Fall 2020

Multiple Imputation Using Influential Exponential Tilting in Case of Non-Ignorable Missing Data

Kavita Gohil

MULTIPLE IMPUTATION USING INFLUENTIAL EXPONENTIAL TILTING IN CASE OF NON-IGNORABLE MISSING DATA

by
KAVITA GOHIL
(Under Direction of Hani M. Samawi)

ABSTRACT

Modern research strategies rely predominantly on three steps, data collection, data analysis, and inference. In research, if the data is not collected as designed, researchers may face challenges of having incomplete data, especially when it is non-ignorable. These situations affect the subsequent steps of evaluation and make them difficult to perform. Inference with incomplete data is a challenging task in data analysis particularly in clinical trials when the data related to the condition under study is missing. Moreover, results obtained from incomplete data are prone to biases. Parameter estimation with non-ignorable missing data is even more challenging to handle and extract useful information. This dissertation proposes a method based on the influential tilting resampling approach to address non-ignorable missing data in statistical inference. This robust approach is motivated by a brief use of the importance resampling approach used by Samawi et al. (1998) for power estimation. The exponential tilting also inspires it for non-ignorable missing data proposed by Kim & Yu (2011). One of the proposed approach bases assumes that the non-respondents' model corresponds to an exponential tilting of the respondents' model. The tilted model's specified function is the influential function of the function of interest (parameter). The other bases of the proposed approach are to use the importance resampling techniques to draw inference about some model parameters. Extensive simulation studies were conducted to investigate the performance of the proposed methods. We provided the theoretical justification, as well as application to real data.

INDEX WORDS: Missing data, Non-ignorable missing data, Exponential tilting, Influence function, Resampling, Mean estimation, Linear model parameter, Multiple imputation, Follow-up data.

MULTIPLE IMPUTATION USING INFLUENTIAL EXPONENTIAL TILTING IN CASE OF NON-IGNORABLE MISSING DATA

by

KAVITA GOHIL

B.D.S., Gujarat University, India, 2008
M.P.H., Tulane University, United States, 2013
M.P.H., Georgia Southern University, United States, 2016

A Dissertation Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PUBLIC HEALTH

MULTIPLE IMPUTATION USING INFLUENTIAL EXPONENTIAL TILTING IN CASE OF NON-IGNORABLE MISSING DATA

by
KAVITA GOHIL

Major Professor:	Hani Samawi
Committee:	Haresh Rochani
	Lili Yu

Electronic Version Approved:
December 2020

ACKNOWLEDGMENTS

Primarily, I would like to express my sincere gratitude to my advisor Dr. Hani Samawi for unstopping support for my masters and my doctoral journey with his exceptional patience, motivation, and unbeatable knowledge. His guidance helped in all the times of my Georgia Southern career. His inspiration helped me to conduct this research and write this thesis. I am very thankful for his time, efforts, guidance, encouragement, and timely response at each instance, without which this dissertation would not have existed.

I am fortunate to have Dr. Haresh Rochani and Dr. Lili Yu to be part of my dissertation committee. Their insights were the most valuable to expand this research's horizons and facilitated me to widen my research perspectives. Without their precious support and guidance, it would not be possible to conduct this research.

My sincere thanks also go to Dr. Karl Peace and Dr. Robert Vogel. They provided me an opportunity to join JPHCOPH and gave me access to the Karl E. Peace Center for Biostatistics, the epicenter of the research facilities.

I am very grateful for my family and friends whose uncompromising support and encouragement took me this far in this journey. Without their sustenance, I couldn't have done this.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

In statistics, missing data frequently occurs in practice and can significantly affect conclusions of the data. Missing data occurs because of nonresponse or no information provided for one or more items or the entire sampling unit. Data often are missing in research such as economics, sociology, political science, and clinical trials. Governments or private entities choose not to or fail to report critical information. Sometimes the information is not available. Moreover, it could be due to the researchers' errors, such as when the data is collected inefficaciously, or errors made during the data entry.

Analysis with incomplete data leads to biased results and can severely affect the inference. In such cases, the reliability and accuracy of the results are misleading.

The primary effects of missing data during analysis are the loss of power when testing hypotheses and bias in parameter estimation. As the proportion of these missing values increases, the study's power reduces, which has severe consequences on its accuracy (Rubin, 1987). In clinical studies, missing data can pose a risk of false conclusions and misdirect the drug development program (Walton, 2009). In social sciences, missing data at the research design stage causes indistinctness in inferences (Kenward, 2017).

For example, in a social science research survey, data obtained from respondents' answers are presented in the following manner:

Source: Missing Data: The hidden Problem, Retrieved from

"https://www.bauer.uh.edu/jhess/documents/2.pdf

Table 1.1

Missing data example, a social science research survey

Case	Age	Gender	Home	Education	Occupation
1	.	Female	No	16	Non-professional
2	22	Male	No	.	Non-professional
3	39	Male	.	20	Professional
4	.	Female	Yes	.	Professional
5	40	.	Yes	16	Non-professional
6	22	Female	No	16	.
7	35	Male	Yes	18	Professional
8	39	Male	Yes	20	Professional

In this recorded dataset, '.' indicates a missing response, and if it is ignored in any variable, the available data would be inadequate. For instance, if the researcher's goal is to predict an association between homeownership and demographic factors such as age and educational background while ignoring the missing data, the researcher will be left with only half of the observations. Therefore, it is obligatory to properly handle this missing data issue to reduce the impact of missing information.

The early attempt dealing with the issue, in the 1900s, was restricted to algorithmic and computational solutions to the deviations from the intended study designs. The most popular initial method, usually done by statistical software, was complete case analysis. This method recalls and analyzes only available observations. However, in the last quarter of the twentieth century, various strategies have come into the picture, like expectation-maximization (A. P. Dempster, 1977), data imputation, and augmentation methods (Rubin, 1987) (Wong, 1987). All these strategies combined with influential computing resources provided a solution to handle this problem (Geert Molenberghs, 2007).

1.2 Missing Data Mechanism

Missing data can occur at any stage of the research due to many causes in several different scenarios. For example, it can happen in longitudinal studies and clinical trials due to the dropout or follow-up loss. In surveys, it can happen if participants refuse to answer a specific question, accidentally skip a question, or do not know the answer.

When we handle missing data, it is vital to understand the underlying mechanisms for its absence. Using missing data methods depends predominantly on the dependencies' nature in these mechanisms (Rubin, 1987). Rubin distinguished three missing data mechanisms:

1) Missing Completely at Random (MCAR),
2) Missing at Random (MAR),
3) Missing Not at Random (MNAR).

The above mechanisms are used to label the associations between measured variables and the probability of missing data. The first two mechanisms are called the "ignorable" missing mechanism, and the last one is called the "non-ignorable" missing mechanism.

1.2.1 Missing Completely at Random (MCAR)

If the missing data is unrelated to both the missing responses and the set of observed responses, it implies that the observed values represent the entire sample. This mechanism is known as missing completely at random (MCAR). For clinical trials and longitudinal studies, the chance of missing data is the same for the individuals in different groups. It is equally likely to occur in any subject in the study.

Examples for MCAR would be a dropped test tube in the lab for a drug trial, which leads to a missing value in the report for that individual. When a machine fails while collecting or

recording data during usage, it gives a missing reading. Another example is a single question skipped accidentally during the survey.

These are all examples of MCAR missing data.

Properly, let $Y = (y_{ij})$ denote a complete $(n \times K)$ rectangular data set without any missing values, with i^{th} row $y_i = (y_{i1}, y_{i2}, ..., y_{ik})$ where y_{ij} is the value of a variable Y_j for subject i. Y can be partitioned into an observed part, labeled as Y_{obs}, and a missing part, Y_{miss}, which yields $Y = (Y_{obs,} Y_{miss})$. Furthermore, we define a matrix of missingness indicators R, which can take the value of 0 or 1, with dimension $(n \times K)$.

The vector of outcomes for a subject partitioned as:

$$Y_i = (Y_i^{obs}, Y_i^{miss}) \quad \begin{cases} Y_i^{Obs},\ R_{ij} = 1\ the\ subset\ of\ observed\ values \\ Y_i^{miss}, R_{ij} = 0\ the\ subset\ of\ missing\ values \end{cases}$$

To better understand the missing types, assume if the missing data mechanism is characterized by the conditional distribution of missing indicator given the data, i.e. $f(R \mid \phi)$, where ϕ denotes the unknown parameters (Thoemmees & Mohan, 2014).

For MCAR mechanism, the distribution of missingness will be independent of data being observed or missing. In other words, the unconditional distribution of missingness $P(R)$ is equal to the conditional distribution of missingness given Y_{obs} and Y_{miss} or simply Y.

$$P(R \mid Y, \phi) = P(R \mid Y_{obs}, Y_{miss}, \phi) = P(R \mid \phi) \qquad \forall Y, \phi$$

In MCAR type, one cannot verify that the observed data is missing only due to complete randomness. However, the examination of homogeneity of means and variances of the data can guide us to believe that the data is missing as MCAR. Little (1998) provided a multivariate test for homogeneity for assessment. For data with MCAR, the analysis remains unbiased. Still, the

loss of power can occur in inference, but the estimated parameters are not biased because of the missing data (Thompson, 2013).

1.2.2 Missing at Random (MAR)

If the missing data is related to the observed data and not on the missing data, it is called Missing at Random (MAR). It is a less restrictive condition than MCAR. There is a systematic relationship between the missing values and the observed data, but not the missing data.

For example, if men are more likely to tell about their weight than women in a survey, then the weight is Missing at Random. Another example is that if both men and women have the same chance of dropout in a clinical trial, but if the dropout rate is higher in men, then the missing data mechanism is MAR. Besides, survey respondents in service occupations are less likely to report income questions in the survey.

More formally, for the MAR mechanism, the conditional probability of missingness, given the observed part Y_{obs} , is equal to the conditional probability of missingness, given both the observed and unobserved part,

$$(Y_{obs,}Y_{miss}) \text{ i.e. } \quad f(R \mid Y, \phi) = f(R \mid Y_{obs,}\phi) \qquad \forall Y_{miss,}\phi$$

$$P(R \mid Y) = P(R \mid Y_{obs,}Y_{miss}) = P(R \mid Y_{obs})$$

In MAR assumption, the missingness is independent of the unobserved portion of Y, given the information about the observed part of Y. Missing at Random does not mean it always produces unbiased results. Still, there are different ways of dealing with this issue to make unbiased estimates (Thompson, 2013).

The above two mechanisms are considered as “Ignorable Missing” because for such data, we can still produce unbiased parameter estimation without model explanation for missingness (Thompson, 2013).

1.2.3 Missing Not at Random (MNAR)

If the probability/likelihood of missing data is systematically related to the missing hypothetical values (the unobserved information), it is called Missing Not at Random (MNAR). MNAR is characterized by the absence of any of the above-mentioned probability equalities or conditional independencies, which implies that,

$$P(R \mid Y_{obs,} Y_{miss,} \phi) \neq P(R \mid Y_{obs})$$

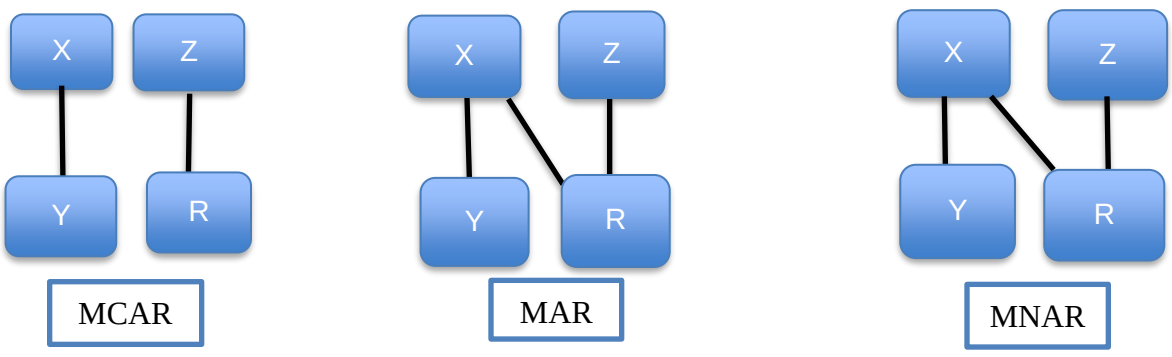
MNAR is the most challenging to deal with as it produces biased results. The bias depends on the correlation between the missing variables. For example, survey respondents with very high incomes are more likely to decline to answer their income questions. In substance abuse trials for abstinence outcomes, people with relapse are more likely to drop out of the study. For the education outcome survey, people with the least education are most likely to skip the question regarding the highest education completed.

MNAR is considered "Non-Ignorable Missing" as in this case, the missing data mechanism itself must be modeled and require some prototype for why the observations are missing and the possible values (Grace-Martin, 2008).

To represent the missing data mechanisms graphically, let us have X - represents the variables that are completely observed, and Y represents a partly missing variable. Let Z represents the causes of missingness unrelated to X and Y, and R represents the missingness indicator (Mohan, 2015).

Figure 1.1

Missing Data Mechanism, Reproduced from Schafer & Graham, 2002



1.3 Methods for Handling Missing Data

Traditionally researchers used a wide variety of techniques to handle missing values. In earlier times, the most common were deletion methods and single imputation methods (Enders, 2004). Deletion techniques include listwise deletion and pairwise deletion in which observations with missing data are discarded, and analysis was done using only complete cases. Those are called Complete Case Analysis (CCS) (Eekhout, n.d.). The single imputation methods include mean/mode substitution, the dummy variable method, and single regression imputations. Such as stochastic regression imputation, hot deck imputation methods, last observation carried forward (LOCF), baseline observation carried forward (BOCF), and worst observation carried forward (WOCF) found in the literature. The regression method consists of constructing a regression model containing missing values used as the response variable. The missing value's replacement is generated by the predicted value derived from the model, and then it is used to impute the missing observations. These regression models depend on the structure of the data like Poisson regression for count variables, logistic regression for binary variables, and linear regression for continuous variables (Raghunathan, Lepkowski, Hoewyk & Solenberger, 2001). The last observation carried forward (LOCF), baseline observation carried forward (BOCF), and worst observation carried forward (WOCF) are used in dropout cases in clinical trials and/or longitudinal studies. These methods are inefficient as they have drawbacks like loss of power, biased coefficient estimates, and underestimated variances (Baraldi & Enders, 2010) (Eekhout et al., 2012).

Due to the weaknesses associated with the above-mentioned traditional methods, researchers came up with model-based approaches to handle the missing data (Graham, 2002). Two widely popular model-based methods, Maximum Likelihood Estimation and Multiple

Imputation are considered "State of the Art" missing data techniques (Schafer & Graham, 2002). These methods are more powerful than traditional methods because not a single piece of data is deleted. These methods are based on the assumptions about the joint distribution of all variables in the model. It produces unbiased estimates with MCAR and MAR data. The major advantage of this method is that with given assumptions, the results obtained using this method can apply to a broader range of contexts with fewer conditions. However, this requires more complex computations (Baraldi & Enders, 2009).

In Maximum Likelihood Estimation, parameter values estimated have the highest probability of producing the samples using complete and incomplete data. It identifies the population parameter values using the highest probability of producing the sample data. In order to obtain the sample data, the log-likelihood function is used to quantify the standardized distance between the observed data point and the parameter of interest. The goal is to minimize this distance. Once the parameters are estimated using the complete data, the missing data are estimated based on those parameters. This method uses the assumption that the observed data are a sample drawn from a multivariate normal distribution. One type of maximum likelihood approach is Expectation-Maximization (EM). The first step is the expectation step, in which parameters are estimated using listwise deletion. Then these estimates are used to create a regression equation that predicts missing data. The second step is the maximization step, which uses those regression equations to fill in the missing values. The steps are repeated with new parameters each time, and new regression equations are determined to fill those missing values. The process is repeated until the convergence is achieved or when the covariance matrix for the subsequent iteration is virtually the same as the preceding one. Disadvantages of using this method are long convergence time when a large portion of data is missing and the complex technique. Another

hindrance is that it also relies on the normality assumption. This method cannot generate values when missingness is present in covariate data (Kang, 2013).

Another model-based method is Multiple Imputation (MI), proposed by Rubin (1987), which is very popular in missing data analysis. In this method, several copies of the dataset are created, and each of them contains different imputed values. Then, analyses are performed on each dataset separately, and they are combined with having a single set of results. This procedure is divided into three phases: the imputation phase, the analysis phase, and the pooling phase. In the imputation phase, the draws are performed to create several complete datasets as needed. In the analysis phase, each data set is analyzed using the standard methods for complete data sets. Lastly, in the pooling phase, the individual analyses' results are combined to get a single estimator and then, consequent inferences are made. MI preserves the advantages of single imputation methods by using standard statistical analysis procedures available for complete data and incorporating data collectors' knowledge. MI eliminates the major problem associated with single imputation by adding uncertainty using multiple data sets. A random draw of imputations increases the estimation's efficiency, and it also contemplates variability due to missing data, and it provides valid inference under the MAR mechanism. MI also allows researchers to study inference sensitivity efficiently as applied to different nonresponse models (Rubin 1987).

Furthermore, regression-based multiple imputation methods include Bayesian least squares, predictive mean matching, and local random residual methods. Other methods for MI include modified propensity score and completion score methods. These model approaches are valid and give unbiased results under the MAR assumption.

Although maximum likelihood and multiple imputations have Bayesian connections, there is a Fully Bayesian (FB) way to handle missing data. Fully Bayesian methods for missing

covariate data involve specifying priors on all the parameters and specifying distributions for the missing covariates. The missing values are then sampled from their full conditional distribution via the Gibbs sampler. Fully Bayesian (FB) approach for missing values is nothing but just incorporating an extra layer in the Gibbs steps. Therefore, Bayesian methods can easily accommodate missing data without extra modeling assumptions or new inference techniques. In this sense, fully Bayesian methods are perhaps the most powerful and more general method for dealing with the missing covariate data.

The methods mentioned above to handle missing data are optimal but require a lot of computation. These methods are only useful when distributional assumptions are correct. If those assumptions are violated, then results are undesirable. A semi-parametric model for missing data is proposed where information about the missing probabilities are used by finding the solution to a set of weighted estimating equations. This approach is called the Weighted Estimating Equation (WEE) method, proposed by Robin et al. (1994). The inference with missing responses is based on the weights, which are inversely proportional to the observed probability.

1.4 Motivation

For the non-ignorable missing data issue, Kim and Yu (2011) and Scharfstein et al. (2014) proposed that the distribution of missing values is related to the observed values' exponential tilted distribution. They used the single imputation regression approach to predict the missing values. This dissertation proposes the influential exponential tilting resampling approach for the missing values to handle non-ignorable missing data problems. Our method is an extension of the exponential tilting approach. This is performed using the exponential tilting probability assignment to the observed data based on the influence function of the statistics under consideration (Samawi et al., 1998). The proposed influential exponential tilting method's

motivation came from a brief use of the importance resampling for power estimation by Samawi et al. (1998).

The subsequent sections provide a literature review in chapter 2, summarizing methods exclusively used to handle non-ignorable missing data in chapter 3. Chapter 4 introduces the influential exponential tilting resampling approach for mean function estimation under non-ignorable missing data. Chapter 5 expands the proposed method to a more robust approach to estimate the linear model parameters under non-ignorable missing data. Then the concluding chapter includes real data examples, discussion, and final remarks.

CHAPTER 2

LITERATURE REVIEW

The effects of missing data during analysis can severely result in loss of power when performing hypotheses testing and biased parameter estimation. It is even worse when missing data is non-ignorable. There are several attempts to handle the non-ignorable missing data during the analysis in the last two decades.

According to Schafer & Graham (2002), model-based approaches, maximum likelihood estimation, and multiple imputations are widely used methods to handle missing data due to their superiority over the traditional missing data techniques for MAR and MCAR data. They attempt to provide unbiased estimates in those cases. These two methods are more powerful than traditional methods because no data needs to be discarded during the analysis. Despite their advantages, these methods are not the perfect solution for handling missing data with an underlying MNAR mechanism. Therefore, it provides a biased parameter estimation. However, the bias tends to be considerably less than the bias that falls out from traditional missing data methods.

Schafer & Graham (2002) indicated that, one must specify a distribution for the missingness and the complete data model to handle missing data without MAR assumption. The missing data framework denotes different factorizations of the full density for modeling incomplete data. Thus, the possible missing data frameworks are the selection model, the pattern mixture model, and the shared parameter model.

The selection model featured by Heckman (1976) encompasses the factorization of the full density. This factorization is the product of the marginal density of the measurement process and the density of the missingness process conditional on the outcome.

The likelihood estimates obtained using the selected model are not computed directly and, therefore, are approximate values. Hence, the parameter values are poorly defined. These models are heavily weighted on non-demonstrated assumptions about the population distribution (Kenward, 1998). According to Laird (1997), these models are considered too complicated for scientific applications and cannot generate any answers. The selection model approaches were used by Wu and Carroll (1988), Diggle and Kenward (1994), Little (1995), Ibrahim et al. (2001), and Stubbendick and Inbrahim (2003). Troxel et al. (1998) propose a selection model, which is valid for non-monotone longitudinal missing data, but it is unmanageable for more than three-time points. The Monte Carlo EM algorithm was used for parametric estimation in selection models with non-ignorable missing response data proposed by Joseph (2011).

In the pattern mixture model, the marginal density is to be factored as the product of the density of the measurement process, which is conditional on the missingness and the marginal density of the missingness process (Rubin, 1987). These models classify individual responses by their missingness group.

Pattern mixture models do not presume robust theories about the missing mechanism. It describes the observed responses in each missing group and then hypothesizes aspects of missing behavior to undetected portions of the data. Thus, pattern mixture models are not extremely sensitive to distribution like selection models, but the estimation of effects is possible by identifying the restrictions, which observed data does not provide. Therefore, they suggested using these models for sensitivity analysis to identify different restrictions to see how the results are changing (Schafer & Graham, 2006). Little used the pattern mixture model approaches to handle missing data (1995) (Little and Wang, 1996) (Hogan and Laird,1997). The main drawback of the pattern mixture model is that one cannot examine the effects of individual

covariates on the marginal distribution of the outcomes in terms of the regression coefficients and computational complexity.

Additionally, it might be possible that pattern mixture models may be intractable for more general patterns of incomplete data (Molenberghs, 2009). Another use of the pattern mixture model is for doing sensitivity analysis using the tipping point approach. In this method, the researcher can specify a subset of observations to derive the pattern mixture model's imputation models. These imputed values can be adjusted by specifying shift and scale parameters for a set of selected observations. That set of selected observations then used for sensitivity analysis with the tipping point approach (Yuan, 2014) (Xu Yan, 2009).

The shared-parameter model uses the same factorization method as the pattern mixture model, with at least one component of the parameter vector shared between both factors (Wu and Carroll, 1988). This model explains the dependency between the measurement and missingness processes through latent variables such as the random effects (Wu and Bailey, 1988; Wu and Carroll, 1988; Creemers et al., 2009). However, this model may fail when the outcomes depend on missing data, such as when varying time residuals cause missingness (Nisha C. Gottfredson, 2014).

When missing data is non-ignorable (MNAR), the maximum likelihood estimation of the data model parameters can give biased results and are based only on the observed data likelihood. Marlin et al. (2003) suggested that to obtain correct maximum likelihood estimates of the data model parameters, and a selection model is needed along with the data model. Most of the time, the parameters of the selection model will also be unknown. The combined data and selection model parameters can be estimated simultaneously by maximizing the full data log-likelihood using the standard EM algorithm (Marlin, 2003) (Zemel).

Another model-based approach introduced by Holman and Glas (2005) is modeling non-ignorable missing data mechanisms with item response theory models. The process is formulated so that the degree to which missing ignorability is violated can be evaluated and used for missing covariate data. In this approach, the distribution of the observed data and the missing data indicator are parameterized by different sets of parameters, which have a common distribution. These distinct parameters are used to find the amount of ignorability (Glas, 2005).

Wang and Fitzmaurice (2006) proposed a simple imputation method for longitudinal studies. It specified two regression models: the marginal mean of the response and the second for the conditional mean of the response given nonresponse patterns. The inference of model parameters is made by using generalized estimating equations. It has a two-step procedure wherein the first step, covariate effects are obtained by solving a generalized estimating equation based on the observed data for imputation. The second step uses the observed and the imputed data to obtain complete longitudinal data and make inferences based on that complete data (Fitzmaurice, 2006).

Harel, in 2008, proposed Outfluence approaches. He introduced a new measure that evaluates the effect of a single missing observation or a group of missing observations, or an incomplete variable, or any combination of these for regression analysis in any parametric settings. In this approach, the outfluence of missing values is calculated by separating the missing values into two categories, like the specific missing value of interest called type B and the rest of the missing values considered type A. For non-ignorable missing, the extended missingness indicator matrix is created for each type, and imputation is done for each type from their predictive distribution. Each missing value has an outfluence measure associated with it, calculated by two-stage multiple imputation for each missing value of the data. The estimated

overall rate of missing information is the sum of the estimated missing information rate due to these missing types, A and B. The outfluence function is calculated, which makes the value of the outfluence function between 0 and 1.

Consequently, if the value of outfluence is close to 0, then that value does not influence the analysis results, and if the value is close to 1, it does have much influence. Moreover, comparing these values will give information about how much influence these missing values have. This measure is analogous to the influence measure in regression analysis but inspects the effect of missing values on a particular analysis and can help in the inference (Harel, 2008; Stratton, 2009).

Cheng (1994) introduced a nonparametric estimation procedure for missing data without modeling the missing mechanism or a joint distribution. Cheng (1994) used kernel regression estimators to estimate the mean function through empirical estimation of the missing pattern, verified under the MAR assumption (Cheng, 1994). Based on his idea, Kim and Yu (2011) proposed a semiparametric approach to estimate the mean in non-ignorable missing data based on the exponential tilted model. The authors assumed a semiparametric logistic regression model for response probability and kernel regression for the missing data. With an exponential tilting model and nonparametric regression, the estimation method became more robust (Yu, 2011).

Kim and Shao (2013) suggested a conditional likelihood approach for handling MNAR, close to the partial likelihood in survival analysis for analyzing censored data under Cox's proportional hazard model. This method utilizes the score function derived from the observed data likelihood. Another variant of this method is a pseudo-likelihood approach by assuming the entire covariate vector as a nonresponse tool. In addition to the methods mentioned above, a few

other approaches to handle MNAR missing data by conditional likelihood are the Callback and capture-recapture experiment (Shao, 2013).

Kim and Shao (2014) suggested some of the naïve approaches which handle the non-ignorable missing data. Out of them, one is the nonresponse instrument method. The nonresponse instrument method is based on partitioning the covariate vector into two parts such that based on that partition, conditional distributions of parameters are identifiable. Thus, it can help identify unknown quantities, giving the observed likelihood of a unique maximum. This unique maximum can help to obtain the maximum likelihood estimate, which eventually maximizing this observed likelihood. Another approach is the Conditional Likelihood approach, which is similar to the partial likelihood in survival analysis for analyzing censored data under Cox's proportional hazard model. Another approach is the Generalized methods of moments (GMM) approach, for which generalized methods of moments are used to construct a set of estimating functions. Based on the vector of observations and parameter space, these estimating functions include the true parameter value. The GMM estimator of unknown parameters obtained by minimizing the estimating functions over the parameter space. The next one is the Latent Variable approach, in which non-ignorable missing is to assume a latent variable related to the study variable. It is assumed that the study variable is observed if and only if the latent variable exceeds a threshold. This approach is applicable in econometrics to explain the self-selection bias and attitude scale (Shao, Statistical methods for handling incomplete data, 2014).

Tang et al. (2014) developed an empirical likelihood for parameters in generalized estimating equations for non-ignorable missing data. They used the exponential tilting model for the MNAR mechanism and proposed modified estimating equations for imputation via the kernel regression method (Tang, 2014).

Linero and Daniels (2018) proposed Bayesian approaches for MNAR outcome data by emphasizing the role of identifying restrictions for likelihood-based perspective and monotone missingness. A nonparametric Bayesian model is used to determine limitations. This nonparametric model is also used to find extrapolation distribution concerning the observed data likelihood. This approach permits putting informative priors on sensitivity parameters and allows simultaneous inference of full data distribution (Daniels, 2018).

The next chapter discusses the specifics of Kim and Yu's (2011) exponential tilting tactic and a brief discussion about the importance resampling approach, from which motivation of the proposed method comes.

CHAPTER 3

METHODS

As non-ignorable missing data is the most difficult to handle than other missing data mechanisms, researchers tried to handle this type of data. They came with various methods to deal with the MNAR missing data, and many of them were mentioned in the previous chapter. The more details described below are Kim and Yu's Exponential tilted model to handle the non-ignorable type of missing data.

3.1 Exponentially Tilting Model

Kim and Yu (2011) used exponential tilting to model the non-ignorable missing data. The proposed method is considered a tilting parameter for determining the amount of departure from the MAR assumption of the response mechanism. Scharfstein et al. (1999) handled the case where the tilting parameter was assumed to be known. Moreover, Kim and Yu (2011) proceeded to estimate the tilting parameter when it was unknown. They used the validation subsample to estimate the tilting parameter and assumed complete responses among the validation subsample elements.

Their model contains one auxiliary variable X and one study variable Y, where the missing values are in Y. The approach is to find a prediction model that fits Y on X. They also defined the response status variable as R. For this method, they proposed an exponential tilting model for non-ignorable missing data. In that method, the nonresponse part of the data is modeled as an exponential tilt for the responding part, and this tilting parameter regulates the amount of departure from the ignorability of the response mechanism.

Let $(x_i, y_i), i = 1, 2, ..., n$ be n independent observations of continuous random variables (X, Y) with joint distribution being $F(x, y)$. In this joint distribution, x_i always observed and y_i is subject to missing. The parameter of interest is $\theta = E(Y)$. If R_i is the response indicator for

$$f_0(y_i \mid x_i) = f_1(y_i \mid x_i) \times \frac{O(x_i, y_i)}{E\{O(x_i, y_i) \mid x_i, R_i = 1\}}, \tag{3.1}$$

where $R_i = 1$ if y_i is observed and 0 otherwise. Then the response mechanism will have Bernoulli distribution with probability π_i.

$$R_i = 1 \mid (x_i, y_i) \sim Bernoulli(\pi_i) \tag{3.2}$$

where $\pi_i = \pi(x_i, y_i)$ and R_i is independent of R_j for any $i \neq j$.

If the conditional density of respondents for the observed part is given by $f_1(y_i \mid x_i)$ where $R_i = 1$ and the conditional density of respondents for the non-observed part is given by $f_0(y_i \mid x_i)$ where $R_i = 0$ then under ignorable missing condition (MAR), $f_1(y_i \mid x_i) = f_0(y_i \mid x_i)$ and more generally probability for observed and non-observed part would be equal and can be written as per below:

$$P(Y_i \in B \mid x_i, R_i = 0) = P(Y_i \in B \mid x_i, R_i = 1) \tag{3.3}$$

It is true for any measurement set *B.*

The following equation attain the consistent estimator of θ

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} (R_i Y_i + (1 - R_i) \hat{\mu}(x_i)) \tag{3.4}$$

where $\hat{\mu}(x_i)$ is a consistent kernel estimator of $\mu(x_i) = E(Y_i \mid x_i)$ when *R=1.*

Under the non-ignorable missing data, $f_1(y_i \mid x_i) \neq f_0(y_i \mid x_i)$ and,

$$P(Y_i \in B \mid x_i, R_i = 0) \neq P(Y_i \in B \mid x_i, R_i = 1) \quad \text{and } \hat{\theta}_1 \text{ is biased.}$$

Thus, one can use

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(R_i Y_i + (1-R_i)\hat{\mu}_0(x_i)) \tag{3.5}$$

where $\hat{\mu}_0(x_i)$ is a consistent weighted kernel estimator of $\mu_0(x_i) = E(Y_i \mid x_i)$ when $R_i = 0$.

The computation of conditional distribution when $R_i = 0$ is as follows:

$$P(Y_i \in B \mid x_i, R_i = 0) = P(Y_i \in B \mid x_i, R_i = 1)\frac{P(R_i = 0 \mid x_i, y_i \in B) / P(R_i = 1 \mid x_i, y_i \in B)}{P(R_i = 0 \mid x_i) / P(R_i = 1 \mid x_i)}.$$

From which the conditional distribution of missing data can be written as the following:

$$f_0(y_i \mid x_i) = f_1(y_i \mid x_i) \times \frac{O(x_i, y_i)}{E\{O(x_i, y_i) \mid x_i, R_i = 1\}} \quad , \tag{3.6}$$

where $O(x_i, y_i) = \frac{\mathrm{P}(R_i = 0 \mid x_i, y_i)}{\mathrm{P}(R_i = 1 \mid x_i, y_i)}$ which is a conditional odd of nonresponse.

If the response probability model is a logistic regression model for some function g(.) as a function of *x*, $r(y_i)$ as a function of Y, and parameter ϕ and is given by,

$$\pi(x_i, y_i) = \mathrm{P}(R_i = 1 \mid x_i, y_i) = \frac{\exp[g(x_i) + \phi r(y_i)]}{1 + \exp[g(x_i) + \phi r(y_i)]}, \tag{3.7}$$

This response probability model in (3.7) is a semiparametric model because in logistic regression, the component accompanying x_i and $g(x_i)$ is unspecified and the component accompanying y_i can be parametrically modeled with parameter ϕ. To simplify the derivation, Kim and Yu (2011) suggested taking $r(y_i) = y_i$.

Under this response model, the odd function is defined as

$$O(x_i, y_i) = \exp\{-g(x_i) - \phi y_i\}$$

Also, the conditional distribution of the missing data is written as,

$$f_0(y_i \mid x_i) = f_1(y_i \mid x_i)\frac{\exp(-\phi y_i)}{E(\exp(-\phi y_i) \mid x_i, R_i = 1)} \tag{3.8}$$

The above equation shows that the nonrespondents' density is an exponential tilting of the respondents' density. The parameter $-\emptyset = \gamma$ is a tilting parameter that examines the amount of departure from the ignorability of the response mechanism.

Thus, the real density of *Y* based on the above model can be given by,

$$f(y_i \mid x_i) = \pi_i . f_1(y_i \mid x_i) + (1-\pi_i) f_0(y_i \mid x_i)$$

$$= \pi_i . f_1(y_i \mid x_i) + (1-\pi_i) f_1(y_i \mid x_i)\frac{\exp(-\phi y_i)}{E(\exp(-\phi y_i) \mid x_i, R_i = 1)}$$

$$= f_1(y_i \mid x_i)\left(\pi_i + (1-\pi_i)\frac{\exp(-\phi y_i)}{E(\exp(-\phi y_i) \mid x_i, R_i = 1)}\right)$$

$$= f_1(y_i \mid x_i)\left(\pi_i + (1-\pi_i)w_i\right), \tag{3.9}$$

where $w_i = \dfrac{\exp(-\phi y_i)}{E\{\exp(-\phi y_i) \mid x_i, R_i = 1\}}$

To find out the departure from the ignorability, the tilting parameter $-\emptyset = \gamma$ needs either to be known using planned missingness or sensitivity analysis. In other cases, it needs to be estimated. Kim and Yu (2011) proposed that it can be estimated using the follow-up studies. The follow-up is done to obtain responses in a subset of the nonrespondents.

Yu's (2011) approach depends on the assumption of the missingness models. This research proposes an approach that is presumably more robust. We use the influential exponential tilting (Yu, 2011) with a resampling method to estimate the model parameters with non-ignorable missing data. This approach uses an influential function to penalize observations that are more influential concerning statistics under consideration, which is in the opposite direction of the possible non-ignorable missingness but rewards those in the same direction. For

this process, the importance-resampling approach is used to impute the missing data by the influential exponential tilting weights, as the resampling distribution. The importance resampling approach is one of the most promising bootstrap methods used to reduce computational efforts. This method uses the variance reduction approach. Next, we discuss briefly the bootstrap and the importance resampling methods.

3.2 Bootstrap & Importance Resampling

3.2.1 Bootstrap Inference

Bootstrap methods are computer-intensive, which involves simulated data sets. The uniform (ordinary) bootstrap resampling method was established by Efron (1979). This method is based on resampling with replacement from the observed sample, and each one has an equal probability to be selected on sample values. Uniform bootstrap resampling described by Efron (1979) and others is an assumption-free method and can be used for inferences. However, it is designed for a complete and continuous set of observations. This initial approach is called the uniform resampling method or uniform bootstrap. This uniform bootstrap involved thousands of simulated datasets. For one sample case, the uniform resampling rules will be applied to each sample separately and independently (Ibrahim, 1991; Samawi et al., 1996; Samawi et al., 1998; Samawi, 2003).

Suppose $\aleph = (Y_{11},\ Y_{12},\ ...,\ Y_{1n})$ is independent and created by random samples drawn from, $f(y)$. Assume that the parameter of interest is the functional $\mu_Y = T(F) = \int m(y)dF(y).$

If S is an estimate of μ_Y based on $\aleph$ that is $S = S(\aleph)$ and can be defined as

$$S = \hat{\mu}_Y = T(\hat{F}_n) = \int m(y)d\hat{F}_n(y)$$

where $\hat{F}_n$ is the empirical distribution.

Furthermore, assume that S is a smooth function of the random samples. Assume another parameter U which is a function of S, that is, $U = U(S)$. Then we can obtain U* for the same function of the data but in resamples $\aleph = (Y_{11}^*, Y_{12}^*, ..., Y_{1n}^*)$ which are drawn from $\aleph$ according to the rules which places probability $\frac{1}{n}$ on each sample value of $\aleph$. Let $u = E(U)$ then the bootstrap estimate (say $\hat{u}$) of u is given by

$$\hat{u} = E(U^* \mid \aleph) \tag{3.10}$$

This expected value is often not computable.

3.2.2 Uniform Resampling Approximation for Bootstrap Estimate

Assume that the probability of selecting Y_{1i} in a resample is

$$P(Y_1^* = Y_{1i} \mid \aleph) = \frac{1}{n}. \tag{3.11}$$

Let $\aleph_1^*$, $\aleph_2^*$, ..., $\aleph_B^*$ denote independent resamples sets of size *B*, each drawn from $\aleph$. To obtain a Monte Carlo approximation to $\hat{u}$ using uniform resampling, let U_b^* denote *U* computed from $\aleph_b^*$. Then, the uniform resampling approximation to the bootstrap estimate $\hat{u}$ is given by

$$\hat{u_B^*} = B^{-1} \sum_{b=1}^{B} \left(U_b^* \right). \tag{3.12}$$

Do and Hall (1991) showed that $\hat{u_B^*}$ is an unbiased approximation to $\hat{u}$, in the sense that $E(\hat{u_B^*} \mid \aleph) = \hat{u}$. Moreover, an approximation of the bootstrap bias of u can be obtained by $\hat{bias}^* = |\hat{u_B^*} - \hat{u}|$, and an approximation of the bootstrap MSE can be obtained by $\hat{MSE}^* = B^{-1} \sum_{b=1}^{B} \left(U_b^* - \hat{u} \right)^2$.

3.2.3 Importance Resampling Approximation for Bootstrap Estimate

In subsequent years, different researchers derived different thoughts to reduce the number of simulated datasets and, thus, computational struggles (D. V. Hinkely, 1989). Among that,

Johns (1988) and Davison (1988) introduced the most promising approach of importance resampling for probability and quantile estimation. In importance resampling, data values are resampled with unequal or tilted probabilities. That makes it more likely for the statistics under consideration to assume a value that is close to the point of interest. Some resampled values y^* may contribute much more to estimate μ than others. Importance sampling aims to sample more frequently from those important values of y^*. This can be achieved by resampling from a distribution that concentrates probability on these values of y^* and then weighing the values of $m(y^*)$ to replicate the approximation if it had been sampled from G. Then importance resampling identity is

$$\mu = \int m(y^*)dG(y^*) = \int m(y^*)\frac{dG(y^*)}{dH(y^*)}dH(y^*)$$

where the support of G includes the support of F.

Importance sampling approximates the above expression using independent resamples $y_j^{**} = (y_{1j}^{**}, y_{2j}^{**}, ..., y_{nj}^{**})$, which are drawn from $\aleph$ according to the rules, which places probability $(g_1, g_2, ..., g_n)$ on each sample value of $\aleph$, respectively. Assume that the probability of selecting Y_{1i} in a resample is

$$P(Y_1^{**} = Y_{1i} \mid \aleph) = g_i; i = 1, 2, ..., n.$$

Let $\aleph_1^{**}$, $\aleph_2^{**}$, ..., $\aleph_B^{**}$ denote independent resamples sets of size *B,* each drawn from $\aleph$. To obtain a Monte Carlo approximation to $\hat{u}$ using importance resampling, let U_b^{**} denote *U* computed from $\aleph_b^{**}$. Then, the importance resampling approximation to the bootstrap estimate $\hat{u}$ is given by

$$\hat{u}_B^{**} = B^{-1}\sum_{b=1}^{B}\left(U_b^{**}\prod_{j=1}^{n}\frac{dF_{jn}}{dG_{jn}}\right).$$

Do and Hall (1991) showed that $\hat{u}_B^{**}$ is an unbiased approximation of $\hat{u},$ in the sense that $E(\hat{u}_B^{**} \mid \aleph) = \hat{u}$. Moreover, an approximation of the bootstrap bias u can be obtained by $\hat{bias}^{**} = \mid \hat{u}_B^{**} - \hat{u} \mid$, and an approximation of the importance resampling variance can be obtained $\hat{Var}^{**} = B^{-1}\sum_{b=1}^{B}\left(U_b^{**2}W_b\right) - (\hat{u})^2$ $W_b = \prod_{j=1}^{n}\frac{dF_{jnb}}{dG_{jnb}}$.

In the following chapters, we introduce the Influential Exponential Tilting Resampling Approach method for parameter estimation in the non-ignorable missing data. The next chapter describes the mean functional estimation method with non-ignorable missing data using an influential exponential tilting approach. Following the mean functional estimation, the subsequent chapter describes the modified Importance Resampling Approach for linear parameter estimation with non-ignorable missing data. Simulation studies were also presented for both functional estimation procedures.

CHAPTER 4

Mean Functional Estimation with Non-Ignorable Missing Data Using Influential Exponential Tilting Resampling Approach

Kim and Yu (2011) used exponential tilting to model the non-ignorable missing data. In this chapter, the exponential tilting approach is extended using the influence function for tilting the assigned probability to the observed responses used by Samawi et al. (1998). The proposed method's advantage is that the tilting based on the influential function depends on the statistics (functional) under consideration. This method is robust compared to other methods as it fixes the tilting parameter for the benchmark assumption, in which different ranges of deviation from missingness at random are considered. The preliminaries for this method are already described in the previous chapter.

The purpose of this work is to propose a method for handling missing data using the influential tilting resampling approach (ITRA), which considers the missing pattern of MNAR assumptions. The proposed ITRA uses an influence function to penalize observations that are more influential concerning the statistics under consideration and in the opposite direction of the possible MNAR missingness but rewards those in the same direction. In this process, importance sampling distribution for the outcome is created, and resampling can be done from that distribution.

Similar to the exponential tilting method proposed by Kim and Yu (2011), ITRA states that the non-responding part's model is an exponential tilting of the responding part. In general, in the exponential tilting approach, the model in (3.7) is derived from

$f_0(y_i \mid x_i) = f_1(y_i \mid x_i) \dfrac{\exp(-\phi y_i)}{E(\exp(-\phi y_i) \mid x_i, R_i = 1)}$ by replacing $r(y)$ with y. Sometimes, (y) serves to

quantify the effect of the observed response on the risk of dropping out (Scharfstein et al., 2014). For our ITRA, we chose *r(y)* to be the influential function for estimating the functional $\mu_Y = T(F)$. The influential function approach considers parameter estimation based on nonparametric estimation of unknown functional. In general, nonparametric estimation consists of estimating a statistical functional $\mu_Y = T(F)$, where we presume that *Y* follows a c.d.f, say *F*. The influence function of a functional (*F*), under some regularity conditions, is defined using Gateaux derivative by

$$L(y) = \underset{\varepsilon \to 0}{Lim}\left[\frac{T\{(1-\varepsilon)+\varepsilon\delta_y\} - T(F)}{\varepsilon}\right] \tag{4.1}$$

where

$$\delta_y(u) = \begin{cases} 0 & \text{if } u < y \\ 1 & \text{if u} \geq \text{y.} \end{cases}$$

For estimating the mean of Y, the influence function, L(y), and its estimate $\hat{L}(y_i)$ are defined by $L(y) = y - \mu_Y$ and $\hat{L}(y_i) = y_i - \bar{y},$ respectively. Using the influential function can be justified because the problem in this work is to estimate a parameter depending on the probability density function of all the responses. This probability density function is partially unknown because only the distribution of the observed data is available. As in the resampling exponential tilting approach (see Samawi et al., 1996; Samawi et al. 1998), we suggest that the distribution of the missing values can be defined as,

$$f_\eta(y_i \mid x_i) = f_0(y_i \mid x_i) = f_1(y_i \mid x_i).\frac{\exp(\eta r(y_i \mid x_i))}{E[\exp(\eta r(y_i \mid x_i)) \mid R_i = 1]}, \tag{4.2}$$

where η is the tilting parameter. The tilting parameter determines the magnitude of the departure from the ignorability of the response mechanism by penalizing observations that are

more influential concerning the statistic under consideration. These influential observations are in the opposite direction of the possible MNAR missingness but reward those in the same direction. In this case, the specified function is chosen as $r(y_i \mid x_i) = L(y_i \mid x_i) / n_1.\sigma_L$, where $\sigma_L = \sqrt{E(L(Y,x)^2)}$. Note that under MAR assumption η=0.

Besides, Huber (1981) showed that $\sqrt{n}[T(F_n) - T(F)] = \sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} L(y_i \mid x_i)] + o_p(1)$ F_n the empirical function F and $o_p(1)$ tends to be 0 as $n \to \infty$. Now by Central Limit Theorem, we have

$$\sqrt{n}[T(F_n) - T(F)] \sim N(0, \tilde{\sigma}^2) \tag{4.3}$$

Where $\tilde{\sigma}^2 = \int L^2(y \mid x) dF(y \mid x)$.

Finally, using the influence function $L(y_i \mid x_i)$ for estimating the population mean is justified because it can produce the same conditional distribution of the nonresponse in (3.8) as follows:

$$f_0(y_i \mid x_i) = f_1(y_i \mid x_i) \frac{\exp(-\phi y_i)}{E(\exp(-\phi y_i) \mid x_i, R_i = 1)}$$

$$= f_1(y_i \mid x_i) \frac{\exp(-\phi n_1.\sigma_L (y_i \pm \mu_Y) / n_1.\sigma_L)}{E(\exp(-\phi n_1.\sigma_L (y_i \pm \mu_Y) / n_1.\sigma_L) \mid x_i, R_i = 1)}$$

$$= f_1(y_i \mid x_i) \frac{\exp(-\frac{n_1.\sigma_L \phi}{n_1.\sigma_L} L(y_i \mid x_i) - \frac{n_1.\sigma_L \phi \mu_Y}{n_1.\sigma_L})}{E(\exp(-\frac{n_1.\sigma_L \phi}{n_1.\sigma_L} L(y_i \mid x_i) - \frac{n_1.\sigma_L \phi \mu_Y}{n_1.\sigma_L}) \mid x_i, R_i = 1)}$$

$$= f_1(y_i \mid x_i) \frac{\exp(-\frac{n_1.\sigma_L \phi}{n_1.\sigma_L} L(y_i \mid x_i))}{E(\exp(-\frac{n_1.\sigma_L \phi}{n_1.\sigma_L} L(y_i \mid x_i)) \mid x_i, R_i = 1)}$$

$$= f_1(y_i \mid x_i) \frac{\exp(\frac{\eta}{n_1.\sigma_L} L(y_i \mid x_i))}{E(\exp(\frac{\eta}{n_1.\sigma_L} L(y_i \mid x_i)) \mid x_i, R_i = 1)} \tag{4.4}$$

where $\eta = -n_1.\sigma_L \phi.$

The proposed approach, ITRA, is similar to Kim and Yu's (2011) approach. However, Kim and Yu (2011) predict the missing observation using the single imputation regression kernel estimates. On the other hand, our proposed approach is using the empirical importance resampling method.

4.1 Semiparametric Approaches to ITRA- 2 stages

For a dataset of size *n* to be used for estimation, we consider the outcome of interest *Y* with a density function *f*. Suppose that we have nonignorable missing data of size (n_2 <n). The following steps are used to perform influential tilting resampling approaches:

1. Determine the tilting parameter (η) by prior information on a benchmark assumption for how data can be MNAR and find the percentage of missing values (**p**) to calculate η which we will discuss later using two-stage resampling (multiple imputations) methods to guess η.

2. Estimate the assumed distribution f_0 of the missing values as an exponential tilted distribution of the observed value as follows:

$$\hat{f}_0(y_i \mid x_i) = \hat{f}_1(y_i \mid x_i).\frac{R_i \exp\left(\eta \hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)}{\sum_{i=1}^{n_1} R_i \exp\left(\eta \hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)} = \hat{f}_1(y_i \mid x_i) w_i(y_i, \eta), \tag{4.5}$$

where n_1 is the size of observed data, $\hat{L}(y|x)$ is the estimate of influential function and

$\hat{\sigma}_L = \sqrt{\frac{1}{n_1}\sum_{i=1}^{n_1} \hat{L}^2(y_i, x_i)}$.

3. Using standard nonparametric resampling methodology to draw n_2 (number of missing values) subsets. Each of them is of size n_1 (number of observed values) using the tilted distribution of the observed data. Denote the imputed (resampling) sets by $\mathbf{y}_1^*, \mathbf{y}_2^*, ..., \mathbf{y}_{n_2}^*$, where $\mathbf{y}_i^* = (y_{i1}, ..., y_{in_1}); i = 1, 2, ..., n_2$.

4. From the i^{th} resampling (imputed subsample) we estimate the i^{th} missing value as,

$$\hat{y}_i^* = \sum_{j=1}^{n} w_{ij} y_{ij}^*; i = 1, 2, ..., n_2 \text{ ,where } w_{ij} = \frac{R_j \exp(\eta \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\eta \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}. \quad (4.6)$$

5. Similar to Kim and Yu (2011) method, we propose estimating μ_Y by

$$\hat{\mu}_Y^* = \frac{1}{n} \sum_{i=1}^{n} (R_i Y_i + (1 - R_i) \hat{y}_i^*) . \quad (4.7)$$

4.1.1 Properties of Semiparametric Approaches (ITRA)

The proposed method ITRA combines multiple imputation techniques, resampling approach, and Kim and Yu's (2011) approach. The desired estimates of the missing values are generated using the tilting resampling method. These estimated values are imputed to generate a complete dataset. The complete data is then used to estimate $\mu_y = T(F(Y, X))$. In the subsequent discussion, we will show the results for estimating the population mean, implying that $\mu_Y = E(Y)$. However, under the above assumption, the parameter of interest μ_Y can be written as $\mu_Y = E[R.E(Y \mid R = 1) + (1 - R).E(Y \mid R = 0)]$

$$= E[\pi(y)(Y \mid R = 1)] + E[(1 - \pi(y))(Y \mid R = 0))]$$

$$= \pi(y) E[Y \mid R = 1] + (1 - \pi(y)) E[Y \mid R = 0] \quad (4.8)$$

where $\pi(y) = \mathrm{P}(R = 1 \mid y)$.

As described in Kim and Yu (2011), the proposed estimator of μ_Y is given by

$$\hat{\mu}_Y^* = \frac{1}{n}\sum_{i=1}^{n}(R_i Y_i + (1-R_i)\hat{y}_i^*), \tag{4.9}$$

$$\text{where } \hat{y}_i^* = \sum_{j=1}^{n} w_{ij} y_{ij}; i = 1, 2, ..., n_2, \text{and } w_{ij} = \frac{R_j \exp(\eta \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\eta \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}. \tag{4.10}$$

Hence using a similar argument as in Kim and Yu (2011) and using the derivation from equation (4.4), we can show that

$$\begin{aligned}
&p\lim_{n\to\infty}\hat{\mu}_Y = p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(R_i Y_i + (1-R_i)\hat{y}_i^*) \\
&= E\left[\pi(Y)Y + (1-\pi(Y))RY\exp\left(\eta L(Y)/(n_1.\sigma_L)\right)/E\left[R\exp\left(\eta L(Y)/(n_1.\sigma_L)\right)\right]\right] \\
&= \pi(y)E(Y \mid R=1) + (1-\pi(y))\frac{E\left[\exp\left(\eta L(Y)/(n_1.\sigma_L)\right)\pi(Y)Y\right]}{E\left[\exp\left(\eta L(Y)/(n_1.\sigma_L)\right)\pi(Y)\right]} \\
&= \pi(y)E(Y \mid R=1) + (1-\pi(y))\frac{E\left[(1-\pi(Y))Y\right]}{E\left[1-\pi(Y)\right]} \\
&= \pi(y)E(Y \mid R=1) + (1-\pi(y))E(Y \mid R=0).
\end{aligned} \tag{4.11}$$

Now, let $q_i^* = \hat{y}_i^* + \frac{R_i}{\pi(y_i)}(Y_i - \hat{y}_i^*)$, then it can be written as

$$q_i^* = Y_i + (\frac{R_i}{\pi(y_i)} - 1)(Y_i - \hat{y}_i^*) \Rightarrow E(q_i^*) = \mu_Y. \tag{4.12}$$

For large sample sizes and using the result in (4.11) we have

$$q_i^* \to q_i = Y_i + (\frac{R_i}{\pi(y_i)} - 1)(Y_i - E(Y_i \mid R=0)). \tag{4.13}$$

Using similar arguments as in Theorem 1 in Kim and Yu (2011) and (4.13), we can show that

$$\sqrt{n}(\hat{\mu}_Y^* - \mu_Y) \to N(0, \sigma_w^2), \tag{4.14}$$

where

$$\sigma_W^2 = Var(q_i) = Var_Y(E_R[q_i]) + E_Y(Var_R[q_i])$$
$$= Var(Y) + E\left[\left(\frac{1}{\pi(X,Y)} - 1\right)(Y - E(Y \mid R=0))^2\right]. \quad (4.15)$$

The variance estimate associated with $\hat{\mu}^*$ is similar to that provided by Kim and Yu (2011) as

$$\hat{\sigma}_W^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{q}_i^{*2} - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{q}_i^*\right)^2, \quad (4.16)$$

where $\hat{q}_i^* = \hat{y}_i^* + \frac{R_i}{\hat{\pi}_i}(y_i - \hat{y}_i^*)$, and $\hat{\pi}_i = \left[1 + \frac{n_2 \exp(\eta \hat{L}(y_i)/(n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} r_j \exp(\eta \hat{L}(y_j)/(n_1.\hat{\sigma}_L))}\right]^{-1}$ is the estimated response probability of (3.9) with fixed known η as in Kim and Yu (2011). Note we can show that

$$\bar{\hat{q}}^* = \frac{1}{n}\sum_{i=1}^{n}\hat{q}_i^* = \hat{\mu}_Y^*.$$

4.1.2. Finding the tuning (tilting) parameter η

This research proposes two ways to guess η; the first is to determine a benchmark assumption for how the data is missing in a nonignorable manner. In other words, we check if the lower values or higher values are missing in the data. After that, find the percentage of missing values (P) to calculate the tilting parameter. If missingness was from above or below, we determine the tilting parameter by

$$\eta = (217.621 + 120.952(\text{P}) - 0.3\sqrt{P})\, sign(\text{direction of missingness})$$

Direction of missingness sign= -1 if the missing from above (smaller values are missing) else sign= +1. This method works in these cases; however, it depends on the benchmark assumption.

If η is guessed using only benchmark assumption, then it's called the nonparametric estimation approach.

However, in absence of a benchmark assumption it is difficult to guess η value. The second method is more flexible by using two-stage multiple imputations. This method allows us to find η without a benchmark assumption as 2 stages provides the room for the sensitivity analysis specifically when a benchmark assumption is not available or difficult to assume.

1- Start with an initial guess of η say η^0

2- In the first stage, with η^0, use the tilting distribution $\hat{f}_0(y_i \mid x_i)$, where

$$\hat{f}_0(y_i \mid x_i) = \hat{f}_1(y_i \mid x_i).\frac{R_i \exp\left(\eta^0 \hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)}{\sum_{i=1}^{n} R_i \exp\left(\eta^0 \hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)}$$

to obtain a resample of size n_2, namely, $\mathbf{y}_0^* = (y_{01},...,y_{0n_2})$. This can be considered a test data that allows us to find the range (-a, a) to obtain η. We are performing sensitivity analysis using η^0 to obtain correct η.

3- For η in (-a, a), (a>0 and the interval is pre-determined and contains η^0)

4- Using the Newton-Raphson method to find the root of the

$$\sum_{i=1}^{n}(1-R_i)(y_{0i} - \hat{y}_i^*(\eta)) = 0$$

where, $\hat{y}_i^*(\eta) = \sum_{j=1}^{n_2} w_{ij}(\eta) y_{0j}; i = 1,2,...,n_2$, and $w_{ij} = \dfrac{R_j \exp(\eta \hat{L}(y_{0j})/(n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\eta \hat{L}(y_{0j})/(n_1.\hat{\sigma}_L))}$.

4.2 Semiparametric Approach to ITRA

The prior information about the benchmark assumption is not always available and hence η is unknown. We propose to use a semiparametric approach to estimate the functional $\mu_Y = T(F)$ using the ITRA method as in Kim and Yu (2011):

1. Estimate η using a follow-up (validation) data (a certain percentage of the missing values) by $\hat{\eta}$ where $\hat{\eta}$ is the solution of

$$\sum_{i=1}^{n}(1-R_i)I_i(y_i - \hat{y}_i^*(\eta)) = 0, \tag{4.17}$$

where I_i is an indicator function that takes the value of one if the observed i^{th} sampling unit belongs to the follow-up sample and takes the value of zero otherwise and $\hat{y}_i^*(\eta)$ is defined similar to equation (4.10).

2. Estimate the assumed distribution $f_{\hat{\eta}}$ of the missing values as an influential exponential tilted distribution of the observed value (including the observed follow-up sample) as follows:

$$\hat{f}_0(y_i \mid x_i) = \hat{f}_1(y_i \mid x_i).\frac{R_i \exp\left(\hat{\eta}\hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)}{\sum_{i=1}^{n} R_i \exp\left(\hat{\eta}\hat{L}(y_i \mid x_i)/(n_1.\hat{\sigma}_L)\right)} = \hat{f}_1(y_i \mid x_i)w(y_i,\hat{\eta}),$$

where $\hat{L}(y|x)$ is the estimated influential function.

3. Using standard semiparametric resampling methodology, draw resampling datasets each of size n_{11} (total observed data), based on the observed data n_{21} (the size of the remaining missing values). Then $n = n_{21} + n_{11}$.

4. From the i^{th} resampling (imputed subsample) $\mathbf{y}_i^{*\prime} = (y_{i1}^*, y_{i2}^*, ..., y_{in_{11}}^*)$ we estimate the i^{th} missing value as,

$$\hat{y}_i^*(\hat{\eta}) = \sum_{i=1}^{n} w_{ij} y_{ij}^*(\hat{\eta}); i = 1, 2, ..., n_{21},$$

$$\text{where } w_{ij} = \frac{R_i \exp(\hat{\eta}\hat{L}(y_{ij}^{**}) / (n_{11}.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\hat{\eta}\hat{L}(y_{ij}^{**}) / (n_{11}.\hat{\sigma}_L))}.$$

5. As similar to Kim and Yu (2011), we propose estimating μ_Y by

$$\hat{\mu}_Y^{**} = \frac{1}{n}\sum_{i=1}^{n}(R_i Y_i + (1 - R_i)\hat{y}_i^*(\hat{\eta})).$$

4.2.1 Estimating the functional $\mu_Y = T(F)$ using ITRA

Kim and Yu (2011) suggested to estimate η using independent survey or using a validation sample, which is a subsample of the nonrespondents. Thus, in either case, the proposed estimator of μ_Y is

$$\hat{\mu}_Y^{**} = \frac{1}{n}\sum_{i=1}^{n}(R_i Y_i + (1 - R_i)\hat{y}_i^*(\hat{\eta})), \tag{4.18}$$

where

$$\hat{y}_i^*(\hat{\eta}) = \sum_{i=1}^{n} w_{ij} y_{ij}^*(\hat{\eta}); i = 1, 2, ..., n_{21}, \text{ where } w_{ij} = \frac{R_i \exp(\hat{\eta}\hat{L}(y_{ij}^*) / (n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\hat{\eta}\hat{L}(y_{ij}^*) / (n_1.\hat{\sigma}_L))}. \tag{4.19}$$

We will consider here the case when a validation sample is randomly selected from the set of nonrespondents and the data collected from all the subjects in the validation sample. Hence,

similar to the method described by Kim and Yu (2011), a consistent estimate of η using a follow-up (validation) data which is called $\hat{\eta}$, and is obtained by solving,

$$\sum_{i=1}^{n}(1-R_i)I_i(y_i - \hat{y}_i^*(\eta)) = 0, \tag{4.20}$$

where I_i is an indicator function that takes the value of one if then observed i^{th} sampling unit belongs to the follow-up sample and takes the value of zero otherwise and $\hat{y}_i^*(\eta)$ is defined in (4.10).

As mentioned in Kim and Yu (2011), the solution, $\hat{\eta}$, of (4.20) exist almost everywhere. Then our proposed semiparametric estimator presented in (4.18) for estimating μ_Y has the following asymptotic properties. The proofs are similar to those provided in Kim and Yu (2011).

1- $\hat{\eta} \to \eta^0$

2- $\sqrt{n}(\hat{\mu}_Y^{**} - \mu_Y) \to N(0, \sigma_S^2)$,

where $\sigma_S^2 = Var(q_i^{**})$, $q_i^{**} = E(Y \mid R_i = 0, \eta^0) + [\frac{I_i(1-R_i)}{P(I_i = 1 \mid R = 0)} + R_i][Y_i - E(Y \mid R_i = 0, \eta^0)]$,

and $E(Y \mid R_i = 0, \eta^0) = p\lim_{n\to\infty} \hat{y}^*(\eta^0)$. With little algebra, we can show that

$$\sigma_S^2 = Var(Y) + \left(\frac{1}{P(I_i = 1 \mid R = 0)} - 1\right) E[(1-R)(Y - E(Y \mid R_i = 0, \eta^0)]$$

Now since,

$$E(Y_i \mid R_i = 0, \eta^0) = p\lim_{n\to\infty} \hat{y}^*(\eta^0) = p\lim_{n\to\infty}\left(\frac{\sum_{j=1}^{n} R_j y_{ij}^* \exp(\eta^0 \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\eta^0 \hat{L}(y_{ij}^*)/(n_1.\hat{\sigma}_L))}\right).$$

$$= \left(\frac{E[RY \exp(\eta^0 L(Y)/(n_1.\sigma_L))]}{E[R\exp(\eta^0 L(Y)/(n_1.\hat{\sigma}_L))]}\right)$$

Then, if the model in (4.1) is true, this implies that $\eta^0 = \eta$ and then

$$E(Y_i \mid R_i = 0, \eta^0) = \left(\frac{E[RY \exp(\eta L(Y)/(n_1.\sigma_L))]}{E[R\exp(\eta L(Y)/(n_1.\sigma_L))]} \right) = \left(\frac{E[(1-R)Y]}{E[(1-R)]} \right) = E(Y \mid R_i = 0).$$

Finally, the variance estimate associated with $\hat{\mu}_Y^{**}$ is similar to that provided by Kim and Yu (2011) as follows: Let $\hat{q}_i^{**} = \hat{y}_i^*(\hat{\eta}) + [\frac{I_i(1-R_i)}{P(I_i = 1 \mid R_i = 0)} + R_i][Y_i - \hat{y}_i^*(\hat{\eta})]$

$$\hat{\sigma}_S^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{q}_i^{**2} - \left(\frac{1}{n}\sum_{i=1}^{n} \hat{q}_i^{**} \right)^2 . \qquad (4.21)$$

4.3 Simulation

To get an insight into the theory, we conducted a simulation study. We generate 2000 samples of size 200 from the model $y_i = 1 + 0.5x_i + e_i$, where $x_i \sim N(2,1)$ and $e_i \sim N(0,1)$. N (2,1). Like Kim and Yu (2011), to generate the missing values we need to generate r_i, we proposed the following models.

M1 (Deleting from below): delete P% of the larger values

M2 (Deleting from above): delete P% of the smallest values

M3 (Linear nonignorable): $\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i]}{1 + \exp[f_0 + f_1 x_i + f_2 y_i]}$, where $(f_0, f_1, f_2) = (-3.4,1,1)$.

M4 (Quadratic in x, nonignorable): $\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 x^2{}_i + f_3 y_i]}{1 + \exp[f_0 + f_1 x_i + f_2 x^2{}_i + f_3 y_i]}$, where

$(f_0, f_1, f_2, f_3) = (-4.1,1,1,1)$

M5 (Quadratic in y, nonignorable): $\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i^2]}{1 + \exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i^2]}$, where

$(f_0, f_1, f_2, f_2) = (-10.1,1,1,1)$

M6 (Interaction x and y, nonignorable): $\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i x_i]}{1+\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i x_i]}$, where

$(f_0, f_1, f_2, f_2) = (-5.2,1,1,1)$

In all the above proposed missing mechanism models, the response rate is approximately 60%. We used a 20% follow-up rate for the semiparametric estimators. Note that models M3 and M4 satisfy the assumed response probability in (3.9). However, the missing mechanism M1, M2, M5, and M6 models do not meet the assumption in (3.9) and are included to test the robustness of the proposed methods suggested by Kim and Yu (2011). We compared our proposed estimators ($\hat{\mu}_Y^*$ and $\hat{\mu}_Y^{**}$) with the complete data estimator $\hat{\mu}_C = \frac{1}{n_1}\sum_{i=1}^{n} R_i y_i$ and the estimator using the ordinary multiple imputation $\hat{\mu}_{mi}$.

Table 4.1 shows that the semiparametric estimator $\hat{\mu}_Y^{**}$ has the smallest relative bias, then comes the two stage semiparametric ($\hat{\mu}_Y^*$) with the second smallest relative bias, compared to ordinary multiple imputation estimator ($\hat{\mu}_{mi}$). The worst is assuming ignorable missingness and using the complete data estimator ($\hat{\mu}_c$). Concerning MSE, our semiparametric estimator performs better than other estimators for all proposed models except that for model M6, the two stage semiparametric estimator has smaller MSE. However, the semiparametric approach using ITRA imputation is more robust than the two stage semiparametric approach.

Table 4.1

Monte Carlo estimation of the relative bias, the variance and mean square error

Missing Mechanism	Estimates	$\hat{\mu}_C$	$\hat{\mu}_{mi}$	$\hat{\mu}_Y^*$	$\hat{\mu}_Y^{**}$
M1	Relative Bias	-0.6470	-0.4760	-0.1426	0.0048
	Variance	0.0345	0.2406	0.0023	0.0203
	MSE	1.7091	1.1470	0.0836	0.0204

Table 4.1 (Continued)

Monte Carlo estimation of the relative bias, the variance and mean square error

Missing Mechanism	Estimates	$\hat{\mu}_C$	$\hat{\mu}_{mi}$	$\hat{\mu}_Y^*$	$\hat{\mu}_Y^{**}$
M2	Relative Bias	0.6548	0.4724	0.1393	-0.0017
	Variance	0.0329	0.1927	0.0023	0.0202
	MSE	1.7178	1.0855	0.0798	0.0202
M3	Relative Bias	-0.6192	-0.1461	0.0054	-0.0001
	Variance	0.0224	0.1677	0.0245	0.0212
	MSE	1.556	0.2532	0.0246	0.0212
M4	Relative Bias	-0.4511	-0.1241	-0.0058	-0.0026
	Variance	0.0222	0.1518	0.0242	0.0214
	MSE	0.8363	0.2133	0.0244	0.0214
M5	Relative Bias	-0.4820	-0.2578	-0.0173	0.0002
	Variance	0.0175	0.1501	0.0044	0.0217
	MSE	0.9469	0.4160	0.0055	0.0217
M6	Relative Bias	-0.4690	-0.1599	-0.0300	0.0016
	Variance	0.0203	0.1576	0.0043	0.0216
	MSE	0.9001	0.2599	0.0079	0.0217

Next, we are extending our method to do parameter estimation for linear models. The following chapter will introduce our method of importance resampling approach for parameter estimation, especially the linear regression slope. The slope of the regression line of non-observed data is obtained based on Importance resampling weights.

CHAPTER 5

Linear Model Parameter Estimation with Non-Ignorable Missing Data Using Importance Resampling Approach

This chapter discusses the methodology for parameter estimation in the linear model in non-ignorable missing data, which is slightly different from the previous chapter's method. We used the linear model parameter estimation via multiple imputations based on influential exponential tilted resampling. To understand this method better, we need to get insight into the impact of non-ignorable missing data in regression analysis.

5.1 Introduction

The causes of missingness in the data are plentiful. Some of which are the result of a particular study design or due to a chance. Certain variables are not collected or recorded from all the subjects. They often refuse to provide some answers. Sometimes, some of the information may be consciously omitted in the case of protecting confidentiality. These can result in different types of missing outcomes described in earlier chapters. In these scenarios, examining the relationship between two variables, called regression analysis, is arduous. The linear model we are investigating is

$$y_i = \beta_0 + \beta_1 x_i + e_i; i = 1, 2, ..., n,$$

where $e_i \sim N(0, \sigma_e^2)$ (iid).

Next, we see how this relationship is affected if data contains missing values. The example below shows how the relationship between two variables fluctuates if we have complete data versus the missing data. This example also illustrates the comparison of doing available case analysis against complete case analysis.

For this example, we simulated the data from a normal distribution, with one variable being independent and the other being the dependent. In addition to that, we generated some missing values based on non-ignorable missing patterns and ran a simple linear regression to assess the relationship between them. Table 5.1 shows how the slopes are affected by missing data and if they use general multiple imputations ignoring the underlying missing mechanism.

Table 5.1

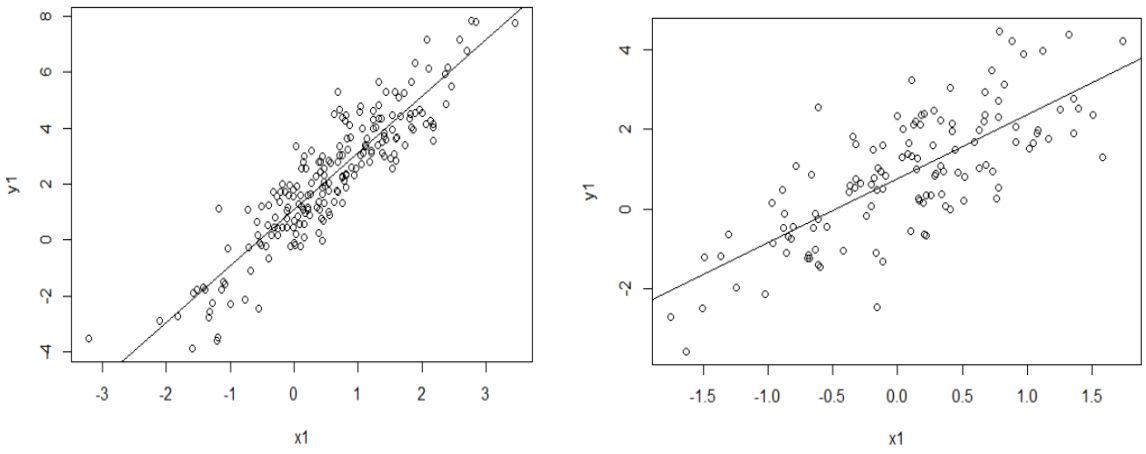
Comparison of complete case analysis vs. missing data vs. multiple imputations

	Complete Case	Missing Data	Multiple Imputation
Intercept	0.9768	0.7117	0.7309
Slope	1.9372	1.5900	1.5962

The below figure shows how the regression is affected by the missing data.

Figure 5.1

Complete case analysis vs. available case analysis



a) Complete data

b) With missing data

The coefficient of determination for the complete data is $R^2 = 0.8251$, while for the data with missing observation became $R^2 = 0.7702$.

The next section mentions our influential tilted resampling approach, which is used to do parameter estimation in the linear regression model.

5.2 Proposed Method

Multiple Imputation (MI) performs better under the assumption of MAR missingness in the case of parameter estimation, and it gives biased results in MNAR missing assumption. This method uses the influential exponential tilted resampling approach (ITRA) to obtain the desired parameter estimation. In this method, we propose an approach in which we are using the probabilities based on ITRA, which are highly influential in estimating the parameter in the presence of missing observations. Similar to the exponential tilting methods proposed by Kim and Yu (2011), assume that response probability has logistic regression model given by,

$$\pi(x_i, y_i) = \mathrm{P}(R_i = 1 \mid x_i, y_i) = \frac{\exp[g(x_i) + \phi r(y_i)]}{1 + \exp[g(x_i) + \phi r(y_i)]}. \tag{5.1}$$

As we discussed in chapter 4, ITRA states that the model for the non-responding part is an exponential tilting of the responding part. Thus, the conditional distribution of nonresponse is

$$f_0(y_i \mid x_i) = f_1(y_i \mid x_i) \frac{\exp(-\phi r(y_i))}{E(\exp(-\phi r(y_i)) \mid x_i, R_i = 1)}, \tag{5.2}$$

and *r(y)* to be the influential function for estimating the functional *T(F)* and parameter ϕ.

The influence function of a functional *T(F)*, where we suppose *Y* follows a c.d.f, say *F*, under some regularity conditions, defined using Gateaux derivative by

$$L(T(F)) = \underset{\varepsilon \to 0}{Lim}\left[\frac{T\{(1-\varepsilon)+\varepsilon\delta_y\} - T(F)}{\varepsilon}\right] \tag{5.3}$$

where $$\delta_y(u) = \begin{cases} 0 & \text{if } u < y \\ 1 & \text{if u} \geq \text{y.} \end{cases}$$

For estimating the slope of the regression line, the influence function estimate is

$$\hat{L}(T(F)) = \frac{x_i[y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i)]}{\frac{1}{n_1}\sum_{i=1}^{n_1} x_i^2}; i = 1, 2, ..., n_1 . \tag{5.4}$$

Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ obtained by using the regression analysis of the available case analysis with size n_1. In resampling the exponential tilting approach (Samawi et al., 1996; Samawi et al. 1998), we suggest that the distribution of the missing values defined as

$$f_\eta(y_i \mid x_i) = f_0(y_i \mid x_i) = f_1(y_i \mid x_i).\frac{\exp(\eta r(y_i \mid x_i))}{E[\exp(\eta r(y_i \mid x_i)) \mid R_i = 1]}, \tag{5.5}$$

where η ($\eta = -\phi$) is the tilting parameter, which determines the magnitude of the departure from the ignorability of the response mechanism. It penalizes observations that are more influential concerning the statistic under consideration and in the opposite direction of the possible MNAR missingness but rewards those in the same direction. In this case, the specified function is chosen as $r(y_i \mid x_i) = L(y_i \mid x_i)/n_1.\sigma_L$, where $\sigma_L = \sqrt{E(L(Y,x)^2)}$. Note that under MAR assumption η=0.

5.2.1 Semiparametric approach to ITRA and estimating the function T(F)

Kim and Yu (2011) suggested estimation of η using an independent survey or a validation sample, which is a subsample of the nonrespondents. Thus, in either case, the proposed estimator of the linear model parameters are as follows:

Define the functional *T(F)* as

$$T(F) = \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 \partial F(y \mid x), \tag{5.6}$$

where $F(y \mid x)$ is consisting of observed part, $F_1(y \mid x)$ and non-observed part $F_0(y \mid x)$, then

$$\begin{aligned} T(F) &= \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 [\pi(x,y)\partial F_1(y \mid x) + (1 - \pi(x,y))\partial F_0(y \mid x)] \\ &= \pi(x,y) \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 \partial F_1(y \mid x) + (1 - \pi(x,y)) \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 \partial F_0(y \mid x) \\ &= \pi(x,y) \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 \partial F_1(y \mid x) \\ &+ (1 - \pi(x,y)) \arg\min_{\beta_0,\beta_1} \int (y - \beta_0 - \beta_1 x)^2 \frac{\exp(\eta L(y_i \mid x_i)/n_1.\sigma_L)}{E[\exp(\eta L(y_i \mid x_i)/n_1.\sigma_L) \mid R_i = 1]} \partial F_1(y \mid x) \end{aligned} \tag{5.7}$$

In the above equation, using the indicator variable *R*, for observed and non-observed value, the above equation's empirical version would be as per below:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*)' &= \hat{\pi} \arg\min_{\beta_0,\beta_1} \sum_{i-1}^{n} R_i (y_i - \beta_0 - \beta_1 x_i)^2 \\ &+ (1 - \hat{\pi}) \arg\min_{\beta_0,\beta_1} \sum_{i-1}^{n} [R_i (y_i - \beta_0 - \beta_1 x_i)^2 + (1 - R_i)(y^*{}_i - \beta_0 - \beta_1 x_i)^2] \frac{\exp(\eta \hat{L}(y_i \mid x_i)/n_1.\hat{\sigma}_L)}{\sum_{i=1}^{n} \exp(\eta \hat{L}(y_i \mid x_i)/n_1.\hat{\sigma}_L))}. \end{aligned} \tag{5.8}$$

A naïve estimate of π is $\hat{\pi} = \frac{m}{n}$, where m is the number of all observed data and $y^*{}_i$ is the missing dependent variable. To find $y^*{}_i$, first, let

$$y_i = \beta_0 + \beta_1 x_i + e_i; i = 1, 2, ..., n, \text{ where } e_i \sim N(0, \sigma_e^2). \tag{5.9}$$

Under missing data (MNAR), the systematic part of the model is

$$\begin{aligned} E(Y \mid x) &= E[R.E(Y \mid R = 1, x) + (1 - R).E(Y \mid R = 0, x)] \\ &= \pi(x,y) E[Y \mid R = 1, x] + (1 - \pi(x,y)) E[Y \mid R = 0, x] \end{aligned} \tag{5.10}$$

If the unobserved part of our dependent variable (y_i^*), is called $m(x) = E[Y \mid R = 0, x]$.

Our goal is to find $m(x) = E[Y \mid R = 0, x]$ of non-observed data, based on the observed data by minimizing the following identity, $E[(Y - m(x))^2 \mid R = 0, x)]$ as follows:

$$\arg\min_{m(x)} \int (y - m(x))^2 \partial F_0(y \mid R = 1, x)$$

$$= \arg\min_{m(x)} \int (y_i - m(x))^2 \frac{\exp(\eta L(y_i \mid x_i) / n_1.\sigma_L)}{E[\exp(\eta L(y_i \mid x_i) / n_1.\sigma_L) \mid R_i = 1]} f_1(y \mid x) \partial y$$

$$\approx \frac{1}{n_1} \sum (y_i - m(x_i))^2 \frac{R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)}{\sum_{i=1}^{n} R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)} \qquad \text{(Empirical Distribution function)}$$

Taking derivative and solving for $m(x)$

$$-2 \frac{1}{n_1} \sum (y_i - \hat{m}(x_i)) \frac{R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)}{\sum_{i=1}^{n} R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)} = 0$$

$$-2 \frac{1}{n_1} \sum R_i (y_i - \hat{m}(x_i)) w(x_i, \eta) = 0$$

where $\hat{w}(x_i, \eta) = \frac{R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)}{\sum_{i=1}^{n} R_i \exp(\eta \hat{L}(y_i \mid x_i) / n_1.\hat{\sigma}_L)}$.

Then

$$y_i^* = \hat{m}(x_i) = \sum_{i=1}^{n} R_i y_i \hat{w}(x_i, \eta) \qquad (5.11)$$

In this approach, a dataset of size *n* to be used for the estimation. We consider the outcome of interest to be *Y*, depending on *X*, our independent variable, and the density function

being *f*. Presume that we have non-ignorable missing data of size (n_2<n). The following steps are used to perform influential exponential tilting resampling approaches. Most of the time, the preliminary information about the benchmark assumption is unknown, and, thus, η is the unknown. Therefore, follow-up (validation) data is used to obtain η value.

5.2.2 Finding the tuning (tilting) parameter η

We are finding η using the follow-up (validation) data obtained from a certain percent of the missing data.

1. Estimate the assumed distribution f_0 of the missing values as an influential exponential tilted distribution of the observed value (including the observed follow-up sample) as follows:

$$\hat{f}_0(y_i \mid x_i) = \hat{f}_1(y_i \mid x_i).\frac{R_i \exp\left(\hat{\eta}\hat{L}(y_i \mid x_i)/(m.\hat{\sigma}_L)\right)}{\sum_{i=1}^{n} R_i \exp\left(\hat{\eta}\hat{L}(y_i \mid x_i)/(m.\hat{\sigma}_L)\right)} = \hat{f}_1(y_i \mid x_i)w(x_i)$$, where $\hat{L}(y|x)$ is the estimated influential function.

2. Using the standard semiparametric resampling methodology, draw n_{22} resamples of size *m* each ($m = n_{11} + n_{21}$, total observed data), say $\mathbf{y}_i^{*\prime} = (y_{i1}^*, y_{i2}^*, ..., y_{im}^*); i = 1, 2, ..., n_{22}$.

3. From the i^{th} resampling (imputed subsample) $\mathbf{y}_i^{*\prime} = (y_{i1}^*, y_{i2}^*, ..., y_{im}^*)$, we estimate the i^{th} missing value as, $\hat{y}_i^{\,*} = \sum_{j=1}^{n} y_{ij}^* \hat{w}_j; i = 1, 2, .., n_{22}$,

where $\hat{w}_j = \dfrac{R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)}{\sum_{j=1}^{n} R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)}$

4. Estimate η using a follow-up (validation) data (a certain percentage of the missing values) by $\hat{\eta}$ where $\hat{\eta}$ is the solution of

$$\sum_{i=1}^{n}(1-r_i)I_i(y_i - \hat{y}_i^*(\eta)) = 0$$

where I_i is an indicator function that takes the value of 1 if then observed i^{th} sampling unit belongs to the follow-up.

5.2.3 Approximation of the variances

As in Kim and Yu (2011) let $\hat{\eta} \to \eta^0$

$$E(Y_i \mid R_i = 0, \eta^0, x) = p\lim_{n\to\infty} \hat{y}^*(\eta^0) = p\lim_{n\to\infty}\left(\frac{\sum_{j=1}^{n} R_j y_{ij}^* \exp(\eta^0 \hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L))}{\sum_{j=1}^{n} R_j \exp(\eta^0 \hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L))}\right).$$

$$= \left(\frac{E[RY\exp(\eta^0 L(Y \mid x)/(m.\sigma_L))]}{E[R\exp(\eta^0 L(Y \mid x)/(m.\hat{\sigma}_L))]}\right)$$

Then, if the model in (4.2) is true, this implies that $\eta^0 = \eta$ and then

$$E(Y_i \mid R_i = 0, \eta^0) = \left(\frac{E[RY\exp(\eta L(Y \mid x)/(m.\sigma_L))]}{E[R\exp(\eta L(Y \mid x)/(m.\sigma_L))]}\right) = \left(\frac{E[(1-R)Y \mid x]}{E[(1-R) \mid x]}\right) = E(Y \mid R_i = 0, x).$$

Now under some regularity conditions, we have

$$\hat{w}_j = \frac{R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)}{\sum_{=1}^{n} R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)} \xrightarrow{p} w_j = \frac{R_j \exp\left(\eta L(y_{ij}^* \mid x_{ij})/(m.\sigma_L)\right)}{\sum_{=1}^{n} R_j \exp\left(\eta L(y_{ij}^* \mid x_{ij})/(m.\sigma_L)\right)}$$

Also, Huber (1981) showed that $\sqrt{n}[T(F_n) - T(F)] = \sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} L(y_i \mid x_i)] + o_p(1)$, where F_n is the empirical function of F and $o_p(1)$ tend to 0 as $n \to \infty$. Now by Central Limit Theorem, we have

$$\sqrt{n}[T(F_n)-T(F)] \sim N(0,\sigma_L^2)$$

where $\sigma_L^2 = \int L^2(y \mid x)dF(y \mid x)$.

The variance of $\hat{\boldsymbol{\beta}}^*$ using the least square approach is given by

$$\text{var}(\hat{\boldsymbol{\beta}}^*) = \pi^2\sigma_e^2(X'X)^{-1} + (1-\pi)^2\sigma_{we}^2(X'WX)^{-1} \quad (5.12)$$

For estimating, the above variances, similar to Kim and Yu (2011), let

$\hat{q}_i^{**} = \hat{y}_i^{**} + [\frac{I_i(1-R_i)}{P(I_i=1 \mid R_i=0)} + R_i][Y_i - \hat{y}_i^{**}]$ then

$$\hat{\sigma}_{we}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{q}_i^{**2} - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{q}_i^{**}\right)^2, \quad (5.13)$$

$$\hat{\sigma}_e^2 = \frac{1}{m-2}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2; \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, i = 1,2,...,m \text{ (Based on the observed data of size } m) \quad (5.14)$$

and

$$\hat{W} = \begin{bmatrix} \hat{w}_1 & 0 & 0 & 0 \\ 0 & \hat{w}_2 & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \hat{w}_n \end{bmatrix}, \text{ where } \hat{w}_j = \frac{R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)}{\sum_{=1}^{n} R_j \exp\left(\hat{\eta}\hat{L}(y_{ij}^* \mid x_{ij})/(m.\hat{\sigma}_L)\right)}.$$

5.3 Simulation

To have insight into the theory we generated in this chapter, we conducted a simulation study, like in chapter 4. We generated 2000 samples of size 200 from the model $y_i = 1 + 2x_i + e_i$ where $x_i \sim (0.5,1)$ and $e_i \sim (0,1)$. As in the previous chapter, to generate missing values using

indicator variable R_i following models are proposed, which are similar to mean functional estimation and Kim and Yu's (2011) approach.

M1 (Deleting from below): delete P% of the larger values

M2 (Deleting from above): delete P% of the smallest values

M3 (Linear nonignorable):

$$\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i]}{1+\exp[f_0 + f_1 x_i + f_2 y_i]}, \text{where}\,(f_0, f_1, f_2) = (-3.4,1,1).$$

M4 (Quadratic in x, nonignorable):

$$\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 x^2{}_i + f_3 y_i]}{1+\exp[f_0 + f_1 x_i + f_2 x^2{}_i + f_3 y_i]}, \text{where}\ (f_0, f_1, f_2, f_3) = (-4.1,1,1,1)$$

M5 (Quadratic in y, nonignorable):

$$\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i^2]}{1+\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i^2]}, \text{where}\ (f_0, f_1, f_2, f_2) = (-10.1,1,1,1)$$

M6 (Interaction x and y, nonignorable):

$$\pi_i = \frac{\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i x_i]}{1+\exp[f_0 + f_1 x_i + f_2 y_i + f_3 y_i x_i]}, \text{where}\ (f_0, f_1, f_2, f_2) = (-5.2,1,1,1)$$

In all the above proposed missing mechanism models, the response rate is approximately 60%. Also, we used a 20% follow-up rate for the semiparametric estimators. We compared our proposed estimator ($\hat{\beta}_1^*$) with a complete data estimator $\hat{\beta}_{1,C}$ and the estimator using the ordinary multiple imputation $\hat{\beta}_{1,mi}$.

From simulation results in table 5.2, we see that our semiparametric estimator $\hat{\beta}_1^*$ has the closest estimate to our simulated dataset's real values for all the models where missingness is included in either right or left tails or randomly spread in the data. It has the smallest bias than

the ordinary multiple imputation estimator ($\hat{\beta}_{1,mi}$) and the available case analysis by ignoring missing data and using the complete data estimator ($\hat{\beta}_{1,C}$). Concerning MSE, the semiparametric estimator outperforms other estimators for these models as well. This analysis shows that our semiparametric approach using ITRA imputation is more robust than the other two approaches for various missingness scenarios.

Table 5.2

Monte Carlo estimation of the slope, bias, variance, and the mean square error (MSE)

		$\hat{\beta}_{1,C}$	$\hat{\beta}_{1,mi}$	$\hat{\beta}_1^*$
Model 1	Estimate	1.8499	1.7214	2.0206
	Bias	-0.1500	-0.2786	0.0206
	Variance	0.0059	0.0074	0.0083
	MSE	0.0285	0.0849	0.0087
Model 2	Estimate	1.8379	1.6789	1.9318
	Bias	-0.1620	-0.3210	-0.0682
	Variance	0.0002	0.0076	0.0001
	MSE	0.0265	0.1102	0.0048
Model 3	Estimate	1.9195	1.4057	2.1326
	Bias	-0.0679	-0.2935	-0.2260
	Variance	0.0010	0.0072	0.0198
	MSE	0.0165	0.3604	0.0375
Model 4	Estimate	1.8680	1.6388	2.0635
	Bias	-0.1319	-0.3612	0.0635
	Variance	0.0180	0.0069	0.0185
	MSE	0.0355	0.1374	0.0226
Model 5	Estimate	1.8446	1.7433	2.0649
	Bias	-0.1554	-0.2567	0.0649
	Variance	0.0038	0.0164	0.0112

Table 5.2 (Continued)

Monte Carlo estimation of the slope, bias, variance, and the mean square error (MSE)

		$\hat{\beta}_{1,C}$	$\hat{\beta}_{1,mi}$	$\hat{\beta}_1^*$
	MSE	0.0279	0.0823	0.0154
Model 6	Estimate	1.8999	1.5231	2.0099
	Bias	-0.1001	-0.4769	0.0099
	Variance	0.0151	0.0114	0.0189
	MSE	0.0251	0.2388	0.0190

CHAPTER 6

Application in Cobb County, GA, Women, Infants, and Children (WIC) Data

6.1 Introduction

This chapter discusses the application of our method to the real data provided by Bindele and Zhao (2018). The authors used this data from the statistical consulting center project at the Department of Mathematics and Statistics of the University of South Alabama. The data source was the Cobb County, GA, Women, Infants, and Children (WIC). The original data consists of six variables: neonatal baby weight (as the response variable), age, body mass index, smoking status, race, and Hispanic ethnicity (as predictors). The data contains a sample of 1499 observations, out of which 22.7% are missing (340 observations).

In this data, it has been suggested that mothers with premature babies may be less likely to disclose their baby's weight. Therefore, it is considered that missing baby weights is non-ignorable. We are interested in illustrating the relationship of mothers' BMI to their baby's weight as they are proportionally related to one another. As we mentioned, mothers with premature babies are less likely to report their baby's weight; thus, this missing data is not random (NMAR). The sample we used has 1499 observations; out of them, 22.7% were missing. Of the 340 non-respondents, 45% were randomly selected for follow-up samples, in which about 153 responded to the follow-up (Zhao, 2018).

6.2 Data Analysis & Results

6.2.1 Mean functional estimation

As mentioned above, out of 1499 observations with having 22.7% (340 observations) missing outcomes, we were able to obtain the follow-up data of 153 observations from the non-

respondents. For the available data, including the follow-up data, the mean of the baby weight came out to be 124.1113, and its standard error came out to be 0.3449. The empirical variance came out to be 0.1187. We applied our proposed method to estimate the mean of the baby weights. Using the proposed formula in chapter 4 for the 153 follow-up data, we estimated the tuning (tilting) parameter η by $\hat{\eta} = -47$. Using the available data, we applied multiple imputations and our semiparametric approach. We obtained the following results shown in table 6.1. The tuning parameter came out to be negative, showing lower values are missing, which are more influential in this case. As we can see, our method's estimation came out to have the lowest value compared to the other two cases, which is assumed to be closer to the real mean birth weights of the data.

Table 6.1.

Results of estimating the mean functional of neonatal baby weight with 95% confidence interval

Method	Estimate	Standard error	Lower bound	Upper bound
$\hat{\mu}_Y^{**}$	124.0003	0.3019	123.4085	124.5917
$\hat{\mu}_{mi}$	124.1710	0.3720	123.4419	124.9001
$\hat{\mu}_C$	124.2252	0.3441	123.5508	124.8996

6.2.2 Linear Model Parameter Estimation

A similar dataset was used to estimate the linear model parameter (slope, β_1) about Baby's weight to their mothers' BMI using our proposed approach of ITRA. We assumed that mothers with lower BMI would have a low birth weight and do not report their baby's weight. Thus, missingness in the baby's weight is considered a non-ignorable type. While considering mothers' BMI as an independent variable, we tried to find the regression coefficient for the baby's weight change. For the available data, including the follow-up data, the estimate came out

to be 0.13826, and its standard error came out to be 0.03729. The empirical variance came out to be 0.0014. We applied the proposed method to estimate the regression coefficient for the baby's weight, and the results are as per below. Linear model parameter value came out to be the lowest as this represents the most influential lowest weights that are missing from the data.

Table 6.2

Results of estimating the beta coefficient of neonatal baby weight in relation to mothers' BMI with a 95% confidence interval

Method	Estimate	Standard error	Lower bound	Upper bound
$\hat{\beta}_1^*$	0.12498	0.0311	0.0640	0.1859
$\hat{\beta}_{1,mi}$	0.1553	0.0385	0.1533	0.1572
$\hat{\beta}_{1,C}$	0.1383	0.0373	0.0651	0.2114

CHAPTER 7

FINAL REMARKS & CONCLUSION

7.1 Final Remarks

We tried to address handling the non-ignorable missing data by using importance resampling and doing parameter estimation through this dissertation. In the non-ignorable missing data, we used the nonparametric and the semiparametric influential tilting resampling approaches to estimate the mean. However, we proposed a semiparametric influential tilting resampling approach for the non-ignorable missing data to estimate the linear model parameter.

Our proposed methods performed very well with smaller variances, biases, and Mean Squared Errors (MSEs) for all six different mean estimation models. That included the non-ignorable missing data with right and left censoring linear, quadratic, and interaction models. Similar to estimating the linear model parameter, the beta coefficient, the proposed method worked very well when the data is missing in a non-ignorable manner, including quadratic and interaction models. This works well as the relationship between independent and dependent variables was identified using the most influential values for the missing observations contributing the most for the association between the two variables.

7.2 Conclusion

As the research field is vastly dependent on the data available to make any inference, incomplete data has the highest impact on making these interpretations. As mentioned in earlier chapters, it is challenging when such missing data is non-ignorable.

This dissertation proposed a method based on the influential tilting resampling approach to handle non-ignorable missing data for inference purposes. This approach is motivated by a brief use of the importance resampling proposed by Samawi et al. (1998) and the exponential

tilting for non-ignorable missing data by Kim & Yu (2011). This anticipated a method of influential exponential tilted resampling approach (ITRA) for desired parameter estimation. The multiple imputation (MI) methods work best for the data missing at random (MAR). However, this is a novel approach to using multiple imputations (MI) incorporating the bootstrap method of importance resampling, which imputes the most influential values for the missing observations obtained by exponential tilted resampling. Thus, this tactic includes the benefits of multiple imputations and bootstrapping services, which exclusively use the resample, that is of the most substantial influence to predict the missing ones.

In preparation for estimating the mean function, this method proved robust despite the data being non-ignorable missing, which is the most difficult situation to address in the analysis. In addition to that, finding an association between two variables and thus estimating the linear model parameter in non-ignorable missing data is the most challenging. Our proposed semiparametric approach of the influential exponential tilted resampling approach (ITRA) worked well for the described models.

7.3 Limitations & Future work

The problem of missing data is critical and one of the most challenging issues in public health data analysis and clinical trials. The well-known multiple imputation (MI) approach is a solution to such a situation, but it has a limitation in the non-ignorable missing data. Our proposed approach has the goodness of two mathematical methods, the multiple imputations (MI) and bootstrapping using the importance resampling techniques. However, this has limitations too.

Survey data and public health datasets do not need to have missing values only in the dependent variables, but missingness could be present in the independent variables as well.

There might be a combination of ignorable and non-ignorable missingness in different variables in the same datasets. Thus, there is a need to explore the robustness of the proposed methods in such situations.

While parameter estimation itself is challenging in non-ignorable missing data, it still needs more investigation to find an association between two variables in such scenarios. It is important to explore the correct relationship between two variables, which has been hindered due to missingness and worsened due to non-ignorable missingness.

REFERENCES

A. P. Dempster, N. M. (1977). Maximum Likelihood from incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.

Amanda N. Baraldi, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 5-37.

Benjamin M. Marlin, S. T. (2003). *Unsupervised Learning with Non-Ignorable Missing Data.*

Cheng, P. (1994). Nonparamteric estimation of mean functionals with data missing at random. *Journal of American Statistical Association*, 81-87.

D. V. Hinkely, S. S. (1989). Importance sampling and the nested bootstrap. *Biometrika*, 435-46.

Daniels, A. R. (2018). Bayesian appoaches for missing not at random outcome data: the role of identifying restrictions. *Statistical Science*, 198-213.

Eekhout, I. (n.d.). *https://www.iriseekhout.com/missing-data/missing-data-methods/deletion-methods/*. Retrieved from https://www.iriseekhout.com/: https://www.iriseekhout.com/missing-data/missing-data-methods/deletion-methods/

Enders, A. N. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 5-37.

Enders, A. N. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 5-37.

Enders, A. N. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 5-37.

Enders, J. L. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement . *Review of Educational Research*, 525-556.

Fitzmaurice, M. W. (2006). A simple imputation method for longitudinal studies with non ignorable non responses. *Biometrical Journal*, 302-318.

Geert Molenberghs, M. G. (2007). *Missing Data in Clinical Studies.* John Wiley & Sons, Ltd.

Glas, R. H. (2005). Modelling non-ignorable missing-datamechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, 1-17.

Grace-Martin, K. (2008). Missing Data Mechanisms: A primer. *The Analysis Factor*.

Graham, J. L. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods*, 147-177.

Hani M. Samawi, G. G. (1998). Power Estimation for Two-Sample Tests Using Importance and Antithetic Resampling. *Biometrical Journal*, 341-354.

Harel, O. (2008). Outfluence - The impact of missing values. *Model assisted statistics and applications*, 161-168.

Hinkley, A. C. (1997). *Bootstrap Methods and their Application* . Cambridge Series in Statistical and Probabilistic Mathematics.

Johns, M. V. (Sep 1988). Importance Sampling for Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, 709-714.

Joseph g. Ibrahin, m.-h. C. (2001). Missing Responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, 551-564.

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of Anesthesiology*, 402-406.

Kenward, J. C. (2017). *Guidlines for Handling Missing Data in Social Science Research.*

Matthew Blackwell, J. H. (2011). Multiple Overimputation: A unified Approach to Measurement Error and Missing Data. *Sociological Methods and Research*, 303-341.

Mohan, F. T. (2015). Graphical Representation of Missing Data Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 631-642.

Molenberghs, J. G. (2009). Missing data methods in longitudinal studies: a review. *NIH*, 1-43.

Niansheng Tang, P. Z. (2014). Empirical likelihood for estimating equations with nonignorably missing data. *Statistica Sinica*, 723-747.

Nisha C. Gottfredson, D. J. (2014). Modeling Change in the Presence of Nonrandomly missing data: evaluating a shared parameter mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, 196-209.

Raymond J. Carroll, D. R. (2006). Measurement error in nonlinear models. In D. R. Raymond J. Carroll, *Measurement error in nonlinear models.*

Rubin, R. J. (1987). *Statistical Analysis with Missing Data.* New York Wiley.

Samawi, H. M. (2003). Importance Resampling Using Logistic weights. *Dirasat*.

Shao, J. K. (2013). *Statistical methods for handling Incomplete Data.* Taylor & Fracis Group.

Shao, J. K. (2014). *Statistical methods for handling incomplete data.* Champan & Hall.

Stephen R Cole, H. C. (2006). Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, 1074-1081.

Stratton, O. H. (2009). Inferences on the Outfluence – How do Missing Values Impact Your Analysis? *Communications in Statistics - Theory and Methods*, 2884-2898.

Thompson, C. (2013). Additional examples of Missingness Mechanisma-foolow up to SON Brown Bag Presentation.

Walton, M. K. (2009). Adderssing and Advancing the Problem of Missing Data. *Journal of Biopharmaceutical Statistics*, 945-956.

Wong, M. A. (1987). The calculation of Posterior Distributionbt Data Augmentation. *Journal of American Statistical Association*, 528-540.

Xu Yan, S. L. (2009). Missing data handling methods in medical device clinical trials. *Journal of Biopharamaceutical Statistics*, 1085-1098.

Yu, J. K. (2011). A semi-parametric estimation of mean functionals with non-ignorable missing data. *Journal of American Statistical Association*, 157-165.

Yuan, Y. (2014). *Sensitivity analysis in multiple imputation for missing data.* SAS Institute Inc.

Zemel, B. M. (n.d.). *Unsupervised Learning with Non-Ignorable Missing Data.* Toronto,Canada: Department of Computer Science.

Zhao, H. F. (2018). Rank Based Estimating Equation with Non-ignorable missing Responses Via Empirical Likelihood. *Statistica Sinica*, 1787-1820.