}_{i,\text{obs}}\mid z_{ig}=1;\zeta_g^{(t)})\alpha_g^{(t)}}{\sum_{g=1}^{G}f(\mathbf{y}_{i,\text{obs}}\mid z_{ig}=1;\zeta_g^{(t)})\alpha_g^{(t)}},\end{aligned}$$

where $f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1;\zeta_g)$ is the marginal density of $\mathbf{y}_{i,\text{obs}}$ derived from $\mathbf{y}_i \mid (z_{ig} = 1) \sim N(\mu_g, \Sigma_g)$.

M-step: Update the parameters by maximizing

$$\begin{aligned}Q(\alpha,\mu,\phi\mid\alpha^{(t)},\zeta^{(t)}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}p_{ig}^{(t)}\\ &\quad\left\{\log\alpha_g+\log f(\mathbf{y}_{i,\text{obs}}\mid z_g=1;\mu_g,\Sigma(\phi_g))\right\}\end{aligned} \tag{5}$$

with respect to $(\alpha,\mu,\phi) \in \Omega$. For unstructured case, the solution is

$$\alpha_g^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}p_{ig}^{(t)}; \tag{6}$$

$$\mu_g^{(t+1)} = \frac{\sum_{i=1}^{n}p_{ig}^{(t)}E(\mathbf{Y}_i\mid\mathbf{y}_{i,\text{obs}},z_{ig}=1;\zeta_g^{(t)})}{\sum_{i=1}^{n}p_{ig}^{(t)}}$$

$$\begin{aligned}\Sigma_g^{(t+1)} &= \frac{1}{\sum_{i=1}^{n}p_{ig}^{(t)}}\sum_{i=1}^{n}p_{ig}^{(t)}E\\ &\quad\left\{\left(\mathbf{Y}_i-\mu_g^{(t+1)}\right)\left(\mathbf{Y}_i-\mu_g^{(t+1)}\right)^{\mathrm{T}}\mid\mathbf{y}_{i,\text{obs}},z_{ig}=1;\zeta_g^{(t)}\right\},\end{aligned}$$

where the conditional expectations can be easily derived from the normality in the conditional distribution of $\mathbf{y}_{i,\text{mis}}$ given $\mathbf{y}_{i,\text{obs}}$ and $z_{ig} = 1$.

Repeat E-step to M-step until the convergence is achieved.

Once the parameters are estimated, we use the fractional imputation method to impute the missing values. Note that the prediction model (=imputation model) for $\mathbf{Y}_{\text{mis}}$ is also GMM in that

$$\begin{aligned}&f(\mathbf{y}_{i,\text{mis}}\mid\mathbf{y}_{i,\text{obs}};\hat\alpha,\hat\zeta)\\ &=\sum_{g=1}^{G}P(z_{ig}=1\mid\mathbf{y}_{i,\text{obs}};\hat\alpha,\hat\zeta)f(\mathbf{y}_{i,\text{mis}}\mid\mathbf{y}_{i,\text{obs}},z_{ig}=1;\hat\zeta),\end{aligned} \tag{7}$$

where

$$P(z_{ig}=1\mid\mathbf{y}_{\text{obs}};\hat\alpha,\hat\zeta)=\frac{f(\mathbf{y}_{i,\text{obs}}\mid z_{ig}=1;\hat\zeta_g)\hat\alpha_g}{\sum_{g=1}^{G}f(\mathbf{y}_{i,\text{obs}}\mid z_{ig}=1;\hat\zeta_g)\hat\alpha_g}:=\hat p_{ig}$$

and $\mathbf{y}_{i,\text{mis}} \mid (\mathbf{y}_{i,\text{obs}}, z_{ig} = 1)$ is also normal. Generating imputed values from (7) involves a two-step procedure. In the first step, imputed values for latent variable z_i are generated using the conditional probability $\hat p_{ig}$. In the second step, imputed values for $\mathbf{y}_{i,\text{mis}}$ are generated from the conditional distribution $f(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}}, z_{ig}^* = 1;\hat\zeta)$ given the imputed value z_{ig}^*. Thus, in fractional imputation, we generate M imputed values from (7) in two steps.

[Step 1] Generate $(M_{i1}^*, M_{i2}^*, \ldots, M_{iG}^*) \sim \text{Multinomial}(M;\hat{\mathbf{p}}_i)$, where $\hat{\mathbf{p}}_i = (\hat p_{i1},\ldots,\hat p_{iG})$.

[Step 2] For each $g = 1,2,\ldots,G$, we generate M_{ig}^* samples of $\mathbf{y}_{i,\text{mis}}$, say $\left\{\mathbf{y}_{i,\text{mis}}^{*(gj)};j=1,\ldots,M_{ig}^*\right\}$, from the conditional distribution $f(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}}, z_g = 1;\hat\zeta_g)$, which is also normal.

Then, the final estimator, say $\hat\theta_{\text{SFI}}$, of θ can be obtained by solving the fractionally imputed estimating equation, given by

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{M_{ig}^*}w_{igj}^*U(\theta;\mathbf{y}_i^{*(gj)})=0, \tag{8}$$

where $w_{igj}^* = \hat p_{ig}/M_{ig}^*$ are the final fractional weights assigned to $\mathbf{y}_i^{*(gj)} = (\mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}}^{*(gj)})$ for $j = 1,\ldots,M_{ig}^*; g = 1,\ldots,G$. The fractional weights satisfy $\sum_{g=1}^{G}\sum_{j=1}^{M_{ig}^*}w_{igj}^* = 1$. By construction,

$$\begin{aligned}\sum_{g=1}^{G}\sum_{j=1}^{M_{ig}^*}w_{igj}^*U(\theta;\mathbf{y}_i^{*(gj)}) &= \sum_{g=1}^{G}\frac{\hat p_{ig}}{M_{ig}^*}\sum_{j=1}^{M_{ig}^*}U(\theta;\mathbf{y}_i^{*(gj)})\\ &\cong\sum_{g=1}^{G}\hat p_{ig}E\{U(\theta;\mathbf{y}_i)\mid\mathbf{y}_{i,\text{obs}},z_{ig}=1;\hat\zeta\}\\ &= E\left\{U(\theta;\mathbf{Y}_i)\mid\mathbf{y}_{i,\text{obs}};\hat\alpha,\hat\zeta\right\}\end{aligned}$$

and the fractionally imputed estimating equation in (8) approximates

$$\frac{1}{n}\sum_{i=1}^{n}E\left\{U(\theta;\mathbf{Y}_i)\mid\mathbf{y}_{i,\text{obs}};\hat\alpha,\hat\zeta\right\}=0.$$

Asymptotic properties of the imputation estimator obtained from (8) will be covered in Section 4. For variance estimation of $\hat\theta_{\text{SFI}}$, we use the grouped jackknife variance estimator described in Shao and Wu (1989).

4. Asymptotic theory

In our proposed fractional imputation method using GMMs in Section 3, we have assumed that the size of mixture components, G, is known. In practice, G is often unknown and we need to estimate it from the sample data. If G is larger than necessary, the proposed mixture model may be subject to overfitting and increase its variance. If G is small, then the approximation of the true distribution cannot provide accurate prediction due to its bias. Hence, we can allow the model complexity parameter G to depend on the sample size n, say $G = G(n)$. The choice of G under complete data has been well explored in the literature. The popular methods are based on Bayesian information criterion (BIC) and Akaike's information criterion (AIC). See Wallace and Dowe (1999), Windham and Cutler (1992), Schwarz (1978), Fraley and Raftery (1998), Keribin (2000), and Dasgupta and Raftery (1998). The alternative way of using SCAD penalty (Fan and Li 2001) is studied in Chen and Khalili (2008) and Huang, Peng, and Zhang (2017).

In this article, under IID setup, we consider using the BIC to select G. For simplicity of presentation, we consider the same values of $\Sigma_g = \Sigma$ across all components to get a parsimonious model. The proposed joint model in (4) can be easily extended to use the group-dependent variance Σ_g, as in Di Zio, Guarnera, and Luzi (2007). Under multivariate missingness, we use the observed log-likelihood function in computing the information criterion, in the sense that

$$\mathrm{BIC}(G) = -2\sum_{i=1}^{n}\log\left\{\sum_{g=1}^{G}\hat{\alpha}_g f(\mathbf{y}_{i,\mathrm{obs}} \mid z_{ig}=1;\hat{\xi}_g)\right\} + (\log n)\phi(G), \tag{9}$$

under the assumption of $\Sigma_g = \Sigma$, where $(\hat{\alpha},\hat{\zeta})$ are the estimators obtained from the proposed method and $\phi(G)$ is a monotone increasing function of G. In (9), $\phi(G) = G + Gp$ if ignoring constant terms. However, our model selection framework and theoretical results can be directly applied to any general penalty function $\phi(G)$. Using GMMs, the observed log-likelihood function is expressed as a closed-form.

In this section, we first establish the consistency of model selection using (9) under the GMM assumption. After that, we establish some asymptotic results when the GMM assumption is violated.

To establish the first part, assume that the true density function of $\mathbf{Y}$ is $f_0(\mathbf{y}) = \sum_{g=1}^{G^o}\alpha_g^o f(\mathbf{y} \mid z_g = 1;\zeta_g^o)$, where (G^o,α^o,ζ^o) are true parameter values. For $\zeta_g^o = (\mu_g^o,\Sigma^o)$, we need the following regularity assumptions:

(A1) The mean vectors for each mixture component is bounded uniformly, in the sense of $\|\mu_g^o\| \le C_1$, for $g = 1,2,\ldots,G^o$.
(A2) $\|\Sigma^o\| \le C_2$. Furthermore, Σ^o is nonsingular.

The first assumption means the first moment is bounded. Assumption (A2) is to make sure that Σ^0 is bounded and nonsingular. Both assumptions are commonly used.

To establish the model consistency, we furthermore make the additional assumptions on the response mechanism:

(A3) The response rate for y_j is bounded below from 0, say $\lim_n n^{-1}\sum_{i=1}^{n}\delta_{ij} > C_3$, for $j = 1,2,\ldots,p$, where $C_3 > 0$ is a constant.
(A4) The response mechanism satisfies the missing-at-random condition in (2).

The following theorem shows that the true number of mixture components can be selected by minimizing BIC(G) in (9) consistently.

Theorem 1. Assume the true density f_0 belongs to the GMMs, satisfying (A1) and (A2). Let $\hat{G}$ be the minimizer of BIC(G) in (9). Under assumptions (A3) and (A4), we have

$$P(\hat{G} = G^o) \to 1,$$

as $n \to \infty$, where G^o is the true number of mixture components.

The proof of Theorem 1 is shown in the supplementary materials. Theorem 1 states that minimizing BIC(G) consistently selects the true mixture components under the assumption that the true distribution is in the GMM.

Now, in the second scenario, the true distribution does not necessary belong to the class of GMMs. Thus, we first establish the following lemma to measure how well GMM can approximate the arbitrary density function. We furthermore make additional assumptions about the true density function f_0. Use E_0 to denote the expectation respect to f_0.

(A5) Assume $f_0(\mathbf{y})$ is continuous and $E_0\|\mathbf{Y}\|^2 < \infty$, where $\|\mathbf{Y}\|^2 = Y_1^2 + \cdots + Y_p^2$.
(A6) Assume $E_0\left\{\partial f(\mathbf{Y})/\partial\alpha\right\} < \infty$ and $E_0\left\{\partial f(\mathbf{Y})/\partial\mu\right\} < \infty$, where $f(\mathbf{y}) = \sum_{g=1}^{G}\alpha_g f(\mathbf{y};\mu_g,\Sigma)$. Moreover, assume $E_0\left\{f(\mathbf{Y})^{-2}\right\} < \infty$.

Assumption (A5) is satisfied for any continuous random variable with bounded second moments. Assumption (A6) is true for any finite GMM and f_0 has a valid moment generating function.

Lemma 1. Under assumptions (A5) and (A6) and missing at random, for any $\epsilon > 0$, there exists $\gamma > 0$ such that for $G = O(\epsilon^{-\gamma})$,

$$\|f_0 - \hat{f}\|_1 = O(\epsilon), \tag{10}$$

$$\mathrm{var}(f_0 - \hat{f}) = O(\epsilon^{-\gamma}n^{-1}), \tag{11}$$

with probability one, where $\hat{f}(\mathbf{y}) = \sum_{g=1}^{G}\hat{\alpha}_g f(\mathbf{y};\hat{\mu}_g,\hat{\Sigma})$ is obtained from the proposed method, and $\|f_0 - \hat{f}\|_1 = \int |f_0(\mathbf{y}) - \hat{f}(\mathbf{y})| f_0(\mathbf{y})\mathrm{d}\mathbf{y}$.

The proof of Lemma 1 is presented in the supplementary materials. If f_0 is a density function of the GMM, then $\gamma = 0$ and by Theorem 1, our proposed BIC(G) can select the true model consistently. For any f_0 satisfying (A5) and (A6), the bias can goes to 0 as $G \to \infty$ from (10). The variance will increase as $G \to \infty$ from (11) for fixed n. There is a trade-off between bias and variance for the divergence case ($\gamma > 0, G \to \infty$).

Using Lemma 1, we can further establish the $\sqrt{n}$-consistency of $\hat{\theta}_{\mathrm{SFI}}$. The following assumptions are the sufficient conditions to obtain the $\sqrt{n}$-consistency.

(A7) $E_0\left\{U^2(\theta;\mathbf{Y}_i)\right\} < \infty$.
(A8) $\gamma \in (0,2)$.
(A9) $\epsilon = O(n^{-1/(2-\Delta)})$, for any $\Delta \in (0,2)$.

Theorem 2. Under assumptions (A5)–(A9), $\gamma + \Delta < 2$ and MAR, we have

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{M_{ig}^*} w_{igj}^* U(\theta;\mathbf{y}_i^{*(gj)}) \cong J_1 + o_p(n^{-1/2}), \tag{12}$$

where $J_1 = n^{-1}\sum_{i=1}^{n} E_0\left\{U(\theta;\mathbf{Y}_i) \mid \mathbf{y}_{i,\text{obs}}\right\}$, if $M = \sum_{i,g}\{M_{ig}\} \to \infty$. Furthermore, we have

$$\sqrt{n}(\hat{\theta}_{\text{SFI}} - \theta_0) \to N(0,\Sigma), \tag{13}$$

for some Σ which is positive definite and θ_0 satisfies $E_0\{U(\theta_0;\mathbf{Y})\} = 0$.

The proof of (12) is shown in the supplementary materials and (13) can be directly derived from (12). From Theorem 2, we have $G = O(n^{\gamma/(2-\Delta)}) = o(n) \to \infty$ with the rate smaller than n. Thus, even under non-Gaussian mixture families, our proposed method still enjoys $\sqrt{n}$-consistency.

5. Extension

In Section 3, we assume that $\mathbf{Y}$ is fully continuous. However, in practice, categorical variables can be used to build imputation models. We extend our proposed method to incorporate the categorical variable as a covariate in the model.

To introduce the proposed method, we first introduce the conditional GMM. Suppose that $(\mathbf{X},\mathbf{Y})$ is a random vector where $\mathbf{X}$ is discrete and $\mathbf{Y}$ is continuous. We further assume that $\mathbf{X}$ is always observed. To obtain the conditional GMM, we assume that $\mathbf{Z}$ satisfies

$$f(\mathbf{Y} \mid \mathbf{X},\mathbf{Z}) = f(\mathbf{Y} \mid \mathbf{Z}), \tag{14}$$

in the sense that $\mathbf{Z}$ is a partition of the sample such that $\mathbf{Y}$ is homogeneous within each group defined by $\mathbf{Z}$. Furthermore, we assume that $f(\mathbf{y} \mid Z_g = 1)$ follows a Gaussian distribution. Combining these assumptions, we have the following conditional GMM

$$f(\mathbf{y} \mid \mathbf{x}) = \sum_{g=1}^{G} \alpha_g(\mathbf{x}) f(\mathbf{y} \mid Z_g = 1), \tag{15}$$

where $\alpha_g(\mathbf{x}) = P(Z_g = 1 \mid \mathbf{x})$ and $f(\mathbf{y} \mid Z_g = 1)$ is the density function of the normal distribution with parameter $\zeta_g = \left\{\mu_g, \Sigma_g\right\}$. We also assume that the identifiability conditions in (15) hold.

Using the argument similar to (7), the predictive model of $\mathbf{y}_{i,\text{mis}}$ under (14) can be expressed as

$$\begin{aligned} &f(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}},\mathbf{x}_i) \\ &= \sum_{g=1}^{G} P(Z_{ig} = 1 \mid \mathbf{y}_{i,\text{obs}},\mathbf{x}_i) f(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}}, z_{ig} = 1), \end{aligned} \tag{16}$$

where $f(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}}, z_{ig} = 1)$ can be derived from $(\mathbf{y}_{i,\text{obs}},\mathbf{y}_{i,\text{mis}}) \mid (z_{ig} = 1) \sim N(\mu_g,\Sigma_g)$. The posterior probability of $z_{ig} = 1$ given the observed data is

$$P(Z_{ig} = 1 \mid \mathbf{y}_{i,\text{obs}},\mathbf{x}_i) = \frac{f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1)P(z_{ig} = 1 \mid \mathbf{x}_i)}{\sum_{g=1}^{G} f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1)P(z_{ig} = 1 \mid \mathbf{x}_i)}.$$

Therefore, the proposed fractional imputation using conditional GMMs can be summarized as follows:

E-step: Using the current parameter values, compute

$$p_{ig}^{(t)} = \frac{f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1;\zeta_g^{(t)})\alpha_g^{(t)}(\mathbf{x}_i)}{\sum_{g=1}^{G} f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1;\zeta_g^{(t)})\alpha_g^{(t)}(\mathbf{x}_i)}.$$

M-step: Update the parameter values by maximizing

$$\begin{aligned} &Q(\alpha,\zeta \mid \alpha^{(t)},\zeta^{(t)}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G} p_{ig}^{(t)}\left\{\log\alpha_g(\mathbf{x}_i) + \log f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1;\zeta_g)\right\}, \end{aligned}$$

respect to (α,ζ).

Repeat E-step to M-step iteratively until convergence is achieved. The final estimator of θ can be obtained by solving the fractionally imputed estimating equation in (8). Note that the proposed method builds the proportion vector of mixture components into a function of auxiliary variable and assumes that the mixture components share the same mean and variance structure. Thus, the proposed method can borrow information across different $\mathbf{X}$ values. Moreover, the auxiliary information is incorporated to build a more flexible class of joint distributions.

Remark 1. If assumption (14) does not hold, we can use, instead of (15),

$$f(\mathbf{y} \mid \mathbf{x}) = \sum_{g=1}^{G} \alpha_g(\mathbf{x}) f(\mathbf{y} \mid \mathbf{x}, Z_g = 1), \tag{17}$$

where $f(\mathbf{y} \mid \mathbf{x}, Z_g = 1)$ is the density function of the normal distribution with mean $B_g\mathbf{x}$ and variance Σ_g. In this case, we can write $\zeta_g = \{B_g,\Sigma_g\}$ and the EM algorithm for parameter estimation can be developed similarly by replacing $f(\mathbf{y}_{i,\text{obs}} \mid z_{ig} = 1;\zeta_g)$ with $f(\mathbf{y}_{i,\text{obs}} \mid \mathbf{x}_i, z_{ig} = 1;\zeta_g)$.

6. Numerical Studies

We consider two simulation studies to evaluate the performance of the proposed methods. The first simulation study is used to check the performance of the proposed imputation method using GMMs under multivariate continuous variables. The second simulation study considers the case of multivariate mixed categorical and continuous variables. To save space, we only present the first simulation study. The second simulation study is presented in the supplementary materials.

In the first simulation study, we consider the following models for generating $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})$.

1. M1: A mixture distribution with density $f(\mathbf{y}) = \sum_{g=1}^{3} \alpha_g f_g(\mathbf{y})$, where $(\alpha_1, \alpha_2, \alpha_3) = (0.3, 0.3, 0.4)$ and $f_g(\mathbf{y})$ is a density function for multivariate normal distribution with mean μ_g and variance
$$\Sigma(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}.$$
Let $\rho = 0.7$ and $\mu_1 = (-3, -3, -3,), \mu_2 = (1, 1, 1), \mu_3 = (5, 5, 5)$.
2. M2: Use the same model as M1 except for $f_2(\mathbf{Y})$, where $f_2(\mathbf{Y})$ is a product of the density for the exponential distribution with rate parameter 1.
3. M3: $Y_{i1} = 1 + e_{i1}$, $Y_{i2} = 0.5Y_{i1} + e_{i2}$, and $Y_{i3} = Y_{i2} + e_{i3}$, where e_{i1}, e_{i2}, e_{i3} are independently generated from $N(0, 1)$, Gamma$(1, 1)$ and χ^2_1 distributions, respectively.
4. M4: Generate (Y_{i1}, Y_{i2}) independently from a Gaussian distribution with mean $(1, 2)$ and variance
$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$
Let $Y_{i3} = Y_{i2}^2 + e_{i3}$, where $e_{i3} \sim N(0, 1)$.

In M1, a GMM with $G = 3$ is used to generate the samples. A non-Gaussian mixture distribution is used in M2 to check the robustness of the imputation methods. M3 and M4 are used to check the performance of the imputation methods under skewness and nonlinearity, respectively.

The size for each realized sample is $n = 500$. Once the complete sample is obtained, for $y_{ij}, j = 2, 3$, we select 25% of the sample independently to make missingness with the selection probabilities equal to π_{ij}, where $\text{logit}(\pi_{i2}) = -0.8 + 0.4y_{i1}$, $\text{logit}(\pi_{i3}) = 0.4 - 0.8y_{i1}$, and $\text{logit}(u) = \exp(u)/\{1 + \exp(u)\}$. Since we assume y_{i1} are fully observed, the response mechanism is missing at random.

The overall missing rate is approximate 55%. For each realized incomplete samples, we apply the following methods:

[Full]: As a benchmark, we use the full samples to estimate parameters.
[CC]: Use the complete cases only to estimate parameters.
[MICE]: Apply multivariate imputation by chained equations (Buuren and Groothuis-Oudshoorn 2011). The predictive mean matching is used as a default.
[MIGM]: Multiple imputation using GMM with 50 components from Kim et al. (2014), where 50 is the upper bound of the size G of mixture components as recommended by Kim et al. (2014). The variance estimators are obtained using Rubin's formula and the confidence intervals are constructed using Wald method.
[PFI]: Parametric fractional imputation method of Kim (2011). We assume that the joint distribution is a multivariate normal distribution and $M = 2000$ imputed values are generated for each missing value.
[SFI]: The proposed semiparametric fractional imputation method using GMMs, where the number of components G is selected using the BIC in (9), where $G \in \{1, \ldots, 10\}$, and $M = 2000$ imputed values are generated for each missing value. The confidence intervals are computed using the group jackknife variance estimator with the group size equal to 10.

Table 1. The coverage rates of the proposed method (SFI) and its most comparable estimator (MIGM) based on $B = 2000$ Monte Carlo samples.

Model	Method	θ_2	θ_3	P_2	P_3
M1	MIGM	94·9	94·9	93·2	95·1
	SFI	94·3	94·2	94·7	94·1
M2	MIGM	95·4	94·8	91·9	94·8
	SFI	95·2	94·7	95·2	94·5
M3	MIGM	96·7	95·1	95·1	93·4
	SFI	96·8	95·2	96·6	95·7
M4	MIGM	95·5	94·9	96·4	89·0
	SFI	94·5	94·6	94·7	94·8

The parameters of interest are the true means and proportions associated with Y_2 and Y_3. Specifically, we are interested in $\theta_2 = E(Y_2), \theta_3 = E(Y_3), P_2 = \text{pr}(Y_2 < c_2)$, and $P_3 = \text{pr}(Y_3 < c_3)$, where $(c_2, c_3) = (-2, -2)$ for M1 and M2, $(c_2, c_3) = (2, 3)$ for M3, $(c_2, c_3) = (2, 5)$ for M4. The estimating functions for θ_j and P_j are given by $U_1(\theta_j; \mathbf{Y}) = Y_j - \theta_j$ and $U_2(P_j; \mathbf{Y}) = I(Y_j < c_j) - P_j$, for $j = 2, 3$.

Figures 1 and 2 present the estimation simulation results based on 2000 Monte Carlo samples for the four parameters under each data generating setup. The estimators using only complete cases (CC) and PFI present large biases across data generating models for some parameters as expected. On the other hand, SFI is comparable to the Full estimator across data generating models and type of parameters. Under M1, MICE is comparable to the Full and SFI estimators for each parameter, however, for skewed-distributions (M2 and M3) or nonlinear distribution (M4), it shows significant biases for the proportions. MIGM shows better overall performances than MICE. It is almost unbiased for each parameter under both M1 and M3. It, however, still shows significant biases for some proportion parameters under M2 and M4.

The coverage rates of the SFI estimator and its most comparable estimator, MIGM, are presented in Table 1. While the SFI estimator shows overall coverage rates around the nominal level 95% across data generating models, MIGM shows about 89% and 92% coverage rates for some proportion parameters under M2 and M4.

Figure 3 presents the histograms of the number of mixture components selected by using the proposed BIC for the SFI method. Under M1, the proposed BIC selected $G = 3$, for most of the Monte Carlo samples, which is the true number of mixture components of the GMM. For non-GMMs which are skewed or have nonlinear mean structure, the proposed BIC selected G greater than 4 or 5 for most of the Monte Carlo samples.

7. Application

In this section, we apply the proposed method in Section 3 to a synthetic data that mimics monthly retail trade survey data at the U.S. Census Bureau. The synthetic data were created by U.S. Census Bureau to be used for one of the contests sponsored by the fifth international conference on establishment surveys. More information can be found in *https://ww2.amstat.org/meetings/ices/2016/contests.cfm*. The sampling scheme is a stratified simple random sample without replacement

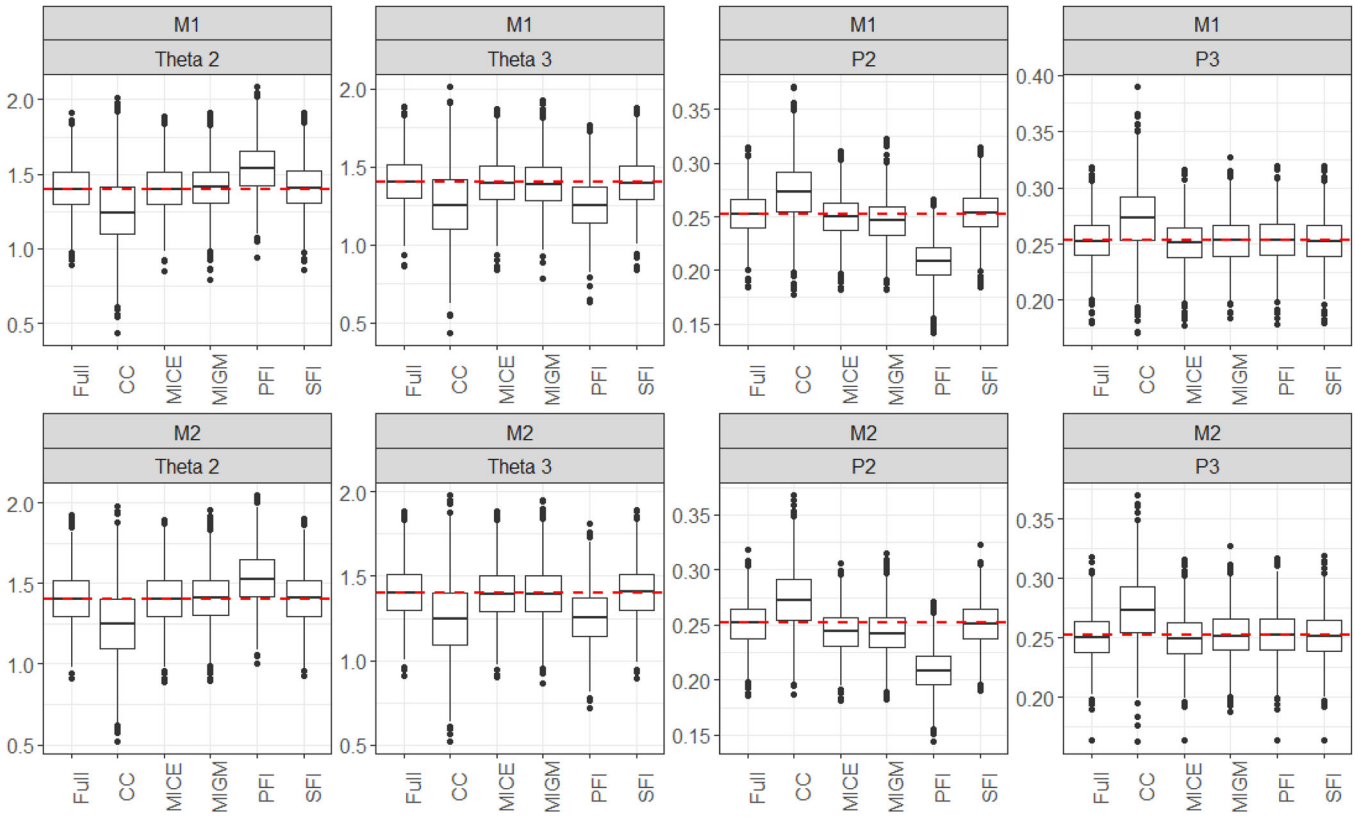

Figure 1. Estimation results from 2000 Monte Carlo samples for the four parameters under data generating models M1 and M2; the dashed lines indicate the true parameter values under each model.

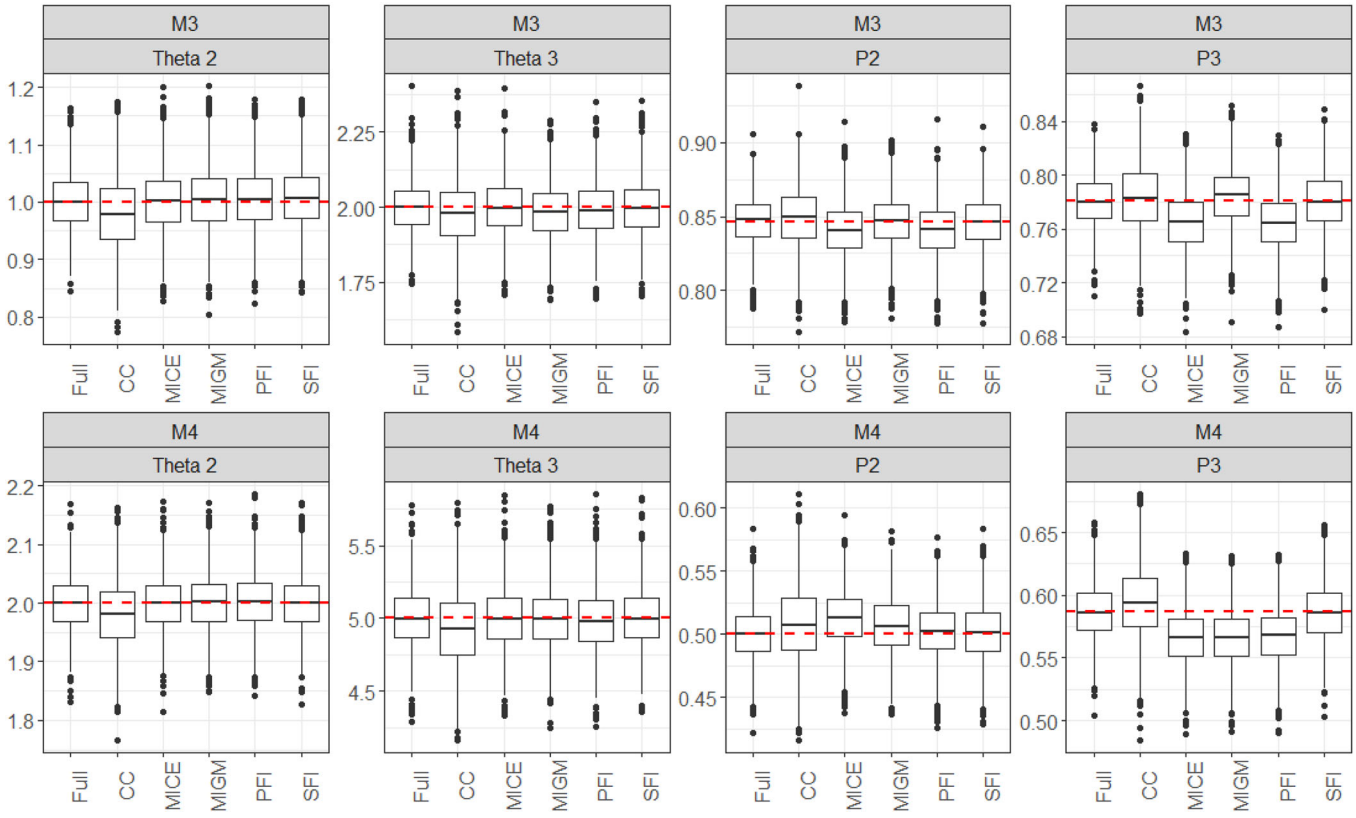

Figure 2. Estimation results from 2000 Monte Carlo samples for the four parameters under data generating models M3 and M4; the dashed lines indicate the true parameter values under each model.

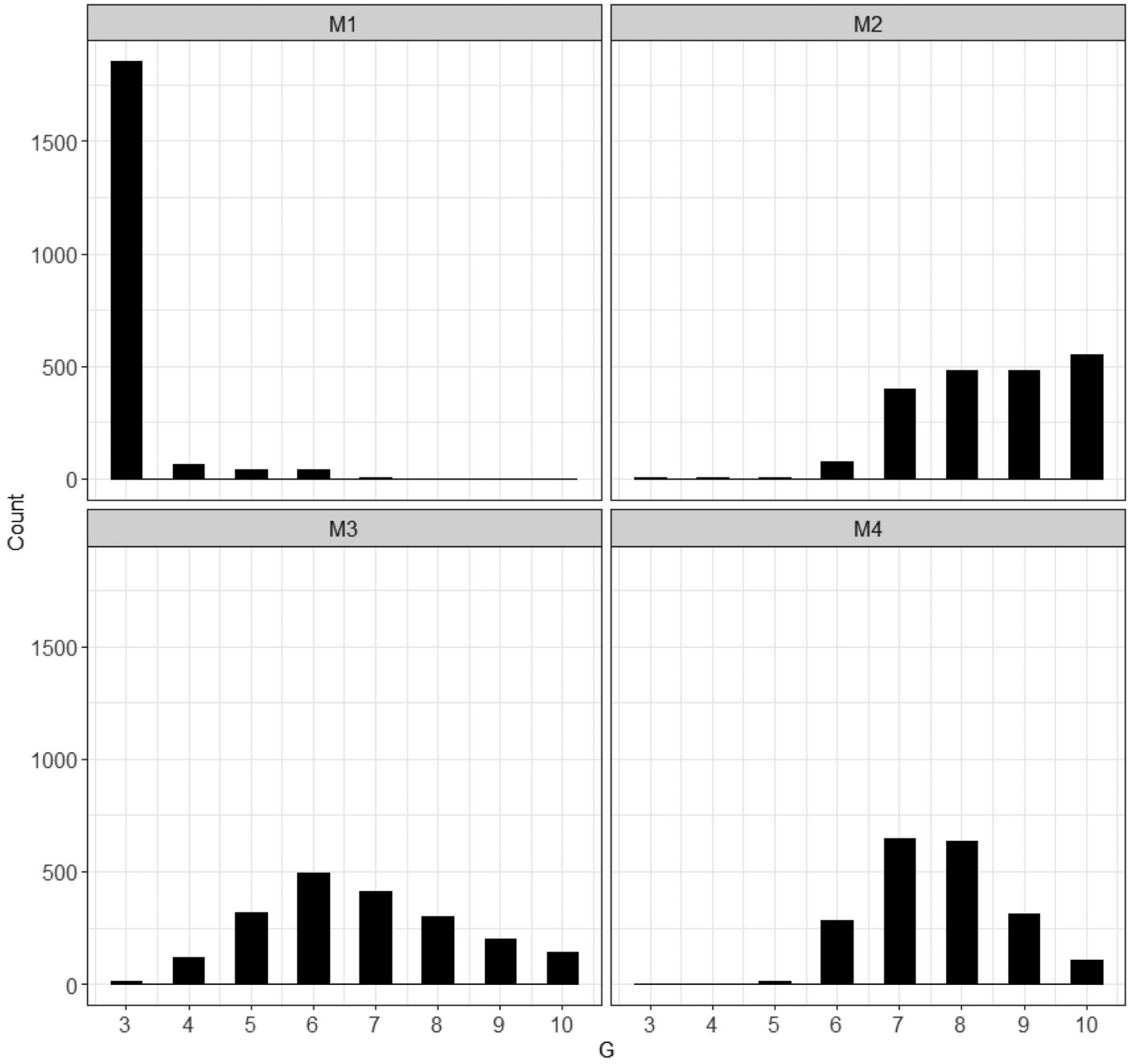

Figure 3. The histograms of selected G values using the proposed BIC.

sample with six strata: one certain (take-all) and five non-certainty strata. The sample sizes are computed using Neyman allocation.

The dataset is incomplete due to item nonresponse, where the current month sales and inventories are subject to missingness and the overall response rate is approximately 71%. We aim to complete the dataset, providing imputed values for the missing cases by using the proposed imputation method. The boxplots in Figure 4 illustrate the distributions of the log-transformed variables across strata for the CC in the data.

Let $\mathbf{Y}$ denote the random vector of seven variables where Y_1 and Y_2 denotes the current month sales and inventories, respectively. The parameters of our interest are the population mean and variance of Y_k, denoted by μ_k and σ_k^2, respectively, for $k = 1, 2$, and the correlation coefficient between the two variables, denoted by ρ. The estimating functions are $U(\mu_k; \mathbf{Y}) = Y_k - \mu_k$, $U(\sigma_k^2; \mathbf{Y}) = (Y_k - \mu_k)^2 - \sigma_k^2$, and $U(\rho; \mathbf{Y}) = (Y_1 - \mu_1)(Y_2 - \mu_2) - \rho\sigma_1\sigma_2$.

As in the simulation study, we compute the estimates of the parameters using only CC, multiple imputation using GMM (MIGM), parameter fractional imputation using a multivariate normal distribution (PFI), and the proposed semiparametric fractional imputation (SFI) for comparison. MICE failed to converge due to high correlations among the survey items. The estimation results are shown in Table 2.

As expected, the estimates using CC only have significant biases for some parameters. Interestingly, we found that MIGM also shows significant biases for the parameters such as variances and correlation coefficient. The MIGM suffers from overfitting problem for this data application which attenuates the correlation structure leading to significant biased estimation. Both PFI and SFI perform properly

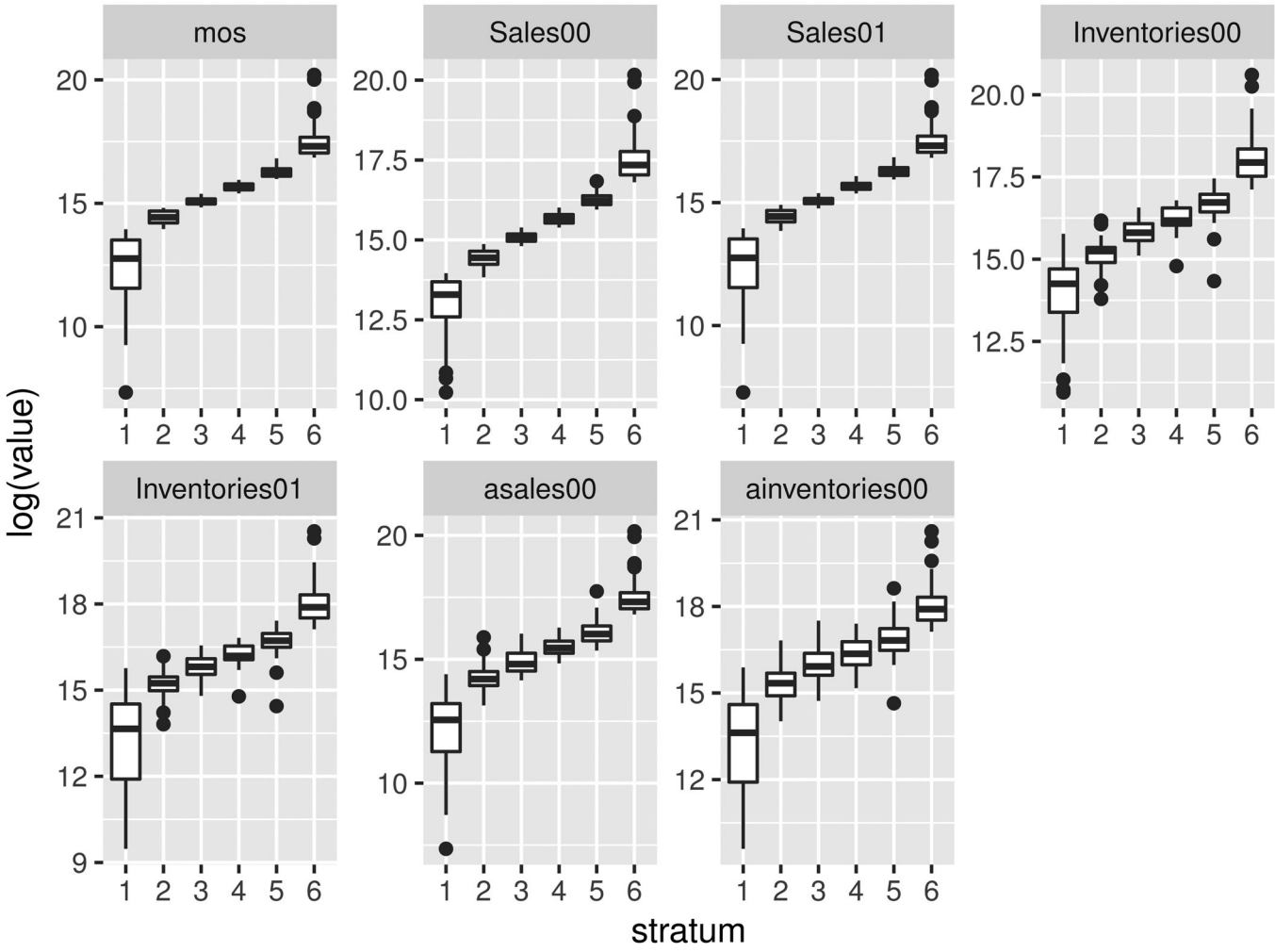

Figure 4. Overview for log-transformed variables of the synthetic monthly retail trade survey data:"mos" is frame measure of size; "Sales00" denotes current month sales for unit (subject to missing); "asales00" is current month administrative data value for sales; "Sales01" means prior month sales for unit; "Inventories00" is current month inventories for unit (subject to missing); "ainventories00" is current month administrative data value for inventories; "Inventories01" is prior month inventories for unit.

Table 2. Imputation results for the synthetic monthly retail trade survey data.

Method	μ_1 ($\times 10^{-6}$)	μ_2 ($\times 10^{-6}$)	σ_1^2 ($\times 10^{-13}$)	σ_2^2($\times 10^{-14}$)	ρ
True	2.30	4.81	4.38	1.19	0.97
CC	2.86 (2.58, 3.14)	5.78 (5.24, 6.32)	6.86 (0.91, 12.80)	1.77 (0.42, 3.12)	0.97 (0.94, 1.00)
MIGM	2.34 (2.13, 2.53)	4.80 (4.30, 5.33)	5.16 (4.28, 6.03)	1.39 (1.18, 1.60)	0.82 (0.71, 0.93)
PFI	2.32 (2.12, 2.51)	4.83 (4.46, 5.20)	4.36 (0.72, 8.00)	1.15 (0.32, 1.99)	0.97 (0.94, 1.00)
SFI	2.29 (2.13, 2.45)	4.76 (4.37, 5.15)	4.37 (0.79, 7.95)	1.17 (0.35, 1.98)	0.97 (0.93, 1.00)

NOTE: Parameter point estimation with 95% confidence intervals and true values are presented.

for all the parameters, but SFI has overall better performance than PFI.

8. Discussion

Fractional imputation has been proposed as a tool for frequentist imputation, as an alternative to multiple imputation. Multiple imputation using Rubin's formula can be biased when the model is uncongenial or the point estimator is not self-efficient (Meng 1994; Yang and Kim 2016). In this article, we have proposed a SFI method using GMMs to handle arbitrary multivariate missing data. The proposed method automatically selects the size of mixture components and provides a unified framework for robust imputation. Even if the group size G increases with the sample size n, the resulting estimator enjoys $\sqrt{n}$-consistency. While the theory is mainly developed under the special case

of $\Sigma_g = \Sigma$, it can be easily extended to the heterogeneous GMM using different Σ_g. In this case, however, the number of parameters can be quite large when G is large. We have also extended the proposed method to incorporate categorical auxiliary variable. The flexible model assumption and efficient computation are the main advantages of our proposed method. An extension to a more robust mixture model is a topic of future research. An R software package for the proposed method is under development.

Supplementary Materials

The online supplementary material contains all technical proofs of the theorems and a simulation study for the case of multivariate mixed categorical and continuous variables.

Acknowledgments

The authors wish to thank the editor, the associate editor, and three anonymous referees for very constructive comments.

Funding

The research of the second author was partially supported by a grant from US National Science Foundation (1733572).

References

Bacharoglou, A. (2010), "Approximation of Probability Distributions by Convex Mixtures of Gaussian Measures," *Proceedings of the American Mathematical Society*, 138, 2619–2628. [1]

Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., and Alzheimer's Disease Neuroimaging Initiative (2015), "Multiple Imputation of Covariates by Fully Conditional Specification: Accommodating the Substantive Model," *Statistical Methods in Medical Research*, 24, 462–487. [1]

Buuren, S., and Groothuis-Oudshoorn, K. (2011), "MICE: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, 45, 1–67. [6]

Chen, H. Y. (2010), "Compatibility of Conditionally Specified Models," *Statistics & Probability Letters*, 80, 670–677. [1]

Chen, J. (1995), "Optimal Rate of Convergence for Finite Mixture Models," *The Annals of Statistics*, 23, 221–233. [3]

Chen, J., and Khalili, A. (2008), "Order Selection in Finite Mixture Models With a Nonsmooth Penalty," *Journal of the American Statistical Association*, 103, 1674–1683. [3,4]

Cheng, P. E. (1994), "Nonparametric Estimation of Mean Functionals With Data Missing at Random," *Journal of the American Statistical Association*, 89, 81–87. [1]

Dasgupta, A., and Raftery, A. E. (1998), "Detecting Features in Spatial Point Processes With Clutter via Model-Based Clustering," *Journal of the American Statistical Association*, 93, 294–302. [4]

Di Zio, M., Guarnera, U., and Luzi, O. (2007), "Imputation Through Finite Gaussian Mixture Models," *Computational Statistics & Data Analysis*, 51, 5305–5316. [2,4]

Elliott, M. R., and Stettler, N. (2007), "Using a Mixture Model for Multiple Imputation in the Presence of Outliers: The 'Healthy for Life' Project," *Journal of the Royal Statistical Society*, Series C, 56, 63–78. [1]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [4]

Fraley, C., and Raftery, A. E. (1998), "How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis," *The Computer Journal*, 41, 578–588. [4]

Huang, T., Peng, H., and Zhang, K. (2017), "Model Selection for Gaussian Mixture Models," *Statistica Sinica*, 27, 147–169. [4]

Judkins, D., Krenzke, T., Piesse, A., Fan, Z., and Haung, W.-C. (2007), "Preservation of Skip Patterns and Covariate Structure Through Semi-Parametric Whole Questionnaire Imputation," in *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 3211–3218. [1]

Keribin, C. (2000), "Consistent Estimation of the Order of Mixture Models," *Sankhya A*, 62, 49–66. [4]

Kim, H. J., Reiter, J. P., Wang, Q., Cox, L. H., and Karr, A. F. (2014), "Multiple Imputation of Missing or Faulty Values Under Linear Constraints," *Journal of Business & Economic Statistics*, 32, 375–386. [1,6]

Kim, J. K. (2011), "Parametric Fractional Imputation for Missing Data Analysis," *Biometrika*, 98, 119–132. [1,2,6]

Kim, J. K., and Fuller, W. (2004), "Fractional Hot Deck Imputation," *Biometrika*, 91, 559–578. [1]

Kim, J. K., and Shao, J. (2013), *Statistical Methods for Handling Incomplete Data*, Boca Raton, FL: CRC Press. [1]

Little, R. J., and Rubin, D. B. (2002), *Statistical Analysis With Missing Data*, New York: Wiley. [1]

Liu, J., Gelman, A., Hill, J., Su, Y.-S., and Kropko, J. (2013), "On the Stationary Distribution of Iterative Imputations," *Biometrika*, 101, 155–173. [1]

Liu, X., and Shao, Y. (2003), "Asymptotics for Likelihood Ratio Tests Under Loss of Identifiability," *The Annals of Statistics*, 31, 807–832. [3]

McLachlan, G., and Peel, D. (2004), *Finite Mixture Models*, New York: Wiley. [1]

Meng, X.-L. (1994), "Multiple-Imputation Inferences With Uncongenial Sources of Input," *Statistical Science*, 9, 538–558. [1,9]

Murray, J. S., and Reiter, J. P. (2016), "Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models With Local Dependence," *Journal of the American Statistical Association*, 111, 1466–1479. [1]

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. [1]

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592. [2]

——— (1996), "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association*, 91, 473–489. [1]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [4]

Shao, J., and Wu, C. F. J. (1989), "A General Theory for Jackknife Variance Estimation," *The Annals of Statistics*, 17, 1176–1197. [3]

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540. [1]

Wallace, C. S., and Dowe, D. L. (1999), "Minimum Message Length and Kolmogorov Complexity," *The Computer Journal*, 42, 270–283. [4]

Wang, D., and Chen, S. X. (2009), "Empirical Likelihood for Estimating Equations With Missing Values," *The Annals of Statistics*, 37, 490–517. [1]

Windham, M. P., and Cutler, A. (1992), "Information Ratios for Validating Mixture Analyses," *Journal of the American Statistical Association*, 87, 1188–1192. [4]

Yang, S., and Kim, J. K. (2016), "A Note on Multiple Imputation for Method of Moments Estimation," *Biometrika*, 103, 244–251. [1,9]