A Comparative Study of Imputation Methods for Estimation of Missing Values of Per Capita Expenditure in Central Java

Y Susianto[1], K A Notodiputro[1], A Kurnia[1] and H Wijayanto[1]

[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University

Email: susianto.sugiharto@gmail.com, khairilnotodiputro@gmail.com, akstk29@gmail.com, hari_ipb@yahoo.com

Abstract. Missing values in repeated measurements have attracted concerns from researchers in the last few years. For many years, the standard statistical methods for repeated measurements have been developed assuming that the data was complete. The standard statistical methods cannot produce good estimates if the data suffered substantially by missing values. To overcome this problem the imputation methods could be used. This paper discusses three imputation methods namely the Yates method, expectation-maximization (EM) algorithm, and Markov Chain Monte Carlo (MCMC) method. These methods were used to estimate the missing values of per-capita expenditure data at sub-districts level in Central Java. The performance of these imputation methods is evaluated by comparing the mean square error (MSE) and mean absolute error (MAE) of the resulting estimates using linear mixed models. It is showed that MSE and MAE produced by the Yates method are lower than the MSE and MAE resulted from both the EM algorithm and the MCMC method. Therefore, the Yates method is recommended to impute the missing values of per capita expenditure at sub-district level.

1. Introduction

Incomplete data problems as a result of missing values in repeated measurements have attracted concerns from researchers in the last few years. For many years, standard statistical methods for repeated measurements have been developed assuming the data was complete. In fact, if the data suffered substantially by missing values then the standard statistical methods cannot produce good estimates. To overcome this problem, proper statistical analysis methods are required.

This problem can be overcome by the complete case analysis with assuming the missing values are missing completely at random (MCAR). However, this approach can substantially reduce the sample size and consequently the estimates become inefficient. Furthermore, Schafer and Graham [1], Nakai and Weiming [2], and Buuren [3] have shown that the complete case analysis under the assumption of missing at random (MAR) produces poor estimates.

In practice, the missing values are usually missing at random (MAR). In this situation the available case analysis can be used. However, as mentioned by Donders *et al.* [4], the available case analysis produces biased estimates unless the missing values are imputed before analysis is carried out. Through the imputation technique, the missing values are replaced by the most likely numbers as if they were observed.

This paper discusses three imputation methods namely the Yates method, expectation-maximization (EM) algorithm, and Markov Chain Monte Carlo (MCMC) method. These three methods were used to estimate the missing values of per capita expenditure data at sub-districts level in Central Java. The performance of these imputation methods are evaluated by comparing the mean square error (MSE) and mean absolute error (MAE) of the resulting estimates using linear mixed models.

2. Methods

2.1. The Data

The national social economy survey (SUSENAS) has been designed by Statistics Indonesia (BPS) to produce social-economic indicators in Indonesia. In 2011 until 2014, SUSENAS was conducted quarterly with rolling samples. As a result of these rolling samples the aggregated data at sub-districts level became incomplete. For a particular quarter, some sub-districts had the observed data but in other sub-districts the data might not be available. In other words, at sub-districts level we had missing data problems especially missing data for repeated measurements.

In this paper, the per capita expenditure data at sub-district level from 2011 until 2014 were imputed. This paper provided a preliminary study of the data imputation applied to five districts in Central Java province, namely: Banyumas, Pati, Semarang, Brebes and kota Semarang. The imputation of the data was conducted separately for each district.

2.2. The Yates Method

This method is a classical approach developed by Yates in 1933 to estimate missing values in experimental design for the purpose of minimizing error variance. This approach consists of 3 stages, namely: estimation of missing values, replacement of the missing values with prediction values, and analysis of the complete data [5].

Since the problem structures were similar to the missing problem in a randomized complete block design then the formula for this design was used.

a). The formula for a single missing value is

$$y_{ij} = \frac{m\,y_{i.} - n\,y_{.j} - y_{..}}{(m-1)(n-1)} \tag{1}$$

Where y_{ij} is the prediction of missing value at the-i^{th} sub-district and the-j^{th} time, $y_{i.}$ is the total observed values at the-i^{th} sub-district containing a missing value, $y_{.j}$ is is the total observed values at the-j^{th} time containing missing values, $y_{..}$ is the total observed values for all observations, m is the number of subjects, and n is the number of times.

b). If number of missing values are more than one, we iterate formula (1) using a starting value

$$y_{\text{initial}(ij)} = \frac{\bar{y}_{i.} + \bar{y}_{.j}}{2} \tag{2}$$

In this formula, $\bar{y}_{i.}$ is the mean values of the-i^{th} subject, $\bar{y}_{.j}$ is the mean values of the-j^{th} time.

2.3. The EM Algorithm

This algorithm is a parametric method to impute missing values based on the maximum likelihood estimation. This algorithm is very popular in statistical literatures and has been discussed intensively by many researchers, such as : Dempster *et al.* [6], Little and Rubin [5,7], Schafer [8], and Watanabe and Yamaguchi [9].

This algorithm uses an iterative procedure to finding the maximum likelihood estimators of parameter vector through two step described in Dempster *et al.* [6], Schafer [8], and Little and Rubin [7] as follows:

a). The Expectation step (E-step)

The E step is the stage of determining the conditional expected value of the full data of log likelihood function $l(\theta|Y)$ given observed data. Suppose for any incomplete data, the distribution of the complete data Y can be factored as

$$\begin{aligned} f(Y|\theta) &= f(Y_{mis}, Y_{obs}|\theta) \\ &= f(\,Y_{obs}|\theta)\, f(Y_{mis}|Y_{obs}, \theta) \end{aligned} \tag{3}$$

where $f(\,Y_{obs}|\theta)$ is the distribution of the data observed Y_{obs} and $f(Y_{mis}|Y_{obs}, \theta)$ is the distribution of missing data given data observed. Based on the equation (3), we obtained log likelihood function

$$l(\theta|Y) = l(\theta|Y_{obs}) + log\, f(Y_{mis}|Y_{obs}, \theta) \tag{4}$$

where $l(\theta|Y)$ is log likelihood function of complete data, $l(\theta|Y_{obs})$ is log likelihood function of observed data, and $f(Y_{mis}|Y_{obs}, \theta)$ is the predictive distribution of missing data given θ.

Objectively, to estimate θ is done by maximizing the log likelihood function (4). Because Y_{mis} not known, the right side of equation (4) can not be calculated. As a solution, $l(\theta|Y)$ is calculated based on the average value $log\, f(Y_{mis}|Y_{obs}, \theta)$ using predictive distribution $f\left(Y_{mis}|Y_{obs}, \theta^{(t)}\right)$, where $\theta^{(t)}$ is temporary estimation of unknown parameters. In this context, an initial estimation $\theta^{(0)}$ be calculated using the complete case analysis. With this approach, the mean value of equation (4) can be expressed

$$\begin{aligned} Q\left(\theta|\theta^{(t)}\right) &= l(\theta|Y_{obs}) + \int log\, f(Y_{mis}|Y_{obs}, \theta) f\left(Y_{mis}|Y_{obs}, \theta^{(t)}\right)\partial Y_{mis} \\ &= \int [l(\theta|Y_{obs}) + \int log\, f(Y_{mis}|Y_{obs}, \theta)]\, f\left(Y_{mis}|Y_{obs}, \theta^{(t)}\right)\partial Y_{mis} \\ &= \int l(\theta|Y) f\left(Y_{mis}|Y_{obs}\,, \theta^{(t)}\right)\partial Y_{mis} \end{aligned} \tag{5}$$

The equation (5) basically a conditional expected value of log likelihood function for complete data $l(\theta|Y)$ given observed data and initial estimate of unknown parameter.

b). The maximization step (M-step)

The M step is to obtained the iteratively estimation $\theta^{(t+1)}$ with maximizes $Q\left(\theta|\theta^{(t)}\right)$ as follow

$$Q\left(\theta^{(t+1)}|\theta^{(t)}\right) \geq Q\left(\theta|\theta^{(t)}\right) \tag{6}$$

Both E and M steps are iterated until convergent.

2.4. The MCMC Method

This method generates pseudo random draws from probability distributions via Markov chains. A Markov chain is a sequence of random variables in which the distribution of each element depends on the value of the previous one. MCMC is a multiple imputation methods be used to imputate the missing values of continuous data set. In application, MCMC assumes that data have a multivariate normal distribution, missing data is MCAR or MAR, and patten of missing data is monotone or arbitrary. Moreover, the inference of MCMC will be robust if the number of missing values are not too large [8].

MCMC consists of the two most popular methods namely Gibbs Sampling and Metropolis-Hastings algorithm. In Gibbs sampling, one draws from the conditional distribution of each component of a multivariate random variable given the other components. In Metropolis-Hastings, one draws from a probability distribution intended to approximate the distribution actually of interest, and then accepts or rejects the drawn value with a specified probability. In this paper, we use a Gibbs sampling algorithm to draw the missing values from the posterior predictive distribution.

In Bayesian inference perspective, information about unknown parameters is expressed in the form of a posterior probability distribution. It is a useful alternative approach to ML with to add a prior distribution for the parameters and compute the posterior distribution of the parameters of interest. Suppose that $\boldsymbol{Y_{mis}}$ and $\boldsymbol{Y_{obs}}$ are the missing and observed values, respectively. The observed data posterior can be express as

$$p(\theta|Y_{obs}) \propto p(\theta)\, p(Y_{obs}|\theta) \tag{7}$$

where $p(\theta)$ is the prior distribution and $p(Y_{obs}|\theta)$ is the observed likelihood function. The problem of incomplete data is the observed data posterior $p(\theta|Y_{obs})$ is intractable and cannot easily be summarized or simulated. To overcome this problem, Y_{obs} is augmented by an assumed value of the Y_{mis}. The resulting complete-data posterior $p(\theta|Y_{obs}, Y_{mis})$ becomes much easier to handle. The observed data posterior is related to the complete-data posterior distribution that would have been obtained if we had observed the missing data Y_{mis}, namely

$$p(\theta|Y_{obs}, Y_{mis}) \propto p(\theta)\, p(Y_{obs}, Y_{mis}|\theta) \tag{8}$$

From equation (7) and (8) can be obtained

$$\begin{aligned} p(\theta|Y_{obs}) &= \int p(\theta, Y_{mis}|Y_{obs})\, dY_{mis} \\ &= \int p(\theta|Y_{obs}, Y_{mis})\, p(Y_{mis}|Y_{obs})\, dY_{mis} \end{aligned} \tag{9}$$

In equation (9), the posterior predictive distribution $p(Y_{mis}|Y_{obs})$ cannot be simulated directly. However, it is possible by create random draws of Y_{mis} from $p(Y_{mis}|Y_{obs})$ using techniques of MCMC. In this regard, we use a Gibbs sampling algorithm to draw the missing values Y_{mis} from $p(Y_{mis}|Y_{obs})$. By assuming that data have a multivariate normal distribution, data augmentation is applied to Bayesian inference with missing data by repeating the following steps

a). The imputation I-step

Given a current guess $\boldsymbol{\theta}^{(t)}$ of the parameter, create random draws of the missing values $\boldsymbol{Y_{mis}}$ from the posterior predictive distribution $\boldsymbol{p(Y_{mis}|Y_{obs})}$

$$Y_{i(mis)}^{(t+1)} \sim p(Y_{i(mis)}|Y_{obs}, \theta^{(t)}) \tag{10}$$

b). The posterior P-step

Then, with conditional to $\boldsymbol{Y_{i(mis)}^{(t+1)}}$, draw a new value of θ from the complete data posterior

$$\theta^{(t+1)} \sim p(\theta|Y_{obs}, Y_{i(mis)}^{(t+1)}) \tag{11}$$

Given starting from a initial values $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{Y_{mis}^{(0)}}$, these two step define a Gibbs sampler. Repeating the Gibbs sampling algorithm with large enough number of iterations, it creates stochastic sequences $\{\boldsymbol{\theta}^{(t)}\}$ and $\{\boldsymbol{Y_{mis}^{(t)}}\}$ whose stationary distribution are $\boldsymbol{p(\theta|Y_{obs})}$ and $\boldsymbol{p(Y_{mis}|Y_{obs})}$, respectively. In regard to MCMC, we use the initial value of the EM algorithm for the posterior mode, and the resulting EM estimates are used to begin the MCMC method. Moreover, we also specify the prior parameter information using one of a noninformative or ridge prior. A noninformative prior is used when no strong prior information is available about $\boldsymbol{\theta}$, it is customary to apply Bayes's theorem with the improper prior which is the limiting form of the normal inverted-Wishart. Meanwhile, a ridge prior is used when the sample covariance matrix is singular or nearly so, either because the data are sparse or because such strong relationships exist among the variables that certain linear combinations of the columns of $\boldsymbol{Y}$exhibit little or no variability [8].

2.5. Evaluation of Missing Values Imputation Results

In this paper, we evaluated the performance of these imputation methods by comparing the mean square error (MSE) of the resulting estimates $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u}$ using linear mixed models (LMM) for each district. For each district and time, we also evaluated the performance by comparing the mean absolute error (MAE) of the resulting estimates $\boldsymbol{y_{ij}}$.

Let $\boldsymbol{y_{ij}}$ denotes the response of repeated mesurement at the t^{th} time on the i^{th} sub-district, $\boldsymbol{i = 1, 2, \dots, m}$ and $\boldsymbol{t = 1, 2, \dots, n}$, the LMM for response vector of $\boldsymbol{y_i}$ is

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in} \end{bmatrix} \quad (12)$$

In matrix notation, equation (12) can be expressed as

$$\boldsymbol{y_i} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1_n}\alpha_i + \boldsymbol{e_i}, \quad i = 1,2,\dots,m \quad (13)$$

Where $\mathbf{X}_i = [\mathbf{1_n} \ \ \mathbf{I}_n]$ is the matrix of time factor fixed effect at the *i*th sub-districts, $\boldsymbol{\beta} = [\mu \, \tau_1 \dots \tau_n \,]$ is the parameter vector of fixed effect, α_i is the random effect of sub-districts, $\mathbf{1_n}$ is an unity vector order n, and $\boldsymbol{e_i} = [e_{i1} \dots e_{in} \,]$ is the vector of error model at i sub-districts. Assuming α_i and e_{it} are independenly distributed with mean 0 and varian σ_α and σ. The variance-covariance matrix of $\boldsymbol{y_i}$**,** is given by $\mathbf{V_i} = \mathbf{V_i}(\boldsymbol{\delta}) = \sigma_\alpha\mathbf{J}_n + \sigma\mathbf{I}_n$ where $\boldsymbol{\delta} = (\delta_1, \ \delta_2)' = (\sigma, \ \sigma_\alpha)'$, $\mathbf{J}_n$ is a unity square matrix order n and $\mathbf{I}_n$ is an identity matrix order n.
Another form of equation (13) is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e} \quad (14)$$

where $\mathbf{y} = (\boldsymbol{y}'_1, \dots, \boldsymbol{y}'_m)'$, $\boldsymbol{u} = (\alpha_1, \dots, \alpha_m)'$, $\mathbf{X} = \mathbf{1_n} \otimes [\mathbf{1_n} \ \ \mathbf{I}_n]$, $\mathbf{Z} = \mathbf{I}_n \otimes \mathbf{1_n}$, and $\boldsymbol{e} = (\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_m)'$. Assuming $\boldsymbol{u}$ and $\boldsymbol{e}$ are independenly distributed with mean **0** and covariance **G** and **R**. The variance-covariance matrix of $\mathbf{y}$ is $\mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \mathbf{ZGZ}' + \mathbf{R}$**.**

Refering to Rao and Molina [10], we have found the empirical best linear unbiased prediction (EBLUP) of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u}$ is

$$\widehat{\boldsymbol{\mu}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{u}} \quad (15)$$

where

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\delta}}_{RE}) = (\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}y} \quad (16)$$

is the best unbiased estimator (BLUE) of $\boldsymbol{\beta}$,

$$\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{u}}(\widehat{\boldsymbol{\delta}}_{RE}) = \mathbf{GZ'V^{-1}}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \quad (17)$$

is the best unbiased prediction (BLUP) of $\boldsymbol{\alpha_i}$, and $\widehat{\boldsymbol{\delta}}_{\boldsymbol{RE}}$ the restricted maximum likelihood (REML) estimator of $\boldsymbol{\delta}$ is obatained iteratively using the Fisher-scoring algorithm, with updateing equation

$$\widehat{\boldsymbol{\delta}}_{RE}\widehat{\boldsymbol{\delta}}_{RE}^{(a+1)} = \widehat{\boldsymbol{\delta}}_{RE}^{(a)} + \left[\boldsymbol{I}\left(\widehat{\boldsymbol{\delta}}_{RE}^{(a)}\right)\right]^{-1} \boldsymbol{s}\left[\widehat{\boldsymbol{\delta}}_{RE}^{(a)}\right] \quad (18)$$

Note that $\boldsymbol{s}\left[\widehat{\boldsymbol{\delta}}_{\boldsymbol{RE}}^{(\boldsymbol{a})}\right]$ is the partial derivative of log-likelihood function

$$l(\boldsymbol{\delta}) = c - \tfrac{1}{2}\log|\mathbf{V}| - \tfrac{1}{2}\log|\mathbf{X'V^{-1}X}| - \tfrac{1}{2}\mathbf{y'Py} \quad (19)$$

with respect to $\boldsymbol{\delta}$, $\boldsymbol{I}\left(\widehat{\boldsymbol{\delta}}_{\boldsymbol{RE}}^{(a)}\right)$ is the matrix of expected second-order derivatives of $-\boldsymbol{l}(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$. In equation (19), $\mathbf{P} = \mathbf{V^{-1}} - \mathbf{V^{-1}X}(\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}}$, c denotes a generic constant, and $\mathbf{Py} = \mathbf{V^{-1}}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$.

The MSE of $\widehat{\boldsymbol{\mu}}$ is given by

$$\begin{aligned} \mathrm{MSE}(\widehat{\boldsymbol{\mu}}) &= E[\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}]^2 \\ &\approx g_1(\hat{\delta}) + g_2(\hat{\delta}) + g_3(\hat{\delta}) \end{aligned} \quad (20)$$

where

$$g_1(\hat{\delta}) = \boldsymbol{Z}'(\mathbf{G} - \mathbf{GZ}'\mathrm{V}^{-1}\mathbf{ZG})\mathbf{Z} \quad (21)$$

is MSE of $\mathbf{X}\boldsymbol{\beta} + \mathbf{B}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ with $\mathbf{B} = \mathbf{ZGZ'V^{-1}}$,

$$g_2(\hat{\delta}) = \mathbf{D}(\mathrm{X}'\mathrm{V}^{-1}\mathrm{X})^{-1}\mathbf{D}' \tag{22}$$

is variance of $\mathbf{D}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ with $\mathbf{D} = (\mathbf{I} - \mathbf{B})\mathbf{X}$, and

$$g_3(\hat{\delta}) = \mathrm{tr}\left[(\partial \mathrm{B}/\partial\delta)\, \mathbf{V}(\partial \mathrm{B}/\partial\delta)'\, \bar{\mathbf{V}}(\hat{\delta})\right] \tag{23}$$

with $\bar{\mathbf{V}}(\hat{\boldsymbol{\delta}})$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\delta}}$.

Then for each district, the MAE of $\boldsymbol{y_{ij}}$ in the jth time is formulated

$$\mathrm{MAE}\ (y_{ij}) = \frac{1}{m}\sum_{i}^{m}\left|y_{ij} - \hat{y}_{ij}\right| \tag{24}$$

3. Results and Discussion

In this section, we showed the results of each imputation method to estimate the missing values of per capita expenditure data at sub-district level of five districts in Central Java province based on Susenas data from 2011 until 2014. After the data being completed then we estimated the mean of per capita expenditures for each district using the three methods. We also calculated the corresponding MSE. Moreover, for each district the mean of per capita expenditures by time as well as the mean absolute error of the resulting estimates of per capita expenditures were calculated.

Table 1 shows the results of each imputation method applied to the first ten sub-district of Banyumas district. The three methods produced different estimates of the missing values. Based on fifteen quarterly missing cells, the mean imputation results of the Yates method tend to be small, whereas the results of the MCMC method tend to be large. Nevertheless, we will evaluate these results by comparing the MSE after using these three methods.

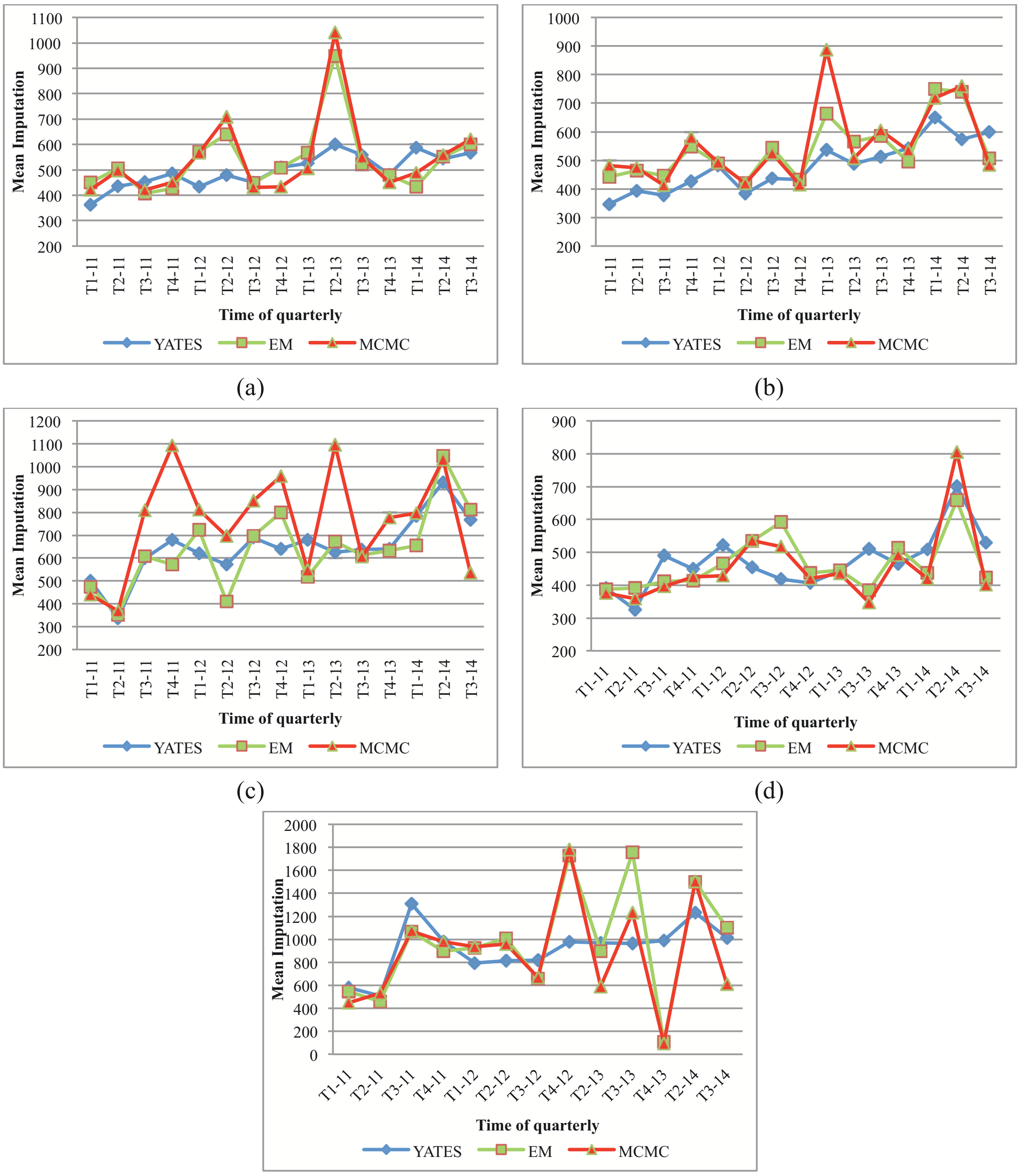
Table 2 shows the estimates of means as well as the corresponding MSE for each districts based on the three different methods of imputation. The results showed that MSE produced by the Yates method was lower than the MSE resulted from both the EM and the MCMC methods.

In Figure 1 the estimates of imputed means of per capita expenditures for each quarter using the three methods of five districts were presented. Generally for these five districts, the mean imputation resulted from the Yates method tended to be small, whereas the MCMC method tended to be large.

Figure 2 showed the estimates of MAE for each quarter in five districts, based on the three different methods of imputation. The results showed that MAE produced by the Yates method was lower than the MAE resulted from both EM and MCMC methods. It is apparent that to impute the missing values of per capita expenditure data at sub-district level we should use the Yates method.

Table 1. Mean Imputation Results for the First Ten Sub-districts in Banyumas District Using the Yates, EM Algorithm, and MCMC Methods

Sub-districts	Yates	EM Algorithm	MCMC
1. Lumbir	379.04	293.31	296.68
2. Wangon	633.46	547.53	547.48
3. Jatilawang	555.85	804.55	804.67
4. Rawalo	338.18	621.57	571.42
5. Kebasen	473.88	395.49	425.25
6. Kemranjen	455.10	549.33	637.50
7. Sumpiuh	397.25	467.16	541.47
8. Tambak	422.88	624.95	581.48
9. Somagede	403.23	261.31	368.17
10. Kalibagor	443.20	575.97	655.74



(a)

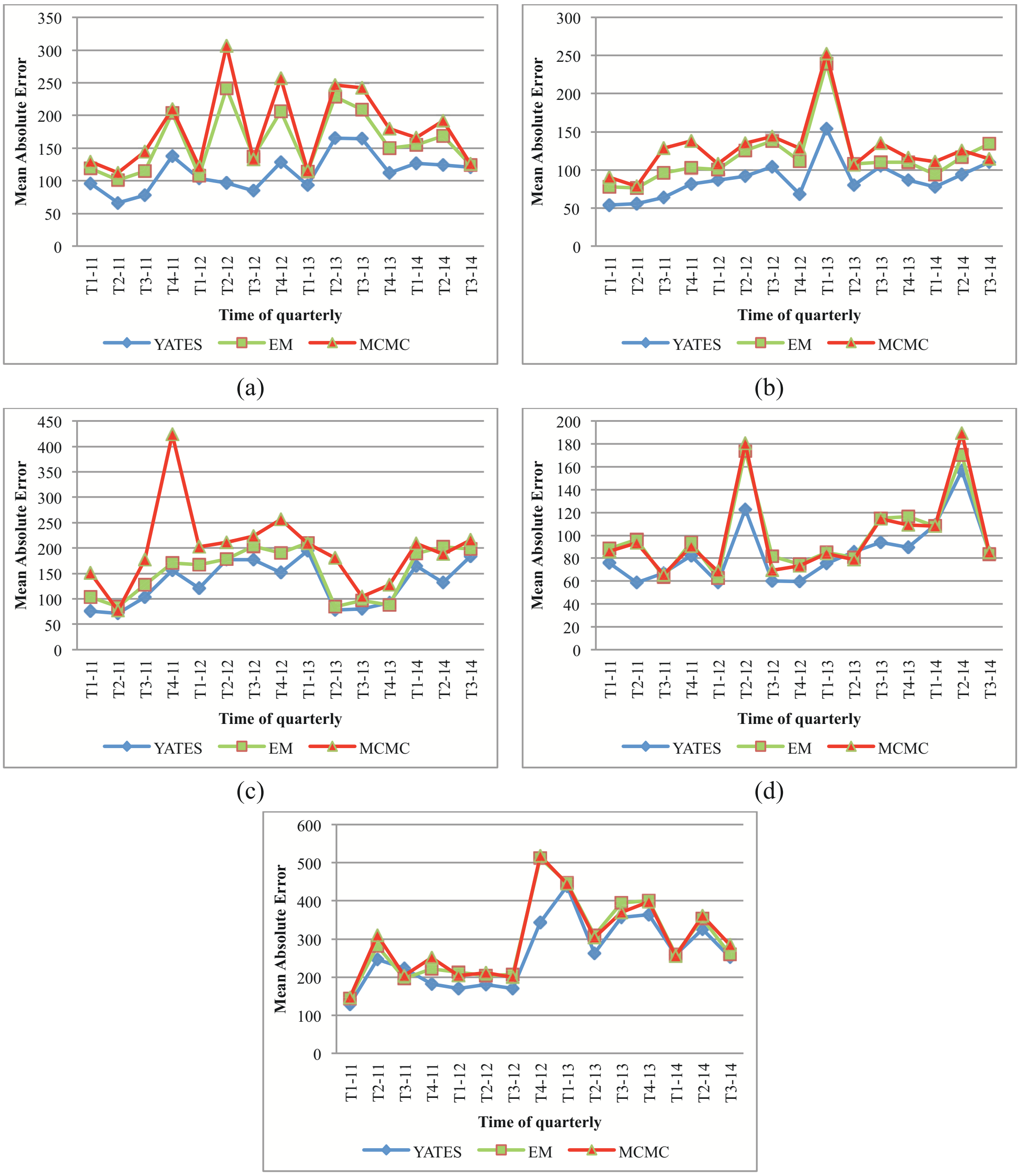
(b)

(c)

(d)

(e)

Figure 1. Comparison of Imputed Means among Yates, EM, and MCMC methods for five districts: (a) Banyumas, (b) Pati, (c) Semarang, (d) Brebes, and (e) Kota Semarang



(a)

(b)

(c)

(d)

(e)

Figure 2. Comparison of Mean Absolute Error among Yates, EM, and MCMC methods for each district: (a) Banyumas, (b) Pati, (c) Semarang, (d) Brebes, and (e) Kota Semarang

Table 2. Comparison of Mean and MSE results Missing Values Imputation between Yates method, EM Algorithm, and MCMC per district

Districts	Yates		EM Algorithm		MCMC	
	Mean	MSE	Mean	MSE	Mean	MSE
1. Banyumas	490.03	3.5×10^4	527.71	5.6×10^4	531.80	6.6×10^4
2. Pati	460.96	2.7×10^4	520.30	3.4×10^4	533.48	3.8×10^4
3. Semarang	635.33	4.9×10^4	632.76	5.7×10^4	769.88	9.7×10^4
4. Brebes	456.01	1.8×10^4	460.80	2.1×10^4	450.62	2.2×10^4
5. Kota Semarang	885.61	15.5×10^4	959.71	18.0×10^4	915.84	18.2×10^4

Note : MSE were calculated based on residual sum of squares divided by their degree of freedom.

4. Conclusion

The results of this study revealed that the MSE produced by the Yates method was lower than the MSE resulted from both the EM algorithm and the MCMC method. These results were consistent with MAE of the Yates method which was also lower than the MAE resulted from the other two methods. Hence, based on those results, we concluded that to impute the missing values of per capita expenditure data at sub-district level we should use the Yates method.

Acknowledgments
We would like to thank Statistics Indonesia for giving us permission to use Susenas data for this paper.

References

[1] Schafer J L and Graham J W 2002 *Our View of the State of the Art* (*Psychological Methods* 7) **2** 147–177

[2] Nakai M and Weiming K 2011 *Int. Journal of Math. Analysis* **5** (1) 1-13

[3] Buuren S V 2012 *Flexible Imputation of Missing Data* (Boca Raton: Taylor & Francis Group, LLC)

[4] Donders A R T, van der Heijden G J M G, Stijnen T, and Moons K G M 2006 *Journal of Clinical Epidemiology* **59** 1087-1091

[5] Little R J A and Rubin D B 1987 *Statistical Analysis with Missing Data* (New York: John Wiley & Son Inc.)

[6] Dempster A P, Laird N M, and Rubin D B 1977 *Journal of the Royal Statistical Society* Series B **39** (1) 1-38

[7] Little R J A and Rubin D B 2002 *Statistical Analysis with Missing Data Second Edition* (Hoboken, New Jersey: John Wiley & Son Inc.)

[8] Schafer J L 1997 *Analysis of Incomplete Multivariate Data* (London: Chapman & Hall/CRC)

[9] Watanabe M and Yamaguchi K 2004 *The EM Algorithm and Related Statistical Models* (New York: Marcel Dekker, Inc.)

[10] Rao J N K and molina I 2015 *Small Area Estimation Second Edition* (Hoboken, New Jersey: John Wiley & Son Inc.)