A Fresh Imputing Method Using Sensible Constraints on Study and Auxiliary Variables: Preliminary Findings

Choukri Mohamed, Stephen A. Sedory and Sarjinder Singh
Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363, USA
E-mail: choukri72@hotmail.com
E-mail: Stephen.sedory@tamuk.edu
E-mail: Sarjinder.Singh@tamuk.edu

Abstract
In this paper, we suggest a new method for imputing missing values by suggesting the use of new sensible constraints on a study variable and auxiliary variables. While we limit ourselves to findings using only one auxiliary variable, extension of these results to multi–auxiliary variables is on the way. The proposed imputing method leads to an estimator which is asymptotically equivalent to the linear regression estimator.

Key Words: Imputation, missing data, auxiliary information.

1. Introduction

Hansen and Hurwitz (1946) were the first to deal with the problem of non-response in mail-surveys. These days that types of non-response occurs in other types of surveys such as web-surveys and telephone surveys. Today, mail surveys or telephone surveys are commonly used by bureaucratic or business organisations because of their low cost. Rubin (1976) dealt with two concepts: missing at random (MAR) and observed at random (OAR). One could refer to Heitjan and Basu (1996) to learn more about MAR and MCAR. Let us consider

$$\bar{Y} = N^{-1} \sum_{i\in\Omega} y_i \tag{1.1}$$

to be the population mean of a study variable y in a finite population $\Omega = \{1,2,..,i,..,N\}$. Assume a simple random and without replacement sample (SRSWOR), s, of size n is taken from Ω to estimate this population mean $\bar{Y}$. Assume that r i.e. the number of responding units out of n sampled units. Let the set of responding units be denoted by A and that of non-responding units be denoted by A^c. For every unit $i \in A$, the value y_i is observed, and for the units $i \in A^c$, the y_i values are missing and imputed values are to be derived. The first choice is to consider dropping the missing $(n-r)$ data values in the set A^c from the sample s of n data values and consider an estimator of the population mean $\bar{Y}$ as:

$$\bar{y}_r = \frac{1}{r} \sum_{i\in A} y_i \tag{1.2}$$

which is the sample mean of the r values in the responding set A. Assuming the data is missing completely at random (MCAR), then applying the concept of double sampling as given in Cochran (1963), it is easy to verify that the sample mean $\bar{y}_r$ in (1.2) is an unbiased estimator of the population mean $\bar{Y}$ with conditional variance, for a given value of r, given by:

$$V(\bar{y}_r) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 \tag{1.3}$$

where $S_y^2 = (N-1)^{-1} \sum_{i \in \Omega} (y_i - \bar{Y})^2$ is the population mean squared error (or population variance) for the study variable. We will indicate the point estimator of the population mean, given by,

$$\bar{y}_{\text{point}} = \frac{1}{n} \sum_{i \in s} d_{\bullet i} \tag{1.4}$$

where $d_{\bullet i}$ are different data values based on different methods of imputation. Under mean method of imputation, the data after imputation take the form:

$$d_{\bullet i} = \begin{cases} y_i & \textit{if} \quad i \in A \\ \bar{y}_r & \textit{if} \quad i \in A^c \end{cases} \tag{1.5}$$

and the point estimator (1.5) becomes

$$\bar{y}_{\text{point}} = \frac{1}{r} \sum_{i=1}^{r} y_i = \bar{y}_r = \bar{y}_m \tag{1.6}$$

which remains the same as the sample mean obtained after discarding the non-responding units in the sample. Thus the mean method of imputation is as precise as the method of dropping the missing data values, that is, $\bar{y}_m = \bar{y}_r$.

2. Proposed Method of Imputation

We propose a new method of imputing missing values as:

$$d_{\bullet i} = \begin{cases} y_i & \textit{if} \quad i \in A \\ \hat{\hat{y}}_i & \textit{if} \quad i \in A^c \end{cases} \tag{2.1}$$

where $\hat{\hat{y}}_i$, $i \in A^c$ is an imputed value by the proposed new method of imputation. We consider a chi-squared type distance function D_f between the imputed mean values $\bar{y}_r$ and the new imputed values $\hat{\hat{y}}_i$ for $i \in A^c$ as:

$$D_f = \frac{1}{2} \sum_{i \in A^c} \frac{(\hat{\hat{y}}_i - \bar{y}_r)^2}{\bar{y}_r} \tag{2.2}$$

We assume that imputation is carried out with the aid of an auxiliary variable, x, such that x_i, the value of x for unit i, is known and positive for every $i \in s = A \cup A^c$. In other words, the data $x_s = \{x_i : i \in \in s\}$ are known. Thus in case of missing data, the sample data values have the following structure:

$$d_{\bullet i} = \begin{cases} (y_i, x_i) & \text{if} \quad i \in A \\ (\text{missing}, x_i) & \text{if} \quad i \in A^c \end{cases} \tag{2.3}$$

We consider minimization of the proposed chi-squared distance D_f defined in (2.2) subject to the following sensible constraint:

$$\frac{1}{(n-r)} \sum_{i \in A^c} \hat{\hat{y}}_i (x_i - \bar{x}_r) = \frac{1}{(r-1)} \sum_{i \in A} y_i (x_i - \bar{x}_r) \tag{2.4}$$

or equivalently,

$$\frac{1}{(n-r)}\sum_{i\in A^c}\hat{\hat{y}}_i(x_i-\bar{x}_r)=s_{xy(r)} \tag{2.5}$$

where

$$s_{xy(r)}=\frac{1}{(r-1)}\sum_{i\in A}(y_i-\bar{y}_r)(x_i-\bar{x}_r)=\frac{1}{(r-1)}\sum_{i\in A}y_i(x_i-\bar{x}_r) \tag{2.6}$$

The constraint (2.4) makes sense in that, if the data is missing completely at random, then the covariance between the study variable and auxiliary variable for the responding units in a sample should be same as the covariance between the imputed values and the auxiliary variable for the non-responding units in a sample. Note the use of $\bar{x}_r$ in the left hand side of equation (2.5); this is somewhat arbitrary, one could also use $\bar{x}_{(n-r)}$. At the same time please note the use of $(n-r)$ and $(r-1)$ in the denominator of non-responding and responding units; this adjustment is necessary to obtain consistent method of imputation. The proposed constraint in (2.4) has been named a "sensible constraint" on both the study variable and the auxiliary variable.

The Lagrange function is given by:

$$L=\frac{1}{2}\sum_{i\in A^c}\frac{(\hat{\hat{y}}_i-\bar{y}_r)^2}{\bar{y}_r}-\lambda\left[\frac{1}{(n-r)}\sum_{i\in A^c}\hat{\hat{y}}_i(x_i-\bar{x}_r)-s_{xy(r)}\right] \tag{2.7}$$

where λ is a Lagrange's multiplier constant. On differentiating (2.7) with respect to $\hat{\hat{y}}_i$ and equating to zero, that is, on setting:

$$\frac{\partial L}{\partial\,\hat{\hat{y}}}=0$$

we have

$$\frac{(\hat{\hat{y}}_i-\bar{y}_r)}{\bar{y}_r}-\frac{\lambda}{(n-r)}(x_i-\bar{x}_r)=0$$

which implies that the adjusted imputed mean values $\hat{\hat{y}}_i$ are given by:

$$\hat{\hat{y}}_i=\bar{y}_r+\lambda\frac{\bar{y}_r}{(n-r)}(x_i-\bar{x}_r) \tag{2.8}$$

On substituting (2.8) in the sensible constraint (2.4), we have

$$\frac{1}{(n-r)}\sum_{i\in A^c}\left[\bar{y}_r+\frac{\lambda\,\bar{y}_r}{(n-r)}(x_i-\bar{x}_r)\right](x_i-\bar{x}_r)=s_{xy(r)}$$

or

$$\sum_{i\in A^c}\left[\bar{y}_r+\frac{\lambda\,\bar{y}_r}{(n-r)}(x_i-\bar{x}_r)\right](x_i-\bar{x}_r)=(n-r)s_{xy(r)}$$

or

$$n\bar{y}_r(\bar{x}_n-\bar{x}_r)+\frac{\lambda\,\bar{y}_r}{(n-r)}\sum_{i\in A^c}(x_i-\bar{x}_r)^2=(n-r)s_{xy(r)}$$

or

$$\lambda=\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{\dfrac{\bar{y}_r}{(n-r)}\sum_{i\in A^c}(x_i-\bar{x}_r)^2} \tag{2.9}$$

Note that:

$$\sum_{i\in A^c}(x_i-\bar{x}_r)^2=\sum_{i\in s}(x_i-\bar{x}_r)^2-\sum_{i\in A}(x_i-\bar{x}_r)^2$$
$$=\sum_{i\in s}\left[(x_i-\bar{x}_n)+(\bar{x}_n-\bar{x}_r)\right]^2-(r-1)s^2_{x(r)}$$
$$=\sum_{i\in s}\left[(x_i-\bar{x}_n)^2+(\bar{x}_n-\bar{x}_r)^2+2(x_i-\bar{x}_r)(\bar{x}_n-\bar{x}_r)\right]-(r-1)s^2_{x(r)}$$
$$=\sum_{i\in s}(x_i-\bar{x}_n)^2+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}$$
$$=(n-1)s^2_{x(n)}+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)} \tag{2.10}$$

where

$$s^2_{x(n)}=(n-1)^{-1}\sum_{i\in s}(x_i-\bar{x}_n)^2 \text{ and } s^2_{x(r)}=(r-1)^{-1}\sum_{i\in A}(x_i-\bar{x}_r)^2 .$$

On substituting (2.10) in (2.9), the value of the Lagrange multiplier λ is given by

$$\lambda=\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{\frac{\bar{y}_r}{(n-r)}\left[(n-1)s^2_{x(n)}+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}\right]} \tag{2.11}$$

On substituting the value of λ from (2.11) into (2.8), the new imputed values are given by:

$$\hat{\hat{y}}_i=\bar{y}_r+\left\{\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{(n-1)s^2_{x(n)}+n(\vec{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}}\right\}(x_i-\bar{x}_r) \tag{2.12}$$

On substituting (2.12) into (2.1), the new method of imputation leads to a new data set given by:

$$d_{i\bullet}=\begin{cases} y_i & i\in A \\ \bar{y}_r+\left(\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{(n-1)s^2_{x(n)}+n(\bar{x}_n-\vec{x}_r)^2-(r-1)s^2_{x(r)}}\right)(x_i-\bar{x}_r) & \text{if } i\in A^c \end{cases} \tag{2.13}$$

With the new imputed valued values in (2.13), the point estimator of population mean $\bar{Y}$ becomes:

$$\bar{y}_{\text{point}}=\frac{1}{n}\sum_{i\in s}d_{\bullet i}=\frac{1}{n}\left[\sum_{i\in A}y_i+\sum_{i\in A^c}\hat{\hat{y}}_i\right]$$
$$=\frac{1}{n}\left[\sum_{i\in A}y_i+\sum_{i\in A^c}\left\{\bar{y}_r+\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{(n-1)s^2_{x(n)}+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}}\right\}(x_i-\bar{x}_r)\right]$$
$$=\frac{1}{n}\left[r\bar{y}_r+(n-r)\bar{y}_r+\sum_{i\in A^c}\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{(n-1)s^2_{x(n)}+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}}(x_i-\bar{x}_r)\right]$$
$$=\frac{1}{n}\left[n\bar{y}_r+\frac{(n-r)s_{xy(r)}-n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{(n-1)s^2_{x(n)}+n(\bar{x}_n-\bar{x}_r)^2-(r-1)s^2_{x(r)}}\left(n\bar{x}_n-n\bar{x}_r\right)\right]$$

$$= \bar{y}_r + \frac{(n-r)s_{xy(r)} - n\bar{y}_r(\bar{x}_n - \bar{x}_r)}{(n-1)s^2_{x(n)} + n(\bar{x}_n - \bar{x}_r)^2 - (r-1)s^2_{x(r)}}(\bar{x}_n - \bar{x}_r)$$

$$= \bar{y}_r + \hat{\beta}_{\text{ch}}(\bar{x}_n - \bar{x}_r) = \bar{y}_{\text{ch}} \text{ (say)} \tag{2.14}$$

where

$$\hat{\beta}_{\text{ch}} = \frac{s_{xy(r)}\left[(n-r) - \frac{n\bar{y}_r(\bar{x}_n - \bar{x}_r)}{s_{xy(r)}}\right]}{s^2_{x(n)}\left[(n-1) + \frac{n(\bar{x}_n - \bar{x}_r)^2}{s^2_{x(n)}} - \frac{(r-1)s^2_{x(r)}}{s^2_{x(n)}}\right]} \tag{2.15}$$

is an estimator of the regression coefficient

$$\beta = S_{xy} \big/ S_x^2 \tag{2.16}$$

where $S_{xy} = (N-1)^{-1} \sum_{i \in \Omega} (y_i - \bar{Y})(x_i - \bar{X})$ and $S_x^2 = (N-1)^{-1} \sum_{i \in \Omega} ((x_i - \bar{X})^2$.

In short using the new imputed values, the point estimator of the population mean $\bar{Y}$ becomes:

$$\bar{y}_{ch(1)} = \bar{y}_r + \hat{\beta}_{\text{ch}}(\bar{x}_n - \bar{x}_r) \tag{2.17}$$

In the next section, we define some notation that is useful in studying the properties of the proposed estimator in (2.17) under the proposed new imputed method of imputation.

3. Notations

Let us define: $\varepsilon_0 = \frac{\bar{y}_r}{\bar{Y}} - 1$, $\varepsilon_1 = \frac{\bar{x}_r}{\bar{X}} - 1$, $\varepsilon_2 = \frac{\bar{x}_n}{\bar{X}} - 1$, $\varepsilon_3 = \frac{s^2_{x(r)}}{S_x^2} - 1$,

$\varepsilon_4 = \frac{s^2_{x(n)}}{S_x^2} - 1$ and $\varepsilon_5 = \frac{s^2_{xy(r)}}{S_{xy}} - 1$ such that: $E(\varepsilon_i) = 0$, $\forall \quad i = 1, 2, 3, 4, 5$

and

$E(\varepsilon_0^2) = \left(\frac{1}{r} - \frac{1}{N}\right)C_y^2$; $\quad E(\varepsilon_1^2) = \left(\frac{1}{r} - \frac{1}{N}\right)C_x^2$; $\quad E(\varepsilon_2^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2$;

$E(\varepsilon_3^2) = \left(\frac{1}{r} - \frac{1}{N}\right)(\lambda_{04} - 1)$; $\quad E(\varepsilon_4^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\lambda_{04} - 1)$; $\quad E(\varepsilon_5^2) = \left(\frac{1}{r} - \frac{1}{N}\right)(\lambda_{22} - 1)$;

$E(\varepsilon_0\varepsilon_1) = \left(\frac{1}{r} - \frac{1}{N}\right)\rho C_y C_x$; $\quad E(\varepsilon_0\varepsilon_2) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_y C_x$;

$E(\varepsilon_0\varepsilon_3) = \left(\frac{1}{r} - \frac{1}{N}\right)C_y\lambda_{12}$; $\quad E(\varepsilon_0\varepsilon_4) = \left(\frac{1}{n} - \frac{1}{N}\right)C_y\lambda_{12}$;

$E(\varepsilon_0\varepsilon_5) = \left(\frac{1}{r} - \frac{1}{N}\right)C_y\frac{\lambda_{21}}{\rho}$; $\quad E(\varepsilon_1\varepsilon_2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2$; $E(\varepsilon_1\varepsilon_3) = \left(\frac{1}{r} - \frac{1}{N}\right)C_x\lambda_{03}$;

$E(\varepsilon_1\varepsilon_4) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x\lambda_{03}$; $\quad E(\varepsilon_1\varepsilon_5) = \left(\frac{1}{r} - \frac{1}{N}\right)C_x\frac{\lambda_{12}}{\rho}$;

$$E(\varepsilon_2\varepsilon_3)=\left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03}\,;\quad E(\varepsilon_2\varepsilon_4)=\left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03}\,;$$

$$E(\varepsilon_2\varepsilon_5)=\left(\frac{1}{n}-\frac{1}{N}\right)C_x\frac{\lambda_{12}}{\rho}\,;\quad E(\varepsilon_3\varepsilon_4)=\left(\frac{1}{n}-\frac{1}{N}\right)(\lambda_{04}-1)\,;$$

$$E(\varepsilon_3\varepsilon_5)=\left(\frac{1}{r}-\frac{1}{N}\right)\left(\frac{\lambda_{13}}{\rho}-1\right)\text{ and }E(\varepsilon_4\varepsilon_5)=\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{\lambda_{13}}{\rho}-1\right)$$

where

$$\lambda_{ab}=\frac{\mu_{ab}}{\mu_{20}^{a/2}\mu_{02}^{b/2}}\,;\quad C_x^2=\frac{S_x^2}{\bar{X}^2}=\frac{\mu_{02}}{\bar{X}^2}\,;\quad C_y^2=\frac{S_y^2}{\bar{Y}^2}=\frac{\mu_{20}}{\bar{Y}^2}\,;\quad \rho=\frac{S_{xy}}{S_xS_y}=\frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}\,;$$

$$\mu_{ab}=(N-1)^{-1}\sum_{i\in\Omega}(y_i-\bar{Y})^a(x_i-\bar{X})^b\,;\text{ with }a,b=0,1,2,3,4\,.$$

In the next, we investigate asymptotic properties of the estimator $\hat{\beta}_{ch}$ of the regression coefficient β and the proposed regression type estimator $\bar{y}_{ch}$ of the population mean $\bar{Y}$.

4 Properties of the Proposed Estimators

We have the following theorems:

Theorem 4.1. The estimator $\hat{\beta}_{ch}$ is a consistent estimator of the regression coefficient β.

Proof. The estimator $\hat{\beta}_{ch}$ of the regression coefficient β, in terms of ε_i, can be written as:

$$\hat{\beta}_{\text{ch}}=\frac{s_{xy(r)}\left[(n-r)-\dfrac{n\bar{y}_r(\bar{x}_n-\bar{x}_r)}{s_{xy(r)}}\right]}{s_{x(n)}^2\left[(n-1)+\dfrac{n(\bar{x}_n-\bar{x}_r)^2}{s_{x(n)}^2}-\dfrac{(r-1)s_{x(r)}^2}{s_{x(n)}^2}\right]}$$

$$=\frac{S_{xy}(1+\varepsilon_5)\left[(n-r)-\dfrac{n\bar{Y}(1+\varepsilon_0)\{\bar{X}(1+\varepsilon_2)-X(1+\varepsilon_1)\}}{S_{xy}(1+\varepsilon_5)}\right]}{S_x^2(1+\varepsilon_4)\left[(n-1)+\dfrac{n\{\bar{X}(1+\varepsilon_2)-\bar{X}(1+\varepsilon_1)\}^2}{S_x^2(1+\varepsilon_4)}-\dfrac{(r-1)S_x^2(1+\varepsilon_3)}{S_x^2(1+\varepsilon_4)}\right]}$$

$$=\frac{S_{xy}(1+\varepsilon_5)\left[(n-r)-\dfrac{n\bar{Y}\ \bar{X}(1+\varepsilon_0)\{\varepsilon_2-\varepsilon_1\}}{S_{xy}(1+\varepsilon_5)}\right]}{S_x^2(1+\varepsilon_4)\left[(n-1)+\dfrac{n\bar{X}^2\{\varepsilon_2-\varepsilon_1\}^2}{S_x^2(1+\varepsilon_4)}-\dfrac{(r-1)(1+\varepsilon_3)}{(1+\varepsilon_4)}\right]}$$

$$=\frac{S_{xy}(1+\varepsilon_5)\left[(n-r)-\dfrac{n\bar{Y}\ \bar{X}(1+\varepsilon_0)\{\varepsilon_2-\varepsilon_1\}(1+\varepsilon_5)^{-1}}{S_{xy}}\right]}{S_x^2(1+\varepsilon_4)\left[(n-1)+\dfrac{n\bar{X}^2\{\varepsilon_2-\varepsilon_1\}^2(1+\varepsilon_4)^{-1}}{S_x^2}-(r-1)(1+\varepsilon_3)(1+\varepsilon_4)^{-1}\right]}$$

$$= \frac{(n-r)S_{xy}(1+\varepsilon_5)\left[1-\frac{n\bar{Y}\ \bar{X}(1+\varepsilon_0)\{\varepsilon_2-\varepsilon_1\}(1+\varepsilon_5)^{-1}}{(n-r)S_{xy}}\right]}{S_x^2(1+\varepsilon_4)\left[(n-1)+\frac{n\bar{X}^2\{\varepsilon_2-\varepsilon_1\}^2(1-\varepsilon_4+\varepsilon_4^2+..)}{S_x^2}-(r-1)(1+\varepsilon_3-\varepsilon_4+\varepsilon_4^2-\varepsilon_3\varepsilon_4...)\right]}$$

$$= \beta(1+\varepsilon_5-\varepsilon_4+\varepsilon_4^2-\varepsilon_4\varepsilon_5+..)\left[1-\frac{n\bar{Y}\ \bar{X}(\varepsilon_2-\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_1-\varepsilon_2\varepsilon_5+\varepsilon_1\varepsilon_5+...\}}{(n-r)S_{xy}}\right]$$

$$\times\left[\begin{array}{l} 1-\frac{n\bar{X}^2\{\varepsilon_2^2+\varepsilon_1^2-2\varepsilon_1\varepsilon_2\}(1-\varepsilon_4+\varepsilon_4^2+...)}{(n-r)S_x^2}+\frac{(r-1)}{(n-r)}(\varepsilon_3-\varepsilon_4+\varepsilon_4^2-\varepsilon_3\varepsilon_4...) \\ +\frac{(r-1)^2}{(n-r)^2}(\varepsilon_3-\varepsilon_4+\varepsilon_4^2-\varepsilon_3\varepsilon_4...)^2+... \end{array}\right]$$

$$= \beta\Big[1+\varepsilon_5-\varepsilon_4+\varepsilon_4^2-\varepsilon_4\varepsilon_5$$

$$-\frac{n\vec{Y}\ \bar{X}}{(n-r)S_{xy}}\left(\varepsilon_2-\varepsilon_1+\varepsilon_0\varepsilon_2-\varepsilon_0\varepsilon_1-\varepsilon_2\varepsilon_5+\varepsilon_1\varepsilon_5+\varepsilon_2\varepsilon_5-\varepsilon_1\varepsilon_5-\varepsilon_2\varepsilon_4+\varepsilon_1\varepsilon_4\right)$$

$$-\frac{n\bar{X}^2(\varepsilon_2^2+\varepsilon_1^2-2\varepsilon_1\varepsilon_2)}{(n-r)S_x^2}+\frac{(r-1)}{(n-r)}(\varepsilon_3-\varepsilon_4+\varepsilon_4^2-\varepsilon_3\varepsilon_4)+\frac{(r-1)^2}{(n-r)^2}(\varepsilon_3^2+\varepsilon_4^2-2\varepsilon_3\varepsilon_4)$$

$$+\frac{(r-1)}{(n-1)}\left(\varepsilon_3\varepsilon_5-\varepsilon_3\varepsilon_4-\varepsilon_4\varepsilon_5+\varepsilon_4^2\right)-\frac{n(r-1)\bar{Y}\ \bar{X}}{(n-r)^2 S_{xy}}(\varepsilon_2\varepsilon_3-\varepsilon_1\varepsilon_3-\varepsilon_2\varepsilon_4+\varepsilon_1\varepsilon_4)\Big]$$

$$= \beta+\beta\left[\varepsilon_5-\varepsilon_4-\frac{n\bar{Y}\ \bar{X}}{(n-r)S_{xy}}(\varepsilon_2-\varepsilon_1)+\frac{(r-1)}{(n-r)}(\varepsilon_3-\varepsilon_4)+O(\varepsilon^2)\right] \qquad (4.1)$$

Taking expected value on both sides of (4.1), we have

$$E\left(\hat{\beta}_{ch(1)}\right)=\beta+O(n^{-1}) \qquad (4.2)$$

which proved the theorem.

Theorem 4.2. The relative bias in the proposed estimator $\bar{y}_{ch(1)}$, to the first order of approximation, is given by

$$RB(\bar{y}_{ch(1)})=-\left(\frac{1}{r}-\frac{1}{n}\right)\frac{\mu_{12}}{\bar{Y}S_x^2}-\frac{1}{r} \qquad (4.3)$$

is a consistent estimator, in terms of r, of the population mean $\bar{Y}$.

Proof. The proposed regression type estimator $\bar{y}_{ch(1)}$, in terms of ε_i, can be approximated as:

$$\bar{y}_{\text{ch}(1)} = \bar{Y}(1+\varepsilon_0) + \left[\beta + \beta\left(\varepsilon_5 - \varepsilon_4 - \frac{n\bar{Y}\,\bar{X}(\varepsilon_2-\varepsilon_1)}{(n-r)S_{xy}} + \frac{(r-1)}{(n-r)}(\varepsilon_3-\varepsilon_4) + O(\varepsilon^2)\right)\right]\bar{X}(\varepsilon_2-\varepsilon_1)$$

$$= \bar{Y}(1+\varepsilon_0) + \beta\,\bar{X}(\varepsilon_2-\varepsilon_1) + \beta\bar{X}\left\{\varepsilon_2\varepsilon_5 - \varepsilon_2\varepsilon_4 - \varepsilon_1\varepsilon_5 + \varepsilon_1\varepsilon_4 - \frac{n\bar{Y}\,\bar{X}}{(n-r)S_{xy}}(\varepsilon_2^2 - 2\varepsilon_1\varepsilon_2 + \varepsilon_1^2)\right.$$

$$\left. + \frac{(r-1)}{(n-r)}(\varepsilon_2\varepsilon_3 - \varepsilon_2\varepsilon_4 - \varepsilon_1\varepsilon_3 + \varepsilon_1\varepsilon_4)\right\} \qquad (4.4)$$

Thus the asymptotic bias in the proposed estimator $\bar{y}_{ch(1)}$ is given by:

$$B(\bar{y}_{ch(1)}) = E(\bar{y}_{ch(1)}) - \bar{Y}$$

$$= \beta\,\bar{X}\left[\left(\frac{1}{n}-\frac{1}{N}\right)C_x\frac{\lambda_{12}}{\rho} - \left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03} - \left(\frac{1}{r}-\frac{1}{N}\right)C_x\frac{\lambda_{12}}{\rho} + \left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03}\right.$$

$$- \frac{n\bar{Y}\,\bar{X}}{(n-r)S_{xy}}\left\{\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2 - 2\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2 + \left(\frac{1}{r}-\frac{1}{N}\right)C_x^2\right\}$$

$$\left. + \frac{(r-1)}{(n-r)}\left\{\left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03} - \left(\frac{1}{n}-\frac{1}{N}\right)C_x\lambda_{03} - \left(\frac{1}{r}-\frac{1}{N}\right)C_x\lambda_{03} + \left(\frac{1}{r}-\frac{1}{N}\right)C_x\lambda_{03}\right\}\right]$$

$$= -\left(\frac{1}{r}-\frac{1}{n}\right)\frac{\mu_{12}}{S_x^2} - \frac{\bar{Y}}{r} \qquad (4.5)$$

The relative bias in the proposed estimator $\bar{y}_{ch(1)}$ is given by

$$\text{RB}(\bar{y}_{\text{ch}(1)}) = \frac{B(\bar{y}_{\text{ch}(1)})}{\bar{Y}} = -\left(\frac{1}{r}-\frac{1}{n}\right)\frac{\mu_{12}}{\bar{Y}S_x^2} - \frac{1}{r}$$

which proves the theorem.

Theorem 4.3. The mean squared error of the proposed estimator, $\bar{y}_{ch(1)}$, to the first order of approximation, is given by:

$$\text{MSE}(\bar{y}_{ch(1)}) = \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)S_y^2(1-\rho^2) \qquad (4.6)$$

Proof. The mean squared error of the proposed estimator $\bar{y}_{ch(1)}$ is given by

$$\text{MSE}(\bar{y}_{ch(1)}) = E\left[\bar{y}_{ch} - \bar{Y}\right]^2 \approx E\left[\bar{Y}\,\varepsilon_0 + \beta\,\bar{X}(\varepsilon_2-\varepsilon_1)\right]^2$$

$$= \bar{Y}^2E(\varepsilon_0^2) + \beta^2\bar{X}^2(E(\varepsilon_2^2) + E(\varepsilon_1^2) - 2E(\varepsilon_1\varepsilon_2)) + 2\beta\bar{Y}\,\bar{X}(E(\varepsilon_0\varepsilon_2) - E(\varepsilon_0\varepsilon_1))$$

$$= \bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)C_y^2 + \beta^2\bar{X}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2 + \left(\frac{1}{r}-\frac{1}{N}\right)C_x^2 - 2\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2\right]$$

$$+ 2\beta\bar{Y}\,\bar{X}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_yC_x - \left(\frac{1}{r}-\frac{1}{N}\right)\rho C_yC_x\right]$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(\beta^2S_x^2 - 2\beta S_{xy}\right)$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)S_y^2\left(1-\rho^2\right)$$

which proves the theorem.

In the next section, we compare the proposed methods of imputations with the mean method of imputation through a simulation study using a real population.

5. Illustrations with a real data set

We use a dataset, FEV.DAT, available on the CD that accompanies the text by Rosner (2006) that contains data on $N = 654$ children from the Childhood Respiratory Disease Study done in Boston.

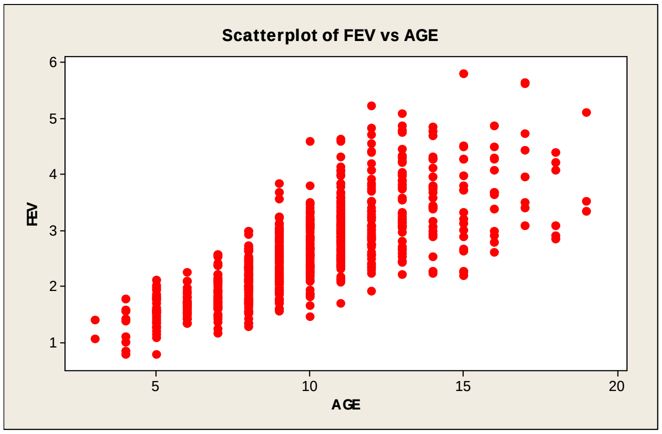

Fig. 5.1. Scatter plot of AGE versus FEV

Among the variables are Age and FEV (forced expiratory volume). We consider the problem of imputing the FEV (=Y) of a child given the AGE (=X) value of that child. To investigate several situations, following Singh (2013), we also used a Box-Cox transformation on the auxiliary variable AGE given by:

$$X = \frac{(\text{AGE})^T - 1}{T} \tag{5.1}$$

for values of T between -4 to +4 with a step of 0.5. We choose simple random and with replacement sample (SRSWR) of $n = 60$ units, which is approximately 9.17% of the total population size N. We let r, the number of respondents, vary between 5 and 50 with a step of 5, which means a response rate of 8.3% to 83.33% with an increase in response rate of 8.33% in every step. We compute the value of approximate percent relative bias in the proposed method of imputation as:

$$RB = -\left[\left(\frac{1}{r} - \frac{1}{n}\right)\frac{\mu_{12}}{\bar{Y}S_x^2} + \frac{1}{r}\right] \times 100\% \tag{5.2}$$

The percent relative efficiency of the proposed method of imputation with respect to the mean method of imputation, is computed as:

$$\text{RE} = \frac{V(\bar{y}_r)}{\text{Min.V}(\bar{y}_{\text{new(ch)}})} \times 100\% \tag{5.3}$$

We wrote FORTRAN codes (see Appendix) to find the values of the percent relative bias (RB) and the percent relative efficiency of the proposed method of imputation over the mean method of imputation. In producing the following output given in Table 5.1 we exclude those cases where the percent relative bias exceed 10%

Table 5.1. Relative Bias (RB) and Relative Efficiency (RE) of the proposed imputation method with one auxiliary variable.

T	ρ	r	RR	RB	RE		T	ρ	r	RR	RB
-4	0.374	10	16.67	-6.07	113.44		0.737	35	58.33	-2.69	131.42
	0.374	15	25	-4.31	112.04		0.737	40	66.67	-2.38	123.89
	0.374	20	33.33	-3.43	110.65		0.737	45	75	-2.15	117.07
	0.374	25	41.67	-2.9	109.28		0.737	50	83.33	-1.95	110.87
	0.374	30	50	-2.55	107.91	0.5	0.761	10	16.67	-9.93	196.21
	0.374	35	58.33	-2.3	106.57		0.761	15	25	-6.63	180.1
	0.374	40	66.67	-2.11	105.23		0.761	20	33.33	-4.97	166.24
	0.374	45	75	-1.96	103.9		0.761	25	41.67	-3.98	154.18
	0.374	50	83.33	-1.84	102.59		0.761	30	50	-3.32	143.6
-3.5	0.43	10	16.67	-6.27	118.56		0.761	35	58.33	-2.85	134.24
	0.43	15	25	-4.43	116.55		0.761	40	66.67	-2.49	125.9
	0.43	20	33.33	-3.51	114.57		0.761	45	75	-2.22	118.42
	0.43	25	41.67	-2.96	112.64		0.761	50	83.33	-2	111.68
	0.43	30	50	-2.59	110.73	1	0.757	15	25	-6.93	178.33
	0.43	35	58.33	-2.32	108.86		0.757	20	33.33	-5.18	164.89
	0.43	40	66.67	-2.13	107.03		0.757	25	41.67	-4.12	153.16
	0.43	45	75	-1.97	105.23		0.757	30	50	-3.42	142.83
	0.43	50	83.33	-1.85	103.45		0.757	35	58.33	-2.92	133.67
-3	0.491	10	16.67	-6.54	125.58		0.757	40	66.67	-2.54	125.5
	0.491	15	25	-4.59	122.66		0.757	45	75	-2.25	118.15
	0.491	20	33.33	-3.62	119.83		0.757	50	83.33	-2.02	111.52
	0.491	25	41.67	-3.03	117.09	1.5	0.743	15	25	-7.21	173.42
	0.491	30	50	-2.64	114.43		0.743	20	33.33	-5.36	161.11
	0.491	35	58.33	-2.36	111.85		0.743	25	41.67	-4.25	150.27
	0.491	40	66.67	-2.15	109.34		0.743	30	50	-3.52	140.65
	0.491	45	75	-1.99	106.91		0.743	35	58.33	-2.99	132.07
	0.491	50	83.33	-1.86	104.54		0.743	40	66.67	-2.59	124.35
-2.5	0.552	10	16.67	-6.88	134.8		0.743	45	75	-2.28	117.38
	0.552	15	25	-4.8	130.58		0.743	50	83.33	-2.04	111.05
	0.552	20	33.33	-3.75	126.55	2	0.722	15	25	-7.47	166.63
	0.552	25	41.67	-3.13	122.7		0.722	20	33.33	-5.53	155.82
	0.552	30	50	-2.71	119.03		0.722	25	41.67	-4.37	146.19
	0.552	35	58.33	-2.41	115.51		0.722	30	50	-3.6	137.55
	0.552	40	66.67	-2.19	112.15		0.722	35	58.33	-3.05	129.76
	0.552	45	75	-2.01	108.92		0.722	40	66.67	-2.63	122.69
	0.552	50	83.33	-1.88	105.83		0.722	45	75	-2.31	116.26
-2	0.611	10	16.67	-7.3	146.27		0.722	50	83.33	-2.05	110.38
	0.611	15	25	-5.05	140.23	2.5	0.696	15	25	-7.69	159.12
	0.611	20	33.33	-3.92	134.6		0.696	20	33.33	-5.68	149.9
	0.611	25	41.67	-3.24	129.32		0.696	25	41.67	-4.48	141.56
	0.611	30	50	-2.79	124.36		0.696	30	50	-3.67	133.99
	0.611	35	58.33	-2.47	119.7		0.696	35	58.33	-3.1	127.08
	0.611	40	66.67	-2.23	115.3		0.696	40	66.67	-2.67	120.75
	0.611	45	75	-2.04	111.15		0.696	45	75	-2.34	114.94
	0.611	50	83.33	-1.89	107.23		0.696	50	83.33	-2.07	109.57
-1.5	0.664	10	16.67	-7.78	159.45	3	0.666	15	25	-7.89	151.7
	0.664	15	25	-5.33	151.1		0.666	20	33.33	-5.81	143.95
	0.664	20	33.33	-4.11	143.47		0.666	25	41.67	-4.57	136.85
	0.664	25	41.67	-3.38	136.46		0.666	30	50	-3.74	130.32
	0.664	30	50	-2.89	130.02		0.666	35	58.33	-3.15	124.29
	0.664	35	58.33	-2.54	124.06		0.666	40	66.67	-2.7	118.71
	0.664	40	66.67	-2.28	118.54		0.666	45	75	-2.36	113.53
	0.664	45	75	-2.07	113.41		0.666	50	83.33	-2.08	108.71
	0.664	50	83.33	-1.91	108.64	3.5	0.635	15	25	-8.06	144.85
-1	0.706	10	16.67	-8.3	173.05		0.635	20	33.33	-5.93	138.39
	0.706	15	25	-5.65	162.05		0.635	25	41.67	-4.65	132.39
	0.706	20	33.33	-4.32	152.21		0.635	30	50	-3.8	126.8
	0.706	25	41.67	-3.52	143.38		0.635	35	58.33	-3.19	121.59

	0.706	30	50	-2.99	135.39		0.635	40	66.67	-2.73	116.71
	0.706	35	58.33	-2.61	128.14		0.635	45	75	-2.38	112.14
	0.706	40	66.67	-2.33	121.52		0.635	50	83.33	-2.09	107.85
	0.706	45	75	-2.11	115.46	4	0.603	15	25	-8.2	138.76
	0.706	50	83.33	-1.93	109.89		0.603	20	33.33	-6.02	133.38
-0.5	0.737	10	16.67	-8.85	185.05		0.603	25	41.67	-4.72	128.32
	0.737	15	25	-5.98	171.49		0.603	30	50	-3.84	123.56
	0.737	20	33.33	-4.54	159.61		0.603	35	58.33	-3.22	119.08
	0.737	25	41.67	-3.68	149.12		0.603	40	66.67	-2.76	114.84
	0.737	30	50	-3.1	139.78		0.603	45	75	-2.39	110.83
			Continued…				0.603	50	83.33	-2.1	107.03

6. Discussion of the Results

For a T value of -4, the value of the correlation coefficient ρ between the study variable Y (=FEV) and the auxiliary variable X(=AGE) is 0.3741, the values of the percent relative bias (RB) lie between -6.01% and -1.84% and that of the percent relative efficiency (RE) lie between 113.44% and 102.39% as the response rate increases from 16.67% to 83.33%. For $T = -3.5$, the value of the correlation coefficient ρ increase to 0.4301, the RB value varies between -6.27% and -1.85% and the value of RE varies between 118.56% and 103.45% as the response rate increases from 16.67% to 83.33%. For $T = -3.0$, the value of the correlation coefficient ρ increase to 0.4906, the RB value varies between -6.54% and -1.86% and the value of RE varies between 125.58% and 104.54% as the response rate increases from 16.67% to 83.33%. For $T = -2.5$, the value of the correlation coefficient ρ increase to 0.5523, the RB value varies between -6.88% and -1.88% and the value of RE varies between 134.80% and 105.83% as the response rate increases from 16.67% to 83.33%. For $T = -2.0$, the value of the correlation coefficient ρ increase to 0.6114, the RB value varies between -7.30% and -1.89% and the value of RE varies between 146.27% and 107.23% as the response rate increases from 16.67% to 83.33%. For $T = -1.5$, the value of the correlation coefficient ρ increase to 0.6637, the RB value varies between -7.78% and -1.91% and the value of RE varies between 159.45% and 108.64% as the response rate increases from 16.67% to 83.33%. For $T = -1.0$, the value of the correlation coefficient ρ increase to 0.7063, the RB value varies between -8.30% and -1.93% and the value of RE varies between 173.05% and 109.89% as the response rate increases from 16.67% to 83.33%. For $T = -0.5$, the value of the correlation coefficient ρ increase to 0.7369, the RB value varies between -8.85% and -1.95% and the value of RE varies between 185.05% and 110.87% as the response rate increases from 16.67% to 83.33%. For $T = 0.5$, the value of the correlation coefficient ρ increase to 0.7612, the RB value varies between -9.93% and -2.00% and the value of RE varies between 196.21% and 111.68% as the response rate increases from 16.67% to 83.33%. For $T = 1.0$, the value of the correlation coefficient ρ decreases to 0.7565, the RB value varies between -6.93% and -2.02% and the value of RE varies between 178.33% and 111.52% as the response rate increases from 25.00% to 83.33%. Note that in this situation if the response rate (RR) is less than 25% then the percent relative bias (RB) in the proposed imputing method remains higher than 10%, so those results are not reported in the table. For $T = 1.5$, the value of the correlation coefficient ρ decreases to 0.7427, the RB value varies between -7.21% and -2.04% and the value of RE varies between 173.42% and 111.05% as the response rate increases from 25.00% to 83.33%. For $T = 2.0$, the value of the correlation coefficient ρ decreases to 0.7218, the RB value varies between -7.47% and -2.05% and the value of RE varies between 166.63% and 110.38% as the response rate increases from 25.00% to 83.33%. For $T = 2.5$, the value of the correlation coefficient ρ decreases to 0.6957, the RB value varies between -7.69% and -2.07% and the value of RE varies between 159.12% and 109.57% as the response rate increases from 25.00% to 83.33%. For $T = 3.0$, the value of the correlation coefficient ρ decreases to 0.6663, the RB value varies between -7.89% and -2.36% and the value of RE varies between 151.70% and 108.71% as the response rate increases from 25.00% to 83.33%. For $T = 3.5$, the value of the correlation coefficient ρ decreases to 0.6351, the RB value varies between -8.06% and -2.09% and the value of RE varies between 144.85% and 107.85% as the response rate increases from 25.00% to 83.33%. For $T = 4.0$, the value of the correlation coefficient ρ decreases to 0.6033, the RB value varies between -8.20% and -2.10% and the value of RE varies between 138.76% and 107.03% as the response rate increases from 25.00% to 83.33%.

In order to have another look at the values of the percent relative bias (RB) and percent relative efficiency (RE) as a function of response rate (RR), we developed the scatter plots shown in Figure 6.1 For each value of the response rate (RR) there are several dots showing the percent relative bias and percent relative efficiency values corresponding to the values of the correlation coefficient ρ between 0.3741 and 0.7612 obtained through the different values of the Box-Cox transformation T.

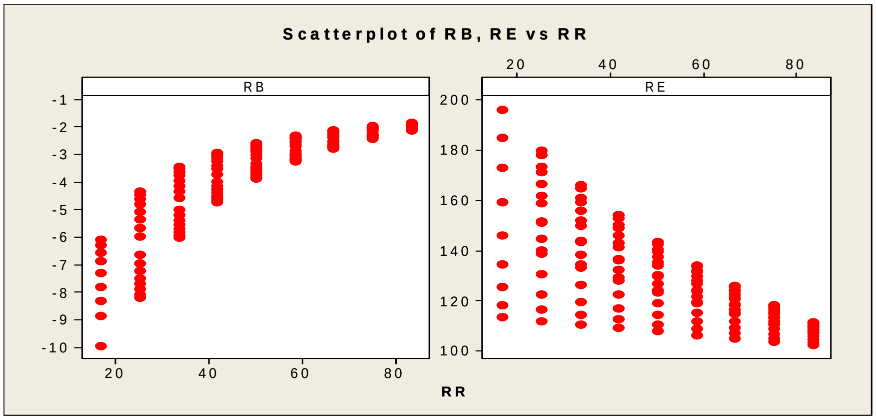

Fig. 6.1. Percent RB and Percent RE versus response rate (RR)

From Fig. 6.1, it is clear that if the response rate (RR) is high, then the absolute value of the percent relative bias (RB) remains close to zero, while at the same time there is adverse effect of having a low value of the percent relative efficiency (RE). If the response rate is low then the absolute value of the percent relative bias remains less than 4%, and the relative efficiency can vary up to 140%, depending on the value of the correlation coefficient. Thus if the response rare (RR) is moderate and value of the correlation coefficient between the study and auxiliary variable is also moderate, the use of proposed imputing method based on sensible constraint is useful. We devote Figure 6.2 to visualizing the RR and RE values as functions of the value of correlation coefficient ρ (=RHO(y,x)) between the study variable and auxiliary variables.

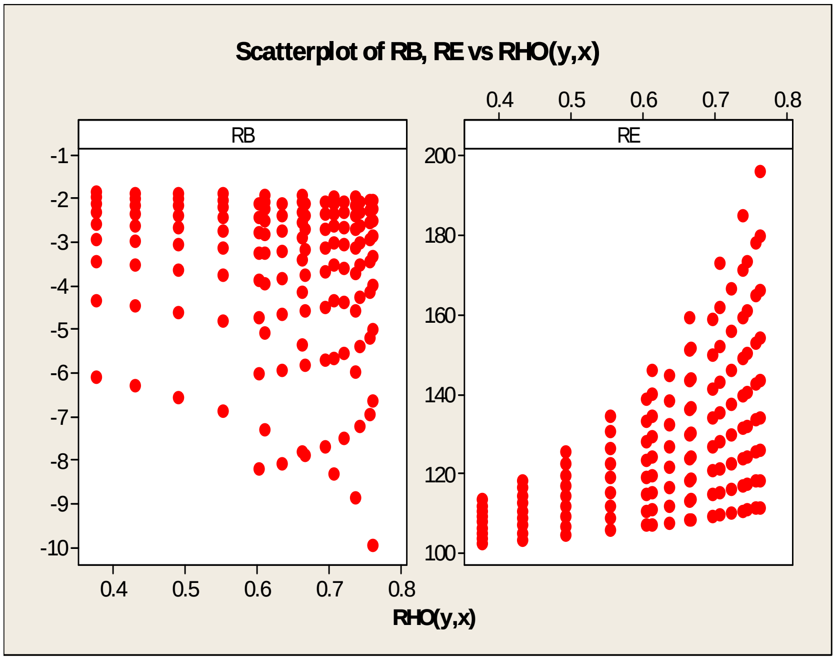

Fig. 6.2. RB and RE values as a function of correlation coefficient with varying RR.

From Fig. 6.2 it is obvious that higher value of the correlation coefficient ρ yield higher values of the percent relative efficiency (RE) but at the same time, may produce more biased estimates. The finding from Fig. 6.2 are not as clear as were those from Fig. 6.1, because of noisy nature of RB when the value of ρ becomes more than 0.6. Fig. 6.3 provides closer look at the behaviour of RB versus the correlation coefficient ρ, with the panel variable being RR.

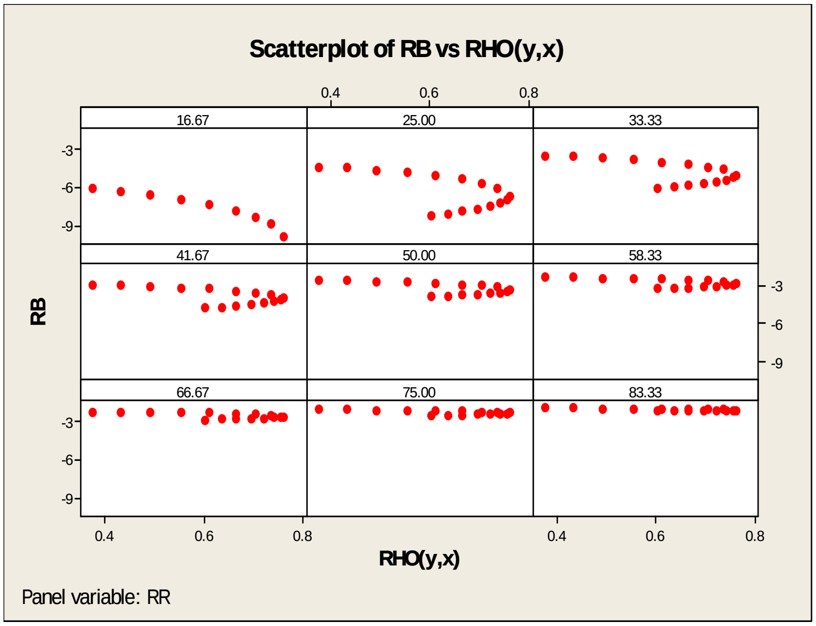

Fig. 6.3. RB as a function of correlation coefficient for different levels of RR.

Figure 6.4 provides a close look at the behaviour of RE versus the correlation coefficient ρ, with the panel variable being RR.

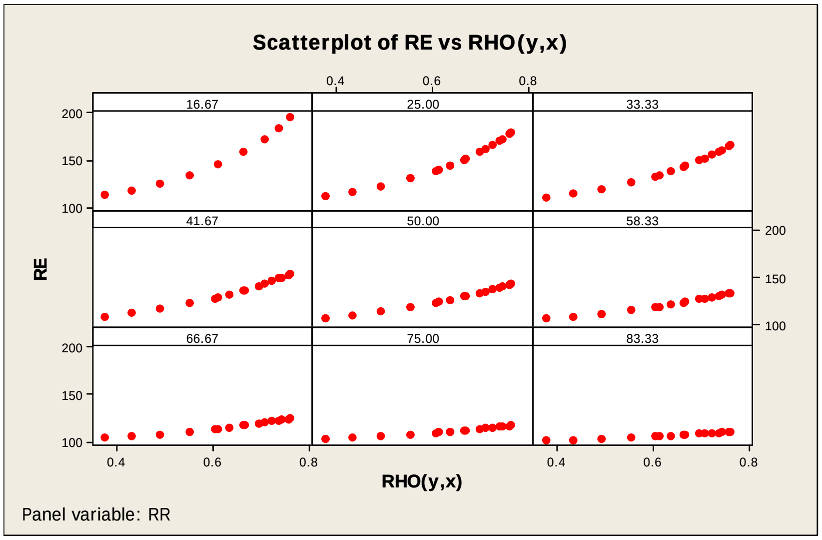

Fig. 6.4. RE as a function of correlation coefficient for different levels of RR.

7. Conclusion

As a result of this paper, we conclude that the proposed imputing method performs better than the mean method of imputation under the assumption of satisfying a set of sensible constraints imposed on the auxiliary variables and the study variable for the responding and non-responding units in a sample. The proposed imputing method is computationally oriented and allows for extension to the use of multi auxiliary variables. This is a first attempt in the literature of imputing methodology, and seems to have broad possibilities for extending to various sampling schemes and different situations. It will be worth mentioning that the proposed 'sensible constraints' are not unique, and it will always be possible to come up with an improved set of "sensible constraint" based on experience. We look forward to developing more such sensible constraints in our future research. In these proceedings paper, we have provided only preliminary results of the main article which will appear somewhere else in future.

8. Appendix

```
! FORTRAN CODES CHOUKRI10.F95
        USE NUMERICAL_LIBRARIES
        IMPLICIT NONE
        INTEGER NP,I,NS,NR,ID(1000)
        REAL Y1(1000), X1(1000),T, SUMY, SUMX, SMU12
        REAL Y(1000),X(1000),V1(1000),V2(1000),V3(1000)
        REAL ANR, ANS,ANP,BIASP,YM,XM,SUMX2,SUMY2,SUMXY
        REAL RHOXY,AMU12,COVXY,VARX,VARY,RE,RR
        CHARACTER*20 OUT_FILE
        CHARACTER*20 IN_FILE
        WRITE(*,'(A)') 'NAME OF THE INPUT FILE'
        READ(*,'(A20)') IN_FILE
        OPEN(41, FILE =IN_FILE, STATUS='OLD')
        WRITE(*,'(A)') 'NAME OF THE OUTPUT FILE'
        READ(*,'(A20)') OUT_FILE
        OPEN(42, FILE=OUT_FILE, STATUS='UNKNOWN')
        READ(41,*)NP
        ANP = NP
        DO 10 I =1, NP
```

```
10        READ(41,*)ID(I),Y1(I),X1(I),V1(I),V2(I),V3(I)
          DO 16 T = -4, 4, 0.5
         DO 11 I= 1, NP
         Y(I) = Y1(I)
11        X(I) = (X1(I)**T-1)/T
          NS = 60
          ANS = NS
          DO 21 NR = 5, 50, 5
          ANR = NR
          SUMY = 0.0
          SUMX = 0.0
          DO 12 I=1, NP
          SUMY = SUMY + Y(I)
12        SUMX = SUMX + X(I)
          YM = SUMY/ANP
          XM = SUMX/ANP
          SMU12 = 0.0
          SUMXY = 0.0
          SUMY2 = 0.0
          SUMX2 = 0.0
          DO 14 I = 1, NP
          SMU12 = SMU12 + (Y(I)-YM)*(X(I)-XM)**2
          SUMXY = SUMXY + (Y(I)-YM)*(X(I)-XM)
          SUMX2 = SUMX2 + (X(I)-XM)**2
14        SUMY2 = SUMY2 + (Y(I)-YM)**2
          AMU12 = SMU12/(ANP-1)
          COVXY = SUMXY/(ANP-1)
          VARX = SUMX2/(ANP-1)
          VARY = SUMY2/(ANP-1)
          RHOXY = COVXY/SQRT(VARX*VARY)
          BIASP = -( (1/ANR-1/ANS)*AMU12/(YM*VARX)+1/ANR )*100
          RE=(1/ANR-1/ANP)*100/((1/ANS-1/ANP)+(1/ANR-1/ANS)*(1-RHOXY**2))
          RR = ANR*100/ANS
          IF(ABS(BIASP).LT.10) THEN
          WRITE(42,101)NP,NS,NR,RR,T,RHOXY,BIASP,RE
101      FORMAT (2X,3(I5,1X),2X,F7.3,2X,F7.3,2X,F9.4,2X,F9.2,2X,F9.2)
         ENDIF
21       CONTINUE
16       CONTINUE
        STOP
        END
```

Acknowledgements

The authors would like to thank the College of Arts & Sciences, Texas A&M University-Kingsville for the summer travel support to present this paper at the JSM 2015 American Statistical Association Conference.

References

Cochran, W.G. (1977). *Sampling Techniques*. John Wiley & Sons, New York.

Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.*, 41, 517-529.

Heitjan, D.F. and Basu S. (1996). Distinguishing 'Missing At Random' and 'Missing Completely At Random'. *The American Statistician*, 50, 207-213.

Rosner, B. (2006). *Fundamentals of biostatistics*. Belmont, CA: Thomson-Brooks/Cole.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63(3), 581 -592.