Imputation Based Treatment Effect Estimators

P. B. Kenfac Dongmezo[1], P. N. Mwita[2] and I. R. Kamga Tchwaket[3]

Abstract

The problem of counterfactual and control group is at the core of impact evaluation. Almost all existing methods aim to find the best control group to compare with the treated group. The aim of this study is to use imputation methods to estimate counterfactual and derive average treatment effect estimators from the data sets completed using the basic definition of treatment effect described in Rubin framework. The estimators obtained are called Imputation Based Treatment Effects estimators. A number of imputation methods are tested, among them there is Maximum likelihood, Multiple Imputation, Linear and Quantile regressions. Using simulations and bootstrap methodology, we found that the best imputation methods (data reconstruction) in the framework of impact evaluation are Quantile regression and Multiple Imputation. We also found that our estimators (taking average) obtained from data imputed are convergent and can perform as well as average treatment effects estimators obtained from classical methods such as Difference in Difference and Propensity Score Matching.

[1] Pan African University Institute for Basic Sciences, Technology and Innovation (PAUISTI), Kenya
[2] Machakos University, .PO Box 136-90100; Machakos, Kenya.
[3] Institut Sous régional de Statistique et d'Economie Appliquée (ISSEA)
P.O Box 294; Yaoundé, Cameroon.

Imputation Based Treatment Effect estimators are then tested on a program (Lalonde data) and the results show that they can perform as well as existing estimators and even better in certain cases especially when there is a shortage in data.

Mathematics Subject Classification: 62M10

Keywords: Average Bias, Estimator, Impact, Imputation, Treatment Effect

1   Introduction

First works on Impact Evaluation (IE) were done by Rubin in 1970's. He developed what is called today "Rubin Framework" in which he explained how to get the effect of a specific action on a unit (Individual, County, Household etc.). Vocabulary used in that framework comes from medical experiments. Hence, what he calls "treatment" can be seen as administration of a drug, receiving a training, a given implemented policy or something else which is supposed to make a different between two groups in the same population. In IE, we are interested in the assessment 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 the values $T_i = 1$ and $Y = Y_{i1}$ while for a non-treated unit $T_i = 0$ and $Y = Y_{i0}$. Therefore, for a given unit, Rubin defined the causal effect of the treatment T by the following quantity $\Delta_i = Y_{i1} - Y_{i0}$. The main problem of IE is, it is not possible to observe at the same time Y_{i1} and Y_{i0} only one of them can be observed since units only have one existence: it is what is called the problem of counterfactual.

Talking about counterfactual, it is impossible to compare units before and after only because parameters can change with time, this has to be controlled. A group or a twin unit is needed which will be close or equal to what would have happened in absence of the treatment. Consequently, all IE methods aims at estimating or

approximating the counterfactual depending on the status of the unit considered. For example, for a treated unit i_0 in the population, IE methods will tend to find a twin in the population of non-treated units which is identical to i_0 in all characteristics except in the treatment status. Constructing the comparison group or building the counterfactual is done by IE method such as have Randomization, Difference in Difference (DID), Propensity Score Matching (PSM), Instrumental Variables (IV) and Regression Discontinuity Design (RDD). Assuming that the treatment was randomly assigned (Randomized experiments), the counterfactual will be a randomized group, with a sample big enough the bias of estimators of Δ will be very small. Using Matching, the counterfactual is a matched group, declared identical to treatment given some characteristics. For DID, it assumes constant change over time of control group and removes that effect before comparison. For IV regression, the causal effect is identified out of the exogenous variation of the instrument which is supposed to be those variables that affect participation in the program, but not outcomes conditional on participation. Finally, for the RDD, the counterfactual are individuals just below the cut-off who did not participate as the assignment to treatment is decided by the threshold of a given variable. For most of these methods, except the Matching, it is difficult to get a single control unit but they provide a control group from which they can compute the average effect of the treatment. In incapacity to get individual effects, they will assume that average effect is in fact the same across all units. That is one of the biggest weakness of current IE methods.

As a solution to overcome that drawback, our research proposes to address the problem of IE as a full missing data problem like some researchers ([1], [2], [3]) did. More than what they did, the study would like to solve it as a full missing data problem. In fact, given that counterfactual is an observation that is impossible to obtain in reality, it can be seen as missing value. So as a missing data problem, to study the missingness process which is quite close to the assignment of treatment process is the first goal. Then, from the missingness process the structure of

missing values is studied: what percentages, how many variables have missing data, which information do we have to impute and derive the appropriate method of imputation taking into account different specificities. As imputation methods, the most popular are Conditional mean imputation, K-nearest neighbour imputation, Fuzzy K-means imputation, Singular value decomposition imputation and Multiple Imputation. The study shall also try Regression modelling imputation especially quantile regression to preserve rank. In addition to those methods, some impact evaluation methods like Matching shall be used also, just as the one developed by [1], [2] and [3] in IE framework. The idea or the challenge is to come up with one imputation method that performs very well as imputation methods in IE framework then from which the average treatment effects can be computed to obtain better estimators of impact than the ones obtained with classical method. That new class of estimators are called Imputation Based Treatment Effect estimators (IB-TE). The main advantage will be to have individual effect which aggregation will lead to average treatment effects. Secondly, with imputation methods, it is possible to have the treatment effect on the population, on the treated units and also on the non-treated unit depending of the type of treatment assigned.

Among imputation methods used in IE framework, the most popular is regression imputation used by [1]. Propensity score matching is also one of the most popular methods used in IE framework and as imputation method. We can also highlight the Smooth Quantile Ratio developed by [3] and extended by [4].

The paper is outlined as follows: Section 2 develops the missingness process in general and then missingness in Impact evaluation setting, a focus will be done on the structure of data in practical IE setting. Section 3 discusses imputation methods in general, imputation method in IE framework and their specificities. Section 4 deals with simulation results in a context of hypothetical data base designed by us. Section 5 discusses an application on a real database drew from an implemented program in Kenya. Section 6 concludes and gives some remarks and

comments.

2   Missingness in Impact Evaluation Framework

This section develops the issue of missingness in IE framework. Recall that, the counterfactual is the side of the world that is not possible to observe. If a unit is treated, it is not possible to observe what would have happened at the same period if it was not treated. In a similar manner, if a unit is not treated, what would have happened if it was treated at the same period is not observed. Instead of looking for a twin or a control group (as counterfactual) as all of the IE methods do, the study considers counterfactual as a missing value. The side of the world that is not observed is a missing value the aim being to impute those missing values and rebuild completely the two parallel worlds.

Considering the problem of IE as a missing value problem implies that the missingness process should be described, then from that the paper shall propose the appropriate imputation method.

2.1   Missingness mechanism in general

First works on missing data were done by Rubin ([5], [6], [7]). Close to that, some researcher like [8], [9], [10]. Most of those works done starts by the missingness mechanism.

Prior to presentation of general imputation methods or how to handle missing data problems, it is good to know why these data are missing. We present different missing data mechanisms, meaning how in our data base missing values appeared? There are 4 main situations where data can be missing:

   ✓   Missingness completely at random (MCAR): the probability of missingness is the same for all unit in the sample. For a given variable X in the data base, the probability for an observation to be missing does not depend on X itself and on other variables of the same data base. The perfect example

will be if the choice is given to respondent to answer to a question or not given a random condition (rolling a dice for example). ([11], [12])

✓ Missingness at random (MAR): Most missingness is not completely at random, as can be seen from the data themselves. Probability can depend on an auxiliary variable in the same survey. Respondent can decide to answer or not, or interviewer may forget to ask some question to respondents. A more general assumption, missing at random, is that the probability a variable is missing depends only on available information. Thus, if for example sex, race, education, and age are recorded for all the people in the survey, then "earnings" is missing at random if the probability of nonresponse to this question depends only on these other, fully recorded variables. ([12], [13])

✓ Missingness that depends on unobserved predictors (NMAR): Missingness is no longer "at random" if it depends on information that has not been recorded and this information also predicts the missing values. There are some underlying unobserved factors that could lead people not to answer a given question and they can differ from a person to another one. Therefore, the probability of missingness is different across unit in our survey. An example is when during a survey a corrupted person is not going to declare his revenue because he knows that if he declares he can be exposed to pursuit so because of corruption (information not recorded) the data will be missing. ([13])

✓ Missingness that depends on the missing value itself: Finally, a particularly difficult situation arises when the probability of missingness depends on the (potentially missing) variable itself. This often happens because of the value itself like people are unlikely to reveal a high income to avoid to be exposed. ([13])

All these types of missingness can happen during survey and can be observed in data base depending on variables and the data collection process. To identify the

type of missingness, the final database ultimate user should be close to the data base constructor or be involved in data collection. The most frequent type of missingness mechanism is MAR. Practically, it's the one which can easily happen.

2.2   Missingness type in IE framework

In IE framework, considering that the missing value is supposed to be the counterfactual, the missingness will be linked to the treatment assignment process. For each type of treatment assignment process, the missingness mechanism will follow. In fact, being treated means that what would have happened in the absence of treatment is missing. In IE settings, we can distinguish three main assignment to treatment processes:

- ✓ **Randomization**: this means that the treatment 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. This process is usually observed in medical sciences when testing a new drug with a placebo. From this selection process, the missingness mechanism is **MCAR**;

- ✓ **Selection under fixed criteria**: 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. Therefore, assuming that there is no other unobservable characteristics that can influence the probability to be in the treatment group, the missingness mechanism is **MAR**. But if there are other factors that can modify the probability to be treated that are not recorded, the missingness mechanism is **NMAR**;

- ✓ **Selection given a single variable** (**RDD**): In this case, given a variable (revenue for example), the decision maker can fix threshold and say below the threshold you are treated and above you are not. Being in the treatment group depends on both observable and unobservable characteristics that determine your score on the selected threshold variable. In this case, the missingness is **NMAR**.

Except the randomization which is not common (only in medical studies), the missingness mechanism will be with a high probability a **NMAR** in the framework of impact evaluation. Let us therefore consider imputation methods according to the fact that missingness is at not at random at all but depends on some unobservable parameters or existing unquantifiable parameters that it is not possible to take into account in our analysis. Meanwhile, having many covariates or predictors in the dataset and including all of them can help to move from the NMAR hypothesis to the **MAR** hypothesis of missingness the aim being to control the missingness process or assignment to treatment process.

2.3 Different structure of data in IE

In practice, assignment process in a program is not clear especially for programs or projects ran in developing countries. In fact, given political issues, corruption and other factors, probabilities are not well mastered therefore the missingness cannot be MCAR but it is for sure MAR. In addition to the issue of assignment process, it is difficult to have data before and after the treatment administration. Usually data are just collected after the treatment because of budget issues, knowing who is treated and who is not.

For the structure of data, this study distinguishes two cases:

Case 1: Data are collected before and after

That case is usually the best situation that an impact evaluator can have because with data before and after, almost all IE methods can be applied in this situation. In the data base, there is a variable representing the treatment (T), the treated group ($T = 1$) and control group ($T = 0$). For each unit, there is a set of covariates (X) which can contain one or many other variables. We also have the potential outcome before (Y_b) and the potential outcome after (Y_a). It is true that even if we have data before and after, the set of covariates can be absent but usually it is available in that case. In summary, the data set is as follows: (Y_b, T, Y_a, X) assuming that the set of covariates does not change in time. With

change, the set looks like: $\left(Y_b, X_b, T, Y_a, X_a\right)$. Applying IE methods and Imputation methods is done without extra difficulties.

Case 2: Data are just collected after the treatment

This case is where data are only collected after the treatment assignment. The probabilities of treatment are not well mastered and there is an important data shortage. It can be difficult to obtain the variable treatment (T) because of lack of information about the background and the reasons for the program. This is usually the case in developing countries. Within this case we can have two sub cases. First one is when there are no covariates collected because of budget shortage therefore the data base will look like: (Y_a, T); Second one if there is the set of covariates, the dataset is: (Y_a, T, X). Applying IE methods here comes with additional difficulties that weaken the power of estimators of impact evaluation. Likewise, applying imputation methods will be more difficult and results can suffer from that.

Imputation methods and IE methods to be used will take into account the structure of the data at our disposition. Of course, within the first case many methods are possible and the results will be more accurate. Although within the second case, our possibilities are limited and the results will be less precise.

3   Customized Imputation Methods

Facing missing data issues, the default methods in most of software is ***Listwise deletion*** or ***Complete case analysis***. If a case has a missing value for one of the variables, simply exclude that case and run analysis. This method is excluded in this context, it is not possible to exclude cases because the classical methods of IE behave a bit like the latter and the consequences are probably a small sample for analysis, therefore a high variance for some estimators like mean. Another default method is called ***Available case analysis*** which recommend for a given variable to

delete all cases with missing values and run analysis. This method will lead to indicators constructed on different sample sizes or on different subset of the database so they cannot be inferred on the population in the same way. Even ***Nonresponse weighting*** is not suitable for our context given that our research would like to catch specificities of effect on different case in our sample. In our work, given our main problem, missing data solutions that retain all the data is preferred therefore Imputation methods.

This section first discusses the classical imputation methods used in the literature. They are classified into two main categories: Methods that don't incorporate random variation and Methods that incorporate random variation. After that, some IE methods, actually few IE methods using partially imputation to perform are presented. Thirdly, a choice is made, according to the problem of IE and the structure of data, on the type of methods suitable for IE imputation. Finally, a short presentation of imputation based treatment effect estimator is given.

3.1 Classical Imputation Methods

As we said earlier, our research focuses on methods that conserve all the cases in the dataset. Delete some case is not an option here. Classical imputation methods are divided into two main groups. Let's assume that our variable of interest with missing observation is Y and the set of covariates without any missing observation is X. To simplify notation, forget about the indexes specifying the case. A missing observation in the set is denoted by Y_m and a non-missing one by Y_{nm}. Of course, the corresponding covariates will be X_m and X_{nm} but it does not mean that they are missing.

3.1.1 Imputation methods that doesn't incorporate random variation

The main characteristic of these methods is that the missing value is replaced by a single estimator of the true value. They are deterministic methods meaning

that there is no randomness in the set of values used for imputation. Running the same method on the same sample will always produce the same imputed values for unit missing with the same characteristics.

Mean Imputation and Conditional Mean Imputation

This method can be applied on any type of dataset, with or without covariates. It recommends to replace the missing value by the mean of the missing variable obtained using the non-missing observations. The user can just replace the missing observations in Y by the marginal mean directly: $Y_m \Box E(Y_{nm})$ or knowing some properties of Y, conditional mean can also be used. The mean of Y given certain existing covariates X in our dataset: $Y_m = E(Y_{nm} / X = x_m)$. For example, if among our covariates, there is a variable sex and our variable of interest is determined by sex, we divide our sample into two groups: male and female, then perform mean imputation in each group. It is the most used method even if it leads to biased estimates and low variance and covariances (generally underestimate variances).

Nearest Neighbours Imputation

To apply this method, a database with a set of covariates is absolutely needed. The first step of this method is to define what is a neighbour using the set of covariates X. To define a neighbour, there is need to define a distance between case. The default distance is the Euclidian distance: $d_{ij}^2 = (X_i - X_j)'(X_i - X_j)$. We can also use the Mahalanobis distance by introducing a transfer matrix in the Euclidian distance. After defining a distance, the user can now decide for a given missing value which case is close to it or not. You can replace the missing observation by the value of the nearest neighbour or by a fixed k nearest neighbour (averaging) or use a value obtained by all the data set weighting each available case by the inverse of the distance between the missing case and all of them (weighting average). The simulation in this study used the Gower distance developed by [14] which aggregate all the distances between two points for each variable in one

single quantity. The distance was included in the package VIM on R by [15].

Last value carried forward

This method recommends to use the last value known about the variable for imputation. It means that if we have another survey, collecting the same information a time before the actual survey, from that survey you take information from the same variable and impute to the missing value in the actual data set. This method assumes the value doesn't change much with time. It can be true for some variables like sex but it is not always true. In our framework of IE, this method cannot be used because one of the main assumptions is that potential outcome changes with case and time no matter the time elapsed.

Regression to perform deterministic Imputation

The method is a model based method. It uses econometric (linear regression model or quantile regression for example) to build a model with available cases of Y and their corresponding covariates. The deterministic part of that model is used to predict the missing values given that all the value of covariates for each of them are known: $Y_m \square f(X_{nm})$. The main advantage of this method is the fact that it uses all information available on different units to predict the missing value. The disadvantages are it overestimates model fit and correlation estimates and weakens variance of the variable Y.

Simple random Imputation (Hot deck imputation)

This method recommends to randomly select a set of available cases among our non-missing observations and impute them to the missing observation. Or for each missing observation, randomly select another one among the set of observed data and impute: $Y_m \square Sample(1, Y_{nm})$. This method is quite simple and looks interesting but for some database and if you want to perform some specific studies, results can be very bad. It doesn't take into account of covariates if they are available, consequently you can have some atypical case for example a 12 year old child with a PhD as school level. This method is suitable if the population is stratified according to some determinant of our variable of interest.

3.1.2   Imputation methods that do incorporate 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 the methods presented here can be repeated then the final imputed value will be the average of the different output obtained during repetition.

Regression to perform random Imputation

This imputation method is almost the same as regression presented in the previous section. It also uses suitable econometric models to build a function of covariates that are going to be used to predict the value of the missing observation. The difference now is the error. After estimation of the coefficients of the regression, we obtain the deterministic part of the model and the error. Knowing the distribution of the error, this method recommends to generate for each predicted value an error and add to the deterministic part to obtain the final predicted value. The result is of the form: $Y_m \square f(X_{nm}) + \varepsilon_m$, with ε_m following a specific distribution determined by the econometric model. The main advantage here is the fact that the variance of the variable is preserved helped by the randomness of predicted values. The drawback is the same, estimation of coefficient comes with some bias because the coefficient that we are using in the model are not the true coefficients but just estimators which of course brings another bias.

Multiple Imputation (MI)

Among imputation methods, Multiple imputation is one the most interesting methods. The main objective of this method is to replace the full set of missing values by different sets of a possible candidate provided (each set) by different methods or by a single method allowing random variation. Multiple Imputation is a simulation procedure and the aim is not to obtain imputed values close enough to the real one but obtain acceptable estimators from the completed data bases ([16], [17]).

Multiple imputation involves three main steps:

a) **For each missing observation, generate** m **imputed values to obtain** m **completed sets of data**. After identifying which variable has missing values, the user should identify the missingness pattern and then decide which imputation methods to use keeping in mind that each should allow for randomness;

b) **Analyse the** m **set of completed data using standard procedures to produce estimators that we want**. In our case, each completed data set will produce a treatment effect given the missingness process and the imputation method chosen;

c) **All the estimators produced from each completed data set are combined to form a single set of final estimates of the parameters of interest.** In this step, the average can be used to obtain the final parameters with a standard deviation and confidence interval.

As advantage, this method can be used with any kind of data and model. It is simulation based therefore any user good in programming can perform it in any software. When data is MAR, Multiple Imputation can lead to consistent, asymptotically efficient, and asymptotically normal estimates. The main drawback is instability of the method. Because of randomness, different users can perform it and obtain totally different results. Even the same user, every time you run the program, you obtain different results hopefully slightly different. In the simulations, the MI method used generates Multivariate Imputations by Chained Equations (MICE). In the MICE procedure, a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data. This means that each variable can be modeled according to its distribution, with, for example, binary variables modeled using logistic regression and continuous variables modeled using linear regression.

Maximum likelihood Imputation (ML)

This method is used to obtain the variance-covariance matrix for the variable in

the model based on all the available data points, and then use the obtained variance-covariance matrix to estimate the regression model ([16], [18]). This method is quite simple if you use an appropriate software, you only need to specify your model of interest and indicate that you want to use ML. Theoretically, the basic idea is the following. Given a set of data with n independent observations and $k+1$ variables $(y_i, x_{i1}, ..., x_{ki})$ and assuming that there is no missing data in that set, the likelihood function is given by:

$$L = \prod_{i=1}^{n} f_i(y_i, x_{i1}, ..., x_{ki}; \theta)$$

Where $f_i(.)$ is the joint probability function of i observations and θ the set of parameters to be estimated. The ML estimates are the values of θ that maximise L. Now, in the specific case of this research, suppose that for some observations i, the first variable Y has missing data that satisfies MAR assumption of missingness. Now the joint probability of the observed data is given by:

$$f_i^*(x_{i1}, ..., x_{ki}; \theta) = \int_y f_i(y_i, x_{i1}, ..., x_{ki}; \theta) dy$$

For each observation's contribution to the likelihood function, we integrate over the variables that have missing data, obtaining the marginal distribution of observing those variables that have actually been observed.

Considering that there are m missing observations in the first variable over n, ordered such that the first $n-m$ lines are completed and the last m have missing data, the likelihood function of the full data set becomes

$$L = \prod_{i=1}^{n-m} f_i(y_i, x_{i1}, ..., x_{ki}; \theta) \prod_{i=n-m+1}^{n} f_i^*(x_{i1}, ..., x_{ki}; \theta)$$

This likelihood function can then be maximized to get ML estimates of θ using several different methods.

There are two main ML methods:

a) **Direct Maximum Likelihood**: implies direct maximization of the multivariate normal likelihood function for the assumed linear model.

b) **The expectation – Maximization (EM) algorithm**: provides estimates of the mean and covariance matrix which can be used to get consistent estimates of the parameters of interest.

For the simulation, the R package MissMech is chosen. Two options are used to perform ML: firstly the program assumes that data follow a multivariate normal distribution then secondly no assumption is made on the distribution but a maximization algorithm is used to obtain the covariance matrix.

3.2 Imputation Based IE Methods

Imputation based on IE methods are developed initially to obtain the impact of a treatment on a given population. Some of them were designed to give an average impact but here they are going to be used as imputation method to obtain cases effects of a treatment. The main hypothesis here is unconfoundedness Unconfoundedness, meaning that given covariates taking the treatment or not is independent of the potential outcome.

Regression imputation seen by [1]

Given a data set (Y, T, X) of data, following the work done by [1], under assumption of unconfoundedness, Hahn defines:

$$E(T_i Y_i / X_i) = E(T_i Y_{i1} / X_i) = E(T_i / X_i) E(Y_{i1} / X_i) = E(T_i / X_i) \beta_1(X_i)$$

and at the same time:

$$E((1 - T_i) Y_i / X_i) = E((1 - T_i) Y_{i0} / X_i) = E((1 - T_i) / X_i) E(Y_{i0} / X_i) = E((1 - T_i) / X_i) \beta_0(X_i)$$

From these two equations, it follows that:

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

The quantity $\hat{\beta}_1(X_i)$ is estimation 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 estimation of the value of potential outcome that unit i would have

taken if it was not treated, in absence of treatment on him (in this case unit i is treated). Therefore, under treatment for any unit treated in the population: $\hat{Y}_{i1} = T_i Y_i + (1-T_i)\hat{\beta}_1(X_i)$ 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 is OLS regression, Non-parametric regression, or even simple sample mean or any other method link to regression methods. At the end of imputation, a completed data set is obtained from which estimations can be done. Hahn proposed a nonparametric method for imputation. In this research, a parametric imputation (OLS regression by quantile if possible) 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.

Propensity score matching Imputation

The matching imputation is based on the calculation of two propensity score function. The first one is computed in the control group $\hat{p}_{i0}$ and the second one $\hat{p}_{i1}$ in the treatment group ([19]). 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. Among the different types of matching, there is one-to-one matching, 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:

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

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 unit 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 imputation done. It is a kind of iterative imputation method.

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 [3]. The method was used to proposed an estimator of the mean difference between two highly skewed distributions. We are going to use it as a quantile imputation method for estimating the distributional impact of a treatment.

Considering the general form developed by [4], they define:

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

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:

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

By doing that, the method replaces or completes the sample by smoothed quantile estimation of missing values. Therefore, from that sample one can compute whatever estimator we want including QTE estimators. Unfortunately, this method doesn't suit the objective of imputation here which is point imputation not quantile imputation.

3.3 Quantile Regression Imputation

This research proposes to use quantile regression as imputation method to achieve the objectives. In fact, neither in the literature on imputation nor in the IE literature, no one has used literally quantile regression as imputation method. The

advantage that IE data can provide is having data before and after the treatment. Assumption that is made here is the following: The effect of the treatment is monotonous. This assumption means that the treatment preserves the rank of the potential outcome after the treatment (in this case not the rank but the quantile). The treatment preserves the quantile of cases after the treatment. It is weaker than the rank preservation or the constancy in the impact of the treatment that many IE methods uses.

Given a potential outcome Y, the cumulative distribution function is defined by $F_Y(y) = P(Y \leq y)$. As reminder, the order τ quantile is given by $q_\tau(Y) = \inf\{y : F_Y(y) \geq \tau\}$. If F_Y is continuous, the classical definition is valid: $\tau = P(Y < q_\tau(Y))$. Basically, quantile regression tries to evaluate how conditional quantiles defined by $q_\tau(Y|X) = \inf\{y : F_{Y|X}(y) \geq \tau\}$ change with change of the covariates X considered as determinant of the potential outcome. In fact, there is no reason to consider that the effect of X is the same over the distribution of the potential outcome.

In the standard quantile regression, the assumption made is that the conditional distribution is a linear function of covariates meaning: $q_\tau(Y|X) = X'\beta_\tau$, for every τ we have a vector of coefficients $\beta_\tau = (\beta_{1\tau}, \beta_{2\tau}, ..., \beta_{p\tau})$ corresponding to a set of covariates with p independent variables. To use the classic linear regression, the quantile regression can be written as following:

$$Y = X'\beta_\tau + \varepsilon_\tau, \quad with \quad q_\tau(\varepsilon_\tau|X) = 0$$

Here, the coefficients of the regression are allowed to change from a given quantile to another one. From that basic model of quantile regression, different variants can be defined:

- **Simple shift quantile regression model** when independent variables do influence only the mean of the potential outcome (not on the Variance for example). In this case, conditional distributions ($F_{Y|X=x}$) are parallel when

x changes. So, $q_\tau(Y|X) = X'\gamma + q_\tau(\varepsilon)$. The coefficients are the following

$\beta_{1,\tau} = \gamma_1 + q_\tau(\varepsilon)$ for the constant and for the rest of variables, $\beta_{k,\tau} = \gamma_k$ for

$k > 1$. There is a homogeneity on the slope for each quantile as the model

looks like a homoscedastic model;

- **Scale quantile regression model** where by independent variables do influence the mean and the variance of the potential outcome. It is a kind of heteroscedastic model where $q_\tau(Y|X) = X'(\gamma + q_\tau(\varepsilon)\theta)$. In this case, the coefficients are $\beta_\tau = \gamma + q_\tau(\varepsilon)\theta$. Impact of covariates is different for different quantiles therefore there is heterogeneity in slopes induced by the parameter θ.

In general, to have a regression model where the coefficients are linked to the covariates and the changes in the dependent variable; it is recommended to use what is called Random coefficient models. Quantile regression is one of them. Those models allow the regression coefficient to change according to changes in the dependent variable and the covariates. The general form is:

$Y = X'\beta_U$, U independent to X and following a uniform distribution on $[0,1]$

The function $u \mapsto x'\beta_u$ is strictly increasing for every x. In that model, U can be seen as unobserved characteristic which affects the position of the unit in the distribution of the potential outcome Y. This model generalised the two models presented early: the Shift model if for $k > 1$ the coefficient does not depend on U meaning $\beta_{k,U} = \beta_k$ and the Scale model if the coefficient is defined as

$\beta_U = \gamma + q_U(\varepsilon)\theta$.

In the case where U is defined by the quantile on Y (independence of U and X is verified, and the increasing property also), the coefficient β_τ is interpreted as follows: a marginal change of covariates X no matter the quantile (independently of U) is reflected on Y by the coefficient β_τ. Therefore, β_τ is the marginal effect of X for units in the τth quantile of the distribution of

unobserved characteristics U .

In this research, the model built using quantile regression or more generally random coefficient models will be used for imputation. In fact, having information before and after the treatment allows us to have the initial quantile of all units. Then, making the assumption that the treatment is monotonous (Monotonicity assumption) helps us to identify which model will be used to impute which missing value after the treatment is assigned to all units.

3.4 Imputation Based Treatment Effect (IB-TE) Estimator

The main contribution of this paper is the development of a new approach to solve the problem of impact evaluation. This new approach combines imputation methods and basic definition of impact evaluation to construct a new class of estimators. As described earlier, the aim of imputation is to estimate counterfactual in a program. From the basic definition of treatment given by [5], $\Delta_i = Y_{i1} - Y_{i0}$, one of the quantities is unknown. If the unit i is a treated unit, Y_{i0} is not observed therefore it is a missing value and has to be imputed using information on characteristics non-treated units of the population. Similarly, if unit i is a non-treated unit, Y_{i1} is not observed and is considered as a missing value that has to be imputed using information on treated unit characteristics of the population. Applying imputation method leads to an estimation of the counterfactual Y_{i0} or Y_{i1} depending on the status of the unit. Let's consider $\tilde{Y}_{i1}$ and $\tilde{Y}_{i0}$ as imputed values of counterfactual. An estimation of the effect defined by Rubin is given by $\hat{\Delta}_i = Y_{i1} - \tilde{Y}_{i0}$ if the unit is treated and $\hat{\Delta}_i = \tilde{Y}_{i1} - Y_{i0}$ if the unit is not treated. From that single unit estimator, the following quantities can be defined:

$$IB\_ATE = E\left(\hat{\Delta}_i\right) = E\left(Y_{i1} - \tilde{Y}_{i0} \mid T = 1\right) + E\left(\tilde{Y}_{i1} - Y_{i0} \mid T = 0\right)$$

$$IB\_ATT = E\left(\hat{\Delta}_i \mid T = 1\right) = E\left(Y_{i1} - \tilde{Y}_{i0} \mid T = 1\right)$$

$$IB\_ATNT = E\left(\hat{\Delta}_i \mid T = 0\right) = E\left(\tilde{Y}_{i1} - Y_{i0} \mid T = 0\right)$$

They are respectively the Imputation Based Average Treatment Effect in the population (IB-ATE), on Treated units (IB-ATT) and on Non-Treated units (IB-ATNT). Each imputation method produces its own IB-TE estimators. As estimators, the properties are studied empirically in the next section. Simulation and Bootstrap are used to study asymptotic bias, asymptotic convergence and average bias is used to compare them together and with classic IE estimators.

4   Simulation and Summary of Results

The first objective of this section is to use simulation to test our hypothesis that imputation methods can lead to better estimators of average impact than IE estimators or at least as good as existing ones. The second objective is to come up with the best imputations method that best complete the data set, those are going to be used in the next step of our research which is to obtain the distributional effects of a program through quantile treatment effect. In this section, description of simulation procedure and parameters is done, then simulation are performed under Random Assignment (MCAR missingness) hypothesis and under Deterministic Assignment (MAR missingness) hypothesis.

4.1 Description of Simulation Procedure

The aim of this simulation is to come up with the best imputation methods, two or three, suitable for impact evaluation framework. Best in terms of imputation of course but also in terms of performance of imputation based estimators of average effect produced by the given imputation methods.

The simulation recreates a hypothetical situation where a treatment (project or

program) has to be assigned in a population with all the parameters being known. For example, assignment process is well known, the potential outcome is known, decision to treat everyone or not to treat everyone can be taken so that computation of the true impact of the project 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 (impact evaluation method), simulation results will tell which imputation method is suitable and why. Imputation methods will be judged at two levels: first the capacity to complete faithfully the initial dataset that can be used to better estimate the true impact of the treatment (compare the impact got from completed data set with the one got from uncompleted data set and the true value simulated) and second, the capacity to reconstruct the exact value of the missing observation (RMSE indicators are used). After imputation, Imputation Based average treatment effects are produced and compare to existing ones in IE framework.

In the simulation process, a data base of 10,000 cases is generated. The potential outcome (Y_b) and the covariates before the treatment assignment or before the program are generated. Since we are in the simulation, a situation where by all units are not treated (Y_{2NT}) and a situation where all units are treated (Y_{2T}) is simulated at the same time. From that, the true average impact of the treatment is computed in the overall population as follows $ATE_{true} = mean(Y_{2T} - Y_{2NT})$.

The next step now is to create the treatment variable (T), by deciding which case is treated and which case is not treated according to the assignment process decided. In this study, two cases are decided: Random assignment leading to MCAR missingness mechanism and Controlled assignment leading to MAR missingness mechanism. If $T = 1$ the case is treated and if $T = 0$ the case is not treated. From this stage of simulation, the true average impact of treatment in the population can be computed by: $ATT_{true} = mean(Y_{2T} - Y_{2NT} | T = 1)$.

The potential outcome in the real world is now generated in the variable Y_a as follows:

- For non-treated ($T = 0$), $Y_a = Y_{2NT}$, the value of Y_{2NT} is just reported when $T = 0$;

- For treated ($T = 1$), $Y_a = Y_{2T}$, the value of Y_{2T} is just reported when $T = 1$.

Using Y_a the potential outcome with missing values are generated (Y_T and Y_{NT}), if you are treated, what would have happened if you were not treated is a missing value and also if you are not treated, what would have happened if you were treated is consider as a missing value.

- ✓ For treatment case, create Y_T as follows: report all observations of potential outcome for treated and for non-treated consider as missing values: $Y_T = Y_a$ if the unit is treated and $Y_T = .$ (*miss*) if the unit is not treated and has to be imputed;

- ✓ For non-treatment case, create Y_{NT} as follows: report all observations of potential outcome for nontreated and for treated consider as missing values: $Y_{NT} = Y_a$ if the unit is not treated and $Y_{NT} = .$ (*miss*) if the unit is treated and has to be imputed.

Knowing that in our simulation we hypothetically have all of them, the aim is to estimate them using imputation methods and on the way, compute the associated imputation based estimators. Those estimators will be compared to the true values and the IE estimators produced by classical methods like Randomization, Difference in Difference, Matching, Instrumental variable regression and Regression Discontinuity Design.

The final data base after simulation of population looks like:

MATRIX OF DATABASE AFTER SIMULATION OF POPULATION

Case N°	Bef treatment		Hypothetical			Aft Treatment		Pot 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									

Y denote the potential outcome;

X a set of covariates collected for identification of each case;

T is the treatment indicator for treated (1) and non-treated (0);

b index means before the treatment;

a index means after 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;

With $Y_{a,i} = Y_{T,i}$ and $Y_{a,j} = Y_{NT,j}$.

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.

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 Imputation Based Average Treatment Effect estimators (IB-ATE), the average bias is computed ($AvrgBias = E(\hat{\theta} - \theta)$) and compared to the one for existing IE methods. Of course, this is done under a bootstrap procedure of 1000 replications.

4.2 Random Assignment Hypothesis Results (MCAR)[4]

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 is fixed to be 40% of the total population (binomial distribution of parameter 1 and 0.4 for the population) and for the simulation we made sure that each sample draw 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.

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 no matter the size of the sample was DID. For example, the average bias for N=50 was 0.5 for DID, -5.8 for PSM and finally -7.1 for RA. Also for N=1500, the average bias was close but the DID bias was still the smallest. At the same time, the standard deviation was always small for the estimators of DID. In consequence, the DID was the best one among the IE methods (see Appendix 2 for more details).

When we look at the results of imputation methods, we found that three imputations methods were indisputably the best for all sample sized. The best one was Quantile regression (QR Imp), the second best was Deterministic Linear Regression (Det LM) and the third one was Multiple Imputation (MI). They

---

[4] See Appendix 2 for an example of bootstrap simulations more detailed on this section

always recorded the smallest RMSE, and that RMSE was decreasing with sample size. Looking at the performance of imputation methods related to impact evaluation estimators, estimators of all of them were asymptotically unbiased except for the PSM as imputation method and the QR Imp. For those two methods, the bias was increasing for large sample size especially for ATT. As example, for N=800, the bias was -148 for PSM ATT and -33 for QR Imp ATT.

For the convergent estimators, the bias was decreasing and tend to 0 as the sample size was increasing. According to the sample size, the best Imputation Based Treatment Effects Estimators (IB-TE) were changing. For N=50, the best was QR for ATE and Deterministic Linear regression for ATT. For N=100, the best one was random linear regression imputation for ATE and k-NN for ATT. For large samples, N=1500, the best one was 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 was the Maximum Likelihood (ML) and Deterministic Linear Regression model (Det ML), methods which was always among the three best IB-TE when changing the sample size. For the case of mean imputation, IB-TE and IE estimators are the same theoretically and empirically. Increasing the sample size (N=2000 see appendix) leads to very good estimators as good as DID estimators.

Table 1: Average bias of treatment effect estimators given the sample size (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 Gen	-7.1	-7.4	-2.6	-2.4	1.508	1.464	0.665	0.662	-0.2277	-0.3652	**0.081**	**-0.0163**
	Mean 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**
	Rand Imp 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
	k-NN 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
	ML Imp 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

Source: Our simulation with R

We now want to compare the IE estimators with IB-TE estimators. Since all of them are asymptotically convergent we compared them using the speed of convergence and the standard deviation in certain cases. For N=50, the DID was better than all IB-TE estimators, average bias of 0.5 for ATE and 0.2 for ATT far smaller than average bias of others IB estimators. As the sample size increased, the average bias of IB-TE 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=200, the average bias of DID was the smallest for ATE and ATT but quite close to the best IB-TE estimators. When increasing again the sample size, from 200 to 500, the average bias was respectively 0.069 and 0.066 for ATE and ATT for the DID method and 0.057 for ATE Random imputation and 0.082 for ML ATT imputation. Imputation performs better than DID for ATE but quite close for ATT. 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 always an IB-TE estimator performing better than DID estimators or as well as.

To summarize, it is true that the DID methods gave better results on average for all sample given the standard deviation and compared to a single method of IB-TE but at the same time for some IB-TE 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 used need information before the program which is not always available. In that case, DID is not applicable and the second-best IE methods is PSM which is less better than IE method especially ML imputation of k-NN no matter the sample size.

4.3 Assignment controlled by Variables (NMAR or MAR)[5]

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 at the same time the IV regression and the sharp RDD. From that subpopulation around the threshold, we have drawn our sample 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 best imputation method.

When looking at the results of the simulations recorded in table 2, all the average bias of IE methods was decreasing meaning that the estimators are asymptotically convergent or unbiased. When looking at the variance of each estimator, it was also decreasing to zero. DID method is almost the best method in general no matter the sample size. Out of the six sample size presented here, DID

---

[5] See Appendix 3 for an example of bootstrap simulations more detailed on this section

was the best 5 times and the second best, for N=100 where the smallest average bias for RA was respectively -0.17 and -0.43 for ATE and ATT and for DID the average bias was -0.48 and 0.74. For the others sample size, the average bias of the DID method was the smallest and decreasing. For example, N=200, the average bias of DID estimators were respectively -0.74 and -1.01 for ATE and ATT, the second-best method was PSM estimators with -1.91 and -1.55 respectively for ATE and ATT. Indubitably, DID estimators were the best again despite IV estimators and RDD estimators.

Table 2: Average bias of treatment effect estimators given the sample size (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**	-1.50	-1.76	-1.22	-1.48
	PSM		-1.54	-1.82	-3.42	-2.15	-1.91	-1.55	-1.680	-0.85	-2.38	-1.87	-2.10	-2.20
	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		-8.40	-8.66	-5.83	-6.09	-5.43	-5.69	-5.200	-5.46	-5.18	-5.45	-5.12	-5.38
	RDD		-4.04	-4.30	-2.10	-2.37	-2.23	-2.49	-0.829	-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	**-1.50**	-1.76	-1.22	-1.48
		Condi	-4.29	**-3.73**	**-0.12**	**-0.50**	-2.54	-3.10	0.740	**0.407**	**-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	**-3.40**	**-3.48**	-4.01	-1.50	-4.00	-3.96	-1.37	-1.14	-3.81	-3.92	-3.03	-2.88
		V2	**-3.49**	-3.90	-3.69	-0.85	-4.07	-4.00	-1.19	-1.06	-3.50	-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	-3.20	-1.32	**-1.20**
	MI	Mice	-4.42	-3.82	-2.61	-2.42	-2.46	-3.10	-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	-1.80	-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	-34.00	5.99	-34.00	4.88	-35.02	5.82	-34.10

Source: Our simulation with R

Looking at the imputation methods, again here the best imputation methods were the same regarding ability to reconstruct the initial data. The methods that recorded the smallest RMSE was Quantile regression (QR Imp), the second best was Deterministic Linear Regression (Det LM) and the third one was Multiple Imputation (MI). If the imputation methods are assessed on their ability to

estimate the treatment effect, given the different sample sizes, all the average bias was decreasing 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 2, 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 others methods, like MI also performed well but not as well as Mean and Random imputation. For example, N=500, the average bias for MI estimators was respectively -1.35 and -0.89 for ATE and ATT while for the best IB-TE estimators it was -0.005 and 0.026. That bias is acceptable since MI is among one of the best imputation methods. Imputation Methods were able to produce acceptable average treatment effect estimators no matter the sample size.

Comparing classical IE estimators with IB-TE 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-TE was DID especially for large sample and because of its smallest variance. For small samples (N=50 and 100), DID was not the best QR Imp and Conditional mean imputation were the best, smallest average bias but with a bigger variance than DID. For N=500 and 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 and the smallest variance even if the others IB-TE estimators were close in terms of average bias and variance. The others IE methods did not perform well compared to DID and IB-TE estimators, especially IV and RDD which were supposed to produce better estimator given the assignment process. For IV, the average bias was decreasing then became constant around 5 when the sample size increased. For RDD, the estimator was asymptotically convergent and perform as well as some IB-TE estimators but still less than the best three.

In summary, except DID which gives asymptotically unbiased estimators with smallest variance, one can always find an imputation method that gives small average bias and small variance than other IE methods. This is to say IB-TE 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-TE is the best solution if the assignment process is not random. PSM, IV and RDD produce convergent estimators but not as good as IB-TE estimators.

4.4 Advantages of IB-TE and Discussions

The first advantage of having IB-TE estimators is related to availability of data. By using bootstrap, we obtained estimators (IB-TE) as good as the one obtained with IE methods. Some of those estimators used only the potential outcome of the treatment after the treatment to produce good estimators (Random imputation estimators). While DID 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-TE estimators are the best ones to use. Another advantage of using IB-TE estimators is the fact that from them we can have all types of treatment effects. Imputation gives the possibility to produce ATE, ATT and ATNT which are average treatment effect on every units, on treated units and also on non-treated units. In addition to that, with IB-TE it is easily possible to obtain case 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 what would have been the impact of the drug the other way round after experience. Instead of starting experience 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 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 different results absolutely but the results can lean on the simulation parameters. By changing the parameters, the results can be in favour of IB-TE or IE methods. For the case of distribution of potential outcome, it does not matter a lot but assumptions made before and after and for the treatment can influence the results. The share of the treated unit 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%. Simulation of covariates is tricky because with simulation anything can happen. Assuming that the program is simulating age and education level, if the user is not careful, simulation can produce a teenager with a PhD level which is quite rare even impossible. In addition, the instrument which is also used as for the threshold is questionable. For all these reasons, IB-TE estimators are tested on a real program in the next section to see how they perform for a real program.

5   Applications

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

Firstly, the Lalonde ([20]) data set is considered for application. Lalonde data ([20]) set contain the treated and control units from 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 used control groups, selected using preintervention variables to compare and obtain the treatment effect on treated.

Based on pre-intervention variables, [21] 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 come up with an average treatment effect on treated of $1794. Later, they used the propensity score method and hey come up with a treatment effect on treated range of $1473 to $1774, quite close to the result of Lalonde with the same dataset according to them.

As reminder, IB-TE estimators combines imputation methods and bootstrap to obtain treatment effects estimators. In this case, we applied our methods on the subset of [20] draw by [21].

Table 3: IB-TE estimators using Lalonde 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	Rand		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	Mean	Gen	1794	767	1794	167	1785,3	252	1785,3	252
	Rand Imp	General	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	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

Source: Our computation in R

For a fixed sample size of n=200 draw from a population of 445 (respecting the share of 41% of treated) and after 1000 replications, results recorded in table 3 are interesting. The bootstrap of mean imputation method led to exactly the same treatment effect on treated as the results of [20] 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 [21] in their work (a range from $1473 to $1774). Also looking at the standard errors, they are smaller than the ones obtain by Lalonde and Dehejia et al, 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 more 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 and as good results as the one obtains by Lalonde. 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.

6   Conclusion

The problem of impact evaluation has always been counterfactual: what would have happened in the absence of the treatment if a unit is assigned to the treatment? and what would have happened in the presence of the treatment if a unit is not assigned to the treatment? So far, the classical impact evaluation methods have been addressing the problem by using a control group which is supposed to be identical to the treatment group in all aspects except the treatment status and the potential outcome. The first consequence of that is the sensitivity of the results to

the quality of the control group. Added to that there is the impossibility to have individual effects since most of those methods give the average treatment effects.

This study, trying to overcome those weaknesses, has proposed a new class of estimators of treatment effects called Imputation Based Treatment Effects estimators (IB-TE). To achieve that objective, we addressed the problem of impact evaluation as a missing data problem. The study considers counterfactual as a missing value then combining imputation methods and bootstrap, it came up with an estimator of treatment effect as good as existing one. We have tested a small number of imputation methods but there are many others and the process of producing this class of estimator is the same no matter the type of imputation method that you chose. The IB-TE estimators were applied under simulations and on real database. We found that results were as good as the classical estimators in some case and in others even better than classical IE estimators. For the application case, we found exactly the same result as Lalonde (1986) and comparing the results with Dehejia et al (1999), IB-TE estimators were better.

The new estimator came with some advantages as 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 a possibility to improve the quality of these estimators by improving the imputation method. Many others imputation methods have to be tested in accordance to structure of data and availability of data. In case of a not well design program or in case of shortage of data, we advise program evaluator to use IB-TE estimators to evaluate treatments.

ACKNOWLEDGEMENTS.

References

[1] J. Hahn, On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects, *Econometrica*, **66**, (1998), 315-331.

[2] G. V. Montes-Rojas, Nonparametric estimation of ATE and QTE: An application of fractile graphical analysis. *Journal of Probability and Statistics*, (2011).

[3] F. Dominici, L. Cope, D. Q. Naiman, and S. L. Zeger, Smooth quantile ratio estimation, *Biometrika*, **92(3)**, (2005), 543–557.

[4] S. Venturini, F. Dominici and G. Parmigiani, Generalized quantile treatment effect: A flexible bayesian approach using quantile ratio smoothing, *Bayesian Analysis*, **10(3)**, (2015), 523–552.

[5] D. Rubin, Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies, *Journal of Educational Psychology*, **66**, (1974), 688-701.

[6] D. B. Rubin, Multiple Imputation for Nonresponse in Surveys, *New York, Chichester*, (1987).

[7] D. B. Rubin, Multiple Imputation after 18+ Years, *Journal of the American Statistical Association*, **91, 434**, (1996), 473-489.

[8] A.A. Afifi, and R.M. Elashoff, Missing observations in multivariate statistics I. Review of the literature, *Amer. Statistical Association* **61**, (1966), 595-604.

[9] H.O. Hartley and R.R. Hocking, The analysis of incomplete data, *Biometrics*, **27**, (December 1971): 783-823.

[10] T. Orchard and M.A. Woodbury, A missing information principle: theory and applications, *Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability*, Theory of Statistics, *Univ. of California Press*, (1972).

[11] A. Briggs, T. Clark, J. Wolstenholme and P. Clarke, (2003), Missing.... presumed at random: cost-analysis of incomplete data, *Health Econ.*, **12**, (2003), 377–392.

[12] P.D. Allison, Missing Data, *Sage University Papers Series on Quantitative Applications in the Social Sciences*, **series no. 07-136**, (2001), Thousand Oaks.

[13] A. Gelman and J. Hill, Missing-data imputation. In Data Analysis Using Regression and Multilevel/Hierarchical Models, *Analytical Methods for Social Research*, pp. 529-544, (2006), Cambridge: Cambridge University Press.

[14] J. C. Gower (1971) A general coefficient of similarity and some of its properties, Biometrics **27**, (1971), 857–874.

[15] A. Kowarik and M. Templ, Imputation with R package VIM, *Journal of Statistical Software*, *74(7)*, (2016), 1-16.

[16] J. L. Schafer, Analysis of incomplete multivariate data, *London: Chapman & Hall/CRC*, (1997).

[17] P. D. Allison, Multiple Imputation for Missing Data: A Cautionary Tale, *Sociological Methods & Research*, **28(3)**, (2000), 301–309.

[18] P. D. Allison, Handling Missing Data by Maximum Likelihood, *SAS Global Forum* (2012) *Statistics and Data Analysis*,1–21.

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

[20] R. Lalonde, Evaluating the Econometric Evaluations of Training Programs, *American Economic Review*, **76**, (1986), 604-620.

[21] R. Dehejia and S. Wahba, Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs*, Journal of the American Statistical Association*, **94(448)**, (1999), 1053-1062.

Appendices

Appendix 1 : R Code to generate the population for simulation

RANDOM ASSIGNMENT

```
rm(list=ls())
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)
##MCAR Process of missingness
T=rbinom(N,1,0.4)
###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
ASSIGMENT DEPENDING ON A SINGLE VARIABLE
```

```
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)
```

Appendix 2: Example of bootstrap simulation for N=800 in case of Random Assignment

Parameters	N (Sample size)	800
	P (Share of treated)	0,4
True values of Impact	ATE	200,0985
	ATT	200,236

IE Methods Results		Mean (RA)		Matching		Diff in Diff		IV Regression		RDD method	
		ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
	IE Methods Results (avrg)	199,8708	199,8708	199,2383	199,6444	200,33	200,33				
	SD	22,5	22,5	22,9	24,4	4,5	4,5				
	B Inf	198,5	198,5	197,8	198,1	200,1	200,1				
	B Sup	201,3	201,3	200,7	201,2	200,6	200,6				

	Imputation Methods		Imp Based Estim								RMSE	
			ATE				ATT				Treated	NonTreated
			Avrg Val	SD	B Inf	B Sup	Avrg Val	SD	B Inf	B Sup		
Imputation Based results	Mean Imputation (1)	Gen Mean	199,8708	22,5	198,5	201,3	199,8708	22,5	198,5	201,3	240,7	192,6
		Conditional	199,9221	22,5	198,5	201,3	199,9347	22,5	198,5	201,3	241,1	192,8
	Random Imputation (2)	General	199,6601	24,5	198,1	201,2	199,8057	28,1	198,1	201,6	339,2	271,7
		Hot deck	199,5951	27,4	197,9	201,3	200,0526	30,7	198,2	202	340,5	271,4
	Deterministic linear regression Imp (3)		200,0818	18,9	198,9	201,3	200,0509	18,8	198,9	201,2	202,9	160,5
	k-NN imputation (4)	V1	200,1565	23,6	198,7	201,6	200,6325	25,7	199	202,2	285,9	225,3
		V2	200,1424	23,6	198,7	201,6	200,7847	26,7	199,1	202,4	285,9	226,2
	Random linear regression Imp (5)		199,8343	21,2	198,5	201,1	199,758	23,6	198,3	201,2	285,5	226,9
	Multiple imputation (6)	Mice	200,0018	19,8	198,8	201,2	200,1799	21	198,9	201,5	221,9	175,6
	Maximu Likelihood Imp (7) (MissMech package)	Normal	199,9148	19,4	198,7	201,1	199,7396	20,2	198,5	201	378,5	300,3
		Dist-free	200,2224	19,8	199	201,5	199,9051	20,7	198,6	201,2	378,6	300,2
	Propensity score matching Imp (8)		51,9894	43,4	49,3	54,7	52,2697	47,8	49,3	55,2	438,23	466,64
	Quantile Regression Imp (10)		207,0142	15,2	206,1	208	167,6476	14,6	166,7	168,6	139,5	105,3

Appendix 3: Example of bootstrap simulation for N=1000 in case of controlled assignment

Parameters	N (Sample size)	**1000**
	P (Share of treated)	**0,4**

		Value
True values of Impact	ATE	**198,334**
	ATT	**198,5964**

		Mean (RA)		Matching		Diff in Diff		IV Regression		RDD method	
		ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT	ATE	ATT
IE Methods Results	IE Methods Results (avrg)	197,1118	197,1118	196,2389	196,4002	**198,104**	**198,104**	193,2146	193,2146	196,4159	196,4159
	SD	19	19	19,1	21,9	**3,8**	**3,8**	18,7	18,7	16,1	16,1
	B Inf	195,9	195,9	195,1	195	**197,9**	**197,9**	192,1	192,1	195,4	195,4
	B Sup	198,3	198,3	197,4	197,8	**198,3**	**198,3**	194,4	194,4	197,4	197,4

		Imp Based Estim								RMSE	
	Imputation Methods	ATE				ATT				Treated	NonTreated
		Avrg Val	SD	B Inf	B Sup	Avrg Val	SD	B Inf	B Sup		
Imputation Based results	Mean Imputation (1) — Gen Mean	197,1118	19	195,9	198,3	197,1118	19	195,9	198,3	241,7	194,7
	Mean Imputation (1) — Conditional	**197,211**	19	196	198,4	197,1434	19	196	198,3	242,2	194,8
	Random Imputation (2) — General	**197,8251**	21,6	196,5	199,2	**197,6858**	24,8	196,2	199,2	342,9	275,2
	Random Imputation (2) — Hot deck	**197,2592**	24,2	195,8	198,8	**197,2424**	28,4	195,5	199	342,6	275,6
	Deterministic linear regression Imp (3)	196,4246	16,1	195,4	197,4	196,3827	16,1	195,4	197,4	**203,3**	**163,1**
	k-NN imputation (4) — V1	195,3001	19,6	194,1	196,5	195,7154	22,5	194,3	197,1	288,4	229,3
	k-NN imputation (4) — V2	195,5732	19,7	194,3	196,8	196,3547	23,5	194,9	197,8	288,4	230,3
	Random linear regression Imp (5)	197,0166	18,3	195,9	198,2	**197,4009**	20,6	196,1	198,7	288,3	229,4
	Multiple imputation (6) — Mice	196,091	16,8	195,1	197,1	196,2681	17,6	195,2	197,4	**222,9**	**178,4**
	Maximu Likelihood Imp (7) (MissMech package) — Normal	196,5308	16,7	195,5	197,6	196,4096	17,6	195,3	197,5	381,7	304,6
	Maximu Likelihood Imp (7) (MissMech package) — Dist-free	196,8137	16,6	195,8	197,8	196,8617	17,5	195,8	197,9	381,8	304,8
	Propensity score matching Imp (8)	26,81655	31,9	24,8	28,8	-0,5065	32,5	-2,5	1,5	442,1	473,6
	Quantile Regression Imp (10)	204,1574	12,6	203,4	204,9	164,5	12,3	163,8	165,3	**138,5**	**106,2**

Appendix 4: Bootstrap with increasing sample size using Lalonde data

	ATT
Lalonde results	1794 usd
Dejejia Results	1473 to 1774 usd

General summary of results			Bootstrap with Lalonde's Data																
			N=50				N=100				N=200				N=400			
			ATE	SD	ATT	SD	ATE	SD	ATT	SD	ATE	SD	ATT	SD	ATE	SD	ATT	SD
IE Meth	Rand		2071,9	2210	2071,9	2210	1754,8	1259	1754,8	1259	1794	767	1794	767	1785,3	252	1785,3	252
	PSM		1694,6	2910,1	2013,5	3061,4	1588,8	1619	1886,1	1647	1632,4	995	1925,8	996	1772,8	458	2039,4	454
IB Results	Mean	Gen	2071,9	2210	2071,9	2210	1754,8	1259	1754,8	1259	1794	767	1794	167	1785,3	252	1785,3	252
	Rand Imp	General	1989,6	2731	2295,5	2449	1733,5	1445	1745,9	1507	1792,4	903	1806,7	967	1773,1	418	1792,3	494
		Hot deck	2663,7	2620,8	2424,1	2022	1776,9	1623	1790,4	1628	1784	1073	1769,1	1067	1812,1	586	1789,5	616
	Det LM		1818,5	1916,4	2328,3	2322,8	1511	1286	1788,5	1312	1576,6	759	1836,8	785	1606,9	246	1816	251
	k-NN	V1	2029,1	2639,1	23,69,3	2986,4	1578	1496	2008	1547	1538,3	923	2003,4	1015	1920,2	343	2377,6	393
		V2	2064,7	2578,7	2458,4	2950,3	1604,9	1492	2075,8	1578,6	1611,7	913	2174,8	993	1853,2	336	2178,2	369
	Rand LM		1848,2	2201,7	2346,4	2743,2	1492,4	1470	1734,1	1573	1591,2	891	1834,1	1010	1586,2	420	1804,6	499
	MI	Mice	1940,8	2479,8	1980,5	2601,6	1702,9	1466	1562,6	1415	1670,5	898	1668,2	892	1631,9	371	1620,8	378
	ML Imp	Normal	1896,2	2948,5	2239,1	2878	1498,3	1495	1746,9	1544	1572,2	927	1797,4	961	1613,4	407	1782,1	459
		Dist-free	1698,6	2755,8	2089,6	2697,1	1482,4	1507	1730,9	1501	1585,1	918	1802,3	940	1629,1	406	1782,8	444