World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:12, 2007

Neural Network Imputation In Complex Survey Design

Safaa R. Amer

***Abstract*—**Missing data yields many analysis challenges. In case of complex survey design, in addition to dealing with missing data, researchers need to account for the sampling design to achieve useful inferences. Methods for incorporating sampling weights in neural network imputation were investigated to account for complex survey designs. An estimate of variance to account for the imputation uncertainty as well as the sampling design using neural networks will be provided. A simulation study was conducted to compare estimation results based on complete case analysis, multiple imputation using a Markov Chain Monte Carlo, and neural network imputation. Furthermore, a public-use dataset was used as an example to illustrate neural networks imputation under a complex survey design.

***Keywords*—**Complex survey, estimate, imputation, neural networks, variance.

I. Introduction

TRADITIONAL methods presented in the statistical literature, outside of survey sampling, have been based on a simple random sample [7]. This assumption is not appropriate when the data were generated using a complex sampling design [24]. As an alternative to standard formulas and techniques used in case of simple random sample, design-based procedures were developed to handle probability sampling. Design-based procedures, which date back to the 1950's, provide accurate inference for complex surveys and account for complex sampling designs [9], [16], [20], [21].

In a complex survey design, characteristics of the population may affect the sample and are used as design variables. Sample design involves the concepts of stratification, clustering, etc. These concepts usually reflect a complex population structure and should be accounted for during the analysis. In design-based inference, the main source of random variation is induced by the sampling mechanism [8]. Furthermore, in complex survey design, the variance is the average squared deviation of the estimate from its expected value, averaged over all possible samples which could be obtained using a given design. Design-based approaches make use of sampling weights as part of the estimation and inference procedures.

In survey sampling, two types of weights are of interest. These weights are sampling weights and nonresponse weights. If a unit is sampled with a specific selection probability then the sampling weight is the inverse of the probability of sample selection. For example, stratified sampling occurs when the population is divided into distinct subpopulations and a separate sample is selected within each subpopulation. From the sample in each subpopulation, a separate mean is computed. The means are weighted to calculate a combined estimate for the entire population. Weighting is used as a nonresponse adjustment for unit nonresponse as well. Nonresponse weight is the reciprocal of the probability that a unit is selected in the sample and responds. A combined weight results from multiplying the response weight times the sampling weight.

Several estimators and their corresponding variances have been introduced in the literature for different sampling designs (e.g. Horvitz-Thompson estimator). Point estimators are usually calculated using survey weights, which may involve auxiliary population information. However, sampling variance estimation is more complicated than parameter estimation [24]. Alternatives to conventional variances, in case of complex survey designs, were proposed to facilitate the variance calculation. Methods like the random group method [25] are based on the replication of the survey design. These methods are simple to apply to nonparametric problems, but lead to imprecise estimates of the variances [24]. Woodruff [38] illustrated the Taylor series linearization method to approximate the variance in complex surveys. In case of Taylor series linearization, in spite of a complex calculation, the linearization theory may be applied if the partial derivatives are known.

II. Neural Network and Complex Survey Design

One major advantage of artificial neural networks (ANN) is their flexibility in modeling many types of nonlinear relationships. Artificial neural networks can be structured to account for complex survey designs and for unit nonresponse as well. In general, sampling weights have been used to adjust for the complex sampling design using unequal sampling [24]. We suggest two different methods to include sampling weights into ANN. The first method is to include the sampling weights in the ANN similar to weighted least squares (WLS). The second method is based on accounting for the sampling design structure in constructing the corresponding ANN.

A. Method based on Weighted Least Squares

Weighted least squares are used in regression in several situations to account for variance. When the deviations from the responses are available; the weights are the reciprocal of the response variance to give observations with smaller error more weight in the estimation procedure. In addition, weights are used when the responses are averaged from samples with

Safaa R. Amer is with NORC at the University of Chicago, Chicago, IL 60063 USA (phone: 312-451-3673; e-mail: amer-safaa@norc.org).

different sizes. Additionally, when the variance is proportional to a covariate, the weight is the inverse of the covariate. In survey sampling, statisticians debated about the relevance of the sampling weights for inference in regression [5]. Part of this debate is based on the idea that weights are needed for estimating population parameters in complex survey sampling and by analogy should be used in regression. Amer, Lesser, and Burton [2] illustrated the similarities between ANN and linear regression. Based on these similarities, we propose including the sampling weights in the network in the same manner it would be incorporated in case of regression. If sampling weights are used in the WLS estimation, point estimates will be similar to design based estimates. However, the standard errors also need to be developed based on the survey design. In case of complex survey design, using the approach of weighted least squares proves to be useful specially when the analyst is not involved in the design stage but is presented with the final weights.

B. Method based on ANN Structure

When the analyst has access to the design variables, an alternative method to WLS is to construct the neural network using the sampling design features. For example, in a stratified sampling design, a separate network could be built and trained using data from a specific stratum in the imputation procedure. We suggest using a separate network for each stratum. These networks are then connected with a binary activation function at the input layer. The binary activation function directs each observation to the corresponding stratum, taking the value 0 when the observation is not in the stratum and 1 when the observation belongs to that stratum. This leads to a different network parameter estimates for each stratum. The disadvantage of using a separate network for each stratum (without connecting the networks or assigning a probability for each stratum) is that it does not provide estimates for the entire sample. Therefore, a suggested solution is to train separate networks for each output and then to combine all strata to account for the full sample. Using a mixture of network models is common in ANN and can be considered as a possible technique to account for the sampling design. Mixture of expert networks are mixture models used to solve complex problems [4].

The solution of these complex problems is achieved by dividing the problem into smaller sub-problems, where each network is considered an expert for a subgroup of the observations. These expert networks are connected together through a gating network, which divides the input space into different subgroups of conditional densities as shown in Fig. 1 [19]. The use of a mixture of expert networks allows the introduction of sampling probabilities and construction of a separate model for each stratum in a stratified sample design. In mixtures of expert network models, we have $y_j = f_j\left(x_j;\theta\right)$ where strata j=1,...,M such that M is the number of strata.

Fig. 1 corresponds to a model representing a stratified sampling design with three strata and a gating network. Each of the networks Net 1, Net 2 and Net 3 serves as a unique network for imputation in each stratum separately. The covariates are represented by the matrix *X*. The matrix is fed into each of the networks. The gating network serves as a portal to synchronize between the different strata. The gating network acts as dummy variables that differentiates between the different strata and assigns a sampling weight to each network. The output node is the sum of the results from the different strata. The neural network model corresponding to such a design can be formulated as $\hat{y} = \sum_{j=1}^{M} \frac{1}{\rho_j(x)} f_j\left(y \mid x\right)$ where $\rho_j(x) = \frac{1}{w_j(x)}$ is the probability at each stratum while $f_j\left(y \mid x\right)$ is the function representative of network j. The sampling weights are $w_j(x) = \frac{N_j}{n_j}$ such that n_j is the sample size from stratum j and N_j is the population size from stratum j. Using a mixture of experts is a convenient way to adjust for complex designs. After the design is taken into account, the network may then be used for imputation.

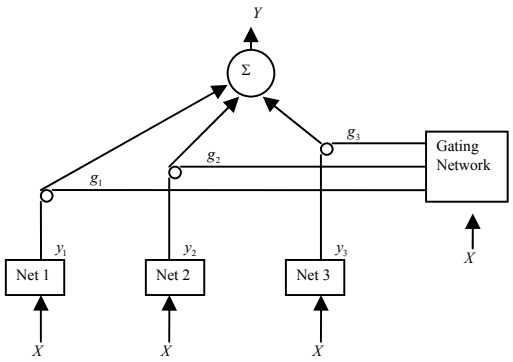

Fig. 1 Mixture Expert Networks for a stratified sampling design

III. Bias/Variance Trade-off in ANN and Inference

To calculate the mean squared error (MSE) in ANN, the expected error rate can be broken down into three main components: bias, variance and noise [17]. Partitioning of the MSE into bias and variance is helpful to understand the variability in ANN.

Let $y = f(x) + \xi$ where $\xi \sim \mathrm{N}\left(0,\sigma^2\right)$. The performance of the network in prediction is based on the MSE where $MSE = Bias^2 + Var$ such that $Var = Var\left[\hat{y}\right] + \sigma^2$. Similar to regression, neural networks attempt to minimize the MSE. In ANN, bias arises when the true function cannot be represented correctly, i.e., under-fitted. However, variance in ANN comes from over-fitting, where the network adapts to the specific data used and cannot be generalized to new observations.

There are several ways to calculate variance and to construct confidence intervals for neural networks. Rivals and Personnaz [30] suggest using Taylor series approximation as a variance estimation procedure. For a nonlinear function $f(\theta)$; the estimate is $f\left(\hat{\theta}\right) \overset{approximately}{\sim} N\left(f(\theta), \sigma^2 u'(X'X)^{-1} u\right)$

where $u' = \left[\frac{\partial f(\theta)}{\partial \theta_1}, \frac{\partial f(\theta)}{\partial \theta_2}, \ldots, \frac{\partial f(\theta)}{\partial \theta_p}\right]$ is the vector of derivatives. At $\hat{\theta}$, the vector of derivatives u would become: $z_o' = \left[\frac{\partial f(X_o,\theta)}{\partial \theta_1}, \frac{\partial f(X_o,\theta)}{\partial \theta_2}, \ldots, \frac{\partial f(X_o,\theta)}{\partial \theta_p}\right]_{\theta=\hat{\theta}}$. As a result, the standard error of a predicted response $f\left(X_o,\hat{\theta}\right)$ is given by $s_{f(X_o,\hat{\theta})} = \sqrt{s z_o' \left(X'X\right)^{-1} z_o}$. Therefore, approximate 100(1-α) % confidence interval on the mean response at an arbitrary location X_o is $f\left(X_o,\hat{\theta}\right) \pm t_{\alpha/2,m-p-1} s\sqrt{z_o' \left(X'X\right)^{-1} z_o}$ and an approximate 100(1-α) % prediction limits on a new observation at x_o is $f\left(X_o,\hat{\theta}\right) \pm t_{\alpha/2,m-p-1} s\sqrt{1+z_o' \left(X'X\right)^{-1} z_o}$.

IV. Imputation

When survey nonresponse is encountered, either nonresponse weighting or imputation may be used to handle the missing data. Imputation is the procedure of filling in the missing values. Imputation can be performed as single imputation, or repeated several times resulting in multiple imputations [11]. One drawback to single imputation is the unaccounted uncertainty attributed to the imputation from the filled-in data. Multiple imputation (MI), as proposed by Rubin [31], replaces the missing value by a vector of imputed values to obtain a number of complete datasets. Regular analysis run on the multiply imputed datasets yields estimates that are subsequently combined to get the final results. The combined estimate from a multiply-imputed dataset is the average of the estimates resulting from the analysis of each completed dataset separately. However, the variance of this estimate is divided into two components, the average within imputation variance and the between imputation component. The total variance is then a weighted sum of these two variance components [23]. Inferences resulting from combining the imputations reflect the uncertainty due to nonresponse. In real data analyses, MI may not result in good performance if it is not applied properly or if the mechanisms generating either the data or the missing values depart substantially from the underlying statistical assumptions [10].

Many single imputation techniques can be repeated several times resulting in multiple imputation [1], [18]. Reference [34] offers an extended review of techniques used for MI. In this paper, the MCMC data augmentation technique will be used as an example. The MCMC procedures are a collection of methods for simulating random draws from the joint distribution of $\left(Y_{mis},\theta|Y_{obs}\right)$ where Y_{mis} are the missing values, Y_{obs} are the observed values, and θ is the distribution parameter. This conditional distribution is assumed to be a multivariate normal distribution [14], [30]. The simulated random draws result in a sequence of values that form a Markov chain [12], [15], [36]. A Markov chain is a sequence of random variables where the distribution of each element depends only on the value of the previous one and the iterative procedure consists of two steps. The first step is an imputation step (I-step), which is a draw Y_{mis} from the conditional predictive distribution $P\left(Y_{mis}|Y_{obs},\theta\right)$ given a value for θ. The second step is a posterior step (P-step), given Y_{mis}, draw θ from its complete data posterior $P\left(\theta|Y_{obs},Y_{mis}\right)$. The goal of MCMC procedure is to sample values from a convergent Markov chain in which the limiting distribution is the joint posterior of the quantities of interest [35]. In practice, the major challenge in using MCMC is the difficulty, for the user, to assess convergence [13]. Overall, multiple imputation is difficult to implement in large datasets, due to the amount of computer memory needed to store the different, multiply-imputed datasets and the time required to run the analysis.

Increased computer power and decreased cost have encouraged more research into the automated edit and imputation techniques. Advances in computer software and increased memory have made the use of both MI and ANN more practical. The type of ANN used in this paper for imputation in each stratum are called feed-forward, where input terminals receive values of explanatory variables *X*, whereas the output provides the imputed variable *Y*. Multilayer feed-forward networks consist of one or more hidden layers. The role of the hidden layer of neurons is to intervene between the external input and the network output. Inputs and outputs are connected through neurons that transform the sum of all received input values to an output value, according to connection weights and an activation function. The connection weights represent the strength of the connection between the neurons. The network weights (parameters) are randomly initialized and are then changed in an iterative process to reflect the relationships between the inputs and outputs. Many linear or nonlinear functions are suitable candidates for an activation function. ANN do not require a model, which is advantageous in large dataset imputation.

Most traditional imputation techniques do not account for sampling design during the imputation procedure [6]. For example, multiple imputation (MI) is considered imperfect because it does not account for survey design. One solution is to run a separate imputation within each sampling subgroup and run a weighted analysis for each imputed dataset. Another solution is to base MI on models that specifically include design characteristics. Binder and Sun [3] suggest that finding accurate methods for imputation may be very difficult under complex survey design and requires a correct model for imputation. Reference [26] states that variance estimation of imputed values under complex survey design has not been solved and needs further research. Remedies such as imputing the nonrespondents with the sample weighted mean have been suggested [37]. In this case, the weighted mean from complete cases is calculated and used for imputation.

In this paper, the focus is on computing weighted estimates for large public use data files and use imputation methods that account for complex surveys. With large sample sizes, we assume the central limit theorem applies thus the sampling

distribution of the parameter estimator is approximately normal.

V. Imputation and Inference Under ANN with a Complex Survey Design

In case of nonresponse, bias needs to be quantified and both estimation and inference procedures are harder to handle [33]. With an increasing rate of nonresponse, when the mean of the nonrespondents differs from respondents, bias increases. Therefore, the mean square error (MSE) is customarily used when comparing different estimates. According to Lee, Rancourt and Särndal [22], there are two reasons why MSE should be considered instead of variance. First, the assumption of obtaining an unbiased estimate after imputation is not usually guaranteed. Secondly, the MSE is a measure of accuracy. In general, Total error = Sampling error + Nonsampling error. Sampling error accounts for most of the variable errors in a survey. Nonsampling error is mostly bias caused by measurement, editing, and/or imputation errors [20].

The estimate of the population parameters from the imputed dataset includes several sources of bias. The first source of bias is from the estimate provided using traditional statistical techniques in complete case analysis. The second source of bias is due to imputation using ANN. The total MSE is defined as $MSE = \left(\sum_g B_g \right)^2 + \sum_v \frac{S_v^2}{m_v}$. In this case, the bias is the sum of the bias expected from a sample survey and the bias from the neural network estimate based on imputed values. However, analytically, the bias cannot be estimated. Therefore, most analysts estimate the variance only. The variance can be divided into several parts where Total variance is $S_{obs}^2 + S_{imp}^2 + 2S_{\text{joint}}$ [32] such that S_{obs}^2 is the Observed sample variance under complex survey design, S_{imp}^2 is the imputation variance, and $S_{\text{joint}} \overset{Asymptotically}{\approx} 0$ [28]

It is necessary to identify the observed and imputed values using ANN in the data file before the analysis. Assume there is a stratified random sample of size *n* with *X* observed for all sampled units. Let $\{x_{thi} : i = 1, \ldots, t\}$ and $\{x_{mhj} : j = t+1, \ldots, n\}$ denote the observed *X* values which correspond to the *t* observed *Y* values and *m* missing *Y* values in strata *h* for $h = 1, \ldots, H$, respectively. The weighted sample mean for the completed data $\bar{y}_{cw}$ can be calculated to estimate the population mean $\bar{Y}$. The standard error of $\bar{y}_{cw}$ can also be calculated to estimate the variability associated with this estimate. The weighted mean can be expressed as $\bar{y}_w = \sum_h W_h \bar{y}_h$ where $W_h = \frac{N_h}{N}$ ϶ $\sum_h W_h = 1$. At each stratum we have $\bar{y}_h = \frac{\sum_{k=1}^{n_h} y_{hk}}{n_h} = \frac{1}{n_h}\left(\sum_{i=1}^{t_h} y_{thi} + \sum_{j=t_h+1}^{n_h} y_{mhj} \right)$ where the observed values are presented by y_{thi} and the imputed values presented by y_{mhj}.

Let $\bar{y}_t = \frac{\sum_h t_h \bar{y}_{th}}{t}$ be mean of the observed data and $\bar{y}_m = \frac{\sum_h m_h \bar{y}_{mh}}{m}$ be mean of the imputed data, then $\bar{y}_{cw} = (1-\pi)\bar{y}_t + \pi \bar{y}_m$ where π is the percent of missing data. The sample variance is expressed as:

$$Var(\bar{y}_w) = \sum_h W_h^2 Var(\bar{y}_h)$$

$$Var(\bar{y}_h) = \frac{1}{n-1}\left[\sum_i (y_{thi} - \bar{y}_h)^2 + \sum_j (y_{mhj} - \bar{y}_h)^2 \right]$$

$$Var(\bar{y}_{cw}) = (1-\pi)^2 Var(\bar{y}_t) + \pi^2 Var(\bar{y}_m)$$

$$Var(\bar{y}_t) = \sum_h \left(\frac{t_h}{t} \right)^2 Var(\bar{y}_{th})$$

$$Var(\bar{y}_m) = \sum_h \left(\frac{m_h}{m} \right)^2 Var(\bar{y}_{mh})$$

A total weighted variance is derived by combining the variances from the complete cases and from the imputation procedure with ANN. The imputed values are given their relative importance depending on the percentage of missing data.

VI. Results

This section contains simulation results as well as results using data from the NHIS under a complex survey design.

A. Simulation

A simulation study was performed to compare the results of imputation using nonlinear ANN to multiple imputation (MI) using Markov chain Monte Carlo (MCMC) method under a complex survey design. Data for this simulation were generated using a stratified simple random design with two strata having equal allocation. For this simulation study, each of the two strata (Z = 1, 2) had three variables X_1, X_2 and Y, and 1000 observations. Using the Matlab software, the X's were generated separately with a normal distribution in each stratum. The Y was generated as a function of the X's with normal random error. The relationship between the X's and Y was simulated to be linear. The linear model was simulated using a linear combination of the X's and the error term. The parameters in this model used for data simulation (α, β_1, and β_2) were set arbitrarily and separately for each stratum. A random number was generated and used to select certain observations to have missing Y values. The number of missing observations represented 10 percent of the sample size in each stratum. In addition, the missing observations were used for evaluating the performance of the imputation techniques.

Two software packages were combined in the analysis of the data to make use of several imputation techniques. SAS was used for MI with MCMC and the Matlab Neural Network toolbox was used for ANN imputation. In case of MI with MCMC, imputation was performed separately in each stratum.

Weighted estimates were calculated from each stratum. The weighted estimates from the imputed data were combined using the formulas presented by Little and Rubin [23]. In ANN imputation, a separate network was used for imputation in each stratum. Online training was used to provide the network with one observation at each pass. The activation function at the hidden layer was chosen as a logistic function whereas the activation function at the output layer is set to be a linear function. Initial parameter values for each network were randomly assigned. Using a gating network, the results were combined based on the weights to yield the final estimates.

Table I shows the weighted results using multiple imputation with MCMC and nonlinear ANN. Results show that both the weighted mean and SE resulting from MI using MCMC and ANN are approximately equal. However, the complete case analysis provides a slightly higher weighted mean and SE.

TABLE I
COMPARISON BETWEEN ANN AND MI USING MCMC IN COMPLEX SURVEY DESIGN

	Sample size	Weighted Mean	SE
Complete cases	1800	39.9713	0.3575
ANN	2000	39.7680	0.3433
MCMC 1	2000	39.7537	0.3432
MCMC 2	2000	39.7814	0.3434
MCMC 3	2000	39.7763	0.3436
MCMC 4	2000	39.7673	0.3434
MCMC 5	2000	39.7680	0.3433
MI (MCMC) combined		39.7693	0.3436

B. Application

The National Health Interview Survey (NHIS), a health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control (CDC), is the principal source of information on the health of the civilian, non-institutionalized, household population of the United States. NCHS-CDC has been releasing microdata files for public use on an annual basis since 1957 [27]. The focus of this application is on the 2001 sample adult core survey, where one adult from each household is randomly sub-sampled to receive a questionnaire. This questionnaire collects basic information on health status, health care services and behavior of adults in the population. The U.S. Census Bureau collects the data for the NHIS by personal interviews. The sample for the last quarter (September-December) of 2001 survey consisted of 8673 adults for the sample adult component. The response rate for the sample adult component was 73.8%.

The NHIS data are obtained through a complex sample design involving stratification, clustering, and multistage sampling designed to represent the civilian, non-institutionalized population of the United States. The respondent weights are further modified by adjusting them to Census control totals for sex, age, and race/ethnicity population using post-stratification. The probability of selection for each person and adjustments for nonresponse and post-stratification are reflected in the sample weights. These weights are necessary for the analysis to yield correct estimates and variance estimation. If the data are not weighted, and standard statistical methods are used, then the estimators are overly biased and the results misleading. Variance estimation is suggested to be calculated using the Taylor series linearization method. For more information about the sampling design, the reader may refer to NCHS 2002.

The NHIS contains demographic information in addition to information about whether the respondent had cardiovascular disease, emphysema, asthma, ulcers, cancer, diabetes, respiratory conditions, liver conditions, joint symptoms, pain. Information is also available on the mental health of respondent (sadness, nervousness, etc.), daily activities, social activities, smoking, and the ability to perform physical tasks. Information on body mass index (BMI = weight/height^2) was also provided. In addition, sampling weights were included. A total of 73 variables were maintained in the dataset used in the imputation.

The BMI was the variable of interest for the imputation procedure approximately 4.5% of the respondents had BMI missing in this dataset. Artificial neural network was used for imputation of the BMI missing values. A feed-forward network with 38 input nodes corresponding to the auxiliary variables in the dataset, a hidden layer with three nodes, and one node at the output layer corresponding to the output (imputed) variable was used for imputation. The number of nodes at the hidden layer was based on multiple trials to minimize the total network error. Results of the weighted BMI mean and standard error after ANN imputation were compared to the results from the weighted analysis using the complete cases only. Table II shows a comparison between the results from running a weighted analysis using the complex survey weights on each of the following: complete cases, imputed cases using ANN, and the full dataset after imputation. Imputation was performed using ANN and using a weighted mean. The weighted mean imputation was chosen for its simplicity. Multiple imputation using MCMC was not applied to this example due to the difficulties of its application and due to the need for a model based approach which is beyond the scope of this paper. The comparison results in Table II show that ANN yield an estimate with higher precision than the complete case analysis where the difference detected in the variance is estimated to be approximately eight percent. This difference in the variance is not trivial and requires further investigation in future research.

TABLE II
IMPUTATION RESULTS

	Sample size	Weighted Mean	SE
Complete cases	8282	26.92	0.071
ANN Imputed cases	391	27.18	0.078
Overall with ANN imputation	8673	26.93	0.065
Overall with weighted mean imputation	8673	26.92	0.068

VII. Conclusion

Design-based inference accounts for the survey design and provides reliable inferences in large samples without requiring any modeling assumptions. Variance, standard error, and tests of significance based on the assumption of independent selections are misleading and not valid for complex samples. The mean may be an acceptable estimate, but the standard error is underestimated. Measures of variability depend on the sample design and are subject to design effects. It is important to incorporate the complex survey design during the imputation procedure and in the inference after imputation.

Multiple imputation (MI) has the capability of providing a variance estimate. However, MI lacks the ability to account for the survey design in case of more complex survey designs such as NHIS. In the simulation study, the design was very simple which provided a design based analysis within each stratum. However, in the real-world data application, the design was more complex and in order to account for the design weights more research needs to be pursued. Artificial neural network represents an alternative imputation technique that requires fewer resources and offers a variance that accounts for the imputation as well as the survey design.

Acknowledgment

The author would like to thank Dr. Virginia Lesser from the department of Statistics and Dr Robert Burton from the department of Mathematics at Oregon State University for guidance and helpful suggestions of earlier drafts of the paper.

References

[1] Paul D. Allison (1999). "Multiple imputation for missing data: A cautionary tale". Available: http://www.ssc.upenn.edu/~allison/MultInt99.pdf

[2] S. Amer, V. Lesser, and R. Burton, "Neural network imputation, a new fashion or a good tool: Linear neural network imputation," *Proceedings of the Survey Research Section, American Statistical Association Meetings*, 2003.

[3] D.A. Binder, W. SUN, "Frequency valid multiple imputation for surveys with a complex design. Proceedings of the Section on Survey Research Methods", *American Statistical Association,,* pp. 281-286, 1996.

[4] C.M. Bishop, *Neural networks for pattern recognition*. Oxford: Clarendon Press, 1995.

[5] K.R.W. Brewer, and R.W. Mellor, "The effect of sample structure on analytical surveys," *Australian Journal of Statistics, 15*, pp. 145-152, 1973.

[6] E.M. Burns, "Multiple imputation in a complex sample survey," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 233-238, 1989.

[7] G. Casella, and R.L. Burger, *Statistical inference*. California: Duxbury press, 1990.

[8] R.L. Chambers, and C.J. Skinner (eds.) *Analysis of survey data*. Chester: Wiley, 2003.

[9] W.G. Cochran, *Sampling techniques*, (3rd Edition). New York: Wiley, 1977.

[10] L.M. Collins, J. L. Schafer, and C-M. Kam, "A comparison of inclusive and restrictive strategies in modern missing data procedures", *Psychological Methods,* 6 (4), pp. 330-351, 2001.

[11] I. P. Fellegi, and D. Holt. "A systematic approach to automatic edit and imputation," *Journal of the American Statistical Association*, 71, pp. 17-35, 1976.

[12] A.E. Gelman, J.B.Carlin, H.S. Stern, and D.B. Rubin. *Bayesian data analysis*, London: Chapman & Hall, 1995.

[13] A.E. Gelman and D.B. Rubin. "Inference from iterative smulation using multiple sequences," *Statistical Science*, 7, pp. 457-472, 1992.

[14] S. Geman, and D. Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 721-741, 1984.

[15] C.J. Geyer. "Practical Markov Chain Monte Carlo," *Statistical Science*, 7(4), 1992.

[16] M.H. Hansen, W.N. Hurwitz, and W.G. Madow. *Sampling survey methods and theory*, Vols. I and II. New York: Wiley, 1953.

[17] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: Data mining, inference, and prediction.* Springer, New York, 2001.

[18] N.J. Horton and S.R. Lipsitz. "Multiple imputation in practice: Comparisons of software packages for regression models with missing variables," *The American Statistician*, 5(3), 2001.

[19] R.A. Jacobs, M.I. Jordan, S.J. Nolman, and G.E. Hinton. "Adaptive mixtures of local experts," *Neural Computation*, 3, pp. 79-87(1991)..

[20] L. Kish. *Survey sampling*, New York: Wiley, 1965.

[21] Kish, L. "The Hundred years' wars of survey sampling," *Statistics in Transition*, 2, pp. 813-830, 1995.

[22] H. Lee, E. Rancourt, and C.E. Särndal. "Variance estimation from survey fata under single imputation," *Survey Nonresponse,* R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, (Eds). New York: John Wiley and Sons, 2002.

[23] Little, Roderick J.A. and Rubin, Donald B. *Statistical analysis with missing data*, New Jersey: John Wiley & Sons, 2002.

[24] S. L. Lohr. *Sampling: Design and analysis*, Duxbury Press, 1999.

[25] P.C. Mahalanobis. "Recent experiments in statistical sampling in the Indian Statistical Institute," *Journal of the Royal Statistical Societ,,* 109, pp. 325-370, 1946.

[26] D.A. Marker, D.R. Judkins, and M. Winglee. "Large-scale imputation for complex surveys." R.M. Groves, D.A.Dillman, J.L. Eltinge, and R.J.A Little, (Eds.) *Survey Nonresponse*, New York: John Wiley and Sons, 2002.

[27] National Center for Health Statistics. *Data file documentation, National Health Interview Survey, 2001* (machine readable file and documentation). National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland, 2002.

[28] E. Rancourt, C.-E. Särndal, and H. Lee. "Estimation of the variance in presence of nearest neighbor imputation," Proceedings of the Section on Survey Research Methods, *American Statistical Association*, pp. 888-893, 1994.

[29] I. Rivals and L. Personnaz. "Construction of confidence intervals for neural networks based on least squares estimation," *Neural Networks*, 13, 463-484 (2000)..

[30] D.B. Rubin. "Formalizing subjective notions about the effect of non-respondents in sample surveys," *Journal of the American Statistical Association*, 77, pp. 538-543, 1977.

[31] C.-E. Särndal, B. Swensson, and J. Wretman. *Model assisted survey sampling*, Springer-Verlag, 1991.

[32] C.-E. Särndal. "Methods for estimating the precision of survey estimates when imputation has been used," *Survey Methodology*, 18, pp. 241-265, 1992.

[33] J.L. Schafer. *Analysis of incomplete multivariate data*. London: Chapman and Hall, 1997.

[34] J. Schimert, J.L. Schafer, T.M. Hesterberg, C. Fraley, and D.B. Clarkson. *Analyzing data with missing values in S-Plus.* Seattle: Insightful Corp, 2000.

[35] A.F.M. Smith and G.O. Roberts. "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods," *Journal of the Royal Statistical Society*, Series B, 5(1), 1992.

[36] Vartivarian, S.L. and Little, R.J. (2003). "Weighting adjustments for unit nonresponse with multiple outcome variables," *The University of Michigan Department of Biostatistics* (Working Paper Series: Working Paper 21.) Available: http://www.bepress.com/umichbiostat/paper21

[37] R.S. Woodruff. "A simple method for approximating the variance of a complicated estimate," *Journal of the American Statistical Association*, 66, pp. 411-414, 1971.