IMPUTATION BASED ESTIMATORS FOR TREATMENT EFFECTS IN IMPACT EVALUATION FRAMEWORK

PAUL BRICE KENFAC DONGMEZO

DOCTOR OF PHILOSOPHY IN MATHEMATICS

(Statistics Option)

PAN AFRICAN UNIVERSITY

INSTITUTE FOR BASIC SCIENCES, TECHNOLOGY

AND INNOVATION

2018

Imputation Based Estimators for Treatment Effects in Impact Evaluation Framework

Paul Brice Kenfac Dongmezo

MS 400-0004/16

A Thesis submitted to Pan African University, Institute for Basic Sciences, Technology and Innovation in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mathematics (Statistics Option)

2018

DECLARATION

This thesis is my original work and no part of it has been presented for another degree award in any other university.

Signature:_____ Date:_____

Paul Brice Kenfac Dongmezo,

Reg Number: MS 400-0004/16.

The thesis has been submitted with our approval as university supervisors.

Signature:_____ Date:_____

Prof. Peter N. Mwita,

Department of Mathematics and Statistics,

Machakos University, Kenya.

Signature:_____ Date:_____

Dr. Ignace Roger Kamga Tchwaket,

Department of Economics,

Sub regional Institute of Statistics and Applied Economics, Cameroon.

DEDICATION

To my beloved wife Ariane

and my lovely children Jason, Chrys and Lyz.

ACKNOWLEDGMENTS

All belong to the Almighty God because of the Gift of Life.

I would like to express first my heartfelt gratitude to my supervisors Prof. Peter N. Mwita and Dr. Kamga Ignace, for their advices and encouragements in the preparation and completion of this thesis. In second hand, my deepest gratitude goes to the African Union Commission for the financial support, without which this study would not have been successful. I express a special acknowledgment to my Ministry, especially my Director Mme Siewe Ariane for the full support during this research. My sincere gratitude to my good friend Suleman Nasiru for his advices and positive criticisms.

I express my overwhelming gratitude to my wife Wakap Tchagang Ariane, and my children Jason, Chrys and Lyz for their prayers and unflinching support. You are my source of inspiration and countless happiness. Thanks to my parents Kenfac Joseph and Nguelemo Dongmo Rosalie, for their support and inspiration, you are part of the reasons of this success in my life. I also acknowledge to my brothers and sisters, Brigitte, Marius, Duchelle, Jerome and Natacha for their support and encouragements.

Finally, many thanks to my friends especially those from FRIMATHS for their support, my colleagues especially Jalira and all those who supported me closely or in prayers and my sisters in law Sidonie and Hadassa.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATE	Average Treatment Effect
ATT	Average Treatment Effect on the Treated
ATNT	Average Treatment Effect on Non-Treated
CDF	Cumulative Distribution Function
DDE	Difference in Difference Estimators
DID	Difference In Difference
EM	Expectation Maximization (Algorithm)
FWER	Family Wise Error Rate
FDR	False Discovery Rate
IB-AQTE	Imputation Based Average Quantile Treatment Effect
IB-ATE	Imputation Based Average Treatment Effect
IB-DTE	Imputation Based Distributional Treatment Effect
IB-MedQTE	Imputation Based Median Quantile Treatment Effect
IB-mQTE	Imputation Based modified Quantile Treatment Effect
IB-QTE	Imputation Based Quantile Treatment Effect
IB-TE	Imputation Based Treatment Effect
IB-TED	Imputation Based Treatment Effect on Distribution

IE	Impact Evaluation
i.i.d	Independent and Identically Distributed
IV	Instrumental Variable
K-S	Kolmogorov-Smirnov
LATE	Local Average Treatment Effect
LLM	Local Linear Matching
LQTE	Local Quantile Treatment Effect
MAR	Missing At Random
MCAR	Missing Completely At Random
MDRC	Manpower Demonstration Research Cooperation
ME	Matching Estimators
MI	Multiple Imputation
MICE	Multivariate Imputation by Chained Equations
ML	Maximum Likelihood
MTP	Multiple Testing Procedure
NMAR	Not Missing At Random
NN	Nearest Neighbor
NSW	National Supported Work (Demonstration)

OLS	Ordinary Least Square
PSM	Propensity Score Matching
QF	Quantile Function
QG	Quantile Group
QR	Quantile Regression
QTE	Quantile Treatment Effect
RCT	Randomized Control Trials
RDD	Regression Discontinuity Design
RE	Randomized Experiments
RMSE	Root Mean Squared Error
RV	Random Variable
SQUARE	Smooth Quantile Ratio
TrQTE	True Quantile Treatment Effect
WC	Wilcoxon rank test

ABSTRACT

The problem of counterfactual has been at the core of impact evaluation framework. Almost all existing methods aim to find the best way to estimate efficiently the counterfactual. A solution for estimation of counterfactual was proposed and investigated in this study. The objective of this research was to use classical imputation methods to estimate counterfactual, then derive treatment effect estimators from the data sets completed using the basic definition of treatment effect described in Rubin framework. The estimators obtained, called Imputation Based Treatment Effects estimators, were theoretically unbiased, convergent and consistent. Using simulation, the results revealed that they were asymptotically unbiased and convergent as well. Results from the data application showed that they performed as well as the classic estimators and sometimes better in cases of shortage in data. To conclude the research, a hypothesis testing procedure was proposed to test the significance of the treatment effect. The results showed that the three approaches proposed were efficient, and could detect any change between two distributions, even slight changes.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Causal inference is one of the most interesting branch of statistics. Quite a number of research questions are causal in nature, therefore, to answer them one needs to learn more about causal inference. It studies the relationship between more than two statistical variables and draws conclusions about causal connection between the variables. It has been used in different fields such as health, economics, psychology, education and so on. Some of the questions which can be addressed are: what is the efficacy of a drug on a given population? And what is the effect of a training on a population individual revenue? To answer these questions, the statistician uses many different methods and one of them is Impact Evaluation (IE). In recent literature, especially in economics and social statistics, policy makers are more interested and strict on evaluation aspect since it is the only scientific and relevant quantitative method to determine whether a program has really had an effect or not.

Early works on IE were done by Rubin in the 1970's. He developed what we call today "Rubin Framework" in which he explained how to get the effect of a specific action on a unit (Individual, County, Household etc.). The vocabulary of his approach comes from medical experiments. Here, what he calls "treatment"

can be seen as administration of a drug, a training, a given policy or something else dividing the sample into two groups. We are interested in evaluation of a treatment denoted by T on another variable called potential outcome denoted by Y in a given population. For a treated unit i in the population, our variables take values $T_i = 1$ and $Y_i = Y_{i1}$; for a non-treated unit in the population, $T_i = 0$ and $Y_i = Y_{i0}$. Therefore, for a given unit i, Rubin (1974) defined the causal effect by: $\Delta_i = Y_{1i} - Y_{0i}$. The main problem with causal inference is that, in most cases, we cannot observe both Y_{i0} and Y_{i1} for a given unit at the same time hence, this problem can be seen as a problem of "missing data" well known and addressed in a classical way in statistic literature. For example, if a unit is treated, we will observe Y_{i1} but not Y_{i0} hence, the second quantity can be considered as missing data, although if unit is not treated, we will observe Y_{i0} but not Y_{i1}. Like previously, the second quantity can be considered as missing data.

To determine the effect of the treatment is simply to estimate and impute those missing data then compute the difference. In the literature, the most used parameter of interest to evaluate the treatment on a population is the Average Treatment Effect on the Treated (ATT). That quantity is given by: $ATT = E(Y_{i1} - Y_{i0} \,|T_i = 1)$. At the same time, one may be interested in the Average Treatment Effect (ATE) on the population given by : $ATE = E(Y_{i1} - Y_{i0})$ or on the Average Treatment effect on Non-Treated ($ATNT$) given by: $ATNT = E(Y_{i1} - Y_{i0} \,|T_i = 0)$. Since this framework was designed in the 1970's by Rubin, many methods have been developed to estimate those parameters and applied to

answer some of the questions previously asked.

Among different existing methods, we have Randomized Experiments (RE), Difference in Difference Estimators (DDE), Matching Estimators (ME), Instrumental Variables (IV) and Regression Discontinuity Design (RDD) for the most popular. As examples of their application, Angrist et al. (2002) used RE to measure the impacts of vouchers for private schooling in Colombia on school results. Chaudhury and Parajuli (2006) assessed the impact of female school stipend on public school enrollment in Punjab in India using DDE. Jalan and Ravallion (2003) used ME to estimate the benefits of an antipoverty program on a given population. To evaluate the effect of delayed primary school enrolment and the role of early childhood nutrition on children, Glewwe and Jacoby (1995) used IV method. Finally, Ravallion (2007) used RDD to evaluate a set of Antipoverty programs in some developing countries. Most of these methods focused on a basic assumption link on how treated and control group are chosen and on homogeneity of effects across all individuals.

The effect of a treatment is an individual thing, but taking the parameters ATT and ATE as the treatment effect on a random unit in the treated group or in the population, is a limited and biased way of estimating the treatment effect. That assumption implies that the effect of the treatment is homogenous in the population. Unfortunately, it is not usually the case, impact varies across units. One of the drawbacks of average treatment estimators is the fact that if there

is heterogeneity in the effects across units, taking an average effect as effect on a random individual is biased since the mean is always sensitive to extreme values. In fact, effect of a treatment on units is always different given observed and unobserved characteristics of units. In the recent literature, researchers have started studying the distributional effect of a given treatment on a population to fill the weaknesses of average treatment estimators. Doksum (1974) introduced what he called a shift function between two distributions to take into account differences that can appear across the distribution of the potential outcome when applying a treatment. Later on, Heckman et al. (1997b) introduced the Quantile Treatment Effect (QTE) which is supposed to be for a given quantile, the difference in outcome across control and treated groups. That parameter is given by: $QTE(\tau) = Y_1(\tau) - Y_0(\tau),\ \tau \in (0,1)$. Estimation of this quantity is one of the recent development in the field of IE. Researchers like Firpo (2007), Yu (2014), Venturini et al. (2015) and many others have proposed different types of estimators both non-parametric and semiparametric. Trying to estimate QTE is a solution to the homogeneity bias previously explained.

Another solution that can be investigated is imputation methods as presented by Hahn (1998). Few researchers have tried to approach the problem of IE as a problem of missing data as highlighted earlier in this section. They proposed some imputation methods suitable for IE framework given some assumptions on data but still, the problem is not really addressed as a missing data problem since the main imputation methods are not associated to those methods. In addition

to the problem of homogeneity that can be addressed in the literature, there are fewer studies done on hypothesis testing procedures around IE framework.

1.2 Statement of the Problem

Basically, the problem of Impact Evaluation (IE) is how to estimate efficiently the counterfactual then derive the estimators of treatment effect. The counterfactual is what would have happened in presence of treatment if the unit is not treated or what would have happened in the absence of treatment if the unit is treated. The classical IE methods estimate the counterfactual based on many assumptions like homogeneity and rank preservation assumption. Homogeneity means that the effect is the same for all units. Consequently, they can not identify the effect of a single unit. Assuming homogeneity also expose the average effect computed to high effect values, the sensitivity to atypical effects is higher. Another assumption usually taken is rank preservation assumption stating that if a unit is in a given quantile, after the treatment it will remain in the same quantile. It is a strong assumption given that units are different and can react differently to a treatment. Other drawback of classical methods is that they do not provide a tool to test if the effect computed is significant or not.

There is a need to develop a new approach of IE problems that can take less assumption and produce better results. The new approach proposed must be based on less assumptions, produce estimators with a smaller bias and give more

possibility in terms of assessing the effects for single units and verification of the significance of the study. A solution to address these needs seems to be an estimation of counterfactual using imputation methods. The only assumption that will be taken is to consider counterfactual as a missing value.

1.3 Significance of the Study

This study is important at two levels. First, many researchers (Hahn, 1998; Gabriel, 2011; Dominici et al., 2005) highlighted the fact that causal inference, in particular IE, can be seen as a problem of missing data. However, few have taken it and addressed it as a full missing data problem. Therefore, this research considered that idea and developed it as a full statistical problem of missing values. It will be interesting to use imputation to estimate counterfactual before applying or developing new estimators based on imputation. If we look at works of Hahn (1998), Gabriel (2011) and Dominici et al. (2005), they used imputation methods as a means to reach their goals without focusing on imputation as the main solution of the problem. In this study, the main focus is to complete the database with robust estimators of the missing potential outcome and then derive estimators of average and quantile treatment effects and their properties.

Secondly, theoretically speaking, undertaking this study can contribute to improve literature on hypothesis testing in the field of IE which is actually growing. In fact, few tests have been proposed specifically in this domain, see for exam-

ple Fisher (1935) and Crump et al. (2008). Also the test proposed here can be extended to the comparison of distributions where the difference is not easily identified. We would therefore, like to enrich the literature in this field.

1.4 Objective of the Study

1.4.1 General Objective

The main objective of this study is to construct a new class of treatment effects estimators called Imputation Based Treatment Effect Estimators ($IB - TE$), investigate their properties and to test the significance of effects computed with those estimators.

1.4.2 Specific Objectives

More specifically, the aim is:

1. To develop an Average Treatment Effect estimator based on Imputation methods and investigate its properties;

2. To derive the Quantile Treatment Effect estimator based on imputation method and investigate its properties;

3. To test the hypothesis "No effect" for presence and significance of Treatment Effects on a population using distribution reconstructed with imputation;

4. To apply the estimators to data: Lalonde data to assess the impact of a training program on the revenue.

1.5 Scope of the Study

In this research, the terms effects and impacts mean the same thing. As highlighted in the first section of this chapter, a treatment is understood as a program, a subsidy, a training or anything else supposed to induce a change on a section of a population. This study focused on single treatment at a time and considered the effect on a continuous potential outcome. As a consequence, distributions of potential outcome are continuous and imputation methods applied here are for continuous distributions.

The major assumption in this study is that counterfactual is a missing data, not a quantity which is not observable as it is in reality. We assume that it is always possible to have a set of covariates X characterising the unit in the population which is helpful in estimation of counterfactual using imputation methods. Another assumption is that from the assignment to treatment we can deduce the missingness mechanism. Under those assumptions and considerations, the study was done properly.

1.6 Thesis Outline

The thesis consists of six chapters including this one as chapter 1. Chapter 2 is the literature review and a review of basic statistics concepts and methods related to this research. Chapter 3 introduces the notion of Imputation Based Treatment Effect Estimators, develops Imputation Based Average Treatment Effect Estimators and studies their properties. Chapter 4 presents Distributional Treatment Effects estimators based on imputation methods. Chapter 5 develops a hypothesis testing procedure based on distribution of a reconstructed potential outcome and Multiple Testing Procedure to test the hypothesis of "No Effects" in the framework of impact evaluation. The concluding remarks and recommendations are presented in chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents what is done in the literature on treatment effects based on imputation methods. It starts with a brief recall of some basic concepts from statistical framework used in this research such as estimators, quantiles and test of hypothesis. Then impact evaluation methods and imputation methods are presented in their simplest form.

2.2 Basic Concepts and Methods

2.2.1 Estimators

Definition 2.1. Let Y be a random variable (r.v) whose distribution depend on a parameter $\theta \in \Theta$ with $\Theta \subseteq \mathbb{R}$ or $\Theta \subseteq \mathbb{R}^p$. Let $Y_\bullet = (Y_1, Y_2, ..., Y_n)$ be a $n-sample$ of Y meaning n random variables independent and identically distributed (i.i.d) with the same density as Y.

A **statistic** is any function T defined as follows:

$$T : \mathbb{R}^n \to \mathbb{R} \ (or \ \mathbb{R}^p)$$

$$y_\bullet = (y_1, ..., y_n) \mapsto T(y_\bullet) = T(y_1, ..., y_n) = t$$

such that $\hat{\theta}_n = T(Y_1, Y_2, ..., Y_n)$ is well defined for every sample size n.

If $\hat{\theta}_n$ is used to estimate one or many parameters of the unknown law of Y, then $\hat{\theta}_n$ is called an **estimator** and its realisations are **estimates**.

Once an estimator is constructed, its properties has to be studied. Among the most important properties there are **Unbiasedness**, **Convergence** and **Consistency**. The definitions and theorems on those properties can be obtain from any statistic book.

2.2.2 Quantiles

A quantile can be defined in different ways, let's give some of the definitions used in literature: theoretical and empirical definitions.

Definition 2.2. Let Y be a random variable, and $Y_\bullet = (Y_1, Y_2, ..., Y_n)$ and $n-$ *sample* i.i.d of the r.v Y. Let Also $Y_{(\bullet)} = (Y_{(1)}, Y_{(2)}, ..., Y_{(n)})$ be the ranked sample associated to the $n-$ *sample* $Y_\bullet$. Let F be the Cumulative Distributive Function (CDF) of the random variable Y. For $p \in]0, 1[$, the theoretical quantile of order p is defined by:

1. The inverse of the CDF taken for the value p: $Q(p) = F^{-1}(p)$;

2. The number $y \in \mathbb{R}$ such that $F(p) = y$.

Definition 2.3. Let Y be a real valued random variable, let $y_\bullet = (y_1, y_2, ..., y_n)$ be a realisation of a $n-$ *sample* of that random variable and let $p \in (0, 1)$ be a probability. Then the p^{th} **empirical quantile** is the smallest number $y_i \in \mathbb{R}$ such that $P(Y > y_i) \leqslant 1 - p$.

More practically, the p^{th} empirical quantile is:

1. the number $y \in \mathbb{R}$ such that $P(Y > y) = 1 - p$;

2. the number $y \in \mathbb{R}$ such that $P(Y \leqslant y) = p$;

3. the number $y \in \mathbb{R}$ satisfying the following equality

$$p \simeq \frac{\#\{y_i : y_i \leqslant y\}}{n}$$

   # being the symbol of the number of elements of the set.

More materials on quantiles such as properties and convergence theorems can be obtain in most of statistics books in case of more information is needed on it.

2.2.3 Hypothesis Testing

Let Y be a random variable and $\mathfrak{L}$ a set of probability law. We assume that there exist two subset of $\mathfrak{L}$, $\mathfrak{L}_0$ and $\mathfrak{L}_1$ such that $\mathfrak{L} = \mathfrak{L}_0 \bigcup \mathfrak{L}_1$ and $\mathfrak{L}_0 \bigcap \mathfrak{L}_1 = \emptyset$.

Definition 2.4. A statistical hypothesis is just the assumption of $\mathcal{L}(Y) \in \mathfrak{L}$ and denoted by $\mathcal{H} = \{\mathcal{L}(Y) \in \mathfrak{L}\}$.

The alternative is a definition of two statistical hypothesis like $\mathcal{H}_0 = \{\mathcal{L}(Y) \in \mathfrak{L}_0\}$ against $\mathcal{H}_1 = \{\mathcal{L}(Y) \in \mathfrak{L}_1\}$. The first hypothesis is called the null hypothesis and the second one the alternative hypothesis.

As example, if $\mathfrak{L} = \{\mathcal{P}(\lambda),\ \lambda \in \mathbb{R}_+\}$ is the set of all Poisson law, then saying that the law of Y is a Poisson law is a statistical hypothesis formulated as follows: $\mathcal{H} = \{\mathcal{L}(Y) \in \mathfrak{L}\}$.

Definition 2.5. A test statistic or an hypothesis test of two alternatives $\mathcal{H}_0$ against $\mathcal{H}_1$ is an application defined as follows: $\psi : \mathbb{R}^n \longrightarrow \{0,\ 1\}$ such that $\psi(Y_\bullet)$ is a random variable such that for all realisation $Y_\bullet = y_\bullet$, there is a decision process given by:

$$\begin{cases} If \quad \psi(y_\bullet) = 0 \quad then \quad \mathcal{H}_0 \ is \ True, \\ \\ If \quad \psi(y_\bullet) = 1 \quad then \quad \mathcal{H}_1 \ is \ True. \end{cases}$$

To simplify notations, in case there is a parameter θ to test and they are two disjoint alternatives θ_0 and θ_1, the formulation of the simplest test of H_0 against H_1 is the following:

$$\begin{cases} H_0 : \theta = \theta_0 \\ \\ H_1 : \theta = \theta_1 \end{cases} \tag{2.1}$$

From this general form, many different hypotheses can be tested: $H_0 = \{\theta = \theta_0\}$ against $H_1 = \{\theta \neq \theta_0\}$; $H_0 = \{\theta \leq \theta_0\}$ against $H_1 = \{\theta < \theta_0\}$ and many others.

In case the problem is to test jointly more than one hypothesis ($m \in \mathbb{N}$ parameters) against more than one alternatives on the same parameters, the appropriate test is the Multiple Testing Procedure. Under definitions and assumptions made in the previous subsection, the formulation of a multiple test using the structure (2.1) is as follows:

$$\begin{cases} H_0 : \{H_{0i}(\theta_i),\ i = 1, ..., m\} \\ \\ H_1 : \{H_{1i}(\theta_i),\ i = 1, ..., m\} \end{cases} \tag{2.2}$$

The hypotheses $\{H_{0i}(\theta_i)\}_{i=1}^m$ are tested simultaneously to be true against at least one of the alternative $\{H_{1i}(\theta_i)\}_{i=1}^m$ being false. In case of acceptance of H_0, it is clear that all the parameters verify their respective null hypothesis, nothing more is required. In case of rejection of H_0, meaning that at least one of null hypothesis is rejected, one may want to know which of the parameters does not satisfy its corresponding null hypothesis. Assuming that each individual hypothesis testing procedure is available, the problem is how to combine all of them in a single test procedure with the corresponding critical values. Specific quantities are used to analyse the results of such tests.

2.2.4 Impact Evaluation Methods

Impact evaluation methods are classified in 5 main groups depending on the assignment process and the structure of the data need to perform them. The five groups are:

1. **Randomized Experiment or Randomization (RE)**: In this method, the selection process is a random experiment. Given a population, the selection of treated units is a sampling experiment with same probability to be selected so that the expected outcome with and without the treatment will be the same. By doing that we ensure that the groups formed are representative of the population. The estimators are obtained by just comparing the relevant quantities in the two group of treated and non treated.

2. **Difference in Difference Estimators (DID)**: Difference in Difference

estimator's method assumes that the potential outcome is available before and after the treatment. In addition to that, there is an unobserved heterogeneity among units in the sample but those factors are time invariant meaning that before and after the treatment, these factors are the same and can be removed. This method estimates the average impact of the treatment as follow:

$$DDE = E(Y_1 - Y_0 \,|T = 1\,) - E(Y_1 - Y_0 \,|T = 0\,). \tag{2.3}$$

It is the average difference of outcome between the two groups after the treatment minus the same difference between the two groups before the treatment so that the bias which is time invariant and additive will disappear.

3. **Matching Estimators**: The last two methods presented in their basic form did not need additional information on units. This method requires more information about units. Matching methods construct a control group which is comparable to treatment group according to some observable characteristics that are present in the data base (covariates X). The basic matching methods uses Mahalanobis distance (Introduced in 1936) to check whether two units are close to each other or not and match them. The main assumption here is what is called in literature - unconfoundedness, meaning that the given covariates taking the treatment or not is independent of the

potential outcome. It can be written as follow

$$Y_{i,1}, Y_{i,0} \perp T_i \mid X_i \tag{2.4}$$

For example, if age is the unique covariate, the outcome which can be the grades in school is independent to the treatment given the age. The average treatment estimator will be the average potential outcome in the treatment group for a specific age minus the average outcome in the control group for the same age then we take the average across different ages.

$$ATE = \int ATE(x)dx, \tag{2.5}$$

where $ATE(x) = E(Y_1 \mid X = x, T = 1) - E(Y_0 \mid X = x, T = 0)$. More developments on this method were done later on, more materials in connection with that can be obtain from Imbens and Wooldridge (2009).

4. **Instrumental Variables (IV) Estimators**: IV method allows for endogeneity (correlation between treatment and errors) in treatment assignment. This method involves finding a variable called Instrument which is highly correlated with treatment assignment but not correlated to unobserved characteristics that affect outcome. With good instrument, IV estimator is unbiased but weak instruments can worsen the bias. The estimators are constructed using a two stage linear regression.

5. **Regression Discontinuity Design (RDD)**: RDD methods are used when

assignment to treatment depends on another variable denoted S (which can be age, income, grades etc.). One may decide a threshold under which a person is treated and not treated above. The method exploits the fact that around the threshold S^* units are comparable therefore the estimator is computed using a difference in the two groups around the threshold.

More developments and material on these methods can be obtained from any impact evaluation book.

2.2.5 Imputation Methods

Imputation is one of the oldest methods in statistics. In this study many imputation methods are used to reconstruct data. Most of them are describe in statistics books and papers. They are classified in two groups:

1. **Imputation Methods without Random Variation**: characterized by the fact that running the same method on the same sample many times will always produce the same imputed values for units missing with the same characteristics (no matter the user). Some of them are as follows:

    - Mean Imputation and Conditional Mean Imputation;
    - Nearest Neighbours Imputation;
    - Last value carried forward;
    - Regression to perform deterministic Imputation;
    - Simple random Imputation (Hot deck imputation).

2. **Imputation Methods with Random Variation**: This group of methods is characterized by the fact that it allows for randomness in the prediction of missing values. Running this method n times in a given sample may produce n different values for a single imputation. Some of them are:

   - Regression to perform random Imputation;
   - Multiple Imputation (MI);
   - Maximum likelihood Imputation (ML).

2.3 ATE Based on Imputation in Literature

Early works on IE using imputation method were done by Hahn (1998), Dominici et al. (2005) and Gabriel (2011). In addition to that, the classic PSM method can be seen as method using imputation. Most of these methods developed use imputation to obtain an average effect.

2.3.1 Regression Imputation seen by Hahn (1998)

Given a data set (Y, T, X), following the work done by Hahn (1998), under the assumption of unconfoundedness, Hahn defines the quantity:

$$E(T_iY_i/X_i) = E(T_iY_{i1}/X_i) = E(T_i/X_i)E(Y_{i1}/X_i) = E(T_i/X_i)\beta_1(X_i) \qquad (2.6)$$

and at the same time the quantity:

$$\begin{aligned} E((1-T_i)Y_i/X_i) &= E((1-T_i)Y_{i0}/X_i) = E((1-T_i)/X_i)E(Y_{i0}/X_i) \\ &= E((1-T_i)/X_i)\beta_0(X_i) \end{aligned} \tag{2.7}$$

From equations (2.6) and (2.7), it follows that for the same unit we can get :

$$\hat{Y}_{i1} = \hat{\beta}_1(X_i) = \frac{\hat{E}(T_iY_i/X_i)}{\hat{E}(T_i/X_i)} \quad and \quad \hat{Y}_{i0} = \hat{\beta}_0(X_i) = \frac{\hat{E}((1-T_i)Y_i/X_i)}{1-\hat{E}(T_i/X_i)} \tag{2.8}$$

The quantity $\hat{\beta}_1(X_i)$ is an estimator of the value of potential outcome that unit i would have taken if it was treated (in this case unit i is not treated). Likewise, $\hat{\beta}_0(X_i)$ is an estimator of the value of potential outcome that unit i would have taken if it was not treated, in absence of treatment on it (in this case unit i is treated). Therefore, under treatment in the population: $\hat{Y}_{i1} = T_iY_i+(1-T_i)\hat{\beta}_1(X_i)$ and under control $\hat{Y}_{i0} = (1-T_i)Y_i + T_i\hat{\beta}_0(X_i)$.

Now, estimation of the mean equation $\hat{\beta}_1(X_i)$ and $\hat{\beta}_0(X_i)$ is the choice of the statistician. Among methods than can be used, there are OLS regression, Non-parametric regression, or even simple sample mean or any other method linked to regression methods. At the end of imputation, a completed data set is obtained from which estimations can be done. Hahn (1998) proposed a nonparametric method for imputation. In this research, a parametric imputation (OLS regression) and a quantile regression imputation to take into account of the distribution of the potential output and try to keep rank or quantile are proposed.

2.3.2 PSM Estimators seen as IB-ATE

The matching imputation is based on the calculation of two propensity score functions. The first one is computed in the control group $\hat{p}_{i0}$ and the second one $\hat{p}_{i1}$ in the treatment group (Rosenbaum and Rubin, 1983). Now the matching exercise shall be done in each group. In the control group as well as in the treated group, the values considered as missing values shall be imputed by the matching algorithm. A unit treated will look for a unit or a group of unit in the control group which have almost the same score to be matched with. Its value of potential outcome if it was not treated will be taken from that matching unit. The same exercise is done for non treated units. Among the different types of matching, there is one-to-one matching (seminal idea of this research), nearest-neighbor (NN) matching, caliper and radius matching, stratification and interval matching, kernel matching and finally local linear matching (LLM).

For example, the Kernel Matching imputation is given by:

$$\hat{Y}_{i0} = \frac{\sum_{k=1}^{N_0^*} K\left(\frac{|p_i - p_k|}{k}\right) Y_{k0}^*}{\sum_{k=1}^{N_0^*} K\left(\frac{|p_i - p_k|}{k}\right)} \quad and \quad \hat{Y}_{i1} = \frac{\sum_{k=1}^{N_1^*} K\left(\frac{|p_i - p_k|}{k}\right) Y_{k1}^*}{\sum_{k=1}^{N_1^*} K\left(\frac{|p_i - p_k|}{k}\right)} \tag{2.9}$$

The quantities N_0^* and N_1^* are the respective numbers of control units and treated units after a given number of imputations. The two numbers vary and express the fact that imputed units are used in the process of imputation. The quantity Y_{kj}^* is the potential outcome of individual k in the group j, thus Y_{kj}^* can be a non-imputed value or an imputed value depending on the number of imputations

done. It is a kind of iterative imputation method.

2.3.3 Smooth Quantile Ratio (SQUARE) Imputation

The intuitive idea behind the SQUARE imputation is to replace empirical quantiles by theoretical quantiles using some assumption on the structure of data and/or the distribution of one of the groups. The SQUARE estimator was first developed by Dominici et al. (2005). The method was used to propose an estimator of the mean difference between two highly skewed distributions. It will be used in this study as a quantile imputation method for estimating the distributional impact of a treatment.

Considering the general form developed by Venturini et al. (2015), they defined:

$$h\left(\frac{Q_1(\tau)}{Q_0(\tau)}\right) = S(\tau,\lambda) = X(\tau,\lambda)\beta, \tag{2.10}$$

with h a chosen function according to the structure of data, S and X are smoothed regression function and λ is the smoothing parameter. If we replace the quantile Q_1 and Q_0 by the empirical quantiles represented by the ordered data $Y_{(i)1}$ in the treatment group and $Y_{(i)0}$ in the control group we get:

$$Y^*_{(i)1} = Y_{(i)0}h^{-1}\left(X(\tau_{(i)},\lambda)\hat{\beta}\right) \quad and \quad Y^*_{(i)0} = \left(Y_{(i)1}\right)^{-1}h^{-1}\left(X(\tau_{(i)},\lambda)\hat{\beta}\right) \tag{2.11}$$

By doing that, the method replaces or completes the sample by smoothed quantile estimation of missing values. Therefore, from that sample, one can compute what-

ever estimator we want including QTE estimators. Unfortunately, this method does not suit the objective of imputation here which is point imputation not a quantile imputation.

2.4 Distributional Treatment Effects in Literature

2.4.1 Drawbacks of ATE

The aim of studying the ways a treatment is distributed in the population comes from the drawbacks of ATE. Compute ATE assumes that it is the treatment effect and as such, means that the effect of the program is homogenous, therefore the impact is made constant across all units. It is a very strong assumption that cannot be possible in real life because units are always different. Consequently, there is heterogeneity in the effects of the program meaning that the average effect of the program can be low or higher because of some extreme values of effects. In fact, as units are different or even if they are almost the same according to the covariates collected, they could not react to the treatment the same way. Some may have a greater impact and others a lower impact. Consequently, the average impact can be affected by some units considered atypical.

To fill the drawback of ATE, to respond to the needs of policy makers and to give more precision about the effect of a given treatment, researchers have started

to work on methods that can inform on how treatment has affected a whole population or a specific sub-group of the targeted population. These methods are called according to Imbens and Wooldridge (2009) "quantile and distributional effects methods". These methods aim to come up with the effect of the treatment in sub-group of the targeted population. For a specific section of the distribution of the potential outcome, or more to give the effect on the whole distribution of potential outcome.

2.4.2 Treatment effect and distribution of potential outcome

First works done in this area of research in statistics were in the seventies. Doksum (1974) and Lehmann (1974) were the first to define the quantile treatment effect. Later on, other researchers like Bitler et al. (2006) estimated the quantile treatment effects in a randomized evaluation of a job training program. Firpo (2007) developed methods for estimating QTE in observational studies given unconfoundedness. Abadie et al. (2002) and Chernozhukov and Hansen (2006) studied quantile treatment effects in instrumental variables settings and many others. All these works were done in the classical context of impact evaluation that is known in the literature.

In the literature, the methods are classified into two groups:

1. Methods using **joint distribution of potential outcome** to estimate dis-

tributional effects. The pioneers of this approach are Hoeffding (1940) and Frechet (1951) with their work on probability distributions. Then, using their results in the aim of assessing distributional impact of a program, Heckman et al. (1997a), Heckman et al. (1993) and Heckman and Smith (1998) found the joint distribution of (Y_0, Y_1) using the marginal distribution of Y_0 and Y_1 in a randomized control experiment, a practical case. Later on, researchers such as Aakvik et al. (1999) used joint distribution to identify treatment effects of discrete outcome assuming heterogeneity in the effects. An improvement of that work can be read in Aakvik et al. (2005). Carneiro et al. (2001) and Carneiro et al. (2003) in their research proposed an approach to bound the distribution using the method common in factor analysis but applied to model counterfactual distributions.

2. Methods using **marginal distribution of potential outcome** to estimate distributional effects. Lehmann (1974) and Doksum (1974) were the first to define the quantile treatment effect as the difference between the same quantile in the distribution after and before the treatment under rank preservation assumption. Abadie (2002) used a new instrumental variable approach that measure program impact on quantiles of the distribution of potential outcome. At the same time, assuming heterogeneity in the effects of the program, the monotonicity assumption developed by Imbens and Angrist (1994) or the uniformity assumption presented in Heckman and Vytlacil (2005), they estimated the Local Quantile Treatment Effect (LQTE) which is a kind of proxy of LATE in the classical IV literature. They

used the identification results in Abadie (2003), see also Imbens and Rubin (1997) for identification of the marginal potential distributions of compliers when no covariates are present, and Abadie (2002) for bootstrap tests of distributional treatment effects in a same framework. Cherrnozhukov and Hansen (2005) also proposed an IV model for quantile treatment effects in the presence of endogeneity and under rank invariance assumption. Later on, more development in this area were done by Chernozhukov and Hansen (2013). In line with IV quantile regression, see also Chetverikov (2013) for their work on estimating the distributional effects of an endogenous treatment that varies at the group level when there are group-level unobservable. Heckman and Vytlacil (2005) also proposed a non-parametric estimators of treatment effects using Marginal Treatment Effect assuming at the same time heterogeneity in choice and response. Later on, Carneiro and Lee (2009) extended that method to the estimation of not only means, but also distributions of potential outcomes. Athey and Imbens (2006) proposed an estimation of quantile treatment effect under the assumption of difference in difference methods meaning data are available before and after the treatment for all units (kind of panel data analysis) and under rank preservation assumption. Firpo (2007) proposed a semi parametric estimator of QTE assuming that selection to treatment is based on observable characteristics. Yu (2014) proposed an estimator of marginal quantile treatment effects meaning conditional quantile on the covariates and rank in the distribution.

The common problem of these two branches of the literature is the counterfactual. In fact, most of these methods suffer from the fact that to estimate marginal distribution or joint distribution of potential outcome, the full set of observations is needed. Given that it is not possible to observe Y_0 and Y_1 at the same time, the previous methods suffer from its incomplete nature, even if the method of estimation of the distribution is good. Despite the precision given by these new methods, it is still difficult, in literature, to obtain individual effect consequently, the true effect on the distribution. So many assumptions are often made before coming up with an acceptable impact. This chapter uses the main result from Chapter 3 to solve the problem of distributional effects from the source as Rubin highlighted, assuming only that counterfactual is a missing value that can be estimated by specific methods according to the assignment process.

2.5 The Hypothesis "No Effects"

In the statistical process, after an estimation, a testing procedure must follow to see whether the estimates are significant or not. The tests of significance of the effects should be performed after estimation of the effect of a treatment. In the literature of IE, few studies have been done in this area.

2.5.1 Testing the treatment effects in the literature

In the framework of impact evaluation, a large part of the recent literature focuses on estimation of the average effect of the treatment under assumptions

of unconfoundedness and/or ignorability, following the seminal work by Rubin (1974) and Rosenbaum and Rubin (1983): homogeneity of effects. Meanwhile, there is a strong evidence of heterogeneity of the effects of a treatment from a subpopulation point of view or across the distribution of the potential outcome (Bitler et al., 2006; Djebbari and Smith, 2008; Jackson and Page, 2013; Bitler et al., 2008). Despite the recent works done to address distributional effect of treatment[1], few studies (Crump et al., 2008; Abadie, 2011) have been done to test rigorously the significance of the treatment effect in a population. Standard tests of the "no-treatment-effect" hypothesis include permutation tests (Pitman, 1937, 1938b,a), the Wilcoxon rank sum test (Wilcoxon, 1945), two-sample t test (Welch, 1938), and Fisher type randomization tests (Fisher, 1935). Most practitioners are aware that these procedure tests differ from "no-effect" hypotheses and are based on different modeling assumptions. Most of these tests are concerned with the distribution and more recently, many tests constructed are related to distributions.

2.5.2 Comparison of distribution

In recent literature, researchers facing comparison of distributions are not only interested in the difference in means and variances but also in the entire distribution (Goldman and Kaplan, 2017). They may want to know first if the distributions are the same, but when they are different, what makes them different and where

[1]See Chapter 3 for literature on it.

are they different. The classical statistical tests used to compare distribution answer two questions: (1) Are the distributions the same or different? (2) Do the distributions differ from the mean or the median, or another specific quantile? Having the answers to those two questions does not tell us how significant the difference in quantiles can be or how different the CDF of the datasets are. The problem highlighted in this paragraph is very common in different areas of research including geography, demography and mostly in the area of economics with program evaluation and other economics experiments. This literature can be used to test the hypothesis "No effect" in the framework of IE.

2.6 Summary

This chapter presented the literature review on the treatment effect estimators starting with elementary statistical concepts such as estimators, quantiles and test of hypothesis. Then, methods of impact evaluation and imputation were presented in their basic form, extended form can be obtain easily from literature. Some seminal works on treatment effect estimators done in the literature very useful for the idea developed in this research were presented as well. The next chapter will introduce the concept of imputation based estimator.

CHAPTER 3

IMPUTATION BASED AVERAGE TREATMENT EFFECTS ESTIMATORS

3.1 Introduction

This chapter introduces the notion of treatment effect estimators based on imputation methods. Imputation Based Average Treatment Effects (IB-ATE) are developed following the basic definition of impact given by Rubin (1974) and their properties are studied here. The intuition is to assume that counterfactual is a missing value that can be imputed using classical imputation methods, then using Rubin's definition of impact, we can derive our estimators.

3.2 Definition and Structural form of IB-ATE

This section presents the framework and assumptions around IB-ATE and conclude with the definition and structural form of IB-ATE.

3.2.1 Framework and Assumptions

There are a number of assumptions made in this research linked to the assignment process and the missingness mechanism. There is also a need to describe the

framework in which this research is done in terms of distributions and hypotheses.

3.2.1.1 Assignment and Missingness Process

For this chapter and throughout the research, there is a parallel between the assignment to treatment process and the missingness mechanism. Since the counterfactual is assumed to be a missing observation, each assignment process leads to missing data. In IE settings, two main assignment to treatment processes can distinguish:

1. **Randomization**: this means that the treated unit or the treated group is randomly selected from a population. All units have the same chance to be selected or not in the treatment group. From this selection process, the missingness mechanism is **MCAR**;

2. **Selection on observable**: in this case, the probability to be selected in the treatment group differ from one unit to another. All units don't have the same probability to be treated. Here the missingness mechanism can be **MAR** or **NMAR** depending on if there are other unobservable characteristics influencing the status of treated or not.

3.2.1.2 Statistical Framework

Let Y_0 and Y_1 be two independent continuous random variables with means μ_0 and μ_1, and standard deviation σ_0 and σ_1. Let also T be a Bernoulli random variable with a given probability p independent to Y_0 and Y_1 taking two possible values 0 and 1. Let's assume that a sample of n observations is drawn from the

random vector (Y_0, Y_1, T) in a population, defined by: $(Y_{0i}, Y_{1i}, T_i)_{i=1}^{n}$. Practically in the framework of impact evaluation, Y_0 and Y_1 cannot be observed in the reality at the same time for a single unit, therefore Y_0 is observed if the unit is not treated $(T = 0)$ and Y_1 is observed if the unit is treated $(T = 1)$, effect of a program assignment decided by T.

From the rubin's framework and definition, we assume that we can observe at the same time for a single unit Y_{i0} and Y_{i1}. It is the hypothetical situation where everyone is not treated in a state and everyone is treated in another state.

For the IB-ATE framework, let $\tilde{Y_0}$ and $\tilde{Y_1}$ be two estimators of Y_0 and Y_1 using a given imputation method or any other methods that can be performed. Let also T be a Bernoulli random variable, the same defined early with all its properties. We assume that a sample of n observations is drawn from the random vector (Y_0^*, Y_1^*, T) in a population, such that; for a given unit i, the different quantities are defined as follows:

$$\begin{cases} Y_{0i}^* = \tilde{Y}_{0i} \; and \; Y_{1i}^* = Y_{1i} & if \;\; T_i = 1 \\ Y_{0i}^* = Y_{0i} \; and \; Y_{1i}^* = \tilde{Y}_{1i} & if \;\; T_i = 0 \end{cases} \tag{3.1}$$

For treated units, the potential outcome if they were not treated is estimated. For non treated units, the potential outcome if they were treated is estimated. In summary, counterfactual is estimated using a given imputation method.

3.2.2 Structural Form of Estimators

Definition 3.1. From Rubin's definition of true effect, under Rubin's framework, assuming that a sample of n observations is drawn from the random vector $(Y_{i0}, Y_{i1})_{i=1}^{n}$ in a subpopulation, the true effect of a program assignment on a specific unit i is given by Δ_i and the true average effect of the program on the whole sample is given by Δ as follows:

$$\Delta_i = Y_{1i} - Y_{0i} \tag{3.2}$$

$$\Delta = \frac{1}{n}\sum_{i=1}^{n}\Delta_i = \frac{1}{n}\sum_{i=1}^{n}(Y_{1i} - Y_{0i}) \tag{3.3}$$

In case the treatment is defined by T but for some reasons the counterfactual can be obtain, the average effect on the real beneficiary of the program (n_1 units) and on the non-beneficiary ($n - n_1$ units) of the program (defined by the random variable T) can as well be defined as follows:

$$\Delta_T = \frac{1}{n_1}\sum_{i=1}^{n}\Delta_i T_i \tag{3.4}$$

$$\Delta_{NT} = \frac{1}{n - n_1}\sum_{i=1}^{n}\Delta_i(1 - T_i) \tag{3.5}$$

Remark 3.1. 1. The main effect or the true effect in the population to estimate is actually the difference of the mean $\mu_1 - \mu_0$ of the two random

variables.

2. The quantity Δ in equation (3.3) is an unbiased estimator of the difference of the mean $\mu_1 - \mu_0$ of the two random variables.

Definition 3.2. In the IB-ATE framework, the Imputation Based Average Treatment Effects are defined as follows:

1. The imputation based effect of the program on the unit i is $\hat{\delta}_i$ and equal to

$$\hat{\delta}_i = Y_{1i}^* - Y_{0i}^* = (Y_{1i} - \tilde{Y}_{0i})T_i + (\tilde{Y}_{1i} - Y_{0i})(1 - T_i) \tag{3.6}$$

2. The Imputation Based Average Treatment Effect on population, treated and non-treated units respectively $\hat{\delta}$, $\hat{\delta}_T$ and $\hat{\delta}_{NT}$ are defined as follows:

$$\hat{\delta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\delta}_i \tag{3.7}$$

$$\hat{\delta}_T = \frac{1}{n_1} \sum_{i=1}^{n} \hat{\delta}_i T_i \tag{3.8}$$

$$\hat{\delta}_{NT} = \frac{1}{n - n_1} \sum_{i=1}^{n} \hat{\delta}_i (1 - T_i) \tag{3.9}$$

Proposition 3.1. Under statistical framework and assumptions of Definitions 3.1 and 3.2,

1. $\hat{\delta}_i$ is an estimator of the true effect Δ_i on the unit i.

2. $\hat{\delta}$ is an estimator of the true effect average treatment effect Δ.

Proof. A quantity or a function being an estimator of a parameter means that using a random variables sample, a realization of the sample used as an argument of estimator will produce a value close to the parameter to be estimated.

1. Since $\hat{\delta}_i$ is a function of $\tilde{Y_0}$ and $\tilde{Y_1}$ which are actually estimators of Y_0 and Y_1 by definition and themselves can be written as a function of Δ_i, therefore $\hat{\delta}_i$ is an estimator of Δ_i.

2. Since $\hat{\delta}$ is a function of $\tilde{Y_0}$ and $\tilde{Y_1}$ which are actually estimators of Y_0 and Y_1 by definition and themselves can be written as a function of Δ, therefore $\hat{\delta}$ is an estimator of Δ.

3.3 Properties of IB-ATE Estimators

After having the structural form of estimators, the properties can be investigated.

3.3.1 Unbiasedness

Proposition 3.2. Let b_0 and b_1 be the bias due to estimation of Y_0 and Y_1 respectively by $\tilde{Y_0}$ and $\tilde{Y_1}$. The estimators $\tilde{Y_0}$ and $\tilde{Y_1}$ are unbiased if $b_0 = 0$ and $b_1 = 0$. Under statistical framework and assumptions of Definitions 3.1 and 3.2 we have the following properties:

1. If $\tilde{Y_0}$ and $\tilde{Y_1}$ are biased estimators of Y_0 and Y_1 then $\hat{\delta}_i$ will be a biased estimator of Δ_i with the bias equal to $b_i = b_{1i} - T_i(b_{0i} + b_{1i})$.

2. If $\tilde{Y_0}$ and $\tilde{Y_1}$ are biased estimators of Y_0 and Y_1 then $\hat{\delta}$ will be a biased estimator of Δ with the bias equal to $B_i = \frac{1}{n}\sum_{i=1}^{n}(b_{1i} - T_i(b_{0i} + b_{1i}))$.

3. $\hat{\delta}_i$ is an unbiased estimator of Δ_i if and only if $\tilde{Y_0}$ and $\tilde{Y_1}$ are unbiased estimators of Y_0 and Y_1.

4. $\hat{\delta}$ is an unbiased estimator of Δ if and only if $\tilde{Y_0}$ and $\tilde{Y_1}$ are unbiased estimators of Y_0 and Y_1.

Proof. The proof of this proposition is based on the expectation calculation of the estimators to evaluate the bias as described in Section 1 of Chapter 2.

1. The expected value of $\hat{\delta}_i$ is given by:

$$
\begin{aligned}
E(\hat{\delta}_i) &= E\left((Y_{1i} - \tilde{Y}_{0i})T_i + (\tilde{Y}_{1i} - Y_{0i})(1 - T_i)\right) \\
&= E\left((Y_{1i} - \tilde{Y}_{0i})T_i\right) + E\left((\tilde{Y}_{1i} - Y_{0i})(1 - T_i)\right) \\
&= T_i E\left((Y_{1i} - \tilde{Y}_{0i})\right) + (1 - T_i)E\left((\tilde{Y}_{1i} - Y_{0i})\right) \\
&= T_i(E(Y_{1i}) - E(\tilde{Y}_{0i})) + (1 - T_i)(E(\tilde{Y}_{1i}) - E(Y_{0i})) \\
&= T_i(Y_{1i} - E(\tilde{Y}_{0i})) + (1 - T_i)(E(\tilde{Y}_{1i}) - Y_{0i}) \\
&= T_i(Y_{1i} - Y_{0i} - b_{0i}) + (1 - T_i)(Y_{1i} + b_{1i} - Y_{0i}) \qquad (as\ \tilde{Y}_{0i}\ and\ \tilde{Y}_{1i}\ are\ biased) \\
&= (1 - T_i + T_i)(Y_{1i} - Y_{0i}) + (1 - T_i)b_{1i} - T_i b_{0i} \\
&= Y_{1i} - Y_{0i} + [b_{1i} - T_i(b_{0i} + b_{1i})] \\
E(\hat{\delta}_i) &= \Delta_i + b_i
\end{aligned}
$$

2. For the sample case we have:

$$
\begin{aligned}
E(\hat{\delta}) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\hat{\delta}_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}E(\hat{\delta}_i) \qquad (\textit{applying result 1. of this proof}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(\Delta_i + b_i) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i + \frac{1}{n}\sum_{i=1}^{n}b_i \\
E(\hat{\delta}) &= \Delta + B
\end{aligned}
$$

3. Computing the expected value of $\hat{\delta}_i$ will give this:

$$
\begin{aligned}
E(\hat{\delta}_i) &= E\left((Y_{1i} - \tilde{Y}_{0i})T_i + (\tilde{Y}_{1i} - Y_{0i})(1 - T_i)\right) \\
&= E\left((Y_{1i} - \tilde{Y}_{0i})T_i\right) + E\left((\tilde{Y}_{1i} - Y_{0i})(1 - T_i)\right) \\
&= T_i E\left((Y_{1i} - \tilde{Y}_{0i})\right) + (1 - T_i)E\left((\tilde{Y}_{1i} - Y_{0i})\right) \\
&= T_i(E(Y_{1i}) - E(\tilde{Y}_{0i})) + (1 - T_i)(E(\tilde{Y}_{1i}) - E(Y_{0i})) \\
&= T_i(Y_{1i} - E(\tilde{Y}_{0i})) + (1 - T_i)(E(\tilde{Y}_{1i}) - Y_{0i}) \\
&= T_i(Y_{1i} - Y_{0i}) + (1 - T_i)(Y_{1i} - Y_{0i}) \qquad (\textit{as } \tilde{Y}_{0i} \textit{ and } \tilde{Y}_{1i} \textit{ are unbiased}) \\
&= (1 - T_i + T_i)(Y_{1i} - Y_{0i}) \\
&= Y_{1i} - Y_{0i} \\
E(\hat{\delta}_i) &= \Delta_i
\end{aligned}
$$

Therefore $\hat{\delta}_i$ is an unbiased estimator of Δ_i.

4. Computing as previously the expected value of δ will give this:

$$\begin{aligned} E(\hat{\delta}) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i\right) \\ &= \frac{1}{n}E\left(\sum_{i=1}^{n}\hat{\delta}_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}E(\hat{\delta}_i) \qquad (\textit{applying proof 3.}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\Delta_i = \Delta \\ E(\hat{\delta}) &= \Delta \end{aligned}$$

Therefore $\hat{\delta}$ is an unbiased estimator of Δ.

3.3.2 Convergence

Proposition 3.3. Under statistical framework and assumptions of Definitions 3.1 and 3.2, the estimator $\hat{\delta}$ is a convergent estimator of Δ if and only if $\tilde{Y_0}$ and $\tilde{Y_1}$ are convergent estimators of Y_0 and Y_1. Mathematically:

$$\lim_{n\to\infty}\ \hat{\delta} = \Delta \Leftrightarrow \lim_{n\to\infty}\ \tilde{Y_0} = Y_0 \textit{ and } \lim_{n\to\infty}\ \tilde{Y_1} = Y_1$$

Proof. From the linearity of the limit of a quantity we have:

$$
\begin{aligned}
\lim_{n\to\infty} \hat{\delta} &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \lim_{n\to\infty} \hat{\delta}_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \lim_{n\to\infty} \Big[(Y_{1i} - \tilde{Y}_{0i})T_i + (\tilde{Y}_{1i} - Y_{0i})(1 - T_i)\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} \Big[(Y_{1i} - \lim_{n\to\infty} \tilde{Y}_{0i})T_i + (\lim_{n\to\infty} \tilde{Y}_{1i} - Y_{0i})(1 - T_i)\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} [(Y_{1i} - Y_{0i})T_i + (Y_{1i} - Y_{0i})(1 - T_i)] \quad (as\ \tilde{Y}_{0i}\ and\ \tilde{Y}_{1i}\ are\ convergent) \\
&= \frac{1}{n}\sum_{i=1}^{n} \Delta_i
\end{aligned}
$$

$$
\lim_{n\to\infty} \hat{\delta} = \Delta
$$

3.3.3 Consistency

Proposition 3.4. Under statistical framework and assumptions of Definitions 3.1 and 3.2, the estimator $\hat{\delta}$ is a consistent estimator of Δ if and only if $\tilde{Y}_0$ and $\tilde{Y}_1$ are consistent estimators of Y_0 and Y_1. Mathematically, the two following conditions have to be verified:

1. $\lim_{n\to\infty} E(\hat{\delta}) = \Delta$ if and only if $\lim_{n\to+\infty} b_0 = \lim_{n\to+\infty} b_1 = 0$.
2. $\lim_{n\to\infty} Var(\hat{\delta}) = 0$ if and only if $\lim_{n\to+\infty} Var(\tilde{Y}_0) = \lim_{n\to+\infty} Var(\tilde{Y}_1) = 0$.

Proof. To prove the proposition, two thing have to be shown:

1. For the first point, see proof of Proposition 3.3;

2. For the variance, it is given by the equation:

$$
\begin{aligned}
Var(\hat{\delta}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i\right) \\
&= \frac{1}{n^2}Var\left(\sum_{i=1}^{n}\hat{\delta}_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n}(Y_{1i}-\tilde{Y}_{0i})T_i + (\tilde{Y}_{1i}-Y_{0i})(1-T_i)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}Var\left((Y_{1i}-\tilde{Y}_{0i})T_i + (\tilde{Y}_{1i}-Y_{0i})(1-T_i)\right)\right) \qquad (the\ \text{cov}=0) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}T_i^2Var(Y_{1i}-\tilde{Y}_{0i}) + (1-T_i)^2Var(\tilde{Y}_{1i}-Y_{0i})\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}T_i^2Var(\tilde{Y}_{0i}) + (1-T_i)^2Var(\tilde{Y}_{1i})\right)
\end{aligned}
$$

As $\tilde{Y}_0$ and $\tilde{Y}_1$ are consistent estimators of Y_0 and Y_1 their variance goes to zero as the sample size increases. For n going to infinity, the variance of $\hat{\delta}$ will go to 0.

3.4 Asymptotic Properties: Simulations

The main objective of this section is to use simulations to test our hypothesis asymptotically that imputation methods can lead to better estimators of average effects than IE estimators or at least as good as existing ones. In the meantime, the theoretical properties such as convergence and consistency are tested. In this section, description of simulation procedure and parameters is done, then simulations are performed under Random Assignment (MCAR missingness) hypothesis and under Deterministic Assignment (MAR missingness) hypothesis.

3.4.1 Algorithm and Assumption

The simulation process recreates an hypothetical situation where a treatment has to be assigned in a population with all the parameters being known. For example, assignment process is well known (T), the potential outcome is known (Y), decision to treat everyone or not to treat everyone can be taken so that computation of the true impact of the project (Δ_i and Δ) can be done easily. In brief, all parameters are mastered and they can be modified to obtain different results according to the objectives fixed. Therefore, for a given assignment process, simulation results will tell which imputation method gives better IB-ATE estimators. Imputation methods will be judged at two levels: first the capacity to estimate counterfactual (using imputation) that lead to better estimators than existing ones and second, the capacity to reconstruct the exact value of the missing observation (RMSE indicators are used). After imputation, IB-ATE are produced and compared to existing ones in IE framework.

The Table 3.1 gives the final structure of the data simulated on which estimations are performed in this simulation.

- Y denotes the potential outcome;
- X a set of covariates collected for identification of each case;
- T is the treatment indicator, 1 for treated units and 0 for non treated units;
- a index meaning after the treatment;

Table 3.1: **Database after simulation of variables.**

Case N	Bef Treat		Hypot			Af Treat		Out With Miss	
	Y_b	X_b	Y_{2T}	Y_{2NT}	T	Y_a	X_a	Y_T	Y_{NT}
1 2 . . .									
i	$Y_{b,i}$	$X_{b,i}$	$Y_{2T,i}$	$Y_{2NT,i}$	**1**	$Y_{a,i}$	$X_{a,i}$	$Y_{T,i}$	.
j	$Y_{b,j}$	$X_{b,j}$	$Y_{2T,j}$	$Y_{2NT,j}$	**0**	$Y_{a,j}$	$X_{a,j}$	.	$Y_{NT,j}$
. . .									
N									

- b index meaning before the treatment;
- Y_{2T} is the hypothetical outcome if everyone is treated;
- Y_{2NT} is the hypothetical outcome if everyone is not treated;
- Y_T is the potential outcome in which non-treated cases are considered as missing values;
- Y_{NT} is the potential outcome in which treated cases are considered as missing values;
- Y_a is the potential outcome which is really observed in the IE framework and it is defined as follows: $Y_{a,i} = Y_{T,i}$ and $Y_{a,j} = Y_{NT,j}$, for i treated and j non treated.

The variable of treatment T is simulated using a binomial distribution[1] with a fixed probability if a RA process is assumed or under specific conditions on X in case the assignment is done on observable characteristics. After simulation of the database, the Rubin's true average treatment effect is given by $\Delta = mean(Y_{2T} - Y_{2NT})$ and the true average treatment on treated unit is given as follows $\Delta_T = mean(Y_{2T} - Y_{2NT}\,|T = 1)$

Under the large class of existing imputation methods, the chosen ones are: Mean imputation, Random imputation, Linear regression imputation (deterministic and random), Nearest Neighbour imputation, Multiple Imputation, Maximum Likelihood imputation, Propensity score matching imputation and finally Quantile regression imputation which is not commonly used. All IE methods presented in Chapter 2 are used as well here for the purpose of comparisons.

To test the performance of imputation methods, the Root Mean Squared Error is computed and comparisons are made among different methods. To test the performance of our computed IB-ATE, the average bias is computed ($AvrgBias = E(\hat{\theta} - \theta)$) and compared to the one for existing IE methods. All this done under a bootstrap procedure of 1000 replications.

3.4.2 Results and comments

The results are presented in two groups depending on assumptions made on the assignment process.

[1]See appendix A1 for more details about distributions and codes of simulation

3.4.2.1 Random Assignment Hypothesis Results (MCAR)

Assuming that the treatment is randomly assigned as in medical experiments for a new drug, the first consequence is the missingness process which is MCAR. A proportion of treated units is fixed to be 40% of the total population (assumption on T as a binomial distribution of parameter 1 and 0.4) and for the simulation we made sure that each sample drawn from the population had the same proportions. The IE methods implemented here were Randomization (RA), PSM and DID. The performance of the IE estimators is evaluated using the average bias and the performance of imputation methods among them is evaluated using the RMSE. Table 3.2 gives a summary of simulation results.

For all purely IE methods, the estimators (ATE and ATT) were asymptotically convergent. The average bias was decreasing as the sample size was increasing. The best method among them no matter the size of the sample was DID. The average bias was the smallest among all the methods for all sample sizes (See Table 3.2, line six). At the same time, the standard deviation was always small for the estimators of DID. Consequently, the DID was the best one among the IE methods [2].

[2]See Appendix A1 for more details on simulations

Table 3.2: **Average bias of IB-ATE estimators (MCAR).**

General summary of results			Average Bias											
			N=50		N=100		N=200		N=500		N=800		N=1500	
			ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
IE Meth	RA		-7.1	-7.4	-2.6	-2.4	1.508	1.464	0.665	0.662	-0.2277	-0.3652	0.081	-0.0163
	PSM		-5.8	-3.5	-2.3	-1.3	1.476	0.237	1.067	1.82	-0.8602	-0.5916	-0.5571	-0.3682
	DID		0.5	0.2	-0.1	0.1	0.066	0.022	0.069	0.066	0.2315	0.094	0.089	-0.0083
IB Results	Mean	General	-7.1	-7.4	-2.6	-2.4	1.508	1.464	0.665	0.662	-0.2277	-0.3652	0.081	-0.0163
		Condi	-7.4	-7.7	-2.3	-2.2	1.581	1.484	0.705	0.654	-0.1764	-0.3013	0.0837	-0.0376
	Rand Imp	General	-8.3	-12.1	-2.3	-1	0.887	1.325	0.057	0.681	-0.4384	-0.4303	0.0742	0.1673
		Hot-deck	-7	-5.4	-2.4	-1.4	0.916	-2.499	0.177	0.53	-0.5034	-0.1834	-0.1004	0.3516
	Det LM		-4.6	-4.5	-2.1	-1.4	0.181	-0.062	0.667	0.646	-0.0167	-0.1851	-0.264	-0.2871
	k-NN	V1	-5.5	-7.3	-2.6	-0.1	0.974	-0.412	0.645	1.235	0.058	0.3965	-0.189	-0.4415
		V2	-5.4	-6.9	-1.9	1.6	0.996	-0.37	0.797	1.679	0.0439	0.5487	-0.1389	-0.3363
	Rand LM		-7.1	-8.7	-1.4	-3.5	0.773	0.884	0.76	0.264	-0.2642	-0.478	-0.2268	-0.4668
	MI-MICE		-6.8	-7.2	-1.9	-0.3	0.705	-0.384	0.655	1.03	-0.0967	-0.0561	-0.4629	-0.3876
	ML Imp	Normal	-4.5	-5.3	-2.1	-1.9	0.624	-0.168	0.615	0.082	-0.1837	-0.4964	-0.0205	-0.2943
		Dist-Free	-6	-5.8	-2.7	-2.4	0.312	-0.363	0.637	0.693	0.1239	-0.3309	-0.2242	-0.3261
	PSM Imp		-9.4	-11.6	-14.2	-15.7	-32.0	-29.9	-90.0	-91.2	-148.1	-148.0	-198.4	-198.5
	QR Imp		2.7	-30.8	5.5	-31.4	6.072	-32.096	7.138	-32.239	6.92	-32.59	6.8	-32.8

For the convergence of estimators, the bias was decreasing and tend to 0 as the sample size was increasing. Depending on the sample size, the best IB-TE were changing. For N=50, the best were QR for ATE and Deterministic Linear regression for ATT. For N=100, the best ones were random linear regression imputation for ATE and k-NN for ATT. For large samples, N=1500, the best ones were ML imputation for ATE (-0.0205 average bias) and mean imputation for ATT (-0.0163 average bias). On average, the best IB-TE estimators no matter the sample size were the Maximum Likelihood (ML) and Deterministic Linear Regression model (Det ML), methods which was always among the three best IB-ATE when changing the sample size. For the case of mean imputation, IB-TE and IE estimators are the same theoretically and empirically.

This led to the comparison of the IE estimators with IB ATE estimators. Here, the speed of convergence and the standard deviation in certain cases were used to compare. As the sample size increased, the average bias of IB-ATE were close to the average bias of DID which is the best method used in IE framework. For example, N=100 the average bias of DID was -0.1 for ATT and it was the same average bias for k-NN method. For N=800, k-NN performed far better than DID for ATE (0.04 against 0.23 for average bias) and also MI performed better than DID for ATT (-0.05 against 0.09 in terms of average bias). For larger sample, there was an IB-ATE estimator performing better than DID estimators or as well as.

In summary, it is true that the DID methods gave better results on average for all sample given the average bias and compared to a single method of IB-ATE but at the same time for some IB-ATE estimators, the average bias was smaller than the average bias of DID even if the standard deviation was bigger. In addition, to implement the DID method, the user needs to collect information before the program which is not always possible. In that case, DID is not applicable. Therefore, the second-best IE methods to use is the PSM which is not as good as the IE method, especially ML imputation or k-NN no matter the sample size.

3.4.2.2 Selection on Observable Results (MAR or NMAR)

Assuming here that the treatment is not randomly assigned but depends on a given variable called instrument (A single variable in this simulation to simplify), the missingness process is MAR. From a population of 25000 units, a proportion of 40% of unit was drawn around a given threshold fixed on the instrument. This was done to be able to apply the IV regression and the sharp RDD at the same time. From that subpopulation around the threshold, we have drawn our data with an increasing sample size making sure that the share of treated in each sample is 40% as in the previous experiment. Finally, repeating what is done in the first case, IE methods and imputation methods were applied in each sample to obtain best IE estimators and imputation method.

When looking at the results of the simulations recorded in Table 3.3, all the average bias of IE methods were decreasing, meaning that the estimators are

asymptotically unbiased and convergent. DID method is always the best method in general no matter the sample size. Out of the six sample size presented here, DID was the best 5 times and the second best for N=100 where the smallest average bias for RA were -0.17 and -0.43 for ATE and ATT respectively and for DID the average bias was -0.48 and 0.74. For the other sample size, the average bias of the DID method was the smallest and decreasing. Indubitably, DID estimators were the best again despite IV estimators and RDD estimators.

If the imputation methods were assessed on their ability to estimate the treatment effect, given the different sample sizes, all the average bias were found to decrease to 0 except for PSM imputation and QR imputation. So, except those two methods, here again IB-TE are asymptotically unbiased. Mean imputation and Random imputation were the best methods among all of them. As it is recorded in Table 3.3, the smallest average bias is either from mean imputation or from random imputation except for N=50 where QR Imp and k-NN were the best IB-TE estimators. The other methods, like MI also performed well but not as well as Mean and Random imputation. Imputation Methods were able to produce acceptable average treatment effect estimators no matter the sample size.

Comparing classical IE estimators with IB-ATE estimators, it was found that with all data available (especially data before the program and a large set of covariates), the only IE method that was as good as the IB-ATE was DID especially for large sample. For small samples (N=50 and 100), DID was not the best, QR

Imp and Conditional mean imputation were the best, smallest average bias. For N=500 and considering ATE, the smallest average bias was recorded for general random imputation -0.005, the second smallest was -0.366 from DID; for ATT, the smallest average bias was -0.026 for general random imputation and the second was 0.028 from hot deck imputation. For large sample strictly greater than 500, DID got the smallest average bias even if the other IB-ATE estimators were close in terms of average bias. The other IE methods did not perform well compared to DID and IB-ATE estimators, especially IV and RDD which were supposed to produce better estimators given the assignment process. For IV, the average bias was decreasing then became constant around 5 as the sample size increased. For RDD, the estimator was asymptotically convergent and performed as well as some IB-TE estimators but still less than the best three.

Table 3.3: **Average bias of IB-ATE estimators (MAR).**

General summary of results			Average Bias											
			N=50		N=100		N=200		N=500		N=800		N=1000	
			ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
IE Meth	RA		-4.93	-5.19	-0.17	-0.43	-2.71	-2.98	0.687	0.42	-1.5	-1.76	-1.22	-1.48
	PSM		-1.54	-1.82	-3.42	-2.15	-1.91	-1.55	-1.68	-0.85	-2.38	-1.87	-2.1	-2.2
	DID		-0.83	-1.10	-0.48	-0.74	-0.74	-1.01	-0.366	-0.63	-0.37	-0.63	-0.23	-0.49
	IV		-8.4	-8.66	-5.83	-6.09	-5.43	-5.69	-5.2	-5.46	-5.18	-5.45	-5.12	-5.38
	RDD		-4.04	-4.3	-2.1	-2.37	-2.23	-2.49	-.829	-1.09	-2.39	-2.66	-1.92	-2.18
IB Results	Mean	Gen	-4.93	-5.19	-0.17	-0.43	-2.71	-2.98	0.687	0.425	-1.5	-1.76	-1.22	-1.48
		Condi	-4.29	-3.73	-0.12	-0.5	-2.54	-3.1	0.74	0.41	-1.43	-1.72	-1.12	-1.45
	Rand Imp	General	-3.54	-5.86	1.39	0.67	-1.70	-1.60	-0.005	-0.026	-1.56	-1.42	-0.51	-0.91
		Hot deck	-6.57	-6.82	0.48	-2.08	-1.72	-2.09	-0.519	0.028	-1.88	-1.47	-1.07	-1.35
	Det LM		-4.88	-3.92	-1.86	-2.84	-2.12	-2.76	-0.84	-1.07	-2.38	-2.68	-1.91	-2.21
	k-NN	V1	-3.4	-3.48	-4.01	-1.5	-4	-3.96	-1.37	-1.14	-3.81	-3.92	-3.03	-2.88
		V2	-3.49	-3.9	-3.69	-0.85	-4.07	-4	-1.19	-1.06	-3.5	-3.49	-2.76	-2.24
	Rand LM		-3.74	-2.47	-1.88	-3.27	-1.68	-1.65	-1.26	-1.95	-2.41	-3.2	-1.32	-1.2
	MI MICE		-4.42	-3.82	-2.61	-2.42	-2.46	-3.1	-1.35	-0.89	-2.73	-3.40	-2.24	-2.33
	ML Imp	Normal	-5.59	-4.46	-1.39	-2.01	-2.19	-2.57	-0.86	-0.77	-2.81	-2.99	-1.8	-2.19
		Dist free	-5.21	-3.87	-2.06	-2.16	-2.41	-2.68	-0.95	-0.96	2.42	-2.51	-1.52	-1.73
	PSM Imp		-11.6	-22.3	-16.7	-33.5	-34.04	-75.02	-89.4	-177.87	-146.0	-196.1	-171.5	-199.1
	QR Imp		0.12	-33.5	3.88	-35.4	5.28	-34	5.99	-34	4.88	-35.02	5.82	-34.10

In summary, except DID which gave asymptotically unbiased estimators with the smallest variance, one can always find an imputation method that gives small average bias and small variance than the other IE methods. This is to say IB-ATE estimators can perform as well as DID estimators but with a bigger variance. In addition, in case of shortage of data (if it is not possible to get data before the program), IB-ATE is the best solution if the assignment process is not random. PSM, IV and RDD produce convergent estimators but not as good as IB-ATE estimators.

3.4.3 Advantages of IB-ATE and Discussions

The first advantage of having IB-ATE estimators is related to availability of data. By using bootstrap, estimators (IB-ATE) were found to be as good as the one obtained with IE methods. Some of those estimators used only the potential outcome after the treatment to produce good estimators (Random imputation estimators). While DID for example needs data before the assignment of the treatment, MI method does not need that to produce a good estimator as DID estimators. When covariates are not available or not enough, IE methods like PSM, IV and RDD cannot be performed but still Random imputation and Mean imputation can be performed. In case of shortage of data in impact evaluation framework, IB-ATE estimators are the best ones to use.

Another advantage of using IB-TE estimators is the fact that from them, all types of treatment effects can be obtained. Imputation gives the possibility to produce

ATE, ATT and ATNT. In addition to that, with IB-ATE it is easily possible to obtain unit effect. For each unit of the program treated, an estimation of the impact of the program can be given. Therefore, the distributional effect across the different subgroups of the population. The last one is very important in medical experiment where the treatment is randomly assigned, one may want to know after the experiment what the impact of the drug would have been if used the other way round. Instead of starting experiment again, IB-TE can give that result without effort. The last advantage is the simplicity of the methods. All of those imputation methods are implemented in R and the only effort to make is the bootstrap program. It is a small price to pay for good estimators in a context of data shortage.

Simulations performed are of course subjected to some simulation choices like the distribution of the potential outcome, the share of treated unit in the sample and the distribution of covariates. This does not mean that changing the parameters of the simulation will lead to totally different results absolutely but the results can lean on the simulation parameters. By changing the parameters, the results can be in favour of IB-ATE or IE methods. The share of the treated units is not an issue because in practice, the treated units are always less than non-treated units. Therefore, it is always possible to complete the sample of treated by non-treated to obtain a share of 40%. For all these reasons, IB-ATE estimators are tested on a real program in the next section to see how they perform for a real data and problem.

3.5 Applications

After simulations, where the results showed that IB-ATE estimators can perform as well as classic treatment effect estimators otherwise better in some cases, the next step is to apply these results to real set of data since simulations are always questionable.

3.5.1 Description of the program and Data

The famous Lalonde (1986) dataset in IE literature is considered for application. Lalonde data set contain the treated and control units of the male sub-sample from the National Supported Work Demonstration. The NSW Demonstration, Manpower Demonstration Research Corporation (MDRC) 1983, was a federally and privately funded program implemented in the mid-1970s to provide work experience for a period of 6-18 months to individuals who had faced economic and social problems prior to enrolment in the program. Those randomly selected to join the program participated in various types of work, such as restaurant and construction work. Preintervention variables where collected by the program to allow Lalonde to use control groups, selected using preintervention variables to compare and obtain the treatment effect on treated.

Based on pre-intervention variables, Dehejia and Wahba (1999) extracted a further subset of Lalonde's NSW experimental data, a subset containing information on RE74 (earnings in 1974). Applying the same method of Lalonde they came up

with an average treatment effect on treated (difference in earning due to training program) of $ 1794. Later, they used the propensity score method and they came up with a treatment effect on treated range of $ 1473 to $ 1774, quite close to the result of Lalonde which had the same dataset.

3.5.2 Results and Comments

As a reminder, IB-TE estimators combines imputation methods and bootstrap to obtain treatment effects estimators. In this case, the methods developed in this chapter were applied on the subset of Lalonde (1986) drawn by Dehejia and Wahba (1999).

Table 3.4: **IB-ATE Estimators using Lalonde's Subsample.**

General summary of results			Bootstrap with subsample of Lalonde's data.							
			n=200				n=400			
			ATE	Sd	ATT	Sd	ATE	Sd	ATT	Sd
IE Meth	RA		1794	767	1794	767	1785.3	252	1785.3	252
	PSM		1632.4	995	1925.8	996	1772.8	458	2039.4	454
IB Results	Gen Mean Imp		1794	767	1794	167	1785.3	252	1785.3	252
	Rand Imp	Gen	1792.4	903	1806.7	967	1773.1	418	1792.3	494
		Hot deck	1784	1073	1769.1	1067	1812.1	586	1789.5	616
	Det LM		1576.6	759	1836.8	785	1606.9	246	1816	251
	k-NN	V1	1538.3	923	2003.4	1015	1920.2	343	2377.6	393
		V2	1611.7	913	2174.8	993	1853.2	336	2178.2	369
	Rand LM		1591.2	891	1834.1	1010	1586.2	420	1804.6	499
	MI with MICE		1670.5	898	1668.2	892	1631.9	371	1620.8	378
	ML Imp	Normal	1572.2	927	1797.4	961	1613.4	407	1782.1	459
		Dist free	1585.1	918	1802.3	940	1629.1	406	1782.8	444

For a fixed sample size of n=200 drawn from a population of 445 (respecting the share of 41% of treated) and after 1000 replications, results recorded in Table 3.4 are interesting. The bootstrap of mean imputation method led to exactly

the same treatment effect on treated as the results of Lalonde (1986) which is $ 1794. IB-TE estimators produce by ML imputation ($ 1797.4), and Random Imputation ($ 1769.1) are closer to the benchmark of Lalonde and better than the results obtain by Dehejia and Wahba (1999) in their work (a range from $ 1473 to $ 1774). Also looking at the standard errors, they are smaller than the ones obtain by Lalonde (1986) and Dehejia and Wahba (1999), implying a smaller confidence interval. For n=200 and R=1000 replication in the bootstrap, the best 3 IB-TE estimators are close to the benchmark and better than those obtain by the propensity score using additional costly data from comparison groups. Increasing the sample size of the bootstrap, results were still the same but with a much smaller confidence interval. As conclusion, instead of spending money and time to find additional control groups and perform propensity score, combining ML imputation or Random imputation with the bootstrap led to better results of average treatment effect on treated than propensity score matching used by Dehejia and Wahba (1999) and results as good as the one obtains by Lalonde (benchmark). On top of these results, IB-TE gives effect on the population and effect on those who were not treated if they have been treated.

3.6 Summary

In this chapter, the notion of imputation based treatment effect estimators in the framework of IE was introduced. Imputation methods were used to estimate counterfactual then derive Imputation Based Average Treatment Effect (IB-ATE)

estimators. The estimators derived are unbiased and convergent if the estimates of counterfactual are unbiased and convergent. Simulations showed that those estimators are asymptotically unbiased and consistent. In addition, they are as good as the existing ones (classical RA, PSM etc. estimators). An application was done to show that even for real data they are effective. The next chapter uses the seminal idea of imputation to propose another class of estimators, thereby assessing the distributional effect of a treatment on a population.

CHAPTER 4

IMPUTATION BASED DISTRIBUTIONAL TREATMENT EFFECTS

4.1 Introduction

This chapter develops Distributional Treatment Effect Estimators based on imputation methods and study their properties. Following the main result of Chapter 3 consisting of estimating counterfactual using imputation methods and based on seminal works done by Doksum (1974) and Lehmann (1974), Imputation Based Distributional Treatment Effects Estimators (IB-DTE) were constructed. They focused on the distributions of the potential outcome before and after the treatment reconstructed using imputation methods to estimate how the treatment is distributed in the population.

4.2 Definition and Structural form of IB-DTE

4.2.1 Framework and Assumptions

Since the results of Chapter 3 are used here as input, the framework and assumptions here are the ones presented in Chapter 3, Subsection 2.2. In summary, the

results obtained here are in a framework where assignment process and missingness process are linked, assignment generates the missingness process. Given the assignment process, the counterfactual is estimated using imputation methods, the same methods used in Chapter 3. The statistical framework concerning the distributions and the variable is the same.

In addition to the assumptions and statistical framework of Chapter 3, the reconstructed distribution using imputation is defined here. In order to define the reconstructed distribution, let us recollect the definition of the potential outcome given by equation (3.7). Under statistical framework of section 3.2.2, under assumption of the same section we define the distribution reconstructed as follows :

$$CoCom = \left(\tilde{Y}_{1,0}, \tilde{Y}_{2,0}, ..., \tilde{Y}_{n_1,0}, Y_{n_1+1,0}, ..., Y_{n,0}\right) \tag{4.1}$$

and

$$TrCom = \left(Y_{1,1}, Y_{2,1}, ..., Y_{n_1,1}, \tilde{Y}_{n_1+1,1}, ..., \tilde{Y}_{n,1}\right) \tag{4.2}$$

as the distribution of the potential outcome of the control group reconstructed in equation (4.1) and the distribution of potential outcome of the treated group reconstructed in equation (4.2) using the same imputation method, with n_1 the number of treated unit in the population.

4.2.2 Structural form of Estimators

Let's define first the true quantile treatment effect seen in Rubin's framework.

Definition 4.1. Under rank preservation assumption, the true quantile treatment effect defined with the idea of Rubin is given by

$$TrQTE_p = Q_1(p) - Q_0(p) = F_1^{-1}(p) - F_0^{-1}(p) \tag{4.3}$$

With F_0 distribution of $(Y_{i0})_{i=1}^n$ and F_1 distribution of $(Y_{i1})_{i=1}^n$, p a given probability. It is basically the difference between the same theoretical quantile of the random variables Y_0 and Y_1, the parameter to estimate in this section.

Going back in the literature, the following definition is widely used to estimate the distributional effects.

Definition 4.2. Let F_{r0} be the restricted CDF of Y_{i0} for units with $T_i = 0$ (distribution of $\{Y_{i0},\ i \in (i/T_i = 0)\}$, n_0 non-treated cases) and F_{r1} be the restricted CDF of Y_{i1} for units with $T_i = 1$ (distribution of $\{Y_{i1},\ i \in (i/T_i = 1)\}$, n_1 treated cases) with $n_0 + n_1 = n$. According to Imbens and Wooldridge (2009), Doksum (1974) and (Lehmann, 1974), under rank preservation assumption, the quantile treatment effect is given by

$$QTE_p = Q_{n_1}(p) - Q_{n_0}(p) = F_{r1}^{-1}(p) - F_{r0}^{-1}(p), \tag{4.4}$$

with p a given probability, the index r stands for restricted.

Remark 4.1. 1. Restricted term is used in this definition because the sample of treated and non-treated are automatically reduced by the assignment process while the true estimator is computed on a full sample assuming

everyone is non-treated first then everyone is treated.

2. This definition assumes rank preservation. Meaning that a unit in a given quantile before the treatment will remain in the same quantile after the treatment has been undertaken.

3. The distributions considered here are somehow incomplete because the counterfactual is not estimated, therefore, distributions are truncated.

4. Would this QTE be the same if considering the initial distribution without taking the one generated by only those who are treated ($T = 1$) and those who are not treated ($T = 0$), meaning only what is observed?

As highlighted in the literature, the measure of distributional effect of a treatment is basically given by the Quantile Treatment Effect (QTE). It is the difference between two quantiles of same order from the restricted distributions before and after the treatment under rank preservation assumption (Doksum, 1974; Imbens and Wooldridge, 2009; Firpo, 2007). At first, this study used the same idea to propose a new form of quantile effect based on imputation overcoming the issue of restriction.

Definition 4.3. Under the statistical framework, the definition of the distributions $CoCom$ and $TrCom$ and under rank preservation assumption, the Imputation Based Modified Quantile Treatment Effect ($IB - mQTE$) is defined by:

$$IB - mQTE_p = Q_{1n,I}(p) - Q_{0n,I}(p) = F_{1I}^{-1}(p) - F_{0I}^{-1}; \tag{4.5}$$

with F_{0I} the empirical CDF of $CoCom$ and F_{1I} the empirical CDF of $TrCom$, n the sample size, p a given probability used to calculate the corresponding quantile. The index I stands for imputed, 0 for before the treatment and 1 for after the treatment.

In addition to this specific definition to this research (based on imputation), 04 others quantiles treatment effect are defined without taking any assumption. In fact, the rank preservation assumption helped to make sure that the quantiles compared are from the same order before and after the treatment, the units compared are the same theoretically. It avoid shifting of units from one quantile to another one. Since imputation helps in tracking each unit before and after the treatment (individual effects), the following definitions do not need rank preservation assumption.

In order to ease the comprehension, the following framework is given. A quantile preceded by the letter G is the quantile in term of group of units such that their value for the random variable is less than the corresponding quantile. As example: $GQ_j(p) = \{Y_{ij}/Y_{ij} \leqslant Q_j(p)\}, \quad j \in \{0,1\}$. Basically, dropping out the rank preservation assumption and since imputation can allow us to have individual effect, the following definitions are based on the comparison of group quantiles using different statistics. Given a structure of quantile (quartiles, deciles or centiles), the process is the following:

(i) Identify the quantile group of belonging for each unit in the distribution of the control group given by the following sample $\left(\tilde{Y}_{1,0}, \tilde{Y}_{2,0}, ..., \tilde{Y}_{n_1,0}, Y_{n_1+1,0}, ..., Y_{n,0}\right)$

and identify the value of the quantile as well which probably correspond to the value of potential outcome of a given unit;

(ii) For each quantile group identified previously, construct the image group in the treatment sample given as follows $\left(Y_{1,1}, Y_{2,1}, ..., Y_{n_1,1}, \tilde{Y}_{n_1+1,1}, ..., \tilde{Y}_{n,1}\right)$ by obtaining their potential outcome after the treatment;

(iii) The average or median change observed in the quantile control group compared to the image group in the treatment sample is the IB-QTE.

Let's p be a probability such that $0 < p < 1$, let $GQ_0(p)$ the p^{th} quantile group and $Q_0(p) = F_0^{-1}(p)$ the value of the p^{th} quantile in the completed distribution of control group ($CoCom$); let's also $GQ_1(p)$ the p^{th} quantile group and $Q_1(p) = F_1^{-1}(p)$ the value of the p^{th} quantile in the completed distribution of treated group ($TrCom$). Let assume that $Q_0(p)$ is attained for the unit i_0 in the $CoCom$ distribution.

Definition 4.4. Under initial statistical framework and the later framework described, we define the following Imputation Based Distributional Treatment Effects:

1. The **Imputation Based Treatment Effect on Distribution (IB-TED)** is given by

$$IB - TED_p = Q_{1n,I}(p) - Q_{0n,I}(p) = F_{1I}^{-1}(p) - F_{0I}^{-1} \tag{4.6}$$

2. The **Imputation Based Quantile Treatment Effect (IB-QTE)** is de-

fined by

$$IB - QTE\left(GQ_{0n,I}(p)\right) = Y^*_{i_{01}} - Q_{0,nI}(p) = Y^*_{i01} - Y^*_{i00} \tag{4.7}$$

with $Y^*_{i1} = \begin{cases} Y_{i1} & if\ the\ unit\ i\ is\ treated \\ \tilde{Y}_{i1} & if\ the\ unit\ i\ not\ treated \end{cases}$ and $Y^*_{i0} = \begin{cases} Y_{i0} & if\ the\ unit\ i\ is\ not\ treated \\ \tilde{Y}_{i0} & if\ the\ unit\ i\ is\ treated \end{cases}$

3. The **Imputation Based Average Quantile Treatment Effect (IB-AQTE)** is defined by:

$$IB - AQTE\left(Q_{0n,I}(p)\right) = E\left(Y^*_{i1}\,|i \in GQ_{0n,I}(p)\right) - E\left(Y^*_{i0}\,|i \in GQ_{0n,I}(p)\right) \tag{4.8}$$

4. The **Imputation Based Median Quantile Treatment Effect (IB-MedQTE)** is defined by

$$IB{-}MedQTE\left(Q_{0n,I}(p)\right) = Med\left(Y^*_{i1}\,|i \in GQ_{0n,I}(p)\right) - Med\left(Y^*_{i0}\,|i \in GQ_{0n,I}(p)\right) \tag{4.9}$$

With Med as the median of the group in bracket in other term the 0.5^{th} quantile.

Remark 4.2. 1. In Definition 4.4, the equation (4.6) gives the global difference between the imputed distribution of potential outcome before and after the treatment. It is the same definition as $IB - mQTE_p$ but without the rank preservation assumption. Equation (4.6) is understood in this work as the effect on global distribution not the quantile effect. This estimator is the answer to the question how much the distribution of the potential

outcome changed after the treatment.

2. In Definition 4.4, the equation (4.7) basically compares two quantile groups by using the quantile value in the $CoCom$ distribution and its corresponding image (potential outcome obtain after treatment) in the $TrCom$ distribution.

3. In Definition 4.4, equations (4.8) and (4.9) are just different ways of comparing two groups: the initial quantile group $\{Y_{i0}^* \,|i \in GQ_{0n,I}(p)\}$ and its corresponding image $\{Y_{i1}^* \,|i \in GQ_{0n,I}(p)\}$ after the treatment using the mean and the median respectively as central tendency of comparison.

4. Given that those estimators are basically group estimators and converge to the true theoretical groups, they don't have a closed form. Therefore, it is difficult to study their theoretical properties so their empirical properties are studied using simulations in the second last section of this chapter.

5. For all the estimators defined, a restriction to a subset defined by covariates can be done to compute estimators only for a subpopulation. As example, the estimator defined in equation (4.6) can be computed like this $IB - mQTE_{p|X=x}$ to restrain the population to those units whose set of covariates is $X = x$.

4.3 Properties of IB-DTE Estimators

A recall of the convergence of sample quantiles and asymptotic normality is given first in this section before getting into the properties of $IB-DTE$ estimators defines with equation (4.4) and (4.5) which have closed forms.

4.3.1 Properties of Theoretical Quantiles

Lemma 4.1. Let Y be a random variable, suppose that $a \leq Y \leq b$, with $a, b \in \mathbb{R}$, then:

$$\mathbb{E}\left(e^{tY}\right) \leq e^{t\mu} e^{\frac{t^2(b-a)^2}{8}}$$

for every $t > 0$, with $\mu = \mathbb{E}(Y)$.

Proof. Let us recall first what is a convex function. A function g is convex if for any $x, y \in \mathbb{R}$ and $\alpha \in [0, 1]$ we have:

$$g\left(\alpha x + (1-\alpha)y\right) \leq \alpha g(x) + (1-\alpha)g(y)$$

Without loss of generality, let us assume that $\mu = \mathbb{E}(Y) = 0$. Since $a \leq Y \leq b$, Y can be written as a convex combination of a and b as follow: $Y = \alpha b + (1-\alpha)a$ with $\alpha = \frac{Y-a}{b-a}$.

Using convexity of $x \mapsto e^{tx}$, $t > 0$, we get:

$$e^{tY} = e^{t(\alpha b + (1-\alpha)a)} \leq \alpha e^{tb} + (1-\alpha)e^{ta} = \frac{Y-a}{b-a}e^{tb} + \frac{b-Y}{b-a}e^{ta}$$

Taking expectation ($\mu = 0$),

$$\mathbb{E}\left(e^{tY}\right) \leq \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta} = e^{g(u)}$$

where $u = t(b-a)$ and $g(u) = -\gamma u + log(1-\gamma-\gamma e^u)$ and $\gamma = \frac{-a}{b-a}$. The function g defined verify $g(0) = g'(0) = 0$ and $g''(u) \leq \frac{1}{4}$, $\forall u > 0$.

Using Taylor theorem, $\exists \xi \in [0, u]$ such that

$$g(u) = g(0) + ug'(0) + \frac{u^2}{2}g''(\xi) = \frac{u^2}{2}g''(\xi) \leq \frac{u^2}{8} = \frac{t^2(b-a)^2}{8}$$

Hence the result:

$$\mathbb{E}\left(e^{tY}\right) \leq e^{g(u)} \leq e^{\frac{u^2}{8}} = e^{\frac{t^2(b-a)^2}{8}} \quad \bullet$$

Lemma 4.2. Chernoff's method: Let Y be a random variable then $\forall \varepsilon > 0$,

$$\mathbb{P}\left(Y > \varepsilon\right) \leq Inf_{t \geq 0} \ e^{-t\varepsilon}\mathbb{E}\left(e^{tY}\right)$$

Proof. For every $t > 0$,

$$\mathbb{P}\left(Y > \varepsilon\right) = \mathbb{P}\left(tY > t\varepsilon\right) = \mathbb{P}\left(e^{tY} > e^{t\varepsilon}\right) \leq e^{-t\varepsilon}\mathbb{E}\left(e^{tY}\right)$$

The last inequality obtained using Markov's inequality [1]. Since the sequence is true for every $t > 0$, then it is true for the Inf, the result follows.

[1] See Appendix A2 for the details theorem of Markov's Inequality.

Lemma 4.3. Hoeffding's Inequality: Let $Y_1, \cdots, Y_n$ be i.i.d random variables such that $\mathbb{E}(Y_i) = \mu$ and $a \leq Y_i \leq b$ this $\forall i \in \{1, 2, ..., n\}$. Then for every $\varepsilon > 0$, we have:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \mathbb{E}(Y_i)\right| \geq n\varepsilon\right) = \mathbb{P}\left(\left|\bar{Y}_n - \mu\right| \geq \varepsilon\right) \leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

Proof. Without loss of generality, let's assume that $\mu = 0$. In the first step we have:

$$\begin{aligned} \mathbb{P}\left(\left|\bar{Y}_n\right| \geq \varepsilon\right) &= \mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) + \mathbb{P}\left(\bar{Y}_n \leq -\varepsilon\right) \\ &= \mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) + \mathbb{P}\left(-\bar{Y}_n \geq \varepsilon\right) \end{aligned} \tag{4.10}$$

Considering the first bloc and applying Chernoff's method (Lemma 4.2) and Markov's inequality we have: $\forall t > 0$,

$$\begin{aligned} \mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) = \mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq n\varepsilon\right) &= \mathbb{P}\left(e^{\sum_{i=1}^{n} Y_i} \geq e^{n\varepsilon}\right) \\ = \mathbb{P}\left(e^{t\sum_{i=1}^{n} Y_i} \geq e^{tn\varepsilon}\right) &\leq e^{-tn\varepsilon}\mathbb{E}\left(e^{t\sum_{i=1}^{n} Y_i}\right) \\ = e^{-tn\varepsilon}\prod_{i=1}^{n}\mathbb{E}\left(e^{tY_i}\right) &= e^{-tn\varepsilon}\left(\mathbb{E}\left(e^{tY_i}\right)\right)^n \end{aligned}$$

The Lemma 4.1 implies that $\mathbb{E}\left(e^{tY_i}\right) \leq e^{\frac{t^2(b-a)^2}{8}}$, then

$$\begin{aligned} \mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) &\leq e^{-tn\varepsilon}\left(E\left(e^{tY_i}\right)\right)^n \\ &\leq e^{-tn\varepsilon}e^{\frac{t^2n(b-a)^2}{8}} = g(t) \end{aligned}$$

The inf of the function g is obtain for $t_0 = \frac{4\varepsilon}{(b-a)^2}$ and $g(t_0) = e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$. So

$$\mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) \leq Inf_t\ g(t) = e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

Then

$$\mathbb{P}\left(\bar{Y}_n \geq \varepsilon\right) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

Following the same steps, it is easy to show that

$$\mathbb{P}\left(-\bar{Y}_n \geq \varepsilon\right) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}},$$

then replacing in equation 4.10, we have the final result:

$$\mathbb{P}\left(\left|\bar{Y}_n\right| \geq \varepsilon\right) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}} + e^{-\frac{2n\varepsilon^2}{(b-a)^2}} = 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

Theorem 4.1. Convergence of Sample Quantile: Let $p \in (0,1)$ a probability, F the CDF of the random variable Y and $Q(p)$ the theoretical quantile of order p. If for every $\varepsilon \in \mathbb{R}_+^*$ we have $F(Q(p)-\varepsilon) < p < F(Q(p)+\varepsilon)$; then the sample quantile $Q_{Y_\bullet}(p) = {F_n}^{-1}(p)$ is a convergent estimators of $Q(p)$. It is a convergence in probability given by :

$$Q_{Y_\bullet}(p) \xrightarrow[n\to+\infty]{P} Q(p) \tag{4.11}$$

Proof. To prove theorem 4.1, since it is convergence in probability, it is enough to show that

$$\mathbb{P}\left(|Q_{Y_\bullet}(p) - Q(p)| \geqslant \varepsilon\right) \xrightarrow[n\to+\infty]{} 0.$$

We have :

$$\begin{aligned}\mathbb{P}\left(|Q_{Y_\bullet}(p) - Q(p)| \geqslant \varepsilon\right) &= \mathbb{P}\left(Q_{Y_\bullet}(p) - Q(p) \geqslant \varepsilon\right) + \mathbb{P}\left(Q_{Y_\bullet}(p) - Q(p) \leqslant -\varepsilon\right) \\ &= \mathbb{P}\left(Q_{Y_\bullet}(p) \geqslant \varepsilon + Q(p)\right) + \mathbb{P}\left(Q_{Y_\bullet}(p) \leqslant Q(p) - \varepsilon\right) \\ &= (1) + (2)\end{aligned}$$

Taking the first block, expand and applying lemma 4.3 gives:

$$\begin{aligned}(1) &= \mathbb{P}\left(Q_{Y_\bullet}(p) \geqslant \varepsilon + Q(p)\right) = \mathbb{P}\left(p \geqslant F_n\left(\varepsilon + Q(p)\right)\right) \\ &= \mathbb{P}\left(p \geqslant \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leqslant \varepsilon + Q(p)\}}\right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}} \geqslant (1-p)\right) \\ &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}}\right) \geqslant (1-p) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}}\right)\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}} - \sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}}\right) \geqslant n(1-p) - \sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_i \geqslant \varepsilon + Q(p)\}}\right)\right) \\ &\leqslant 2e^{-2n(n(1-p) - nF(\varepsilon + Q(p))^2} \xrightarrow[n\to\infty]{} 0 \qquad (Hoeffing's\ inequality)\end{aligned}$$

Performing the same development in the second block gives:

$$
\begin{aligned}
(2) &= \mathbb{P}\left(Q_{Y_\bullet}(p) \leqslant Q(p) - \varepsilon\right) \\
&= \mathbb{P}\left(p \leqslant F_n(Q(p) - \varepsilon)\right) \\
&= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}} \geqslant p\right) \\
&= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}}) \geqslant p - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}})\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}} - \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}}) \geqslant np - \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{\{Y_i \leqslant Q(p)-\varepsilon\}})\right) \\
&\leqslant 2e^{-2n(np - nF(Q(p)-\varepsilon))^2} \xrightarrow[n\to\infty]{} 0 \qquad (Hoeffing's\ inequality)
\end{aligned}
$$

In conclusion, (1) goes to 0 as n increases, (2) goes to 0 as n increases then adding both of then we obtain the result which is

$$
\mathbb{P}\left(|Q_{Y_\bullet}(p) - Q(p)| \geqslant \varepsilon\right) \xrightarrow[n\to+\infty]{} 0.
$$

Theorem 4.2. Asymptotic Normality: If the density function f of Y exists and is continuous and strictly positive then for every $p \in (0,1)$, the estimator $Q_{Y_\bullet}(p)$ of the theoretical quantile $Q(p) = Q_Y(p)$ is asymptotically unbiased and normal (convergence in distribution) and the parameters of normal distribution are given by

$$
\lim_{n\to\infty} L\left(\sqrt{n}\left(Q_{Y_\bullet}(p) - Q_Y(p)\right)\right) = \mathcal{N}\left(0, \frac{p(1-p)}{f^2(Q_Y(p))}\right) \tag{4.12}
$$

or

$$\sqrt{n}\left(Q_{Y_\bullet}(p) - Q_Y(p)\right) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(Q_Y(p))}\right) \tag{4.13}$$

Proof. To show asymptotic normality expressed in 4.13, it is enough to show that

$$\frac{f\left(Q_Y(p)\right)}{\sqrt{p(1-p)}}\sqrt{n}\left(Q_{Y_\bullet}(p) - Q_Y(p)\right) \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,1) \tag{4.14}$$

Let

$$Z_n = \sqrt{n}\left(Q_{Y_\bullet}(p) - Q_Y(p)\right) = \sqrt{n}\left(Y_{n:k} - Q_Y(p)\right)$$

where $k = np + O(n^{1/2})$ as in the definition of a quantile of a specific order. Assuming without loss of generality that k is the rank of the value of Y giving the quantile $Q_{Y_\bullet}(p)$, for every $x \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\left(Z_n \leqslant x\right) &= \mathbb{P}\left(\sqrt{n}\left(Y_{n:k} - Q_Y(p)\right) \leqslant x\right) \\
&= \mathbb{P}\left(Y_{n:k} \leqslant Q_Y(p) + n^{-1/2}x\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}_{\left\{Y_i \leqslant Q_Y(p)+n^{-1/2}x\right\}} \geqslant k\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}_{\left\{Y_i \leqslant Q_Y(p)+n^{-1/2}x\right\}} - nF\left(Q_Y(p) + n^{-1/2}x\right) \geqslant k - nF\left(Q_Y(p) + n^{-1/2}x\right)\right) \\
P\left(Z_n \leqslant x\right) &= \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_{ni} \geqslant t_n\right) \qquad (*)
\end{aligned}$$

Where

$$V_{ni} = \mathbb{1}_{\left\{Y_i \leqslant Q_Y(p)+n^{-1/2}x\right\}} - F\left(Q_Y(p) + n^{-1/2}x\right), \qquad i = 1, 2, ..., n$$

and

$$t_n = \frac{1}{\sqrt{n}}\left(k - nF\left(Q_Y(p) + n^{-1/2}x\right)\right)$$

Now since $V_{ni} = 1 - F\left(Q_Y(p) + n^{-1/2}x\right)$ with probability $F\left(Q_Y(p) + n^{-1/2}x\right)$ and $V_{ni} = -F\left(Q_Y(p) + n^{-1/2}x\right)$ with the probability $1 - F\left(Q_Y(p) + n^{-1/2}x\right)$; And $F\left(Q_Y(p) + n^{-1/2}x\right) = F\left(Q_Y(p)\right) + n^{-1/2}xf(Q_Y(p)) + O(n^{-1/2})$ (Taylor expansion assuming that F is differentiable), it follows that:

$$\mathbb{E}\left(V_{ni}\right) = 0 \; and \; \mathbb{E}\left(V_{ni}^2\right) = p(1-p) + O(n^{-1/2})$$

and

$$\begin{aligned} t_n &= \frac{1}{\sqrt{n}}\left(k - nF\left(Q_Y(p)\right) - n^1 n^{-1/2} x f(Q_Y(p)) - nO(n^{-1/2})\right) \\ &= \frac{k}{\sqrt{n}} - \frac{nF\left(Q_Y(p)\right)}{\sqrt{n}} - xf(Q_Y(p)) + \frac{1}{\sqrt{n}}O(n^{1/2}) \\ &= \frac{np - np}{\sqrt{n}} - xf(Q_Y(p)) + O(1) = -xf(Q_Y(p)) + O(1) \end{aligned}$$

Therefore, using the central limit theorem and Slutsky theorem[2] in (*) we may conclude that as $n \to +\infty$,

$$\mathbb{P}\left(Z_n \leqslant x\right) = \mathbb{P}\left(\frac{1}{\sqrt{np(1-p)}}\sum_{i=1}^{n} V_{ni} \geqslant \frac{1}{\sqrt{p(1-p)}}t_n\right) \xrightarrow{n\to+\infty} 1 - \Phi\left(\frac{-xf(Q_Y(p))}{\sqrt{p(1-p)}}\right)$$

and $1 - \Phi\left(\frac{-xf(Q_Y(p))}{\sqrt{p(1-p)}}\right) = \Phi\left(\frac{xf(Q_Y(p))}{\sqrt{p(1-p)}}\right)$ as Φ is the CDF of a normal distribution of mean 0 and variance 1. Which implies finally that equation 4.14 is true hence the result.

[2]See Appendix A2 for the formulation and proof of these two theorems.

4.3.2 Convergence of IB-DTE

Proposition 4.1. Convergence of QTE and $IB-mQTE$: Let $p \in (0,1)$ a probability, if for every $\varepsilon \in \mathbb{R}_+^*$ we have $F_0(Q_0(p)-\varepsilon) < p < F_0(Q_0(p)+\varepsilon)$ and $F_1(Q_1(p)-\varepsilon) < p < F_1(Q_1(p)+\varepsilon)$ then the sample quantiles treatment effects QTE_p and $IB-mQTE_p$ are convergent estimator of $TrQTE_p$, meaning that:

1. For QTE_p :

$$QTE_p \xrightarrow[n\to+\infty]{P} TrQTE_p \tag{4.15}$$

2. For $IB-mQTE_p$:

$$IB-mQTE_p \xrightarrow[n\to+\infty]{P} TrQTE_p \tag{4.16}$$

Proof. Basically, Definitions 4.1, 4.2 and 4.3 as well as Theorem 4.1 are used to obtain the result.

1. Using definition 4.1, $TrQTE_p = Q_1(p) - Q_0(p)$ difference of two theoretical quantile and from definition 4.2, $QTE_p = Q_{n_1}(p) - Q_{n_0}(p) = F_{r_1}^{-1}(p) - F_{r_0}^{-1}(p)$ difference of two sample quantile, combined with the results obtain from theorem 4.1, we get the following convergence:

$$Q_{n_1}(p) \xrightarrow[n\to+\infty]{P} Q_1(p) \tag{4.17}$$

and

$$Q_{n_0}(p) \xrightarrow[n\to+\infty]{P} Q_0(p) \tag{4.18}$$

Using additivity of convergence in probability, computing (4.17) − (4.18) we get

$$Q_{n_1}(p) - Q_{n_0}(p) \xrightarrow[n\to+\infty]{P} Q_1(p) - Q_0(p)$$

hence the result

$$QTE_p \xrightarrow[n\to+\infty]{P} TrQTE_p$$

2. By definition 4.1, $TrQTE_p = Q_1(p) - Q_0(p)$ difference of two theoretical quantile and by definition 4.3, $IB{-}mQTE_p = Q_{1n,I}(p) - Q_{0n,I}(p)$, difference of two sample quantile with the distribution Y_0 and Y_1 modified by imputation, combined with the results of theorem 4.1, follows this convergence result:

$$Q_{1n,I}(p) \xrightarrow[n\to+\infty]{P} Q_1(p) \tag{4.19}$$

and

$$Q_{0n,I}(p) \xrightarrow[n\to+\infty]{P} Q_0(p) \tag{4.20}$$

Using additivity of convergence in probability, computing (4.19) − (4.20) leads to

$$Q_{1n,I}(p) - Q_{0n,I}(p) \xrightarrow[n\to+\infty]{P} Q_1(p) - Q_0(p)$$

hence the result follow

$$IB - mQTE_p \xrightarrow[n\to+\infty]{P} TrQTE_p.$$

4.3.3 Asymptotic Normality of IB-DTE

Proposition 4.2. Asymptotic Unbiasedness and Normality: If the density function f_0 and f_1 of Y_0 and Y_1 exist and are continuous and strictly positives then for every $p \in (0,1)$, the estimators QTE_p and $IB - mQTE_p$ of the theoretical quantile treatment effect $TrQTE_p$ are asymptotically unbiased and normal. Mathematically we have:

1. For QTE_p :

$$QTE_p \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(TrQTE_p, \frac{p(1-p)}{n_1 {f_1}^2(Q_1(p))} + \frac{p(1-p)}{n_0 {f_0}^2(Q_0(p))}\right) \tag{4.21}$$

2. For $IB - mQTE_p$:

$$IB - mQTE_p \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(TrQTE_p, \frac{p(1-p)}{n {f_1}^2(Q_1(p))} + \frac{p(1-p)}{n {f_0}^2(Q_0(p))}\right) \tag{4.22}$$

Proof. This proof uses the Definitions 4.1, 4.2, 4.3 and the result of theorem 4.2.

1. From definition 4.1 we have, $TrQTE_p = Q_1(p) - Q_0(p)$ difference of two theoretical quantile and from definition 4.2 we have, $QTE_p = Q_{n_1}(p) - Q_{n_0}(p) = F^{-1}_{r_1}(p) - F^{-1}_{r_0}(p)$ difference of two sample quantile.

   Using previous definitions and the results obtain from theorem 4.2, we get the following convergence in distribution:

$$Q_{n_1}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_1(p),\ \frac{p(1-p)}{n_1 f^2(Q_1(p))}\right) \tag{4.23}$$

and

$$Q_{n_0}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_0(p),\ \frac{p(1-p)}{n_0 f^2(Q_0(p))}\right) \tag{4.24}$$

Then, using additivity of convergence in distribution and properties of normal distribution, computing (4.23) − (4.24) the result is the following:

$$Q_{n_1}(p) - Q_{n_0}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_1(p) - Q_0(p),\ \frac{p(1-p)}{n_1 f^2(Q_1(p))} + \frac{p(1-p)}{n_0 f^2(Q_0(p))}\right)$$

hence the result (4.22).

2. From definition 4.1, $TrQTE_p = Q_1(p) - Q_0(p)$ is the difference of two theoretical quantile and by definition 4.3, $IB - mQTE_p = Q_{1n,I}(p) - Q_{0n,I}(p)$, which is the difference of two sample quantile with the distribution Y_0 and Y_1 modified by imputation. Using previous definitions and the results obtain from theorem 4.2, we get the following convergence in distribution:

$$Q_{1n,I}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_1(p),\ \frac{p(1-p)}{nf^2(Q_1(p))}\right) \tag{4.25}$$

and

$$Q_{0n,I}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_0(p),\ \frac{p(1-p)}{nf^2(Q_0(p))}\right) \tag{4.26}$$

Then, using additivity of convergence in distribution and properties of normal distribution, computing (4.25) − (4.26) the result is the following:

$$Q_{1n,I}(p) - Q_{0n,I}(p) \xrightarrow[n\to+\infty]{d} \mathcal{N}\left(Q_1(p) - Q_0(p),\ \frac{p(1-p)}{nf^2(Q_1(p))} + \frac{p(1-p)}{nf^2(Q_0(p))}\right)$$

hence the result (4.23).

Remark 4.3. The convergence and asymptotic normality are proved only for estimators given with equation 4.4 and 4.5 because they have a closed form. For estimators defined in equations 4.6, 4.7, 4.8 and 4.9 not defined with a closed form, the properties are studied empirically in the next section using simulations and bootstrap procedure.

4.4 Asymptotic Properties: Simulations

Empirical properties of estimators defined previously are tested here. This section is much more important for estimators defined in Definition 4.4 since they don't have a closed form to be used to show their properties theoretically. Simulations are performed under random assignment (MCAR missingness) hypothesis and under deterministic assignment (MAR or NMAR missingness) hypothesis.

4.4.1 Algorithm and Assumptions

The framework, hypotheses and assumption defined in Chapter 3, Section 4.1 are still the same used here. An hypothetical situation where all the parameters are mastered is created. As example, from the simulated data, the true quantile $TrQTE_p$ can be easily computed so that the average bias will be computed directly. Here again a bootstrap procedure of 1000 replications is done to obtain the standard deviations and a confidence interval if needed. Simulation is done

using two hypothesis, under MCAR assumption (RA) then under MAR assumption (Selection on observable or unobservable). For each missingness assumption, a second hypothesis is done regarding the rank preservation. The percentage of treated is assumed here again to be 40% and the corresponding quantiles are decile.

As in the first chapter, each imputation method generates an IB estimator. Imputation methods used to complete the sample are mean imputation (general and conditional), random imputation (general and hot deck), deterministic linear regression (Det LM), k-Nearest Neighbor (using two different distances $V1$ and $V2$), random linear regression, multiple imputation using multiple chained equations (MI-MICE), maximum likelihood imputation assuming normal distribution and no distribution and finally the quantile regression imputation. The classical QTE in the literature is computed as well with the sample reduced. All those estimators are compared to the true QTE by computing the average bias.

4.4.2 Results and Comments

Given the density of results, six sample sizes (N=50, 100, 200, 500, 1000, 2000), decile as corresponding quantiles, 12 different imputation methods generating 12 different $IB-DTE$, only one table is presented for each assumption chosen. The summary given in the comments is from all the results and the table presented is just to visualize how the results looks like. Only the sample size $N = 2000$ is presented here.

4.4.2.1 Random Assignment Hypothesis Results (MCAR)

RA under rank preservation assumption: Effect on the whole distribution ($IB - mQTE_p$): Under rank preservation assumption, the simulations show that the bias of all estimators (Classic QTE and $IB - mQTE_p$) is decreasing too slowly and will probably never get to zero. Therefore, it is clear that all estimators are biased and the best one will be the one with the smallest bias and smallest variance (small and convergent variance). From simulation results, it is clear that $IB - mQTE_p$ estimators are far better than Classic QTE in almost all the cases except few cases related to sample size and small quantiles. In fact, for small quantiles (1st and 2nd decile) classic QTE performs as well as the $IB - mQTE_p$, it is even better for the 1st decile than all $IB - mQTE_p$ estimators especially for small sample ($N < 200$). Out of those specific cases highlighted, $IB - mQTE_p$ estimators are better than classic QTE. For example, the k-NN imputation which is among the best $IB - mQTE_p$ gives the best results for large sample ($N > 200$) and no matter the quantile selected.

The k-NN $IB - mQTE_p$ is always among the three best estimators in our simulations. For small samples, PSM $IB - mQTE_p$ and k-NN $IB - mQTE_p$ are sharing the first and the second position in term of estimators with the smallest bias. For example, for the sample size of 50 and the 8th decile, the bias for classic QTE is 4.6 while for PSM $IB - mQTE_p$ it is 2.5 and for k-NN it is 6.5. For the same sample size and for the 5th decile (the median), the bias is 3.5 for PSM $IB - mQTE_p$ and 7 for k-NN.

Table 4.1: **Average bias of $IB - DTE$ Estimators (MCAR & Rank preservation assumption)**

Summary of results			Average Bias of Estimators for N=2000 ($\hat{\theta}$-$TrueQTE$)								
			1st dec	2nd dec	3rd dec	4th dec	5th dec	6th dec	7th dec	8th dec	9th dec
Class QTE			7.87	16.84	20.6	20.72	17.27	18.28	12.60	7.46	1.31
IB-DTE Results	Mean Imp	Gen	24.85	23.82	16.70	14.13	15.04	8.35	8.23	4.42	-1.48
		Condi	24.85	23.82	16.52	13.93	14.81	8.12	8.00	8.42	-1.48
	Rand Imp	Gen	23.91	22.59	17.08	14.57	15.32	8.82	8.69	10.16	-0.58
		Hot deck	23.93	22.28	16.65	14.19	15.13	8.51	8.97	10.38	-0.59
	Det LM		23.97	22.12	16.51	13.70	14.56	8.01	8.05	9.55	-1.33
	k-NN Imp	V1	21.41	20.23	14.94	12.65	13.78	6.79	7.06	8.69	-1.38
		V2	20.99	19.83	14.54	12.17	13.17	6.52	6.69	8.11	-2.20
	Rand LM		22.96	21.97	16.37	13.64	14.63	8.20	8.11	9.45	-0.63
	MI-MICE		23.49	22.24	16.44	13.73	14.70	8.30	8.45	10.21	-0.81
	ML Imp	Normal	22.93	21.98	16.74	14.07	14.52	8.07	8.43	9.63	-0.59
		Dist free	22.91	21.63	16.19	13.86	14.92	8.01	8.48	9.95	-0.68
	QR Imp		24.85	23.80	19.26	17.25	18.10	1.73	-1.03	-0.33	-11.67

RA without rank preservation assumption: $IB-QTE$: Without assuming rank preservation, each unit in a quantile before the treatment will be followed, after the treatment and the two groups will be compared to get the true quantile treatment effect based on the distribution of the potential outcome before the treatment. In other words, units in the quantile will be grouped to form the comparison group after treatment: this is how we define the true Quantile Treatment Effect ($IB - QTE$). From the result of simulation, outcomes are a bit mitigated at first sight, they show globally that $IB - QTE$ are as good as the classic QTE. For small quantiles no matter the sample size, classic QTE are slightly better than most of $IB-QTE$ estimators but for quantile above median, $IB - QTE$ are far better. For large samples, in general classic QTE are better than $IB - QTE$ for extreme quantile (1st decile and 9th decile quantile) but for the other ones in between they are not better. The $IB - QTE$ that are good on average are ML $IB - QTE$, k-NN $IB - QTE$ and MI $IB - QTE$ compared to classic QTE.

Table 4.2: **Average bias of $IB-DTE$ Estimators (MCAR & Without rank preservation assumption)**

Summary of results			Average Bias of Estimators for N=2000 ($\hat{\theta}$-$TrueQTE$)								
			1st dec	2nd dec	3rd dec	4th dec	5th dec	6th dec	7th dec	8th dec	9th dec
Class QTE			6.95	15.74	19.99	19.71	16.70	17.69	12.09	7.47	1.05
$IB-DTE$ Results	Mean Imp	Mean	31.62	24.70	10.63	27.88	13.86	23.40	-6.69	5.80	-9.84
		Condi	32.00	24.86	18.48	26.49	24.55	15.67	6.88	5.57	-10.09
	Rand Imp	Gen	29.68	37.76	17.95	19.55	5.88	3.68	18.42	-9.90	-1.77
		Hot deck	34.78	42.73	35.28	28.11	13.13	-2.14	-11.25	-5.17	-26.18
	Det LM		33.03	11.48	19.82	26.35	14.91	3.71	5.98	7.10	0.98
	k-NN Imp	V1	41.93	24.49	13.38	7.81	24.29	-8.62	-6.65	0.14	-14.73
		V2	30.11	28.91	25.93	9.77	6.74	-0.96	-18.50	0.66	-15.96
	Rand LM		46.58	31.65	23.79	28.91	-10.99	7.57	-3.50	16.04	-14.71
	MI-MICE		22.06	20.04	0.75	16.25	15.74	3.98	3.29	-12.14	-1.81
	ML Imp	Normal	48.73	36.32	22.54	15.04	8.24	-1.54	-1.78	1.05	-25.22
		Dist free	39.18	35.46	26.02	26.80	20.66	3.40	15.15	-3.17	16.43
	QR Imp		53.38	54.78	42.21	42.02	50.39	38.82	-33.35	-39.53	-41.79

4.4.2.2 Selection on Observable Results (MAR or NMAR)

Considering that the assignment process depends on a given variable or combinations of some given variables. When a threshold is established, the population is divided into two parts and the group below the threshold is treated while the group above is not (in he neighborhood of the threshold) then they are compared with each other.

MAR or NMAR under rank preservation assumption: Effect on the whole distribution ($IB - mQTE_p$) Assuming NMAR or MAR in simulation and under rank preservation assumption, the simulations results show that for small samples ($N < 200$), IB-TED produced using deterministic and random imputations are as good as classical imputations for big quantiles but better for small quantiles. Taking for example the $N = 50$, the bias of the IB-TED for deterministic linear regression is smaller than the bias due to classic QTE for the first eight deciles and only bigger for the 9th decile. For the bigger sample size ($N > 200$), the k-NN IB-TED is the best method for almost all the quantiles. In the case of random assignment, on average the best imputation methods are deterministic linear regression IB-TED for small sample and k-NN IB-TED for large samples.

Table 4.3: **Average bias of $IB-DTE$ Estimators (MAR or NMAR & Rank preservation assumption)**

Summary of results			Average Bias of Estimators for N=2000 ($\hat{\theta}$-$TrueQTE$)								
			1st dec	2nd dec	3rd dec	4th dec	5th dec	6th dec	7th dec	8th dec	9th dec
Class QTE			-3.48	-9.18	-18.62	-2.41	-19.22	-18.44	-10.65	-11.21	-1.38
$IB-DTE$ Results	Mean Imp	Mean	-1.10	-2.00	-9.29	-11.86	-10.95	-17.64	-17.76	-17.92	-27.29
		Condi	-1.10	-2.00	-9.11	-11.70	-10.77	-17.43	-17.57	-17.92	-7.29
	Rand Imp	Gen	-3.32	-3.94	-9.43	-11.45	-10.95	-17.63	-17.41	-15.40	-25.94
		Hot deck	-3.01	-3.05	-8.68	-11.21	-10.43	-17.38	-17.01	-15.26	-26.36
	Det LM		-1.10	-2.38	-7.95	-10.49	-9.92	-16.88	-17.13	-15.23	-26.20
	k-NN Imp	V1	-1.26	-2.28	-7.27	-10.13	-9.62	-16.25	-16.06	-14.42	-24.77
		V2	-1.00	-1.69	-7.01	-9.74	-8.96	-15.79	-15.64	-13.97	-24.21
	Rand LM		-2.60	-3.44	-8.83	-11.00	-10.19	-16.71	-16.54	-14.75	-25.32
	MI-MICE		-1.43	-2.39	-7.78	-10.55	-10.15	-17.05	-17.05	-15.32	-26.15
	ML Imp	Normal	-2.40	-3.29	-8.53	-10.98	-10.23	-16.63	-16.62	-14.56	-25.08
		Dist free	-2.24	-3.08	-8.33	-10.90	-10.12	-16.76	-16.96	-14.72	-25.91
	QR Imp		-1.10	-1.90	-6.77	-6.33	5.23	-0.03	-0.39	0.36	-10.75

MAR or NMAR without rank preservation assumption: $IB-QTE$

Assuming that the assignment process is NMAR or MAR, the true QTE called here IB-QTE is estimated using imputation method. The idea here is to follow each quantile after the treatment and obtain the treatment effect on that quantile. Comparison between IB-QTE and classic QTE shows that results are quite clear. In some cases, classic QTE is better while in others it is IB-QTE which is better. The constance is that IB-QTE is far better than classic QTE no matter the sample size for middle deciles (3^{rd}, 4^{th}, 5^{th}, 6^{th} and 7^{th} deciles). Therefore, under NMAR or MAR and without rank preservation assumption, IB-QTE estimators are better in estimating the middle decile of the distributional effect of a treatment. The chosen estimators are k-NN IB-QTE and ML IB-QTE for small samples and k-NN IB-QTE, ML IB-QTE and MI IB-QTE for large samples.

Table 4.4: **Average bias of $IB-DTE$ Estimators (MAR or NMAR & Without rank preservation assumption)**

Summary of results			Average Bias of Estimators for N=2000 ($\hat{\theta}$-$TrueQTE$)								
			1st dec	2nd dec	3rd dec	4th dec	5th dec	6th dec	7th dec	8th dec	9th dec
Class QTE			-3.48	-9.18	-18.62	-22.41	-19.22	-18.44	-10.65	-11.21	-1.38
$IB-DTE$ Results	Mean Imp	Mean	6.38	-0.56	-23.61	-13.83	-16.39	-13.45	-24.32	-19.60	-34.68
		Condi	6.46	-0.59	-1.11	-10.18	-2.01	-17.34	-11.22	-19.08	-35.04
	Rand Imp	Gen	17.88	12.14	-5.68	-6.32	-21.33	-6.99	-33.78	-4.49	-34.95
		Hot deck	-4.20	22.30	10.06	-3.83	-14.73	-16.24	-7.91	-18.99	-34.50
	Det LM		3.33	8.18	-11.63	-9.56	-0.65	-19.67	-15.66	-26.31	-37.19
	k-NN Imp	V1	6.27	-5.50	13.79	-2.79	-30.11	-10.85	-11.73	1.86	-27.25
		V2	-7.66	11.14	-22.65	-13.51	2.43	-8.63	-6.93	-16.00	-35.99
	Rand LM		10.03	-6.00	-2.21	0.41	-25.83	-19.05	-12.79	-19.43	-41.44
	MI-MICE		0.49	-6.05	-11.32	-15.94	-1.80	-18.93	-10.38	-18.32	-28.97
	ML Imp	Normal	11.43	14.82	-18.93	-5.81	-22.88	-27.47	-22.20	-43.63	-59.66
		Dist free	21.90	6.17	-17.36	-10.63	-9.92	-32.59	-8.04	-47.82	-52.11
	QR Imp		39.79	38.18	16.34	21.93	23.10	-42.58	-35.77	-47.03	-55.38

4.5 Applications

After simulations, where the results showed that IB-TED and IB-QTE estimators can perform as well as classic quantile treatment effect estimators otherwise better in some cases, the next step is to apply these results to real set of data since simulation are always questionable.

4.5.1 Description of the program and Data

Here, the Lalonde (1986) data set is considered for application. This database is used to have a benchmark of comparison of our results with the results obtain by the classic estimators and Firpo (2007). The structure and detail on the database are the same as explain in Chapter 3 Section 5.1. The only difference now is the problem: we would like to assess how the training had affected the distribution of the revenue of people using quantiles of distributions.

4.5.2 Results and Comments

Applying our estimators to Lalonde data, the three best IB distributional effects are considered and compared to classic QTE without completing data and to Firpo (2007)'s results on the same data set. Combining bootstrap to imputation methods and applying the empirical quantile shows that estimation of the median impact using IB-TED estimators is closer to the ATE estimator than the Firpo's result and classical QTE.

Table 4.5: **IB-DTE Using Lalonde's Subsample**

Summary of results			Treatment Effect estimators for each decile (N=350)								
			1st dec	2nd dec	3rd dec	4th dec	5th dec	6th dec	7th dec	8th dec	9th dec
Class QTE			0	0	943.0	1169.5	1093.5	1452.3	1802.5	2273.1	3197.8
Class QTE & bootstrap			0	4.6	929.0	1324.9	1184.3	1493.5	2019.2	2248.1	3121.7
Firpo's QTE			0	0	711.0	21.0	1927.0	3879.0	4517.0	6027.0	5530.0
$IB-DTE$ Results	Mean Imp: Gen		1765.3	1858.3	1822.2	1805.2	1805.2	1805.2	1805.2	1746.2	1789.0
	Rand Imp	Gen	1692.0	1742.3	1774.3	1771.6	1810.8	1828.8	1846.2	1864.5	1871.5
		Hot deck	1715.8	1726.9	1787.8	1789.9	1800.0	1814.0	1829.8	1873.1	1906.3
	Det LM		1553.2	1526.0	1526.9	1532.9	1554.4	1571.8	1618.6	1714.6	1789.0
	k-NN Imp	V1	1782.2	1793.8	1776.3	1733.2	1702.4	1693.2	1747.6	1788.9	1811.8
		V2	1724.9	1771.5	1744.8	1603.9	1630.1	1665.2	1685.7	1732.0	1779.2
	Rand LM		1482.1	1527.8	1564.2	1603.9	1630.1	1665.2	1685.7	1732.0	1779.2
	MI-MICE		1675.8	1631.7	1619.9	1601.9	1597.8	1598.8	1599.5	1642.4	1707.1
	ML Imp	Normal	1455.4	1521.7	1546.8	1584.5	1610.1	1638.7	1669.8	1708.4	1757.6
		Dist free	1494.8	1526.8	1569.0	1588.2	1623.2	1648.3	1668.6	1706.8	1771.1
	PSM Imp		1282.6	1259.2	1213.0	1164.6	1014.9	620.4	362.5	287.6	220.7

In fact, looking at classic QTE estimators under rank preservation assumption, the effect of the program is increasing with the deciles. For the first and second deciles, the effect is 0 then for the 3rd decile effect is $ 943. The median effect is $ 1093.5 which is far from the Lalonde ATE benchmark ($ 1794). Adding bootstrap on it did not change the result much, the largest effect being for the 8th and 9th deciles respectively $ 2273 and $ 3197 which is basically explosive and too much. Analysis of Firpo (2007)'s results show also that the effect is increasing with deciles after the 4th decile. The effect is 0 for the 1st and 2nd deciles which is not likely to happen practically, then $ 711 for the 3rd decile, $ 21 for the 4th decile meaning that some effect were probably negative. Then comes the median effect which is $ 1927 quite close to $ 1794 the ATE which is good, but after that the effect becomes explosive. The effect is $ 3879 for the 6th decile, $ 4517 for the 7th decile to end at $ 5530 for the 9th which is again not likely to happen. Firpo (2007)'s method may perform well only for median.

Now if we take one of our best IB-TED estimator (k-NN) under rank preservation assumption, the effect is quite uniform across deciles with an average difference of $ 30. The tendency is the following: effect of the program is larger for the tail of the distribution of potential outcome ($ 1793.8 for the 2nd decile and $ 1811.8 for the 9th decile) and quite stable and close to the ATE for the middle deciles ($ 1702.4 for median effect and $ 1693.2 for 6th decile). This pattern of results is almost the same for all IB-TED estimators computed using imputation methods. In conclusion, explosiveness of Firpo's results and classic QTE show

that estimators constructed using non parametric approach and using incomplete data set are not convergent practically. They basically express the fact that the effect of training program on earning increases with the deciles meaning that the more you earned before the program, the more the program will have an effect on you. This is counter intuitive given that theoretically effect of the training is most likely to be greater for those who were earning less. From the result using IB-TED, the effect is more stable across the distribution of earnings and bigger for people earning less and people earning more (tail of the distribution) which is more likely to happen than the explosive effect of the training.

4.6 Summary

In this Chapter, seminal results of Chapter 3 were used to derive the Imputation Based Distributional Treatment Effect ($IB-DTE$). As recall, it is an estimation of the distribution of the effects of a treatment by estimators derived using imputation methods as a means of estimation of counterfactual. Six new estimators of the distributional treatment effects convergent theoretically and empirically were derived. The comparison made between those estimators and the existing ones in the literature showed that $IB-DTE$ perform better in terms of unbiasedness and convergence. Also, using real data, they are more stable than existing ones, small variance therefore small confidence interval.

After deriving estimators $IB-ATE$ in Chapter 3 and $IB-DTE$ in Chapter 4, the next chapter proposes a testing procedure to determine if the effects computed

are significant or not.

CHAPTER 5

TESTING THE HYPOTHESIS "NO EFFECTS" WITH POTENTIAL OUTCOME RECONSTRUCTED USING IMPUTATION

5.1 Introduction

In this chapter, three approaches for testing the hypothesis "No Effects" using Multiple Testing Procedure (MTP) are presented. Inspired from works done by Crump et al. (2008) and Kaplan and Goldman (2013), the idea here is to compare the distributions (reconstructed by imputation or not) of the potential outcome before and after the treatment in a pointwise manner to identify in which section they are different, otherwise the effect is significant. Since MTP needs pointwise test to be performed upstream, three new pointwise tests are presented in this chapter to achieve our goals.

5.2 Proposed solution: Multiple Testing Procedure

In our research, the marginal distributions of potential outcome before and after the treatment are used to test the significance of the treatment effect. Hypothesis of randomization is assumed if not, we consider that imputation can be used to reconstruct the full distribution of the two potential outcomes[1] or if not the restricted distribution can be used for the same purpose. The test can be applied in any of those configuration. Following the work done by Goldman and Kaplan (2017), the focus is on a Multiple Testing Procedure (MTP) across quantile or CDF to identify which subgroup of the population in term of their value for the potential outcome is not responding very well to the treatment assigned. Analysis starts when Average Treatment Effect (ATE) is computed using different methods (even IB approach). Comparing the distributions before and after the treatment says globally if the distribution has changed or not, meaning if the effect of the treatment was significant or not but does not quantify the effect. It does not tell us if the effect was the ATE computed prior to the test. But in case the interest is in the value of effect, the approach which is proposed in this research section is to shift the distribution of the potential outcome before the treatment by ATE, $Y_0 + ATE$ then compare the distribution of $Y_0 + ATE$ and Y_1. If the distributions are the same at a given level of significance, then the treatment was significant at this level. If the distributions are not the same, then a

[1] Basic idea of chapter 3

multiple testing procedure is required to identify which section of the distribution does not respond favorably to the test. Another way of doing it is to compare the distribution of Y_0 to the distribution Y_1 assuming that the later is the first one plus the ATE computed (taking into account heterogeneity of effects). If they are the same, there is no effect but if they are not the same there is an unknown effect that has to be computed using different approaches especially imputation-based approach[2].

Following pioneer work of Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1939, 1948) and MTP works done by Goldman and Kaplan (2017)including their extension of K-S test to MTP, this research proposes three MTP approach to test the significance of treatment effect in Randomized Control Trial (RCT) and in other assignment process (with data reconstructed using imputation if necessary). In the first case, unconfoundedness and ignorability is assumed to perform a test of comparison of the distribution. In the second case, after using imputation methods to reconstruct all the sample data to eliminate the issue of selection bias, the same MTP is applied without any assumption made and for all assignment process including the framework of Regression Discontinuity Design (RDD). There is a need to test in a jointly manner the hypotheses because doing it independently at a specific threshold then combining the critical area will lead to a test in which the significance is higher than the initial threshold and this is not the objective.

[2]The approaches include IB-ATE, RA, PSM etc.

5.3 Three New Approaches of MTP for the Hypothesis "No Effects"

This section formalizes the approach presented in this research testing generally the hypothesis "No effects" in the framework of impact evaluation. Since the MTP is used to achieve our goals, there is a need to define properly pointwise procedure that can be performed individually then use it to perform the MTP test. The added value here is not strictly a new MTP procedure but new pointwise tests helping in the MTP process. Three pointwise tests used here are described in the next sections:

- MTP using CDF function, basically an extension of K-S test to multiple hypothesis;
- MTP using empirical quantile function;
- MTP using mean of potential outcome comparison across quantile groups.

5.3.1 Statistical Framework

Let's assume that we have two independent potential outcomes, Y_0 before the treatment and Y_1 after the treatment. Their corresponding CDFs are respectively F_0 and F_1. The tests performed here are two sample two sided tests. Given that the MTP approach relies on the fact that a set of probability should be provided in the beginning, a pointwise test should be performed for each single hypothesis

in the multiple test formulation. Therefore, for each approach presented below, a pointwise test is presented for obtention of p-values necessary.

5.3.2 MTP using CDF (CDF)

The hypothesis defined here is as follows:

$$\begin{cases} H_0 : \{H_{0r} : F_0(r) = F_1(r),\ with\ r \in \mathbb{R}\} \\ H_1 : \{H_{1r} : F_0(r) \neq F_1(r),\ with\ r \in \mathbb{R}\} \end{cases} \tag{5.1}$$

The corresponding pointwise test which is $H_{0r},\ \forall r \in \mathbb{R}$ should be performed and the p-value saved for the MTP algorithm later. An estimator of $F_j(r)$ is given by:

$$\hat{F}_j(r) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_{ji} \leqslant r\}}; \quad j \in (0, 1) \tag{5.2}$$

with $\mathbb{1}_{\{.\}}$ as indicatrice function taking 1 if the expression inside the brackets is true and 0 otherwise, n_j the sample size of the empirical distribution Y_j. The quantity $\hat{F}_j(r)$ is a proportion and as a proportion, the pointwise test $H_{0r} : F_0(r) = F_1(r),\ \forall r \in \mathbb{R}$ can be seen as a test of comparison of two proportions from two different samples. To obtain the p-values, the proportions

$$\hat{F}_0(r) = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbb{1}_{\{Y_{0i} \leqslant r\}}\ and\ \hat{F}_1(r) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{1}_{\{Y_{1i} \leqslant r\}} \tag{5.3}$$

will be compared using a classical test for comparison of proportions.

Given $r \in \mathbb{R}$, the statistic computed for this test is as follows:

$$Z = \frac{\hat{F}_0(r) - \hat{F}_1(r)}{\sqrt{\hat{F}(r)\left(1-\hat{F}(r)\right)\left(\frac{1}{n_0}+\frac{1}{n_1}\right)}}; \quad with\ \hat{F}(r) = \frac{\sum_{i=1}^{n_0} \mathbb{1}_{\{Y_{0i} \leq r\}} + \sum_{i=1}^{n_1} \mathbb{1}_{\{Y_{1i} \leq r\}}}{n_0 + n_1} \tag{5.4}$$

It is a $Z - statistic$ following a normal distribution, $\hat{F}(r)$ is the proportion of values less than r in the pooled sample of the two distributions. The p-value will be obtained by reading the table of a normal distribution of mean 0 and standard deviation 1.

5.3.3 MTP using Quantile Function (QF)

The definition of hypothesis in this MTP using quantile function is given below.

$$\begin{cases} H_0 : \{H_{0\tau} : F_0^{-1}(\tau) = F_1^{-1}(\tau),\ with\ \tau \in (0,1)\} \\ H_1 : \{H_{1\tau} : F_0^{-1}(\tau) \neq F_1^{-1}(\tau),\ with\ \tau \in (0,1)\} \end{cases} \tag{5.5}$$

Since to test this joint hypothesis the p-values for the pointwise test are needed, it is important to test first the hypothesis $H_{0\tau} : F_0^{-1}(\tau) = F_1^{-1}(\tau)$ for every τ considered. Basically, to obtain the p-value for the comparison of two sample quantiles, a bootstrap procedure is needed.

The procedure here to obtain the p-values for a pointwise test for each quantile follows works done by Wilcox et al. (2013). They basically used bootstrap procedure to test whether two quantiles are equal or not. For a given quantile, they generated several samples from the two sample distributions (independently) at

the same time. For each row of generation, they computed the quantile in each distribution and made the difference. The difference of quantile in each row is stored in a vector. At the end, they tested if the differences stored in the vector is significantly different from 0 or not to obtain the p-value using classical test of comparison of means.

To summarize, assuming that we have an empirical distribution of Y_0 and Y_1, here are the steps to undertake to obtain the p-values for the test in the difference of quantiles of order τ:

- **Step 1**: Generate two sub samples, one for Y_0 and the second one for Y_1 of a size big enough to keep the global structure of the distribution and small enough to allow randomness in case of multiple draw (a size of 3/4 of the initial sample size should be big enough depending on the sample size);

- **Step 2**: In each row samples generated, compute the quantile of order τ (even all the quantile of interest) which should be estimators of $F_0^{-1}(\tau)$ and $F_1^{-1}(\tau)$ respectively for the distributions of Y_0 and Y_1. Then compute the difference of the two quantiles generated as follows: $\hat{F}_{1s}^{-1}(\tau) - \hat{F}_{0s}^{-1}(\tau)$;

- **Step 3**: Repeat steps 1 and 2 a thousand times as a bootstrap procedure and save the results of the difference in a vector of size 1000. Compute the mean and the standard deviation to get the average difference in quantiles and the standard errors;

- **Step 4**: Using classical test of the equality of mean, test if the mean of the vector is equal to zero. The p-value obtained from that test is the p value

that we are looking for.

Practically during simulations, this will be done for a given number of quantile $\tau \in (0,1)$ and at the end the number of p-values generated will be equal to the number of quantile tested. The bootstrap will actually generate a matrix $qx1000$, q being the number of quantile tested.

5.3.4 MTP using Quantile Groups (QG)

Before presenting the hypothesis definition of the MTP here, some definitions have to be given. The quantile group is understood here as all the units such that their value for a random variable is between two consecutive quantiles of same structure. To simplify the mathematical structure, let's assume without loss of generality that for a given sample size n, the different probabilities to compute the quantile are given by:

$$\tau = \frac{t \times i}{n} \tag{5.6}$$

with $i \in \{1, ..., n\}$ and t a coefficient such that $t \times j < n$.

For example, if $n = 450$ and we want to have deciles as quantiles, then $t = \frac{n}{10} = \frac{450}{10} = 45$, $i = 1, ..., 9$ and the corresponding quantile probability will be $\tau = \frac{t \times i}{n} = \left\{\frac{45 \times i}{n};\ i = 1, ..., 9\right\} = \{0.1, 0.2, ..., 0.9\}$.

Let us now define the quantile group. Given a random variable Y with a CDF

function F, the quantile group τ is defined as follows:

$$G_\tau = G_i = \left\{ Y_j : F^{-1}\left(\frac{t \times (i-1)}{n}\right) \leqslant Y_j < F^{-1}\left(\frac{t \times i}{n}\right) \right\} \tag{5.7}$$

with $\tau = \frac{t \times i}{n}$. It is basically all the units which their values are between two consecutive quantiles of same structure. By default, the first group will be all values less than the first quantile.

The MTP procedure will come from the following hypothesis formalization:

$$\begin{cases} H_0 : \{H_{0\tau} : mean(G_{0\tau}) = mean(G_{1\tau});\ with\ \tau \in (0,1)\} \\ H_1 : \{H_{1\tau} : mean(G_{0\tau}) \neq mean(G_{1\tau});\ with\ \tau \in (0,1)\} \end{cases} \tag{5.8}$$

It is basically MTP based on a comparison of means, test performed between the same quantile group of two distributions. The pointwise test will be the student t-test of comparison of mean. Therefore, before performing MTP test, the p-values will come from a test of comparison of mean formalized as follows for a given τ:

$$\begin{cases} H_{0\tau} : mean(G_{0\tau}) = mean(G_{1\tau}) \\ H_{1\tau} : mean(G_{0\tau}) \neq mean(G_{1\tau}) \end{cases} \tag{5.9}$$

The t-statistic using pooled variance (cause hypothetically from the same distri-

bution) is given by:

$$t_\tau = \frac{mean(G_{0\tau}) - mean(G_{1\tau})}{\sqrt{\frac{1}{n_{G_{0\tau}}} + \frac{1}{n_{G_{1\tau}}}} \times \sqrt{\frac{(n_{G_{0\tau}}-1)S^2_{G_{0\tau}} + (n_{G_{1\tau}}-1)S^2_{G_{1\tau}}}{n_{G_{0\tau}}+n_{G_{1\tau}}-2}}} \tag{5.10}$$

with $S^2_{G_{0\tau}}$ as the variance of the quantile group of order τ for the first random variable Y_0 and $S^2_{G_{1\tau}}$ as the variance of the quantile group of order τ for the second random variable Y_1. The quantities $n_{G_{j\tau}}$ are numbers of elements of each group, in other words the subsample sizes.

The t-statistic when the groups are not pooled is given as follows:

$$t_\tau = \frac{mean(G_{0\tau}) - mean(G_{1\tau})}{\sqrt{\frac{S^2_{G_{0\tau}}}{n_{G_{0\tau}}} + \frac{S^2_{G_{1\tau}}}{n_{G_{1\tau}}}}} \tag{5.11}$$

5.4 Power of MTP Approaches: Simulations

This section is dedicated to the implementation of the approaches developed in the previous section. We would like to see if the pointwise method can detect any change in a specific distribution. The distributions used in this simulation is a normal distribution. Any other distribution might give same or different results.

5.4.1 Algorithm and assumption

5.4.1.1 Intuition behind the Test of "No effects"

The main aim of this section is to apply the three MTP presented previously to compare two distributions. As we said earlier, if the distributions are the same, the conclusion is basically the same as any other simple testing procedure (classical K-S, Wilcoxon Rank test). If the test says the distributions are not the same, here comes the relevance of the MTP which is supposed to tell us which sections of the distribution makes them different. The three approaches will be compared to see which of them is the most powerful. As reminder, a pointwise test breaking the distributions into sections before comparison is more accurate than the classical test of comparison of the whole distribution.

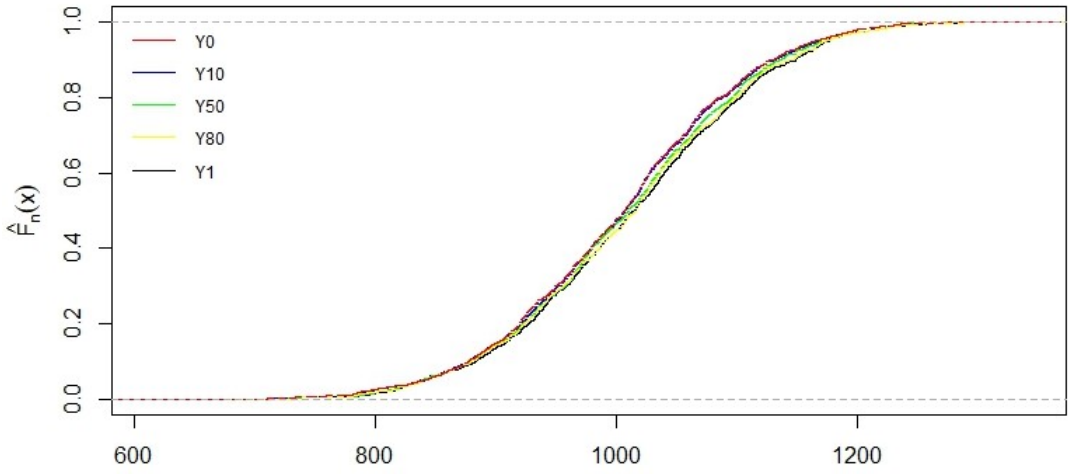

Figure 5.1: **Potential outcome CDF, 1% amplitude change at different percentage of points change.**

In order to stay in the framework of impact evaluation, the first assumption will be that the initial distribution is Y_0 then after a specific treatment randomly

assigned, the new distribution is Y_1. The test of the hypothesis "No effects" is positive if the distributions are the same and negative if the distributions are different. Secondly, if the assignment is not random, the test will be performed as well but using the restricted distributions then the completed distributions with imputation like in Chapter 3 and 4.

5.4.1.2 Algorithm of Simulations

To achieve our goals, a specific distribution is chosen and modified gradually. The modification is done by adding a given value seen here as the treatment effect (ATE) result of a specific treatment assignment. The first stage as a calibration stage is to compare Y_0 with itself, just to see how different tests respond to a perfect match between the two distributions. Second step is to compare Y_0 with $Y_0 + ATE$, with $Sh\%$ of the points in distribution randomly changed, $Sh \in \{10, 20, ..., 90\}$; then just $Y_0 + ATE$ with the whole distribution changed. The graphs, p-values and hypotheses decisions will be saved for analysis.

The simulation steps are as follows:

- **Step 1** : Simulate a specific distribution Y_0 representing the distribution of the potential outcome before the treatment. To calibrate, perform K-S test, Wilcoxon test and different MTP approach presented up between Y_0 and itself and keep the results.

- **Step 2** : Simulate the distribution Y_1 equal to Y_0 with a percentage Sh of point change in the distribution, $Sh \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. The change is such that the new distribution is the old one plus an amplitude

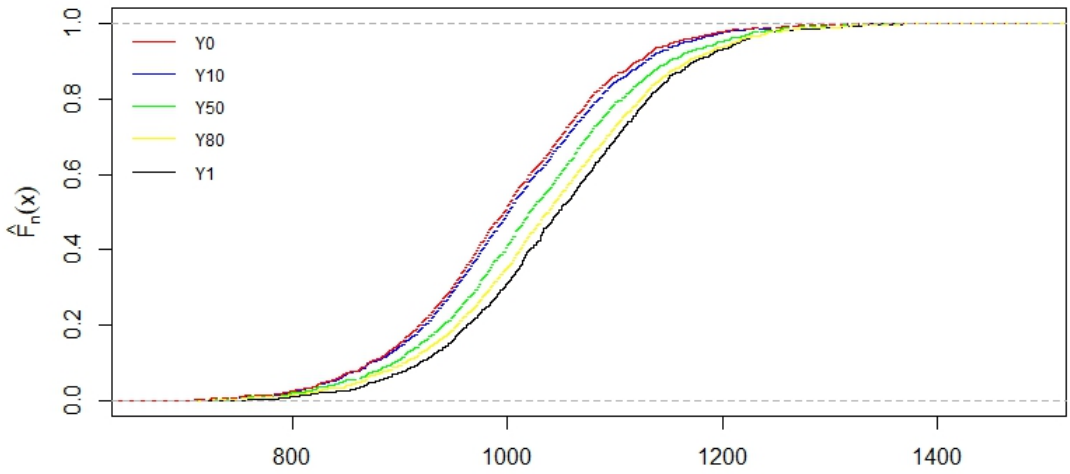

Figure 5.2: **Potential outcome CDF, 5% amplitude change at different percentage of points change.**

ATE ($ATE = a\,\%\ of\ mean\ of\ Y_0$) representing the effect of a treatment. Different amplitudes are chosen to see how the test reacts to small, medium and big changes in the distribution.

- **Step 3** : Compare Y_0 to $Y_1 = Y_0 + ATE\,(Sh\,\%\ of\ obs\ changed\ only)$ using the K-S test, WC test and the MTP test each time and store the results and graphs.

- **Step 4** : Analyze the results and draw the conclusions for the simulation process.

The ATE is considered as the amplitude change in the distribution of the initial potential outcome. An amplitude change corresponding to a given percentage of the mean of the initial distribution was considered: 1% (small changes, Figure 5.1), 5% (medium changes, Figure 5.2) and 25% (big changes, Figure 5.3) in the

simulation.

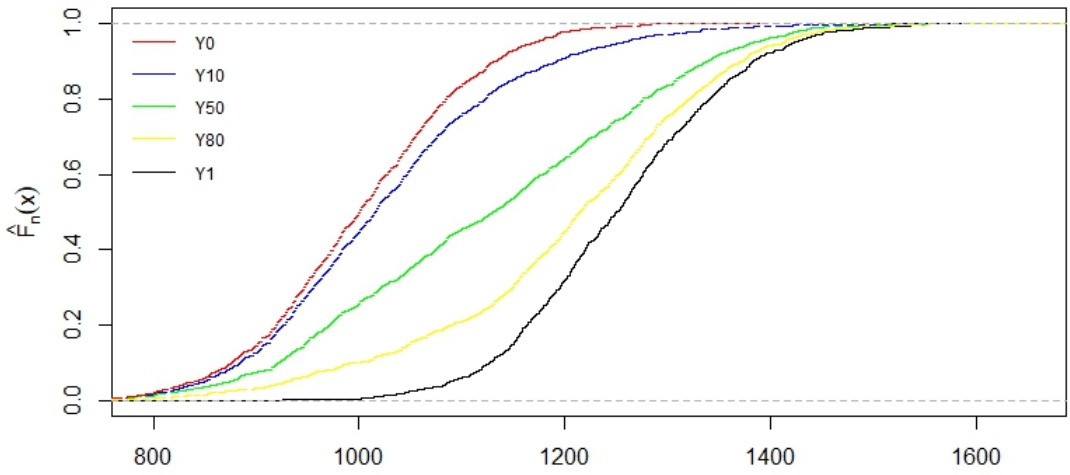

Figure 5.3: **Potential outcome CDF, 25% amplitude change at different percentage of points change.**

5.4.2 Results and comments

The results are presented in three steps. Having the simulated data[3], the first step is to test the difference between the initial potential outcome distribution Y_0 and the potential outcome changed using a specific amplitude and a percentage change in the points of the distribution (consider here as the potential outcome of the treated units Y_1) using K-S and WC tests. The second step is to perform the pointwise test on a certain number of points of the distribution (10 deciles lets say for all of the procedure) and save the p-values. The last step will be the MTP using the vector of p-values generated during the pointwise.

[3]See Appendix A3 for details on simulations choices.

5.4.2.1 Test of comparison of distributions (K-S and WC tests)

After simulation of the potential outcome Y_0 and the different distribution of treated effect, a classical test comparing distributions is performed to see whether the distributions are the same or not. In other words, whether an amplitude of a specific amount and a percentage of points changed in the distribution can be detected by a classical test comparing distributions. Table 5.1 summarizes the results of the test performed with amplitudes 1%, 5% and 25%.

Table 5.1: **Classical test of comparison of distributions**

P-values (Test Y_0 with Y_1)	ATE=1%		ATE=5%		ATE=25%	
	KS Test	WC Test	KS Test	WC Test	KS Test	WC Test
Sh0 (Y_0, Cali)	1	1	1	1	1	1
Sh10 ($Y10$)	1	0.83	0.95	0.29	0.005	0.0005
Sh20 ($Y20$)	1	0.65	0.20	0.03	0	0
Sh30 ($Y30$)	0.99	0.52	0.043	0.0013	0	0
Sh40 ($Y40$)	0.95	0.38	0.002	0	0	0
Sh50 ($Y50$)	0.89	0.27	0	0	0	0
Sh60 ($Y60$)	0.65	0.19	0	0	0	0
Sh70 ($Y70$)	0.43	0.13	0	0	0	0
Sh80 ($Y80$)	0.29	0.078	0	0	0	0
Sh90 ($Y90$)	0.26	0.049	0	0	0	0
Sh100 (Y_1)	0.18	0.028	0	0	0	0

As we can see in Table 5.1, for a small change of amplitude 1%, both tests do not reject the null hypothesis of equality of the two distributions from a share of 5% up to a share of 90% of the distribution. At 100% change, only the WC test was able to detect that the distributions are different. K-S and WC are not suitable to detect infinitesimal changes. In the second round of simulation, the amplitude of change is 5%. After changing 30% of the point in the distribution, the tests detected that the distributions are not the same. Meaning that for an amplitude

of 5% if very few points (less than 30%) in the distribution are changed, K-S test and WC test cannot detect the change. Finally, an amplitude of 25% change was simulated and for all the share of change recorded in the Table 5.1, the test detected that distributions are not equal (all p-values are less than 5%). But changing 1% of the point in the distribution using that amplitude (out of the box), change was not detected by the test.

From an impact evaluation point of view, for some treatments the K-S and WC test are not able to detect the treatment effect (Small amplitude of change and/or when a few sections of the population are affected) therefore the significance cannot be tested. Even if they detect, they cannot tell you at which section of the distribution the change is or where the difference appears. That is why it is important to go for pointwise tests and MTP procedures.

5.4.2.2 Pointwise test results

As highlighted in the previous section, to run a MTP, a vector of p-values representing the pointwise test done for each of hypothesis is needed. An example of the pointwise test done before MTP procedure is given in the Table 5.2. The table depicts the results of a pointwise test at three levels of amplitude and at specific shares of change for each level. Generally, a p-value less than 5% implies that the null hypothesis (equality of distribution at a specific point) is rejected. Three pointwise tests were applied to the simulated data at each amplitude of ATE, for each amplitude a percentage of point change in the distribution was

Table 5.2: **Pointwise test results**

Summary Deciles		P-values of the pointwise test for each decile								
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th
1% (Sh50%)	CDF	1	1	0.60	0.48	0.35	0.50	0.45	0.37	0.71
	QF	0	0	0	0	0	0	0	0	0
	QG	0.42	0.01	0	0	0	0	0	0	0.01
5% (Sh20%)	CDF	1	0.56	0.75	0.36	0.03	0.13	0.13	0.07	0.31
	QF	0	0	0	0	0	0	0	0	0
	QG	0.38	0.01	0	0	0	0	0	0	0
25% (Sh20%)	CDF	0.71	0.29	0.02	0.01	0	0	0	0	0
	QF	0	0	0	0	0	0	0	0	0
	QG	0.07	0	0	0	0	0	0	0	0

chosen. At 1% of amplitude and 50% of points changed in the distribution, only quantile function (QF) method detected that the distributions are different at all the points selected. From all the outputs of the tests, it is the most sensitive pointwise test method. The method is able to detect any small change in the distribution. The method based on CDF is less sensitive than the previous one. The method is not able to detect the differences between two distributions across the point (see Table 5.2 at 1% none, at 5% the 5th decile and at 25% most of the points), rejecting the null hypothesis for some point of the distribution and not rejecting for others. The last method, quantile group (QG) is closer to the quantile function method than the CDF method, but less sensitive as the first one. It is suitable for small changes as well.

5.4.2.3 MTP test results

Using the p-values generated in the pointwise test procedure, it is now possible to run a MTP procedure and come up with the adjusted p-values, the False Discovery Rate and the number of rejected pointwise test in the joint procedure.

Table 5.3 presents the MTP procedure for different amplitudes changes and at different percentage change in the point of distribution (mostly less than 50% points changes), especially when the K-S and WC tests were not able to conclude in terms of difference between the two distributions.

Table 5.3: **MTP results with K-S and WC failing**

Summary Deciles		MTP Tests (Adjusted P-values and Rejections)										
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Rej	FDR
1% (Sh50)	CDF	0.89	0.89	0.89	0.89	0.89	0.89	0.91	1	1	0	0
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0	0	0	0	0	0.001	0.01	0.01	0.42	8	0
5% (Sh20)	CDF	0.3	0.3	0.31	0.31	0.54	0.54	0.72	0.85	1	0	0
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0	0	0	0	0	0	0	0.01	0.38	8	0
25% (Sh20)	CDF	0	0	0	0	0	0.01	0.03	0.33	0.71	7	0.01
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0.07	0	0	0	0	0	0	0	0.07	8	1

Like in the pointwise test, the results are a bit mitigated. For an ATE of 1%, the three tests give different results: the CDF MTP procedure says the distributions are the same but the quantile function (QF) procedure says strictly that they are not the same while the quantile group (QG) says there is a section of the distribution where they are the same (null hypothesis not rejected for the 9th decile section). This pattern in the results is almost the same for 5% and 25% of amplitude change of ATE except that for 25% which is a clear change. The CDF MTP can at least show that there is a clear difference for some points (first seven deciles). For small percentage of point changes in the distribution, the MTP methods developed here reject drastically the hypothesis of equality of distribution: they are strong methods (except CDF) even if very few points are changed in the distribution.

In the case where almost all the points in the distribution (90%) are changed using the same amplitude, Table 5.4 summarized the results for the cases where the K-S and WC were not able to detect the change (1%) and for the other cases where they were able but without a clear position in the distribution. Here, the pattern of results is almost the same. The MTPs were able to detect the difference even for the smallest amplitude of change.

Table 5.4: **MTP results with K-S and WC not failing**

Summary Deciles		MTP Tests (Adjusted P-values and Rejections)										
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Rej	FDR
1% (Sh90)	CDF	0.51	0.51	0.51	0.51	0.58	0.61	0.61	0.91	1	0	0
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0	0	0	0	0	0	0	0	0.09	8	0
5% (Sh50)	CDF	0	0	0	0	0.01	0.01	0.12	0.63	1	6	0.01
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0	0	0	0	0	0	0	0	0.01	9	0
25% (Sh50)	CDF	0	0	0	0	0	0	0	0	0.05	8	1
	QF	0	0	0	0	0	0	0	0	0	9	0
	QG	0.07	0	0	0	0	0	0	0	0.	9	0

Table 5 shows that for a 1% amplitude change in the distribution, when almost all the points are changed, the CDF MTP is not able to detect the change between Y_0 and Y_1, no rejection and FDR null. The quantile function approach is clear, there is a significant difference between the distributions while the quantile group approach detects an equality at the 9th decile in the two distributions (8 rejections).

For the amplitude of 5%, the CDF says there are 6 rejections and for 3 points in the distribution there is equality. The results for the two other methods are the same: 9 rejections just like classical K-S and WC test.

Lastly, there is no doubt that for the amplitude of 25%, changing only 50% of the points in the distribution, the tests rejected jointly the hypothesis of equality of the two distributions. Only the CDF approach presented one point where the distributions can be almost the same at 5% threshold.

5.4.2.4 General comments and remarks

Generally, from the simulation activities and from our analysis of outputs, we noticed that K-S and WC test reacted more quickly when the amplitude change is bigger (25% change in the distribution vs 1%). In addition, when the number of points changed in the distribution is increasing, the tests are more accurate. As a result, those tests cannot detect accurately small amplitude changes especially when they appear on a small section (few points, less than 20% of sample size) of the distribution.

Significance of impact on distribution tested using pointwise test and MTP depend on the amplitude of the effect (1% vs 25%). For small effects the result is obvious and clear with CDF test (test approach is not able to detect small changes) but the two others are accurate for very small amplitudes. Quantile function and quantile group methods are highly sensitive to infinitesimal changes in the distribution. They will be suitable for micro impact that can lead to important spillover with time. Small changes (1% or less than 5%) are the most difficult to detect but important in the framework of impact evaluation. Sometimes they are the ones that are important to detect. For anticipated large impact, the CDF approach is better.

5.5 Summary

In this chapter, a new approach for comparing two distributions was developed with the aim of testing the hypothesis "No effect" in the framework of impact evaluation. Using MTP approach, the aim was to show that testing the significance of the effect of a treatment using the classical tests of comparison of distribution K-S and WC, will not give accurate results especially when the distributions differ at few points and for small amplitude change. The approach presented here is a point wise test joined to a MTP procedure. The three approaches were effective and gave satisfactory results during simulations and application. The CDF approach can detect big amplitude change while QF and QG approaches are able to detect any infinitesimal or very small changes between two distributions.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This chapter presents the conclusions of this work. Some limitations and recommendations on this work are given as well for future studies in the same area of research. Before that, a brief recall of the objectives and a summary of results obtained in this research are given.

6.2 Conclusions

Trying to overcome the weaknesses linked to estimation of counterfactual, heterogeneity of treatment effects (not always taken into account during estimation) and reduction of sample size in the literature of impact evaluation, the objective of this research was to build a statistical theory on impact evaluation estimators based on imputation methods, from estimators derivation and properties to hypotheses testing procedure. Imputation methods were checked to assess whether they can be used to derive efficient impact evaluation estimators. At the end of the study, the response is yes. The construction of estimators more efficient than the existing ones in the literature was done, their properties were studied and a test procedure was developed around them.

In Chapter 3, the notion of imputation was introduced to estimate counterfactual. Imputation Based Average Treatment Effect Estimators (IB-ATE) were derived and their theoretical and asymptotic properties such as unbiasedness, convergence and consistency were shown. The IB-ATE estimators were applied under simulations and on real database. It was observed from the results that they were as good as good as the classical estimators in some case and in others better than classical IE estimators. For applications on real data, almost similar results were found as Lalonde (1986) and comparing the results with Dehejia and Wahba (1999), IB-ATE estimators were better. This new class of estimators came with some advantages like the possibility to have case effects and possibility to perform better than classics IE estimators in a context of shortage in data. Also, there is always a way to improve the quality of these estimators by improving the related imputation method.

In Chapter 4, following the seminal idea of Chapter 3 consisting of using imputation methods to estimate counterfactual, Imputation Based Distributional Treatment Effect Estimators (IB-DTE) were derived and their theoretical and asymptotical properties were studied. IB-DTE estimators are convergent and asymptotically normal. Simulations showed that they perform better than existing estimators in the literature. These estimators came with less assumptions made like rank preservation assumption and a better and stable performance during simulation and on real dataset compared to existing ones in the literature

such asFirpo (2007).

Lastly, in Chapter 5, a testing theory was developed around the hypothesis "No effect" using estimators of the Chapters 3 and 4. Given that the literature on hypothesis testing in the impact evaluation framework is actually growing, a MTP was proposed in this study to respond to the question: was the treatment effect significant? Three pointwise approaches were proposed. The simulations results showed that they were able to detect any difference between two distributions when the classical tests like K-S and WC could not. As a result, in case of a treatment assignment supposed to modify the initial potential outcome, the MTP developed were used and changes were detected, even the smallest one due to treatment. Considering especially the case of small amplitude changes and/or for small number of points change where the classical test of comparison of distributions are not able to detect, two of the three procedures presented in this research were accurate for that (Quantile Function and Quantile Group approaches).

In summary, the research conducted was conclusive, and very helpful and practical thus, enriching the literature on impact evaluation.

6.3 Recommendations

In this study, the counterfactual is estimated using imputation methods implying indirectly that the performance of the IB estimators are pegged on the quality of the results produced by imputation methods. With good estimates of counterfactual, the results of treatment effect estimators will be better. A direct consequence of this limitation appears practically. The user of this method should be absolutely familiar with imputation methods in statistics. A package can be implemented on a software to perform this method but it is the user who, given the structure of the data, has to specify which imputation method is suitable or unsuitable for the data.

The tests performed in Chapter 5 were done mainly with simulated data, assuming imputation is done upstream. Another interesting study to do is to compare the results of the test with data reconstructed using imputation methods vs reduced sample in the classical theory of impact evaluation. It is expected that imputation will lead to more precise results than reduced samples.

Widely in this area and linked to IB estimators, future works after this study are many. One may want to explore how to design imputation based estimators in a context of multiple treatments assigned jointly. Taking the case of a qualitative response variable, it will be interesting to investigate how IB estimators will look like. Lastly, applications were done using the famous data set of Lalonde (1986)

in order to make comparison of the results of this research with the results of the literature. Applying these results to real and actual data to solve a specific problem is as well a good way forward.

REFERENCES

Aakvik, A., Heckman, J. J., and Vytlacil, E. (1999). Semiparametric program evaluation lessons from an evaluation of a norvegian training program. *mimeo Departement of Economics, University of Chicago*.

Aakvik, A., Heckman, J. J., and Vytlacil, E. (2005). Estimating treatment effects for discrete outcomes when response to treatment vary: An application to norvegian vocational rehabilitation programs. *Journal of Econometrics*, 125:15–51.

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, 97:284–292.

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263.

Abadie, A. (2011). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the american Statistical Assoiation*, 97.

Abadie, A., Angrist, J. D., and Imbens, G. W. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70:91–117.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear Difference-in Differences models. *Econometrica*, 74:431–497.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2006). What mean impact miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96:988–1012.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2008). Distributional impacts of the self-sufficiency project. *Journal of Public Economics*, 92(3-4):748–765.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2001). Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Economic policy Review*, 8:273–301.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44:361–422.

Carneiro, P. and Lee, S. (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*, 149(2):191–208.

Chaudhury, N. and Parajuli, D. (2006). Conditional cash transfer and female schooling: The impact of the female school stipend program on public school enrollments in Punjab, Pakistan. *World Bank Policy Research Working Paper, No. 4102*.

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.

Chernozhukov, V. and Hansen, C. (2013). Quantile models with endogeneity. *Annual Review of Economics*, 5(1):57–81.

Cherrnozhukov, V. and Hansen, C. (2005). An IV model of Quantile Treatment Effects. *Econometrica*, 73:245–261.

Chetverikov, D. (2013). Testing regression monotonicity in econometric models. *mimeo, UCLA*.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008). Nonparametric tests for treatment effect heterogeneity. *The Review of Economics and Statistics*, 90(3):389–405.

Dehejia, R. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062.

Djebbari, H. and Smith, J. (2008). Heterogenous impacts in PROGRESSA. *Journal of Econometrics*, 145(1):64–80.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics*, 2:267–277.

Dominici, F., Cope, L., Naiman, D. Q., and Zeger, S. L. (2005). Smooth Quantile Ratio Estimation. *Biometrika*, 92(3):543–557.

Firpo, S. (2007). Efficient semiparametric estimation of Quantile Treatment Effects. *Econometrica*, 75(1):259–276.

Fisher, R. A. (1935). *The Design of Experiments*, volume 35. Oliver and Boyd, London, 1st edition.

Frechet, M. (1951). Sur les tableaux de corrélation sont les marges sont données. *Annals University Lyon*, 14:53–77.

Gabriel, V. M.-R. (2011). Nonparametric estimation of ATE and QTE: An application of Fractile Graphical Analysis. *Journal of Probability and Statistics*, 2011:1–23.

Glewwe, P. and Jacoby, H. G. (1995). An economic anaysis of delayed primary school enrollment in a low income country. *The Review of Economics and Statistics*, 77(1):156–169.

Goldman, M. and Kaplan, D. M. (2017). Fractional order statistic approximation for nonparametric conditional quantile inference. *Journal of Econometrics*, 196(2):331–346.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of Average Treatment Effects. *Econometrica*, 66:315–331.

Heckman, J. J., Ichimura, H., and (HIT), P. E. T. (1997a). Matching as an econometric evaluation estimator: Evidence from evaluationg a job training programme. *The review of Economic Studies*, 64:605–654.

Heckman, J. J., Smith, J., and Clements, N. (1997b). Making the most out of programme evaluation and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535.

Heckman, J. J. and Smith, J. A. (1998). Evaluating the welfare state, econometric and economic theory in the 20th century: The Ragnar Frisch centennial symposium, s. strom eds. *Cambridge: Cambridge University Press*, pages 241–318.

Heckman, J. J., Smith, J. A., Jensen, K., and Madsen, P. K. (1993). Assessing the case for randomized evaluation of social programs, in mesuring labour market measures: Evaluating the effects of active labour market policies. *Copenhagen: Danish ministry of labour*, pages 35–96.

Heckman, J. J. and Vytlacil, E. J. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73:669–738.

Hoeffding, W. (1940). Masstabinvariate korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts fur Angewandte Mathematik des Universitat Berlin*, 5:179–233.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica*, 62:467–475.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 64:555–574.

Imbens, G. W. and Wooldridge, J. M. (2009). Recents developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Jackson, E. and Page, M. E. (2013). Estimating the distributional effects of education reforms: A look at project STAR. *Economics of Education Review*, 32:92–103.

Jalan, J. and Ravallion, M. (2003). Estimating the benefit incidence of an antipoverty program by Propensity-Score Matching. *Journal of Business & Economic Statistics*, 21:19–30.

Kaplan, D. M. and Goldman, M. (2013). Comparing distributions by multiple testing accross quantiles or CDF values. *Working papers 16-19, Department of Economics, University of Missouri*.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto degli attuari*, 4(1):83–91.

Lalonde, R. (1986). Evaluating the econometric evaluation of training programs. *American Econometric Review*, 76:604–620.

Lehmann, E. L. (1974). Nonparametrics: Statistical methods based on ranks. *San Francisco: Holden-Day*.

Pitman, E. J. G. (1937). Significance tests which may be applied to sample from any population. *Journal of Royal Statistic Society*, 4:119–130, 225–232.

Pitman, E. J. G. (1938a). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika*, 30:391–421.

Pitman, E. J. G. (1938b). Significance tests which may be applied to samples from any population. *Biometrika*, 29:322–335.

Ravallion, M. (2007). *Handbook of Developement Economics*, volume 4. Elsevier.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Psychology*, 66:688–701.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Mathematique de l'Université de Moscow*, 2(2):3–16.

Smirnov, N. V. (1948). Table for estimating the goodness-of-fit of empirical distributions. *Annals of Mathematical Statistics*, 19(2):279–281.

Venturini, S., Dominici, F., and Parmigiani, G. (2015). Generalized Quantile Treatment Effect: A flexible Bayesian approach using Quantile Ratio Smoothing. *Bayesian Analysis*, 10(3):523–552.

Welch, B. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3/4):350–362.

Wilcox, R., Erceg-Hum, D. M., Clark, F., and Carlson, M. (2013). Comparing two independent groups via the lower and upper quantiles. *Journal of Statistical computation and Simulation*, 84(7):1543–1551.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometries*, 1:83–90.

Yu, P. (2014). Marginal Quantile Treatment Effects. *mimeo, Department of Economics, University of Auckland*.

APPENDICES

Appendix A1

Population and Data Simulations Random Assignment

```
N=10000
#####Generate data base Potential outcome and covariates before and after
###Covariates (Sex, Education level, Age, CSP respectively)
Xb1=rbinom(N,1,0.55)
Xb2=rbinom(N,3,0.25)
Xb3=round(runif(N,15,60))
Xb4=rbinom(N,2,1/3)

###Potential outcome before (Normal distribution then chi fat tail)
Yb=11*Xb1+12*Xb2+13*Xb3+14*Xb4+rnorm(N,500,250)

###Potential outcome in two worlds (Normal distribution then chi fat tail)
Y2T=Yb+runif(N,250,500)
Y2NT=Yb+runif(N,100,250)

###Generation of treatment variable, Two cases (Random Assigment
or Missingness process)
##MCAR Process of missingness
T=rbinom(N,1,0.4)

##MAR Process of missingness
#Tresh=Xb1+Xb2+Xb3+Xb4
#TR=quantile(Tresh, prob = 0.35, type = 5)
#T=rep(0,n)
#for (i in 1:n){if (Tresh[i]<=TR) {T[i]=1}}
#T

###Generation of potential outcome after treatment
Ya=Y2NT
for (i in 1:N){if (T[i]==1) {Ya[i]=Y2T[i]}}
Ya

###Generation of potential outcome after treatment with missings
Yt=Ya
for (i in 1:N){if (T[i]==0) {Yt[i]="NA"} }
Yt=as.numeric(Yt)
Yt

Ynt=Ya
for (i in 1:N){if (T[i]==1) {Ynt[i]="NA"} }
Ynt=as.numeric(Ynt)
Ynt

###Generation of data base in a data frame format and in a matrix format
DataF=data.frame (Ind=1:N,Yb,Xb1,Xb2,Xb3,Xb4,Y2T,Y2NT,T,Ya,Yt,Ynt)
#DataF
```

```
DataMatF=data.matrix(DataF)
#DataMatF

DataTreat=subset(DataF, T==1)
DataControl=subset(DataF, T==0)

SEval1=DataTreat[sample(nrow(Eval1), n*0.4), ]
SEval0=DataControl[sample(nrow(Eval0), n*0.6), ]
FullData=rbind(SEval1, SEval0)

SEval1=Eval1[sample(nrow(Eval1), n*0.4), ]
SEval0=Eval0[sample(nrow(Eval0), n*0.6), ]
FullData=rbind(SEval1, SEval0)

Xb1=FullData$Cvt1
Xb2=FullData$Cvt2
Xb3=FullData$Cvt3
Xb4=FullData$Cvt4
Yb=FullData$POb
Y2T=FullData$PO2T
Y2NT=FullData$PO2NT
T=FullData$Tr
Ya=FullData$POa
Yt=FullData$POt
Ynt=FullData$POnt
IV=FullData$Inst
Data=data.frame (Ind=1:length(IV),Yb,Xb1,Xb2,Xb3,Xb4,
    Y2T,Y2NT,T,Ya,Yt,Ynt,IV)
DataMat=data.matrix(Data)

###Computation of true average impact of the treatment (ATE)
TrueImpact=mean(Data$Y2T-Data$Y2NT)
TrueImpact
sd(Data$Y2T-Data$Y2NT)

###Computation of true average impact of the treatment
    of treated units (ATT)
sum((Data$Y2T-Data$Y2NT)*T)/sum(T)
```

MAR Assignment

```
N=25000
ID=1:N
Cvt1=rbinom(N,1,0.55)
Cvt2=rbinom(N,3,0.25)
Cvt3=round(runif(N,15,60))
Cvt4=rbinom(N,2,1/3)

POb=11*Cvt1+12*Cvt2+13*Cvt3+14*Cvt4+rnorm(N,500,250)
PO2T=POb+runif(N,250,500)
PO2NT=POb+runif(N,100,250)
#Inst=13*Cvt3+14*Cvt4
Inst=runif(N,10,6000)
TR=quantile(Inst, prob = 0.5, type = 5)
Tr=rep(0,N)
for (i in 1:N){if (Inst[i]<=TR) {Tr[i]=1}}
```

```
POa=PO2NT
for (i in 1:N){if (Tr[i]==1) {POa[i]=PO2T[i]}}
POt=POa
for (i in 1:N){if (Tr[i]==0) {POt[i]="NA"} }
POt=as.numeric(POt)
POnt=POa
for (i in 1:N){if (Tr[i]==1) {POnt[i]="NA"} }
POnt=as.numeric(POnt)
PoPData=data.frame (ID,POb,Cvt1,Cvt2,Cvt3,Cvt4,PO2T,
    PO2NT,Tr,POa,POt,POnt,Inst )
PoPDataMat=data.matrix(PoPData)

EvalData=subset(PoPData, Inst>=quantile(Inst, prob = 0.3, type = 5)
& Inst<=quantile(Inst, prob = 0.7, type = 5))
Eval1=subset(EvalData, Tr==1)
Eval0=subset(EvalData, Tr==0)

save(EvalData,file="EvalData.Rda")
save(Eval1,file="Eval1.Rda")
save(Eval0,file="Eval0.Rda")
load("EvalData.Rda")
load("Eval1.Rda")
load("Eval0.Rda")

TrueATE=mean(EvalData$PO2T-EvalData$PO2NT)
TrueATT=sum((EvalData$PO2T-EvalData$PO2NT)*EvalData$Tr)/sum(EvalData$Tr)
TrueATE
TrueATT

#ivreg(EvalData$POa ~ EvalData$Tr+EvalData$Cvt1+EvalData$Cvt2
+EvalData$Cvt3+EvalData$Cvt4 | EvalData$Inst+EvalData$Cvt1
+EvalData$Cvt2+EvalData$Cvt3+EvalData$Cvt4)
```

Some functions

```
####Function to test performance of Imputation methods
RMSE=function(a,b){
  val=sqrt(mean((a-b)^2))
  return(val) }

####Impute missing values by the mean
mean.imp=function(a){
  m=mean(a, na.rm=TRUE)
  for (i in 1:length(a)){if (is.na(a[i])==1) {a[i]=m} }
  return(a)}

####Impute missing values by the conditional mean,
 condition on the second parameter
cmean.imp=function(a,b){
  U=unique(b)
  V=U
  for (i in 1:length(U)){
    c=0
    s=0
    for (j in 1:length(b)){if (b[j]==U[i] && is.na(a[j])==0 ) {
      s=s+a[j]
```

```
      c=c+1} }
    V[i]=s/c }

  for (k in 1:length(a)){
    for(j in 1:length(U)){if((is.na(a[k])==1)
    &&(b[k]==U[j])){a[k]=V[j]}}}
  return(a)}

####Random imputation: replace NA by a value randomly chosen
random.imp<-function(a){
  missing<-is.na(a)
  n.missing<-sum(missing)
  a.obs<-a[!missing]
  imputed<-a
  imputed[missing]<-sample(a.obs, n.missing, replace=TRUE)
  return(imputed) }

####Impute NA value by results from a regression
reg.imp=function(a,a.impute){
  ifelse(is.na(a), a.impute, a)}
```

Impact evaluation methods and Imputation methods codes

Impact evaluation methods

```
#### Randomization Method [1]
##ATE and ATT at the same time, Difference in mean in the two groups
mean(Yt, na.rm=TRUE)-mean(Ynt, na.rm=TRUE)

#### Propensity Score Matching Method [2]
mylogit2 <- glm(T~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4),
            family=binomial)
X <- mylogit2$fitted

##ATE
result21 <- Match(Y=Ya, Tr=T, X=X, estimand = "ATE", M=1)
summary(result21)

##ATT
result22 <- Match(Y=Ya, Tr=T, X=X, estimand = "ATT", M=1)
summary(result22)


#### Difference in Difference Method [3] (ATE is not computed
and not possible to compute, limit of method)
##ATT typically (only on treated),
mean(Yt-Yb, na.rm=TRUE)-mean(Ynt-Yb, na.rm=TRUE)

#### Instrumental Variable Method [4]

#### Regression Discontunuity Design Method [5]
```

Imputation methods

```
####Imputation Method 1: Mean Imputation
###General mean imputation
##Function for imputation
Yt1=mean.imp(Yt)
Ynt1=mean.imp(Ynt)

mean(Yt1)-mean(Ynt1)
sum((Yt1-Ynt1)*T)/sum(T)

RMSE(Y2T,Yt1)
RMSE(Y2NT,Ynt1)

###Conditional mean imputation
Yt11=cmean.imp(Yt, Xb1)
Ynt11=cmean.imp(Ynt, Xb1)

mean(Yt11)-mean(Ynt11)
sum((Yt11-Ynt11)*T)/sum(T)

RMSE(Y2T,Yt11)
RMSE(Y2NT,Ynt11)

####Imputation Method 2: Random Imputation
        (See VIM package Hot deck Imp)
####(Bootstrapping and conserve positive impact
        for acceptable results)
##Use the function to impute Yt and Ynt (loop here)
Yt2=random.imp(Yt)
Ynt2=random.imp(Ynt)

mean(Yt2)-mean(Ynt2)
sum((Yt2-Ynt2)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt2)
RMSE(Y2NT,Ynt2)

##Hot deck imputation with domain
Data2=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
#hotdeck(Data2, ord_var= c("Xb1", "Xb2", "Xb3", "Xb4"),
            domain_var = c("Yt", "Ynt"))
Imp.base=hotdeck(Data2)

Yt22=Imp.base$Yt
Ynt22=Imp.base$Ynt

mean(Yt22)-mean(Ynt22)
sum((Yt22-Ynt22)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt22)
RMSE(Y2NT,Ynt22)


####Imputation Method 3: Regression to perform deterministic imputation

##Function for imputation
```

```
lm.imp.1=lm(Yt~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4))
        #Regression model for treated units
summary(lm.imp.1)

lm.imp.2=lm(Ynt~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4))
        #Regression model for non treated units
summary(lm.imp.2)

pred.1=predict(lm.imp.1, Data)
        #Predict using deterministic part of reg model (treated)
pred.2=predict(lm.imp.2, Data)
        #Predict using deterministic part of reg model (non treated)

##Imputation using deterministic part of reg model
Yt3=reg.imp(Yt, pred.1)
Ynt3=reg.imp(Ynt, pred.2)

mean(Yt3)-mean(Ynt3)
sum((Yt3-Ynt3)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt3)
RMSE(Y2NT,Ynt3)

####Imputation Method 4: k-Nearest Neighbour Imputation (Package VIM)
##V1
Data4=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
Imp.data=kNN(Data4, k=1)
Imp.data
Yt4=Imp.data$Yt
Ynt4=Imp.data$Ynt

mean(Yt4)-mean(Ynt4)
sum((Yt4-Ynt4)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt4)
RMSE(Y2NT,Ynt4)

##V2
sampImp <- kNN(Data4, dist_var = c("Xb1", "Xb2", "Xb3", "Xb4"),
                      k = 1, numFun = mean)

Yt44=sampImp$Yt
Ynt44=sampImp$Ynt

mean(Yt44)-mean(Ynt44)
sum((Yt44-Ynt44)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt44)
RMSE(Y2NT,Ynt44)

####Imputation Method 5: Regression with random prediction
##Function for imputation
lm.imp.3=lm(Yt~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4))
        #Regression model for treated units
summary(lm.imp.3)
```

```
lm.imp.4=lm(Ynt~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4))
        #Regression model for non treated units
summary(lm.imp.4)

pred.3=rnorm(n, predict(lm.imp.3, Data), summary(lm.imp.3)$sigma)
        #Random prediction of reg model (treated)
pred.4=rnorm(n, predict(lm.imp.4, Data), summary(lm.imp.4)$sigma)
         #Random prediction of reg model (non treated)

##Imputation using random generated predicted values
Yt5=reg.imp(Yt, pred.3)
Ynt5=reg.imp(Ynt, pred.4)

mean(Yt5)-mean(Ynt5)
sum((Yt5-Ynt5)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt5)
RMSE(Y2NT,Ynt5)

####Imputation Method 6: Multiple Imputation (Package ...)
Data6=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))

##Multiple imputation with MICE Package
imp=mice(Data6, m=5)
Yt6=(complete(imp, 1)$Yt+complete(imp, 2)$Yt+complete(imp, 3)$Yt
        +complete(imp, 4)$Yt+complete(imp, 5)$Yt)/5
Ynt6=(complete(imp, 1)$Ynt+complete(imp, 2)$Ynt+complete(imp, 3)$Ynt
        +complete(imp, 4)$Ynt+complete(imp, 5)$Ynt)/5

mean(Yt6)-mean(Ynt6)
sum((Yt6-Ynt6)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt6)
RMSE(Y2NT,Ynt6)

##Multiple imputation with mi Package
mdf <- missing_data.frame(Data6)
impu66 <- mi(mdf)
chain=complete(impu66,5)

Yt66=(chain[[1]]$Yt+chain[[2]]$Yt+chain[[3]]$Yt+chain[[4]]$Yt
        +chain[[5]]$Yt)/5
Ynt66=(chain[[1]]$Ynt+chain[[2]]$Ynt+chain[[3]]$Yt+chain[[4]]$Yt
        +chain[[5]]$Yt)/5

mean(Yt66)-mean(Ynt66)
sum((Yt66-Ynt66)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt66)
RMSE(Y2NT,Ynt66)

####Imputation Method 7: Maximum Likelihood Imputation (Package ...)
Data7=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
Dat7=data.matrix(Data7)

##Imputation using ML method and package missmech
```

```
Res7=Impute(Dat7, mu = NA, sig = NA, imputation.method = "dist.free",
                resid = NA) #method can also be normal

Yt7=Res7$yimp[,1]
Ynt7=Res7$yimp[,2]

mean(Yt7)-mean(Ynt7)
sum((Yt7-Ynt7)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt7)
RMSE(Y2NT,Ynt7)

#method can also be normal
Res77=Impute(Dat7, mu = NA, sig = NA, imputation.method = "normal",
                    resid = NA)

Yt77=Res77$yimp[,1]
Ynt77=Res77$yimp[,2]

mean(Yt77)-mean(Ynt77)
sum((Yt77-Ynt77)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt77)
RMSE(Y2NT,Ynt77)

##Imputation using ML and package norm
#s=prelim.norm(Dat7)
#th=em.norm(s)
#imp.norm(s, th, Dat7)

####Imputation Method 8: Propensity score matching Imputation
        (Package matching)
mylogit <- glm(T~factor(Xb1) + factor(Xb2) + Xb3 + factor(Xb4),
        family=binomial)
X <- mylogit$fitted
result <- Match(Y=Ya, Tr=T, X=X, estimand = "ATE", M=1)
summary(result)
Yt8=result$mdata$Y[1:n]
Ynt8=result$mdata$Y[(n+1):(2*(n))]

mean(Yt8)-mean(Ynt8)
sum((Yt8-Ynt8)*T)/sum(T) #Effect on treated units

RMSE(Y2T,Yt8)
RMSE(Y2NT,Ynt8)

####Imputation Method 9: Hahn's Impact evaluation method Imputation
        (Package np for non parametric regression)

#data.object <- data.frame(Yt, Xb1 = factor(Xb1), Xb2 = factor(Xb2),
                Xb3,  Xb4 = factor(Xb4))
#bw <- npudensbw(dat = data.object)
#bw
#bw <- npregbw(Yt ~ Xb1 + Xb2 +Xb3 + Xb4)
#bw
```

```
####Imputation Method 10: Quantile regression Imputation
           (Package quantreg, decile)

##Quantile model with decile coeficients
rq.imp1=rq(Yt~Xb1+Xb2+Xb3+Xb4, c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6,
                0.7, 0.8, 0.9), Data)
rq.imp2=rq(Ynt~Xb1+Xb2+Xb3+Xb4, c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6,
                0.7, 0.8, 0.9), Data)
summary(rq.imp1)
summary(rq.imp2)
p.rq1=predict(rq.imp1, Data)
p.rq2=predict(rq.imp2, Data)
p.rq1
p.rq2

##Imputation using quantile models coeficients
Yt10=Yt
for (i in 1:length(Yb)) {
  if (Yb[i]<=quantile(Yb, 0.1) && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,1][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.1) && Yb[i]<=quantile(Yb, 0.2)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,2][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.2) && Yb[i]<=quantile(Yb, 0.3)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,3][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.3) && Yb[i]<=quantile(Yb, 0.4)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,4][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.4) && Yb[i]<=quantile(Yb, 0.5)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,5][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.5) && Yb[i]<=quantile(Yb, 0.6)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,6][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.6) && Yb[i]<=quantile(Yb, 0.7)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,7][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.7) && Yb[i]<=quantile(Yb, 0.8)
        && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,8][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.8) && is.na(Yt[i])==1)
            { Yt10[i]=p.rq1[,9][i] } }

Ynt10=Ynt
for (i in 1:length(Yb)) {
  if (Yb[i]<=quantile(Yb, 0.1) && is.na(Ynt[i])==1)
        { Ynt10[i]=p.rq2[,1][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.1) && Yb[i]<=quantile(Yb, 0.2)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,2][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.2) && Yb[i]<=quantile(Yb, 0.3)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,3][i] } }
for (i in 1:length(Yb)) {
```

```
  if (Yb[i]>quantile(Yb, 0.3) && Yb[i]<=quantile(Yb, 0.4)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,4][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.4) && Yb[i]<=quantile(Yb, 0.5)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,5][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.5) && Yb[i]<=quantile(Yb, 0.6)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,6][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.6) && Yb[i]<=quantile(Yb, 0.7)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,7][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.7) && Yb[i]<=quantile(Yb, 0.8)
        && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,8][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.8) && is.na(Ynt[i])==1)
        { Ynt10[i]=p.rq2[,9][i] } }

mean(Yt10)-mean(Ynt10)
sum((Yt10-Ynt10)*T)/sum(T)

RMSE(Y2T, Yt10)
RMSE(Y2NT, Ynt10)
```

Appendix A2

Theorems and Proofs

Theorem (**Chebyshev's Inequality**): Let U be a non-negative random variable with a finite mean $\mu = \mathbb{E}(U)$. Then, for every $t > 0$,

$$\mathbb{P}(U > t\mu) \leqslant \frac{1}{t}$$

Proof. Note that :

$$\begin{aligned}\mu = \mathbb{E}(U) &= \int_0^{+\infty} u d\mathbb{P}(U \leqslant u) \\ &= \int_0^{t\mu} u d\mathbb{P}(U \leqslant u) + \int_{t\mu}^{+\infty} u d\mathbb{P}(U \leqslant u) \\ &\geqslant \int_{t\mu}^{+\infty} u d\mathbb{P}(U \leqslant u) \\ &\geqslant t\mu \int_{t\mu}^{+\infty} d\mathbb{P}(U \leqslant u) = t\mu \mathbb{P}(U > t\mu)\end{aligned}$$

and hence dividing both sides by $t\mu$, we obtain the result: $\frac{1}{t} \geqslant \mathbb{P}(U > t\mu)$. Assuming that f is the density of U, positive and integrable function, the same result can be obtain using f.

Theorem (**Markov's Inequality**): Let U be a non negative random variable with finite r^{th} moment μ_r =, let f be its density, a non negative integrable function, for some $r > 0$. Then, for every $\epsilon > 0$, we have:

$$\mathbb{P}(U > \varepsilon) \leqslant \frac{1}{\varepsilon^r} \mu_r$$

Proof. Let $V = U^r$, so that $\mathbb{E}(V) = \mu_r$. Then note that $[U > \epsilon] \Leftrightarrow [V > \epsilon^r]$. Applying Chebyshev's Inequality we get the following result:

$$\begin{aligned}\mathbb{P}(U > \varepsilon) &= \mathbb{P}(U^r > \varepsilon^r) \\ &= \mathbb{P}(V > \varepsilon^r)\end{aligned}$$

Using the density f, we get:

$$\begin{aligned}\mu_r = \int_0^{+\infty} u^r f(u)du &= \int_0^{\varepsilon} u^r f(u)du + \int_{\varepsilon}^{+\infty} u^r f(u)du \\ &\geqslant \int_{\varepsilon}^{+\infty} u^r f(u)du \\ &\geqslant \varepsilon^r \int_{\varepsilon}^{+\infty} f(u)du = \varepsilon^r \mathbb{P}\left(U > \varepsilon\right)\end{aligned}$$

hence dividing both sides by ϵ^r the result follows.

Theorem (**Central Limit Theorem convergence form**): Let X_k, $k \geq 1$, be independent random varibles such that $P(a \leq X_k \leq b) = 1$ for some finite scalars $a < b$. Also let $\mathbb{E}(X_k) = \mu_k$, $Var(X_k) = \sigma_k^2$, $T_n = \sum_{k=1}^{n} X_k$, $\xi_n = \sum_{k=1}^{n} \mu_k$ and $S_n^2 = \sum_{k=1}^{n} \sigma_k^2$.

Then

$$Z_n = \frac{T_n - \xi_n}{S_n} \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,1)$$

if and only if

$$S_n \xrightarrow[n\to+\infty]{} +\infty$$

.

Proof. It is a double implication proof.

- First suppose that $S_n \xrightarrow[n\to+\infty]{} +\infty$. Then note that

$$|X_k - \mu_k|^3 = |X_k - \mu_k| \left(X_k - \mu_k\right)^2 \leqslant (b-a) \left(X_k - \mu_k\right)^2$$

which implies that

$$E\left(|X_k - \mu_k|^3\right) \leqslant (b-a)\,\sigma_k^2$$

Therefore

$$\rho_n = \sum_{k=1}^{n} \frac{E\left(|X_k - \mu_k|^3\right)}{S_n^3} \leqslant \frac{b-a}{S_n} \xrightarrow[n\to+\infty]{} 0$$

and the result follows from Liapounov Theorem.

- Now let's supposed that $S_n \xrightarrow[n\to+\infty]{} S < +\infty$ and write

$$Z_n = \frac{T_n - \xi_n}{S_n} = \frac{X_1 - \mu_1}{S_n} + \sum_{k=2}^{n} \frac{X_k - \mu_k}{S_n}$$

Then if $Z_n \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,1)$ each of the terms of the right hand side must also converge to $\mathcal{N}(0,1)$ random variable; this is absurd since $\frac{X_1-\mu_1}{S_n}$ is a bounded random variable. Therefore $S_n \xrightarrow[n\to+\infty]{} +\infty$.

The fundamental theorem of integral calculus: Let f be a differentiable function define on $\mathbb{R}$, for $b > a$ elements of $\mathbb{R}$ we have :

$$f(b) - f(a) = \int_a^b g(x)dx$$

where $g(x) = f'(x)$ and f' is the derivative or the first derivative of f.

The derivative of f' or the second derivative of f is f'' or noted $f^{(2)}$ and the chain rule applies to define the n^{th} derivative $f^{(n)}(x)$ for every $n > 1$ element of $\mathbb{N}$ (whenever it exist). In some cases, the common notation is f'' and f''' for $f^{(2)}$

and $f^{(3)}$ respectively are used.

Let $f(x)$, $a \leq x \leq b$, be a continuous functions and having continuous derivatives up to $(k+)th$ order for some $k \geq 0$. Then for every $x \in [a, b]$ and $x_0 \in [a, b]$

$$f(x) = f(x_0) + \sum_{j=1}^{k} \frac{(x - x_0)^j}{j!} f^{(j)}(x_0) + R_k(x, x_0)$$

where

$$R_k(x, x_0) = \frac{(x - x_0)^{k+1}}{(k+1)!} f^{(k+1)}\left(hx_0 + (1-h)x\right)$$

for some $0 < h < 1$. This is known as the **Taylor Expansion** (up to the $k - th$ order) with a remainder term. Puting $k = +\infty$ leads to the **Taylor Series Expansion**.

Slutsky Theorem: Let X_n and Y_n be sequences of random variables such that $X_n \xrightarrow[n\to+\infty]{d} X$ and $Y_n \xrightarrow[n\to+\infty]{d} c$ where c is a constant. Then, it follows that:

1. $X_n + Y_n \xrightarrow[n\to+\infty]{d} X + c$,
2. $Y_n X_n \xrightarrow[n\to+\infty]{d} cX$,
3. $\frac{X_n}{Y_n} \xrightarrow[n\to+\infty]{d} \frac{X}{c}$ if $c \neq 0$.

Proof. Let F denote the distribution function of X, then for every $\epsilon > 0$ and

every $x \in \mathbb{R}$ a point where F is continue, we may write:

$$\begin{aligned}\mathbb{P}\left(X_n + Y_n \leqslant x\right) &= \mathbb{P}\left(X_n + Y_n \leqslant x; |Y_n - c| \leqslant \varepsilon\right) + \mathbb{P}\left(X_n + Y_n \leqslant x; |Y_n - c| > \varepsilon\right) \\ &\leqslant \mathbb{P}\left(X_n + Y_n \leqslant x; |Y_n - c| \leqslant \varepsilon\right) + \mathbb{P}\left(|Y_n - c| > \varepsilon\right) \\ &\leqslant \mathbb{P}\left(X_n \leqslant x - c + \varepsilon\right) + \mathbb{P}\left(|Y_n - c| > \varepsilon\right)\end{aligned}$$

Therefore,

$$\begin{aligned}\limsup_{n\to+\infty} \mathbb{P}\left(X_n + Y_n \leqslant x\right) &\leqslant \lim_{n\to+\infty} \mathbb{P}\left(X_n \leqslant x - c + \varepsilon\right) + \lim_{n\to+\infty} \mathbb{P}\left(|Y_n - c| > \varepsilon\right) \\ &= P\left(X_n \leqslant x - c + \varepsilon\right) = F(x - c + \varepsilon) \qquad (1)\end{aligned}$$

Also

$$\begin{aligned}\mathbb{P}\left(X_n + Y_n \leqslant x\right) &\geqslant \mathbb{P}\left(X_n + Y_n \leqslant x; |Y_n - c| \leqslant \varepsilon\right) \\ &\geqslant \mathbb{P}\left(X_n \leqslant x - c - \varepsilon; |Y_n - c| \leqslant \varepsilon\right) \\ &= \mathbb{P}\left(X_n \leqslant x - c - \varepsilon\right) - \mathbb{P}\left(X_n \leqslant x - c - \varepsilon; |Y_n - c| > \varepsilon\right) \\ &\geqslant \mathbb{P}\left(X_n \leqslant x - c - \varepsilon\right) - \mathbb{P}\left(|Y_n - c| > \varepsilon\right)\end{aligned}$$

Therefore

$$\begin{aligned}\liminf_{n\to+\infty} \mathbb{P}\left(X_n + Y_n \leqslant x\right) &\geqslant \lim_{n\to+\infty} \mathbb{P}\left(X_n \leqslant x - c + \varepsilon\right) - \lim_{n\to+\infty} \mathbb{P}\left(|Y_n - c| > \varepsilon\right) \\ &\geqslant \mathbb{P}\left(X_n \leqslant x - c - \varepsilon\right) = F(x - c - \varepsilon) \qquad (2)\end{aligned}$$

Putting (1) and (2) together we have :

$$\begin{aligned} F(x-c-\varepsilon) &\leqslant \liminf_{n\to+\infty} P\left(X_n+Y_n \leqslant x\right) \\ &\leqslant \limsup_{n\to+\infty} P\left(X_n+Y_n \leqslant x\right) \leqslant F(x-c+\varepsilon) \end{aligned}$$

Letting $\varepsilon \to 0$, it follows that $\lim_{n\to+\infty} P\left(X_n+Y_n \leqslant x\right) = F(x-c)$ and the proof of 1) is completed.

To prove 2), note that for $x \geq 0$ without loss of generality

$$\begin{aligned} \mathbb{P}\left(X_nY_n \leqslant x\right) &= \mathbb{P}\left(X_nY_n \leqslant x; \left|\frac{Y_n}{c}-1\right| \leqslant \varepsilon\right) + \mathbb{P}\left(X_nY_n \leqslant x; \left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \\ &\leqslant \mathbb{P}\left(X_n \leqslant \frac{x}{c(1-\varepsilon)}; \left|\frac{Y_n}{c}-1\right| \leqslant \varepsilon\right) + \mathbb{P}\left(\left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \\ &\leqslant \mathbb{P}\left(X_n \leqslant \frac{x}{c(1-\varepsilon)}\right) + \mathbb{P}\left(\left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \end{aligned}$$

Therefore, for $x \geq 0$

$$\begin{aligned} \limsup_{n\to+\infty} P\left(X_nY_n \leqslant x\right) &\leqslant \lim_{n\to+\infty} \mathbb{P}\left(X_n \leqslant \frac{x}{c(1-\varepsilon)}\right) + \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \\ &\leqslant F\left(\frac{x}{c(1-\varepsilon)}\right) \qquad (3) \end{aligned}$$

Similarly, for $x \geq 0$

$$\begin{aligned} \mathbb{P}\left(X_nY_n \leqslant x\right) &\geqslant \mathbb{P}\left(X_n \leqslant \frac{x}{c(1+\varepsilon)}; \left|\frac{Y_n}{c}-1\right| \leqslant \varepsilon\right) \\ &\geqslant \mathbb{P}\left(X_n \leqslant \frac{x}{c(1+\varepsilon)}\right) - \mathbb{P}\left(X_n \leqslant \frac{x}{c(1+\varepsilon)}; \left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \\ &\geqslant \mathbb{P}\left(X_n \leqslant \frac{x}{c(1+\varepsilon)}\right) - \mathbb{P}\left(\left|\frac{Y_n}{c}-1\right| > \varepsilon\right) \end{aligned}$$

Therefore, for $x \geq 0$

$$\liminf_{n\to+\infty} \mathbb{P}\left(X_n Y_n \leqslant x\right) \geqslant \lim_{n\to+\infty} \mathbb{P}\left(X_n \leqslant \frac{x}{c(1+\varepsilon)}\right) - \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{Y_n}{c} - 1\right| > \varepsilon\right)$$
$$\geqslant F\left(\frac{x}{c(1+\varepsilon)}\right) \qquad (4)$$

From (3) and (4) put together, we get

$$F\left(\frac{x}{c(1+\varepsilon)}\right) \leqslant \liminf_{n\to+\infty} P\left(X_n Y_n \leqslant x\right) \leqslant \limsup_{n\to+\infty} P\left(X_n Y_n \leqslant x\right) \leqslant F\left(\frac{x}{c(1-\varepsilon)}\right)$$

Letting $\varepsilon \to 0$, it follows that $\lim_{n\to+\infty} P\left(X_n Y_n \leqslant x\right) = F\left(\frac{x}{c}\right)$ hence 2) is proved.

Use same tricks and same development to prove 3).

Population Simulated Population RA

```
####CODE FOR POPULATION IN MCAR MODE (random assignment)
rm(list=ls())
N=10000
ID=1:N

##Generating covariates usefull for imputation
Cvt1=rbinom(N,1,0.55)
Cvt2=rbinom(N,3,0.25)
Cvt3=round(runif(N,15,60))
Cvt4=rbinom(N,2,1/3)

##Generating potential outcome before
POb=11*Cvt1+12*Cvt2+13*Cvt3+14*Cvt4+rnorm(N,500,250)

##Generating hypothetical potential outcome situation where
    everyone is treated and everyone is not treated
PO2T=POb+runif(N,250,500)
PO2NT=POb+runif(N,100,250)

##Generating treatment variable (Random assignment)
Tr=rbinom(N,1,0.46)

##Generating potential outcome after given that some are
    treated and others not
POa=PO2NT
for (i in 1:N){if (Tr[i]==1) {POa[i]=PO2T[i]}}
```

```
##Generating potential outcome of treated with non
    treated as missing
POt=POa
for (i in 1:N){if (Tr[i]==0) {POt[i]="NA"}}
POt=as.numeric(POt)

##Generating potential outcome of not treated with
    treated as missing
POnt=POa
for (i in 1:N){if (Tr[i]==1) {POnt[i]="NA"} }
POnt=as.numeric(POnt)

PoPData=data.frame (ID, POb, Cvt1, Cvt2, Cvt3, Cvt4, PO2T,
    PO2NT, Tr, POa, POt, POnt)
PoPDataMat=data.matrix(PoPData)

EvalData=PoPData
Eval1=subset(EvalData, Tr==1)
Eval0=subset(EvalData, Tr==0)

save(EvalData,file="EvalData.Rda")
save(Eval1,file="Eval1.Rda")
save(Eval0,file="Eval0.Rda")

##### In the bootstrap loop
n=50

load("EvalData.Rda")
load("Eval1.Rda")
load("Eval0.Rda")

SEval1=Eval1[sample(nrow(Eval1), n*0.46), ]
SEval0=Eval0[sample(nrow(Eval0), n*0.54), ]
FullData=rbind(SEval1, SEval0)

Xb1=FullData$Cvt1
Xb2=FullData$Cvt2
Xb3=FullData$Cvt3
Xb4=FullData$Cvt4
Yb=FullData$POb
Y2T=FullData$PO2T
Y2NT=FullData$PO2NT
T=FullData$Tr
Ya=FullData$POa
Yt=FullData$POt
Ynt=FullData$POnt
Data=data.frame (Ind=1:length(Yb), Yb, Xb1, Xb2, Xb3, Xb4, Y2T,
                 Y2NT, T, Ya, Yt, Ynt)
DataMat=data.matrix(Data)

##### True QTE in the Population
QTEpop=0
for (i in 1:9){
  QTEpop[i]=quantile(PO2T, i/10)-quantile(PO2NT, i/10)
  }
```

Population MAR and NMAR

```
####CODE FOR POPULATION IN MAR or NMAR MODE
rm(list=ls())
N=10000
ID=1:N

##Generating covariates usefull for imputation
Cvt1=rbinom(N,1,0.55)
Cvt2=rbinom(N,3,0.25)
Cvt3=round(runif(N,15,60))
Cvt4=rbinom(N,2,1/3)

##Generating potential outcome before
POb=11*Cvt1+12*Cvt2+13*Cvt3+14*Cvt4+rnorm(N,500,250)

##Generating hypothetical potential outcome situation
    where everyone is treated and everyone is not treated
PO2T=POb+runif(N,250,500)
PO2NT=POb+runif(N,100,250)

##Generating threshold variable
Inst=runif(N,10,6000)
TR=mean(Inst)
Tr=rep(0,N)

##Generating treatment variable (Assignment on purpose)
for (i in 1:N){if (Inst[i]<=TR) {Tr[i]=1}}

##Generating potential outcome after given that some are
    treated and others not
POa=PO2NT
for (i in 1:N){if (Tr[i]==1) {POa[i]=PO2T[i]}}

##Generating potential outcome of treated with non
    treated as missing
POt=POa
for (i in 1:N){if (Tr[i]==0) {POt[i]="NA"} }
POt=as.numeric(POt)

##Generating potential outcome of not treated with
    treated as missing
POnt=POa
for (i in 1:N){if (Tr[i]==1) {POnt[i]="NA"} }
POnt=as.numeric(POnt)

##Generating full data set of population
PoPData=data.frame (ID, POb, Cvt1, Cvt2, Cvt3, Cvt4, PO2T, PO2NT,
         Tr, POa, POt, POnt, Inst )
PoPDataMat=data.matrix(PoPData)

EvalData=PoPData
Eval1=subset(EvalData, Tr==1)
Eval0=subset(EvalData, Tr==0)

save(EvalData,file="EvalData.Rda")
save(Eval1,file="Eval1.Rda")
```

```
save(Eval0,file="Eval0.Rda")

##### In the bootstrap loop
n=50

load("EvalData.Rda")
load("Eval1.Rda")
load("Eval0.Rda")

SEval1=Eval1[sample(nrow(Eval1), n*0.46), ]
SEval0=Eval0[sample(nrow(Eval0), n*0.54), ]
FullData=rbind(SEval1, SEval0)

Xb1=FullData$Cvt1
Xb2=FullData$Cvt2
Xb3=FullData$Cvt3
Xb4=FullData$Cvt4
Yb=FullData$POb
Y2T=FullData$PO2T
Y2NT=FullData$PO2NT
T=FullData$Tr
Ya=FullData$POa
Yt=FullData$POt
Ynt=FullData$POnt
Data=data.frame (Ind=1:length(Yb), Yb, Xb1, Xb2, Xb3, Xb4,
        Y2T, Y2NT, T, Ya, Yt, Ynt)
DataMat=data.matrix(Data)

##### True QTE in the Population
QTEpop=0
for (i in 1:9){
  QTEpop[i]=quantile(PO2T, i/10)-quantile(PO2NT, i/10)
}
QTEpop
```

Bootstrap code to generate IB-DTE

```
####CODE FOR BOOTSTRAP
rm(list=ls())

n=2000

QTEclass=matrix(nrow =1000, ncol = 9)

ImpT1=0
ImpNT1=0
QTE1=matrix(nrow =1000, ncol = 9)
RMSEt1=0
RMSEnt1=0

ImpT11=0
ImpNT11=0
QTE11=matrix(nrow =1000, ncol = 9)
RMSEt11=0
RMSEnt11=0
```

```
ImpT2=0
ImpNT2=0
QTE2=matrix(nrow =1000, ncol = 9)
RMSEt2=0
RMSEnt2=0

ImpT22=0
ImpNT22=0
QTE22=matrix(nrow =1000, ncol = 9)
RMSEt22=0
RMSEnt22=0

ImpT3=0
ImpNT3=0
QTE3=matrix(nrow =1000, ncol = 9)
RMSEt3=0
RMSEnt3=0

ImpT4=0
ImpNT4=0
QTE4=matrix(nrow =1000, ncol = 9)
RMSEt4=0
RMSEnt4=0

ImpT44=0
ImpNT44=0
QTE44=matrix(nrow =1000, ncol = 9)
RMSEt44=0
RMSEnt44=0

ImpT5=0
ImpNT5=0
QTE5=matrix(nrow =1000, ncol = 9)
RMSEt5=0
RMSEnt5=0

ImpT6=0
ImpNT6=0
QTE6=matrix(nrow =1000, ncol = 9)
RMSEt6=0
RMSEnt6=0

ImpT77=0
ImpNT77=0
QTE77=matrix(nrow =1000, ncol = 9)
RMSEt77=0
RMSEnt77=0

ImpT7=0
ImpNT7=0
QTE7=matrix(nrow =1000, ncol = 9)
RMSEt7=0
RMSEnt7=0

ImpT8=0
ImpNT8=0
```

```
QTE8=matrix(nrow =1000, ncol = 9)
RMSEt8=0
RMSEnt8=0

ImpT10=0
ImpNT10=0
QTE10=matrix(nrow =1000, ncol = 9)
RMSEt10=0
RMSEnt10=0


for (k in 1:1000) {

  ##Sampling of size
  load("EvalData.Rda")
  load("Eval1.Rda")
  load("Eval0.Rda")
  SEval1=Eval1[sample(nrow(Eval1), n*0.46), ]
  SEval0=Eval0[sample(nrow(Eval0), n*0.54), ]
  FullData=rbind(SEval1, SEval0)

  Xb1=FullData$Cvt1
  Xb2=FullData$Cvt2
  Xb3=FullData$Cvt3
  Xb4=FullData$Cvt4
  Yb=FullData$POb
  Y2T=FullData$PO2T
  Y2NT=FullData$PO2NT
  T=FullData$Tr
  Ya=FullData$POa
  Yt=FullData$POt
  Ynt=FullData$POnt
  Data=data.frame (Ind=1:length(T), Yb, Xb1, Xb2, Xb3,
                Xb4, Y2T, Y2NT, T, Ya, Yt, Ynt)
  DataMat=data.matrix(Data)

  ##Classic QTE

  for (j in 1:9){
    QTEclass[k,j]=quantile(Yt, j/10, na.rm = TRUE)-
                quantile(Ynt, j/10, na.rm = TRUE)  }

  ##Imp Based estimators
  ##Mean imputation
  Yt1=mean.imp(Yt)
  Ynt1=mean.imp(Ynt)
  ImpT1=ImpT1+Yt1
  ImpNT1=ImpNT1+Ynt1
  RMSEt1[k]=RMSE(Y2T,Yt1)
  RMSEnt1[k]=RMSE(Y2NT,Ynt1)
  for (j in 1:9){
    QTE1[k,j]=quantile(Yt1, j/10, na.rm = TRUE)-
            quantile(Ynt1, j/10, na.rm = TRUE)  }

  Yt11=cmean.imp(Yt, Xb1)
  Ynt11=cmean.imp(Ynt, Xb1)
```

```
ImpT11=ImpT11+Yt11
ImpNT11=ImpNT11+Ynt11
RMSEt11[k]=RMSE(Y2T,Yt11)
RMSEnt11[k]=RMSE(Y2NT,Ynt11)
for (j in 1:9){
  QTE11[k,j]=quantile(Yt11, j/10, na.rm = TRUE)-
             quantile(Ynt11, j/10, na.rm = TRUE) }

##Random Imputation
Yt2=random.imp(Yt)
Ynt2=random.imp(Ynt)
ImpT2=ImpT2+Yt2
ImpNT2=ImpNT2+Ynt2
RMSEt2[k]=RMSE(Y2T,Yt2)
RMSEnt2[k]=RMSE(Y2NT,Ynt2)
for (j in 1:9){
  QTE2[k,j]=quantile(Yt2, j/10, na.rm = TRUE)-
          quantile(Ynt2, j/10, na.rm = TRUE) }

Data2=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
Imp.base=hotdeck(Data2)
Yt22=Imp.base$Yt
Ynt22=Imp.base$Ynt
ImpT22=ImpT22+Yt22
ImpNT22=ImpNT22+Ynt22
RMSEt22[k]=RMSE(Y2T,Yt22)
RMSEnt22[k]=RMSE(Y2NT,Ynt22)
for (j in 1:9){
  QTE22[k,j]=quantile(Yt22, j/10, na.rm = TRUE)-
          quantile(Ynt22, j/10, na.rm = TRUE) }

##Regression to perform deterministic imputation
lm.imp.1=lm(Yt~Xb1 + Xb2 + Xb3 + Xb4)
          #Regression model for treated units
lm.imp.2=lm(Ynt~Xb1 + Xb2 + Xb3 + Xb4)
          #Regression model for non treated units
pred.1=predict(lm.imp.1, Data)
          #Predict using deterministic part
                  of reg model (treated)
pred.2=predict(lm.imp.2, Data)
          #Predict using deterministic part
                  of reg model (non treated)

Yt3=reg.imp(Yt, pred.1)
Ynt3=reg.imp(Ynt, pred.2)
ImpT3=ImpT3+Yt3
ImpNT3=ImpNT3+Ynt3
RMSEt3[k]=RMSE(Y2T,Yt3)
RMSEnt3[k]=RMSE(Y2NT,Ynt3)
for (j in 1:9){
  QTE3[k,j]=quantile(Yt3, j/10, na.rm = TRUE)-
          quantile(Ynt3, j/10, na.rm = TRUE) }

##Nearest neigbour imputation (VIM Package)
Data4=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
Imp.data=kNN(Data4, k=1)
```

```
  Yt4=Imp.data$Yt
  Ynt4=Imp.data$Ynt
  ImpT4=ImpT4+Yt4
  ImpNT4=ImpNT4+Ynt4
  RMSEt4[k]=RMSE(Y2T,Yt4)
  RMSEnt4[k]=RMSE(Y2NT,Ynt4)
  for (j in 1:9){
    QTE4[k,j]=quantile(Yt4, j/10, na.rm = TRUE)-
            quantile(Ynt4, j/10, na.rm = TRUE) }

sampImp <- kNN(Data4, dist_var = c("Xb1", "Xb2", "Xb3", "Xb4"),
         k = 1, numFun = mean)
  Yt44=sampImp$Yt
  Ynt44=sampImp$Ynt
  ImpT44=ImpT44+Yt44
  ImpNT44=ImpNT44+Ynt44
  RMSEt44[k]=RMSE(Y2T,Yt44)
  RMSEnt44[k]=RMSE(Y2NT,Ynt44)
  for (j in 1:9){
    QTE44[k,j]=quantile(Yt44, j/10, na.rm = TRUE)-
            quantile(Ynt44, j/10, na.rm = TRUE) }

  ##Regression with random prediction
  lm.imp.3=lm(Yt~Xb1 + Xb2 + Xb3 + Xb4)
        #Regression model for treated units
  lm.imp.4=lm(Ynt~Xb1 + Xb2 + Xb3 + Xb4)
        #Regression model for non treated units
  pred.3=rnorm(n, predict(lm.imp.3, Data),
        summary(lm.imp.3)$sigma)
                #Random prediction of reg model (treated)
  pred.4=rnorm(n, predict(lm.imp.4, Data),
        summary(lm.imp.4)$sigma)
        #Random prediction of reg model (non treated)
  Yt5=reg.imp(Yt, pred.3)
  Ynt5=reg.imp(Ynt, pred.4)
  ImpT5=ImpT5+Yt5
  ImpNT5=ImpNT5+Ynt5
  RMSEt5[k]=RMSE(Y2T,Yt5)
  RMSEnt5[k]=RMSE(Y2NT,Ynt5)
  for (j in 1:9){
    QTE5[k,j]=quantile(Yt5, j/10, na.rm = TRUE)-
            quantile(Ynt5, j/10, na.rm = TRUE)  }

  ##Multiple imputation using package MICE
  Data6=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
  imp=mice(Data6, m=5)
  Yt6=(complete(imp, 1)$Yt+complete(imp, 2)$Yt
    +complete(imp, 3)$Yt+complete(imp, 4)$Yt
    +complete(imp, 5)$Yt)/5
  Ynt6=(complete(imp, 1)$Ynt+complete(imp, 2)$Ynt
  +complete(imp, 3)$Ynt+complete(imp, 4)$Ynt
  +complete(imp, 5)$Ynt)/5
  ImpT6=ImpT6+Yt6
  ImpNT6=ImpNT6+Ynt6
  RMSEt6[k]=RMSE(Y2T,Yt6)
  RMSEnt6[k]=RMSE(Y2NT,Ynt6)
```

```
for (j in 1:9){
  QTE6[k,j]=quantile(Yt6, j/10, na.rm = TRUE)-
      quantile(Ynt6, j/10, na.rm = TRUE) }

##Maximum likelihood using MissMech package (check)
Data7=data.frame(cbind(Yt, Ynt, Xb1, Xb2, Xb3, Xb4))
Dat7=data.matrix(Data7)

Res77=Impute(Dat7, mu = NA, sig = NA,
      imputation.method = "normal", resid = NA)
Yt77=Res77$yimp[,1]
Ynt77=Res77$yimp[,2]
ImpT77=ImpT77+Yt77
ImpNT77=ImpNT77+Ynt77
RMSEt77[k]=RMSE(Y2T,Yt77)
RMSEnt77[k]=RMSE(Y2NT,Ynt77)
for (j in 1:9){
  QTE77[k,j]=quantile(Yt77, j/10, na.rm = TRUE)-
          quantile(Ynt77, j/10, na.rm = TRUE) }

Res7=Impute(Dat7, mu = NA, sig = NA,
      imputation.method = "dist.free", resid = NA)
              #method can also be normal
Yt7=Res7$yimp[,1]
Ynt7=Res7$yimp[,2]
ImpT7=ImpT7+Yt7
ImpNT7=ImpNT7+Ynt7
RMSEt7[k]=RMSE(Y2T,Yt7)
RMSEnt7[k]=RMSE(Y2NT,Ynt7)
for (j in 1:9){
  QTE7[k,j]=quantile(Yt7, j/10, na.rm = TRUE)-
      quantile(Ynt7, j/10, na.rm = TRUE) }

##Propensity score imputation
mylogit <- glm(T~Xb1 + Xb2 + Xb3 + Xb4, family=binomial)
X <- mylogit$fitted
result <- Match(Y=Ya, Tr=T, X=X, estimand = "ATE", M=1)
Yt8=result$mdata$Y[1:n]
Ynt8=result$mdata$Y[(n+1):(2*(n))]
ImpT8=ImpT8+Yt8
ImpNT8=ImpNT8+Ynt8
RMSEt8[k]=RMSE(Y2T,Yt8)
RMSEnt8[k]=RMSE(Y2NT,Ynt8)
for (j in 1:9){
  QTE8[k,j]=quantile(Yt8, j/10, na.rm = TRUE)-
  quantile(Ynt8, j/10, na.rm = TRUE) }

##Quantile regression imputation
rq.imp1=rq(Yt~Xb1+Xb2+Xb3+Xb4, c(0.1, 0.2, 0.3, 0.4,
      0.5, 0.6, 0.7, 0.8, 0.9), Data)
rq.imp2=rq(Ynt~Xb1+Xb2+Xb3+Xb4, c(0.1, 0.2, 0.3, 0.4,
      0.5, 0.6, 0.7, 0.8, 0.9), Data)
p.rq1=predict(rq.imp1, Data)
p.rq2=predict(rq.imp2, Data)
Yt10=Yt
for (i in 1:length(Yb)) {
```

```
  if (Yb[i]<=quantile(Yb, 0.1) && is.na(Yt[i])==1)
          { Yt10[i]=p.rq1[,1][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.1) && Yb[i]<=quantile(Yb, 0.2)
           && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,2][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.2) && Yb[i]<=quantile(Yb, 0.3)
           && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,3][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.3) && Yb[i]<=quantile(Yb, 0.4)
          && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,4][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.4) && Yb[i]<=quantile(Yb, 0.5)
          && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,5][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.5) && Yb[i]<=quantile(Yb, 0.6)
          && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,6][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.6) && Yb[i]<=quantile(Yb, 0.7)
          && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,7][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.7) && Yb[i]<=quantile(Yb, 0.8)
          && is.na(Yt[i])==1) { Yt10[i]=p.rq1[,8][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.8) && is.na(Yt[i])==1)
          { Yt10[i]=p.rq1[,9][i] } }
Ynt10=Ynt
for (i in 1:length(Yb)) {
  if (Yb[i]<=quantile(Yb, 0.1) && is.na(Ynt[i])==1)
           { Ynt10[i]=p.rq2[,1][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.1) && Yb[i]<=quantile(Yb, 0.2)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,2][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.2) && Yb[i]<=quantile(Yb, 0.3)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,3][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.3) && Yb[i]<=quantile(Yb, 0.4)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,4][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.4) && Yb[i]<=quantile(Yb, 0.5)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,5][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.5) && Yb[i]<=quantile(Yb, 0.6)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,6][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.6) && Yb[i]<=quantile(Yb, 0.7)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,7][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.7) && Yb[i]<=quantile(Yb, 0.8)
          && is.na(Ynt[i])==1) { Ynt10[i]=p.rq2[,8][i] } }
for (i in 1:length(Yb)) {
  if (Yb[i]>quantile(Yb, 0.8) && is.na(Ynt[i])==1)
          { Ynt10[i]=p.rq2[,9][i] } }

ImpT10=ImpT10+Yt10
```

```
  ImpNT10=ImpNT10+Ynt10
  RMSEt10[k]=RMSE(Y2T, Yt10)
  RMSEnt10[k]=RMSE(Y2NT, Ynt10)
  for (j in 1:9){
    QTE10[k,j]=quantile(Yt10, j/10, na.rm = TRUE)-
        quantile(Ynt10, j/10, na.rm = TRUE) } }

#Summary results

##True QTE and Classic QTE
QTEpop=0
for (i in 1:9){
  QTEpop[i]=quantile(PO2T, i/10)-quantile(PO2NT, i/10)}
QTEpop
colMeans(QTEclass)

#Summary mean imputation
BQTE1=0
for (i in 1:9){
  BQTE1[i]=quantile(ImpT1/1000,i/10)-quantile(ImpNT1/1000,i/10)}
BQTE1
colMeans(QTE1)

data1=cbind(1:n,ImpNT1/1000,ImpT1/1000)
data1=data1[order(data1[,2],decreasing=FALSE),]
GQTE1=0
for (i in 1:9){
  GQTE1[i]=data1[n*i/10,3]-data1[n*i/10,2]}
GQTE1

BQTE11=0
for (i in 1:9){
  BQTE11[i]=quantile(ImpT11/1000, i/10)-quantile(ImpNT11/1000, i/10)}
BQTE11
colMeans(QTE11)

data11=cbind(1:n,ImpNT11/1000,ImpT11/1000)
data11=data11[order(data11[,2],decreasing=FALSE),]
GQTE11=0
for (i in 1:9){
  GQTE11[i]=data11[n*i/10,3]-data11[n*i/10,2]}
GQTE11

#Summary random imputation imputation
BQTE2=0
for (i in 1:9){
  BQTE2[i]=quantile(ImpT2/1000, i/10)-quantile(ImpNT2/1000, i/10)}
BQTE2
colMeans(QTE2)

data2=cbind(1:n,ImpNT2/1000,ImpT2/1000)
data2=data2[order(data2[,2],decreasing=FALSE),]
GQTE2=0
for (i in 1:9){
  GQTE2[i]=data2[n*i/10,3]-data2[n*i/10,2]}
GQTE2
```

```
BQTE22=0
for (i in 1:9){
  BQTE22[i]=quantile(ImpT22/1000,i/10)-quantile(ImpNT22/1000,i/10)}
BQTE22
colMeans(QTE22)

data22=cbind(1:n,ImpNT22/1000,ImpT22/1000)
data22=data22[order(data22[,2],decreasing=FALSE),]
GQTE22=0
for (i in 1:9){
  GQTE22[i]=data22[n*i/10,3]-data22[n*i/10,2]}
GQTE22

#Summary linear regression deterministic imp
BQTE3=0
for (i in 1:9){
  BQTE3[i]=quantile(ImpT3/1000, i/10)-quantile(ImpNT3/1000, i/10)}
BQTE3
colMeans(QTE3)

data3=cbind(1:n,ImpNT3/1000,ImpT3/1000)
data3=data3[order(data3[,2],decreasing=FALSE),]
GQTE3=0
for (i in 1:9){
  GQTE3[i]=data3[n*i/10,3]-data3[n*i/10,2]}
GQTE3

#Summary nearest neigbour imputation
BQTE4=0
for (i in 1:9){
  BQTE4[i]=quantile(ImpT4/1000, i/10)-quantile(ImpNT4/1000, i/10)}
BQTE4
colMeans(QTE4)

data4=cbind(1:n,ImpNT4/1000,ImpT4/1000)
data4=data4[order(data4[,2],decreasing=FALSE),]
GQTE4=0
for (i in 1:9){
  GQTE4[i]=data4[n*i/10,3]-data4[n*i/10,2]}
GQTE4

BQTE44=0
for (i in 1:9){
  BQTE44[i]=quantile(ImpT44/1000,i/10)-quantile(ImpNT44/1000,i/10)}
BQTE44
colMeans(QTE44)

data44=cbind(1:n,ImpNT44/1000,ImpT44/1000)
data44=data44[order(data44[,2],decreasing=FALSE),]
GQTE44=0
for (i in 1:9){
  GQTE44[i]=data44[n*i/10,3]-data44[n*i/10,2]}
GQTE44

#Summary regression with random imputation
```

```
BQTE5=0
for (i in 1:9){
  BQTE5[i]=quantile(ImpT5/1000, i/10)-
    quantile(ImpNT5/1000, i/10)}
BQTE5
colMeans(QTE5)

data5=cbind(1:n,ImpNT5/1000,ImpT5/1000)
data5=data5[order(data5[,2],decreasing=FALSE),]
GQTE5=0
for (i in 1:9){
  GQTE5[i]=data5[n*i/10,3]-data5[n*i/10,2]}
GQTE5

#Summary multiple imputation using mICE
BQTE6=0
for (i in 1:9){
  BQTE6[i]=quantile(ImpT6/1000, i/10)-
    quantile(ImpNT6/1000, i/10)}
BQTE6
colMeans(QTE6)

data6=cbind(1:n,ImpNT6/1000,ImpT6/1000)
data6=data6[order(data6[,2],decreasing=FALSE),]
GQTE6=0
for (i in 1:9){
  GQTE6[i]=data6[n*i/10,3]-data6[n*i/10,2]}
GQTE6

#Summary maximum likelihood
BQTE77=0
for (i in 1:9){
  BQTE77[i]=quantile(ImpT77/1000, i/10)-
    quantile(ImpNT77/1000, i/10)}
BQTE77
colMeans(QTE77)

data77=cbind(1:n,ImpNT77/1000,ImpT77/1000)
data77=data77[order(data77[,2],decreasing=FALSE),]
GQTE77=0
for (i in 1:9){
  GQTE77[i]=data77[n*i/10,3]-data77[n*i/10,2]}
GQTE77

BQTE7=0
for (i in 1:9){
  BQTE7[i]=quantile(ImpT7/1000, i/10)-
        quantile(ImpNT7/1000, i/10)}
BQTE7
colMeans(QTE7)

data7=cbind(1:n,ImpNT7/1000,ImpT7/1000)
data7=data7[order(data7[,2],decreasing=FALSE),]
GQTE7=0
for (i in 1:9){
  GQTE7[i]=data7[n*i/10,3]-data7[n*i/10,2]}
```

```
GQTE7

#Summary propensity score imutation
BQTE8=0
for (i in 1:9){
  BQTE8[i]=quantile(ImpT8/1000, i/10)-quantile(ImpNT8/1000, i/10)}
BQTE8
colMeans(QTE8)

data8=cbind(1:n,ImpNT8/1000,ImpT8/1000)
data8=data8[order(data8[,2],decreasing=FALSE),]
GQTE8=0
for (i in 1:9){
  GQTE8[i]=data8[n*i/10,3]-data8[n*i/10,2]}
GQTE8

#Summary quantile regression imputation
BQTE10=0
for (i in 1:9){
  BQTE10[i]=quantile(ImpT10/1000,i/10)-quantile(ImpNT10/1000,i/10)}
BQTE10
colMeans(QTE10)

data10=cbind(1:n,ImpNT10/1000,ImpT10/1000)
data10=data10[order(data10[,2],decreasing=FALSE),]
GQTE10=0
for (i in 1:9){
  GQTE10[i]=data10[n*i/10,3]-data10[n*i/10,2]}
GQTE10

#Matrices outputs
Results1=rbind(BQTE1,BQTE11,BQTE2,BQTE22,BQTE3,BQTE4,
        BQTE44,BQTE5,BQTE6,BQTE77,BQTE7,BQTE8,BQTE10)
Results1

Results2=rbind(colMeans(QTE1),colMeans(QTE11),
    colMeans(QTE2),colMeans(QTE22),colMeans(QTE3),
    colMeans(QTE4),colMeans(QTE44),colMeans(QTE5),
    colMeans(QTE6),colMeans(QTE77),colMeans(QTE7),
    colMeans(QTE8),colMeans(QTE10))
Results2

Results3=rbind(GQTE1,GQTE11,GQTE2,GQTE22,GQTE3,GQTE4,
    GQTE44,GQTE5,GQTE6,GQTE77, GQTE7, GQTE8, GQTE10)
Results3
```

Appendix A3

Code for generating data, classical tests of comparison of distribution and MTP approaches implemented.

```
####CODE FOR POPULATION IN MCAR MODE (random assignment)
rm(list=ls())
N=1000

####Poutential outcome before treatment
Id=1:N
Y0=rnorm(N,1000,100)
Y1=Y0 + rnorm(N,250,25)
Y1Sav=Y1
M=cbind(Id, Y0, Y1)

####Effect of different treatment and potential
        outcome after treatment (six PO)
Index=sample(M[,1],0.1*N, F)
M10=M
M10[Index,2]=M10[Index,3]
Y10=M10[,2]

Index=sample(M[,1],0.2*N, F)
M20=M
M20[Index,2]=M20[Index,3]
Y20=M20[,2]

Index=sample(M[,1],0.3*N, F)
M30=M
M30[Index,2]=M30[Index,3]
Y30=M30[,2]

Index=sample(M[,1],0.4*N, F)
M40=M
M40[Index,2]=M40[Index,3]
Y40=M40[,2]

Index=sample(M[,1],0.5*N, F)
M50=M
M50[Index,2]=M50[Index,3]
Y50=M50[,2]

Index=sample(M[,1],0.6*N, F)
M60=M
M60[Index,2]=M60[Index,3]
Y60=M60[,2]

Index=sample(M[,1],0.7*N, F)
M70=M
M70[Index,2]=M70[Index,3]
Y70=M70[,2]
```

```
Index=sample(M[,1],0.8*N, F)
M80=M
M80[Index,2]=M80[Index,3]
Y80=M80[,2]

Index=sample(M[,1],0.9*N, F)
M90=M
M90[Index,2]=M90[Index,3]
Y90=M90[,2]

####Graphs of potential outcome
hist(Y0)
ecdf(Y0)

par(mfrow=c(2,2))

plot(ecdf(Y0), col="red", xlab = '', ylab = '',
    main = 'Empirical Cumluative Distribution\nPotential Outcome')
lines(ecdf(Y10), lty=2, col="blue")
mtext(text = expression(hat(F)[n](x)), side = 2, line = 2.5)
legend ('topleft', legend=c("Y0", "Y10") ,
    lty=1, col=c('red', 'blue'), bty='n', cex=.75)

plot(ecdf(Y0), col="red", xlab = '', ylab = '',
    main = 'Empirical Cumluative Distribution\nPotential Outcome')
lines(ecdf(Y50), lty=2, col="blue")
legend ('topleft', legend=c("Y0", "Y50") , lty=1,
    col=c('red', 'blue'), bty='n', cex=.75)

plot(ecdf(Y0), col="red", xlab = 'Sample Quantiles of Outcomes',
    ylab = '', main = '')
lines(ecdf(Y80), lty=2, col="blue")
mtext(text = expression(hat(F)[n](x)), side = 2, line = 2.5)
legend ('topleft', legend=c("Y0", "Y80") , lty=1,
    col=c('red', 'blue'), bty='n', cex=.75)

plot(ecdf(Y0), col="red", xlab = 'Sample Quantiles of Outcomes',
    ylab = '', main = '')
lines(ecdf(Y1), lty=2, col="blue")
legend ('topleft', legend=c("Y0", "Y1") , lty=1,
    col=c('red', 'blue'), bty='n', cex=.75)

####KS and Wilcox tests
ks.test(Y0,Y0)
wilcox.test(Y0,Y0)

ks.test(Y0,Y10)
wilcox.test(Y0,Y10)

ks.test(Y0,Y20)
wilcox.test(Y0,Y20)

ks.test(Y0,Y30)
wilcox.test(Y0,Y30)
```

```
ks.test(Y0,Y40)
wilcox.test(Y0,Y40)

ks.test(Y0,Y50)
wilcox.test(Y0,Y50)

ks.test(Y0,Y60)
wilcox.test(Y0,Y60)

ks.test(Y0,Y70)
wilcox.test(Y0,Y70)

ks.test(Y0,Y80)
wilcox.test(Y0,Y80)

ks.test(Y0,Y90)
wilcox.test(Y0,Y90)

ks.test(Y0,Y1)
wilcox.test(Y0,Y1)

####MTP section

#Y1=Y20
Y1=Y50

##Point Wise Test CDF
M=min(max(Y0), max(Y1))
m=max(min(Y0), min(Y1))

st=(M-m)/10

r=m+st
test=t.test(Y0<r, Y1<r)
P1=test$p.value

for (i in 2:9){
  r=m+i*st
  test=t.test(Y0<r, Y1<r)
  P1=c(P1, test$p.value)}
P1

##Point Wise Test for quantile function

DQ1=0; DQ2=0; DQ3=0; DQ4=0; DQ5=0; DQ6=0; DQ7=0; DQ8=0; DQ9=0
s=0.85

for (i in 1:1000) {
  SY0=sample(Y0, s*N, F)
  SY1=sample(Y1, s*N, F)
    DQ1[i]=quantile(SY1, 1/10)-quantile(SY0, 1/10)
    DQ2[i]=quantile(SY1, 2/10)-quantile(SY0, 2/10)
    DQ3[i]=quantile(SY1, 3/10)-quantile(SY0, 3/10)
    DQ4[i]=quantile(SY1, 4/10)-quantile(SY0, 4/10)
    DQ5[i]=quantile(SY1, 5/10)-quantile(SY0, 5/10)
    DQ6[i]=quantile(SY1, 6/10)-quantile(SY0, 6/10)
```

```
    DQ7[i]=quantile(SY1, 7/10)-quantile(SY0, 7/10)
    DQ8[i]=quantile(SY1, 8/10)-quantile(SY0, 8/10)
    DQ9[i]=quantile(SY1, 9/10)-quantile(SY0, 9/10)}

test=t.test(DQ1, mu=0)
P3=test$p.value
for (i in 2:9){
  test=t.test(get(paste("DQ", i, sep = "")), mu=0)
  P3=c(P3, test$p.value)}
P3

##Point Wise Test for quantile groups
G01=0
G11=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 1/10)){G01=c(G01, Y0[i])}
  if (Y1[i] < quantile(Y1, 1/10)){G11=c(G11, Y1[i])}
}
G01=G01[2:length(G01)]
G11=G11[2:length(G11)]

G02=0
G12=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 2/10) && Y0[i] >=
        quantile(Y0, 1/10)){G02=c(G02, Y0[i])}
  if (Y1[i] < quantile(Y1, 2/10) && Y1[i] >=
        quantile(Y1, 1/10)){G12=c(G12, Y1[i])}}
G02=G02[2:length(G02)]
G12=G12[2:length(G12)]

G03=0
G13=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 3/10) && Y0[i] >=
        quantile(Y0, 2/10)){G03=c(G03, Y0[i])}
  if (Y1[i] < quantile(Y1, 3/10) && Y1[i] >=
        quantile(Y1, 2/10)){G13=c(G13, Y1[i])}}
G03=G03[2:length(G03)]
G13=G13[2:length(G13)]

G04=0
G14=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 4/10) && Y0[i] >=
        quantile(Y0, 3/10)){G04=c(G04, Y0[i])}
  if (Y1[i] < quantile(Y1, 4/10) && Y1[i] >=
        quantile(Y1, 3/10)){G14=c(G14, Y1[i])}}
G04=G04[2:length(G04)]
G14=G14[2:length(G14)]

G05=0
G15=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 5/10) && Y0[i] >=
        quantile(Y0, 4/10)){G05=c(G05, Y0[i])}
```

```
  if (Y1[i] < quantile(Y1, 5/10) && Y1[i] >=
        quantile(Y1, 4/10)){G15=c(G15, Y1[i])}}
G05=G05[2:length(G05)]
G15=G15[2:length(G15)]

G06=0
G16=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 6/10) && Y0[i] >=
        quantile(Y0, 5/10)){G06=c(G06, Y0[i])}
  if (Y1[i] < quantile(Y1, 6/10) && Y1[i] >=
        quantile(Y1, 5/10)){G16=c(G16, Y1[i])}}
G06=G06[2:length(G06)]
G16=G16[2:length(G16)]

G07=0
G17=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 7/10) && Y0[i] >=
        quantile(Y0, 6/10)){G07=c(G07, Y0[i])}
  if (Y1[i] < quantile(Y1, 7/10) && Y1[i] >=
        quantile(Y1, 6/10)){G17=c(G17, Y1[i])}}
G07=G07[2:length(G07)]
G17=G17[2:length(G17)]

G08=0
G18=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 8/10) && Y0[i] >=
        quantile(Y0, 7/10)){G08=c(G08, Y0[i])}
  if (Y1[i] < quantile(Y1, 8/10) && Y1[i] >=
        quantile(Y1, 7/10)){G18=c(G18, Y1[i])}}
G08=G08[2:length(G08)]
G18=G18[2:length(G18)]

G09=0
G19=0
for (i in 1:1000){
  if (Y0[i] < quantile(Y0, 9/10) && Y0[i] >=
        quantile(Y0, 8/10)){G09=c(G09, Y0[i])}
  if (Y1[i] < quantile(Y1, 9/10) && Y1[i] >=
        quantile(Y1, 8/10)){G19=c(G19, Y1[i])}}
G09=G09[2:length(G09)]
G19=G19[2:length(G19)]

test=t.test(G01, G11)
P2=test$p.value
for (i in 2:9){
  test=t.test(get(paste("G0", i, sep = "")), get(paste("G1", i, sep = "")))
  P2=c(P2, test$p.value)}
P2

##MTP test with prob P1, P2 and P3 at 5% (sgof)
install.packages("sgof")
install.packages("poibin")
library(poibin)
```

```
library(sgof)

mtp1=BH(P1, alpha=0.05)
summary(mtp1)
mtp1$Adjusted.pvalues
mtp1$Rejections
mtp1$FDR

mtp3=BH(P3, alpha=0.05)
summary(mtp3)
mtp3$Adjusted.pvalues
mtp3$Rejections
mtp3$FDR

mtp2=BH(P2, alpha=0.05)
summary(mtp2)
mtp2$Adjusted.pvalues
mtp2$Rejections
mtp2$FDR
```

Appendix A4

1. Kenfac D. P. B., Mwita N. P. and Kamga T. I. R (2017), "Imputation Based Treatment Effect Estimators", Journal of Statistical and Econometric Methods, vol. 6, no. 3, 2017, 23-64, ISSN: 1792-6602 (print), 1792-6939 (online) Scienpress Ltd, 2017.

2. Kenfac D. P. B., Mwita N. P. and Kamga T. I. R (2018), "Distributive and Quantile Treatment Effects: Imputation Based Estimators Approach", Journal of Statistical and Econometric Methods, vol. 7, no. 2, 2018, 43-69, ISSN: 2241-0384 (print), 2241-0376 (online) Scienpress Ltd, 2018.

3. Kenfac D. P. B., Mwita N. P. and Kamga T. I. R (2018), "Performance of imputation methods towards increasing percentage of missing values", Proceedings of 1st International Conference, Machakos University, Page

1008-1025, April 2018.

4. Kenfac D. P. B., Mwita N. P. and Kamga T. I. R (2018-2019), "Testing the Hypothesis "No effects" using Marginal Distributions of Potential Outcome and Multiple Testing Procedure": In review at Journal of King Saud University - Science.