REVISTA INVESTIGACIÓN OPERACIONAL VOL. XX, NO. X, xxx-xxx, 2

FORTHCOMING 62B05-3-24-01

ESTIMATION OF DIFFERENCES-IN-DIFFERENCES IN THE PRESENCE OF NON-RESPONSE IN A RANDOM SAMPLE

Carlos N. Bouza-Herrera[1]
MATCOM. Universidad de La Habana, La Habana, Cuba

ABSTRACT
The estimation of Differences-in-Differences **(**DiD) using a simple random sampling with replacement (SRSWR) sample is developed. The cases in which the sample s is selected and subsamples s_1 and s_0 are determined deterministically or randomly are considered, as well as the case in which non-responses are present. Different alternative models are developed. Their behaviour in a real-life problem is discussed.

KEYWORDS: Difference-in-Difference estimation, simple random sampling, expected error, non-responses, subsampling rules.

MSC: 62D05, 62P12.

RESUMEN
Se desarrolla la estimación de las Diferencias existentes entre Diferencias (Differences-in-Differences **,** DiD) usando muestreo simple aleatorio con reemplazo. Los casos en los que la muestral s es seleccionada y submuestras s_1 y s_0 son determinadas ,deterministamente o aleatoriamente, son considerados., así como el caso en que hay no respuestas. Diferentes modelos alternativos son desarbolados . Su desempeño en problemas de la vida real es discutido.

PALABRAS CLAVE: estimación de las Diferencias entre Diferencias, muestreo símple aleatorio , error esperado, no respuestas, reglas de submuestreo

1. INTRODUCTION

Econometricians are using frequently Difference-in-Differences (DiD) methods for developing evaluations of the impact of new policies. The studies are based on evaluating the effect of them by comparing how they increase the response. These methods allow estimating the causal effects of a policy, program or treatment. See as examples Ashenfelter (1978) , Card (1990). Recent discussion on DiD methods are provided by Abadie- Cattaneo (2018), Wing et al. (2018). Its application is generating a growing literature . See for example Aggarwal-Hsu (2014), Distelhorst et al. (2016), Conyon et al. . (2019), Holm (2018), Kumar et al. (2016) and He-Zhang (2018).
The basic idea of DiD is that a group of similar units is observed. It is divided in two subgroups. In a subgroup the policy is implemented and in the other one no. They are denominated treatment and control groups respectively. DiD methods have been used not only in economics but in studies on management. The evaluation looks for estimating the effect in the response of the observed units to the policy . It is non randomly implemented and will be denominated in the sequel as “treatment”.

[1] bouza@matcom.uh.cu

The goal is to compare the treatment and control groups. The observed difference is possibly due to the policy. Non controlled factors may be present. They may be observable or not. The factors may be affecting the level of the outcomes both in the treatment and control groups. The aims of the decision maker is to establish if the observed difference is the effect of treatment.
The use of sampling modelling is not considered in the studies. This paper considers a survey sampling study where is needed to develop inferences of DID. A randomly selected sample s is partitioned into two subsamples and the treatment is assigned to one of them. A variable Y is evaluated in the selected units before and after applying the treatment. The difference of the estimated means is estimated in both groups and the difference between them is estimated. The question is if the observed DiD sustains that the effect of the treatment is significatively different for zero.
The next section presents the estimation of DiD of a simple individual-level DID model is presented. Simple random sampling with replacement (SRSWR) is used for selecting the sample. Estimators of the population DID are developed. Two alternatives are studied. The cases in which s is selected and it is s_1 and s_0 are determined deterministically or randomly are considered. Section 3 is devoted to study the case in which non-responses are present. The subsampling rules of Hansen-Hurwitz (1946), Srinath (1971) and Bouza (1981) are used for determining alternative models. Finally, a real-life problem is analysed. The data comes from a study of the effect of an after heart-stroke new treatment.

2. ESTIMATION OF DIffERENCES-IN-DIffERENCES

Differences-in-Differences (DID) is widely used in applied economics.
a simple individual-level DID model is given by

$$Y_{ijt} = \alpha\delta_{jt} + \theta_j + \gamma_t + \nu_{jt} + \varepsilon_{ijt}$$

Y_{ijt} is the outcome of individual i belonging to the j-th group.

$$\theta_j \text{ is the group effect}$$

γ_t is a fixed time effect
ν_{jt} is the interaction of the j-th group with the time error term .
ε_{ijt} is an individual random error term
The development of inferences is rather complicated due to the fact that errors are affected by the possible existence of intra-group and serial correlations. Underestimation of DID`s standard errors is expected if these effects are not considered, see Bertrand et al. (2004). Still, there is as yet no unified approach to dealing with this problem.
We are going to consider a problem arising in many survey sampling studies where is needed to develop inferences of DID.
Take a finite population $U = \{u_1, \dots, u_N\}$. Consider that the interest is establishing the effect of a certain treatment on the behaviour of a variable Y. A sample s of size n is to be selected. It is partitioned into subsamples s_1 of size n_1, the individuals assigned to the treatment, and a subsample s_0 of size n_0, a control group. The variable is measured in two different moments. For example, before introducing the changes (treatment) and after. Take for example a new teaching method. A test is developed at the beginning of the course to both groups at the end of it. The question of the teacher is on the improvements of the marks. Similarly, are the studies of a physician establishing the difference of parameters in persons recovering from a heart stroke, the biologist in evaluating a new treatment pest control, the engineering introducing technological changes in some factories, etc. These problems particular cases where the decision maker aims to estimate DiD. Consider only a group and fixing:

$$t = \begin{cases} 1 \text{ if the treatment is applied to the individual} \\ \quad\quad\quad\quad 0 \text{ otherwise} \end{cases}$$

The response of an individual may be modelled as

$$Y_{itq} = \mu_{tq} + \gamma_t I_q + \varepsilon_{itq},$$

$$q = \begin{cases} a \text{ if the measurement is made in the second moment } (after) \\ b \text{ if the measurement is made in the first moment } (before) \end{cases}$$

$$I_q = \begin{cases} 1 \text{ if } q = a \\ 0 \text{ if } q = b \end{cases}$$

γ_t is the effect of the treatment. The difference in the sample of units assigned to the treatment is

$$\widehat{D}_1 = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} Y_{i1a} - \sum_{i=1}^{n_1} Y_{i1b}\right) = \frac{1}{n_1}\sum_{i=1}^{n_1} d_{i1}$$

If simple random sampling with replacement (SRSWR) is used for selecting the sample

$$E(Y_{i1q}) = \mu_{1q} + \gamma_1 I_q.$$

Hence

$$E(\widehat{D}_1) = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} E(Y_{i1a}) - \sum_{i=1}^{n_1} E(Y_{i1b})\right) = \mu_{1a}+\gamma_1 - \mu_{1b} = D_1 + \gamma_1$$

Note that γ_1 is the effect of the treatment. The variance of the estimator is

$$V(\widehat{D}_1) = V\left(\frac{1}{n_1}\sum_{i=1}^{n_1} d_{i1}\right) = \frac{\sigma_{1b}^2 + \sigma_{1a}^2 - 2\sigma_{1ab}^2}{n_1}$$

where

$$\sigma_{1q}^2 = E(Y_{1q} - \mu_{1q})^2; q = a, b$$

$$\sigma_{1ab} = E(Y_{1a} - \mu_{1a})(Y_{1b} - \mu_{1b})$$

These parameters are estimable by

$$s_{1q}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(Y_{i1q} - \bar{y}_{1q})^2, \bar{y}_{1q} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{i1h}; q = a, b$$

$$s_{1ab} = \tfrac{1}{n_1-1}\textstyle\sum_{i=1}^{n_1}(Y_{i1a} - \bar{y}_{1a})(Y_{i1b} - \bar{y}_{1b}) \quad .$$

Hence

$$s^2(\widehat{D}_1) = \frac{s_{1b}^2 + s_{1a}^2 - 2s_{1ab}^2}{n_1}$$

is an unbiased estimator of $V(\widehat{D}_1)$.

$\widehat{D}_1$ is a mean, therefore under suitable conditions it is distributed $N(D_1 + \gamma_1, V(\widehat{D}_1))$. Then, we may test the validity of $H_1: E(\widehat{D}_1) > 0$ using T-tests .

A similar analysis of the case t=0 derives that

$$\widehat{D}_0 = \frac{1}{n_0}\left(\sum_{i=1}^{n_0} Y_{i0a} - \sum_{i=1}^{n_0} Y_{i0b}\right) = \frac{1}{n_0}\sum_{i=1}^{n_0} d_{i0}$$

$$E(\widehat{D}_0) = \mu_{0a} - \mu_{0b} + \gamma_0 = D_0 + \gamma_0$$

γ_0 is a residual effect due to the elapsed time between the first and second moments. The sampling error is

$$V(\widehat{D}_0) = V\left(\frac{1}{n_0}\sum_{i=1}^{n_0} d_{i0}\right) = \frac{\sigma_{0b}^2 + \sigma_{0a}^2 - 2\sigma_{0ab}^2}{n_0}$$

where

$$\sigma_{0q}^2 = E(Y_{0q} - \mu_{0q})^2; q = a, b$$

$$\sigma_{0ab} = E(Y_{0a} - \mu_{0a})(Y_{0b} - \mu_{0b})$$

The components of the error are estimated by

$$s_{0q}^2 = \frac{1}{n_0 - 1}\sum_{i=1}^{n_1}(Y_{i0q} - \bar{y}_{0q})^2, \bar{y}_{0q} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{i0q}; q = a, b$$

$$s_{0ab} = \frac{1}{n_0 - 1}\sum_{i=1}^{n_1}(Y_{i0a} - \bar{y}_{0a})(Y_{i0b} - \bar{y}_{0b})$$

and the variance is unbiasedly estimated by

$$s^2(\widehat{D}_0) = \frac{s_{0b}^2 + s_{0a}^2 - 2s_{0ab}^2}{n_0}$$

The difference between the effect of the treatment is estimated by the DID-estimator $\widehat{\Delta} = \widehat{D}_1 - \widehat{D}_0$. The decision maker is generally interested in evaluating the difference of $\Delta = E(\widehat{D}_1) - E(\widehat{D}_0)$. The subsamples permit to estimate unbiasedly Δ.

The following lemma fixes the relevant results derived above:

Lemma 1. If SRSWR is used for selecting the samples to be assigned to the control and the treatment $\widehat{\Delta} = \widehat{D}_1 - \widehat{D}_0$ is unbiased for Δ and its error is

$$V(\widehat{\Delta}) = \frac{\sigma_{1b}^2+\sigma_{1a}^2-2\sigma_{1ab}^2}{n_1} + \frac{\sigma_{0b}^2+\sigma_{0a}^2-2\sigma_{0ab}^2}{n_0}.$$

Proof.

The results follow from the discussion developed previously and the independence of s_0 and s_1 □

Under mild condition is possible to develop inferences using these results.

Consider the validity of $H_1: E(\widehat{\Delta}) > 0$. The variance is estimated unbiasedly using

$$s^2(\widehat{\Delta}) = s^2(\widehat{D}_1) + s^2(\widehat{D}_0)$$

The inferences may use the asymptotic normality of the estimators for performing tests. Accepting it that the observed effects are not due to the treatment a normal test may be used for testing $H_1: \Delta > 0$. The T-Student tests statistic is

$$T = \frac{\widehat{D}_1 - \widehat{D}_0}{\sqrt{\frac{s_{1b}^2 + s_{1a}^2 - 2s_{1ab}^2}{n_1} + \frac{s_{0b}^2 + s_{0a}^2 - 2s_{0ab}^2}{n_0}}}$$

The distribution is a $T(n_1 + n_0 - 2)$, which is roughly approximated by a N(0,1) even for moderate values of $n_1 + n_0$.

Accepting $H_1: E(\widehat{\Delta}) > 0$ means that the treatment has a significant positive effect.

Note that the experimenter may select the sample s and perform a Bernoulli experiment with probability of success P. For each $i \in s$ the experiment is performed and i is assigned to the treatment group if the result is a success. Otherwise i is assigned to the control group. In such cases the subsample sizes are random and the conditional expectations and variances are

$$E(\widehat{D}_j|n_j) = D_j + \gamma_j; j = 1,0$$

$$V(\widehat{D}_j) = V\left(E(\widehat{D}_j|n_j)\right) + E\left[V\left(\frac{1}{n_j}\sum_{i=1}^{n_j} d_{ij}\,|n_j\right)\right] = E\left(\frac{1}{n_j}\right)(\sigma_{jb}^2 + \sigma_{ja}^2 - 2\sigma_{jab}^2); j = 1,0$$

because the first term is zero.

Using the approximation developed by Stephan (1945)

$$V(\widehat{D}_j) \cong (\sigma_{jb}^2 + \sigma_{ja}^2 - 2\sigma_{jab}^2)\left(\frac{1}{Q_j} + \frac{1}{Q_j^2}\right), Q_j = \begin{cases} nP \text{ if } j = 1 \\ n(1-P) \text{ if } j = 0 \end{cases}$$

Then

$$E(V(\widehat{\Delta})) = \sum_{j=0}^{1}(\sigma_{jb}^2 + \sigma_{ja}^2 - 2\sigma_{jab}^2)\left(\frac{1}{Q_j} + \frac{1}{Q_j^2}\right)$$

These results are fixed in the following lemma.

2. THE CASE OF NON-RESPONSES

In real life applications the experimenter selects a samples s_0 and s_1 and in the second visit some non-responses may be present. Then both populations are stratified as follows $U_0 = U_{01} \cup U_{02}$ *and* $U_1 = U_{11} \cup U_{12}$. Denote

$$U_{hj} = \begin{cases} \{u_t \in U_h \text{ that responds in the 2nd visit}\} \text{ if } j = 1 \\ \{u_t \in U_h \text{ that does not respond in the 2nd visit}\} \text{ if } j = 2 \end{cases}, \quad h = 0,1$$

The sampled units belonging to U_{h1} give response to the variable of interest in both visits and those in U_{h2} report it only in the first visit. The sample s is going to be denoted in the sequel as $s_h = s_{h1} \cup s_{h2}$, $\|s_{hj}\| = n_{h1}, h = 0,1$ Without losing in generality take

$$s_{h1} = \{u_t \in s_h = s_h | 1 \leq t \leq n_{h1}\}, s_{h2} = \{u_t \in s_h = s_h | n_{h+1} + 1 \leq t \leq n_h\}$$

The existence of missing observations determines that only n_{h2*} units respond. The need of obtaining information from the stratum of the no-respondents determines selecting a subsample form s_{h2}. This problem was treated in the seminal paper of Hansen-Hurwitz (1946). They proposed a subsampling rule. Srinath (1971) and Bouza (1981) have proposed alternative rules. A unified notation is that the subsample size is $n_{h2*} = \theta n_{h2}, \theta \leq 1$, see Singh (2003).
The response-data allows computing

$$\frac{\sum_{i=1}^{n_{h1}} Y_{i1b}}{n_{h1}}; h = 0,1$$

$$\frac{\sum_{i=1}^{n_{h2*}} Y_{i1a}}{n_{h2*}}; h = 0,1$$

Considering the stratum of nonresponses, the subsample means in the second visits are:

$$\bar{y}_{jha} = \frac{\sum_{i=1}^{n_h} Y_{ijha}}{n_h}, h = 0,1, w_j = \frac{n_j}{n}$$

As some missing data is observed a subsample of size n_{h2*} is selected among the n_{h2} non respondents and is calculated

$$\bar{y}_{jh*} = \frac{\sum_{i=1}^{n_{h2*}} Y_{iha}}{n_{h2*}}$$

Take

$$\bar{y}_{ha*} = \sum_{j=1}^{2} w_j \bar{y}_{ha} + w_2(\bar{y}_{2ha*} - \bar{y}_{2ah})$$

Noting that. $E(\bar{y}_{2ha*} - \bar{y}_{2ah}|n_{h2}) = 0$ because $\sum_{j=1}^{2} w_j \bar{y}_{ha} = \bar{y}_{ha}$. In addition, $E(\bar{y}_{ha*}|n_{h2}) = \bar{Y}_{ha}$
Therefore, defining $\bar{d}_h = \bar{y}_{ha*} - \bar{y}_{hb}$

$$E(\bar{d}_h|n_{h2}) = E(\bar{y}_{ha*} - \bar{y}_{hb}|n_{h2}) = \mu_{ha} + \gamma_h - \mu_{hb} = D_h + \gamma_h; \; h = 0,1$$

The estimator sustains the unbiasedness. Hence, the sampling error is the variance. Note that the first term in the proposed estimator of the mean of is the sample mean. Then, for a fixed s

$$\bar{y}_{ha*} = \sum_{j=1}^{2} w_j \bar{y}_{ha} + w_2(\bar{y}_{2ha*} - \bar{y}_{2ah})$$

and

$$V(\bar{y}_{ha*}|n_{h2}) = V\left(\sum_{j=1}^{2} w_j \bar{y}_{ha} + w_2(\bar{y}_{2ha*} - \bar{y}_{2ah})|n_{h2}\right) = V(\bar{y}_{ha}|n_{h2}) + w_2^2 V(\bar{y}_{2ha*} - \bar{y}_{2ah}|n_{h2})$$

because the cross product is equal to zero, see Singh (2003), Bouza (2013). $V(\bar{y}_{ha}|n_{h2})$ and

$$V(\bar{y}_{2ha*} - \bar{y}_{2ah}|n_{h2}) = \frac{\sigma_{h2a}^2}{n_{ha*}} .$$

Therefore

$$E\big(V(\bar{y}_{ha*}|n_{h2})\big) = \frac{\sigma_{ha}^2}{n_h} + E\left(\frac{\frac{n_{h2}^2}{n_h^2}}{n_{ha*}}\right)\sigma_{h2a}^2 = \frac{\sigma_{ha}^2}{n_h} + E\left(\frac{n_{h2}}{\theta n_h^2}\right)\sigma_{h2a}^2$$

The estimation of a difference under non responses has been studied by Bouza-Ajgaonkar (1993). For the difference in h=0,1

$$E\left(V\left(\bar{d}_h|n_{h2}\right)\right) = \frac{\sigma_{ha}^2}{n_h} + E\left(\frac{n_{h2}}{\theta n_h^2}\right)\sigma_{h2a}^2 + \frac{\sigma_{hb}^2}{n_h} - \frac{2\sigma_{hab}}{n_h} \quad \text{(A)}$$

because

$$E\big(Cov\,(\bar{y}_{ha*}, \bar{y}_{hb}|n_{h2})\big) = \sigma_{hab} + E\left[E\left(\bar{y}_{hb}\big(w_2^2(\bar{y}_{2ha*} - \bar{y}_{2ah}|n_{h2})\big)\right)\right]$$

and the conditional expectation is equal to zero due to the fact that $E(\bar{y}_{2ha*}|n_{h2}) = \bar{y}_{2ah}$. Then is proved the following statement:
Lemma 2. Consider samples s_0 and s_1 are selected using simple random sampling with replacement and $s_h \in U_h = U_{h1} \cup U_{h2}, U_{h1} \cap U_{h2} = \emptyset, h = 0,1$ where

$$U_{hj}=\begin{cases}\{u_t \in U_h \text{ that responds in the 2nd visit}\} \text{ if } j=1 \\ \{u_t \in U_h \text{ that does not respond in the 2nd visit}\} \text{ if } j=2\end{cases},\quad h=0,1$$

The non-respondents subsample size is determined by $n_{h2*}=\theta n_{h2}$ where

$$\theta=\begin{cases}\frac{1}{K},\ K>1\ (Hansen-Hurwitz\text{´}s\ rule) \\ \frac{n_2}{Hn+n_2},H>0\ (Srinath\text{´}s\ rule) \\ \theta=\frac{n_{h2}}{n_h}\ (Bouza\text{`}s\ rule)\end{cases}.$$

a) The DiD´s estimator

$$\hat{\Delta}_{nr|\theta}=\left[\left(\frac{\sum_{i=1}^{n_{11}} Y_{i1a}}{n_{11}}+\frac{\sum_{i=1}^{n_1} Y_{i1a}}{n_{12*}}\right)-\frac{\sum_{i=1}^{n_1} Y_{i1b}}{n_1}\right]-\left[\left(\frac{\sum_{i=1}^{n_{01}} Y_{i0a}}{n_{01}}+\frac{\sum_{i=1}^{n_0} Y_{i0a}}{n_{02*}}\right)-\frac{\sum_{i=1}^{n_0} Y_{i0b}}{n_0}\right]$$

is unbiased for Δ.

b) The expected error of $\hat{\Delta}_{nr|\theta}$ *is*

$$E\left(V(\hat{\Delta}_{nr|\theta}|n_{h2})\right)=\begin{cases}\frac{\sigma_{ha}^2}{n_h}+\frac{KW_2\sigma_{h2a}^2}{n_h}+\frac{\sigma_{hb}^2}{n_h}-\frac{2\sigma_{hab}}{n_h}(Hansen-Hurwitz\text{´}s\ rule) \\ \frac{\sigma_{ha}^2}{n_h}+\frac{(H+W_2)\sigma_{h2a}^2}{n_h}+\frac{\sigma_{hb}^2}{n_h}-\frac{2\sigma_{hab}}{n_h}(Srinath\text{´}s\ rule) \\ \frac{\sigma_{ha}^2}{n_h}+\frac{\sigma_{h2a}^2}{n_h}+\frac{\sigma_{hb}^2}{n_h}-\frac{2\sigma_{hab}}{n_h}(Bouza\text{`}s\ rule)\end{cases}$$

Proof:

a) Due to the unbiasedness of $\hat{D}_1$ and $\hat{D}_0$ follows that $E\left(\hat{\Delta}_{nr|\theta}\right)=\Delta$ for both rules.

b) Substituting θ in the corresponding expression of $E\left(V\left(\bar{d}_h|n_{h2}\right)\right)$is derived the second result.□

3. A STUDY OF THE BEHAVIOR OF THE PROPOSALS.

3.1. The real-life problem.

The use of surgery is needed in the cases of patients with some kind of disease, as those related with coronary artery or valvular diseases and with symptoms-signs of cardiac function impairment, see Castellanos et al. (2019). The study dealt with the evaluation of the recovering of patients in terms of pulmonary flux, mitral flux-gram, maximal cardiac frequency and myocadiac efficiency 1,7. The effect of a change in the common protocol was the interest of the researchers. The levels of the treatment based on carvedilol, clopidogrel, ASA and atorvastatin where changed. The 180 patients in the study underwent cardiac surgery (revascularization and valve surgery) without postoperative atrial fibrillation and free of diabetes. Bias was eliminated by allocating patients based on random sequence generation. A measurement was made in a first visit to the cardilog and echocardiograms and ergograms were made. After six months a new visit was to be made. In the control group 44,4% did not assist to the second visit. In the treatment group that percent was 62,2%. The physicians contacted the non-respondents and programmed a second and third visit for obtaining all the data. Then the relative sizes of the non-response strata were 0,444 and 0,622 respectively.

3.2. The full-response model.

Samples of size 30,40 and 50 were selected from this population. for each sample s_h , h=1,…, 1000. The accuracy of the method was evaluated computing

$$\vartheta_n=\frac{1}{1000}\sum_{h=1}^{1000}\frac{|\hat{\Delta}_h-\Delta|}{\Delta};n=30,40,50$$

Table 1: the full response case: Accuracy of $\widehat{\Delta}$ in the efficiency measures

Efficiency Measures	n=30	n=40	n=50
Pulmonary Flux 081	0,0544	0,0477	0,0431
Mitral Flux-Gram 86	0,0466	0,0461	0,0411
Maximal Cardiac Frequency 144	0,7936	0,7811	0,7806
Myocardiac Efficiency 1,7.	0,0855	0,0838	0,0812

Table 1 suggests that the estimator is very accurate for all the measures. The increase of the sample sizes has not a significative effect in reducing the estimator`s error.

The use of the developed estimator for testing was evaluated using the test statistic

$$T = \frac{\left(\widehat{D}_1 - \widehat{D}_0\right) - \Delta}{\sqrt{\frac{s_{1b}^2 + s_{1a}^2 - 2s_{1ab}^2}{n_1} + \frac{s_{0b}^2 + s_{0a}^2 - 2s_{0ab}^2}{n_0}}}$$

was computed for each sample s_h , h=1,…, 1000 and was tested

$$H_0: E\left(\widehat{D}_1\right) - E\left(\widehat{D}_0\right) = \Delta \; vs \; H_1: \left(\widehat{D}_1\right) - E\left(\widehat{D}_0\right) \neq \Delta$$

The experiment was evaluated using $1 - \alpha = 0,95$ and computing

$$\hat{\gamma}_n = \frac{1}{1000}\sum_{h=1}^{1000} L_{1hn};\; L_{1hn} = \begin{cases} 1 \; if \; H_0 \; is \; accepted \\ 0 \; otheerwise \end{cases}; n = 30, 40,50.$$

Table 2. Estimated probability of accepting that Δ was the population DiD for $1 - \alpha = 0,95$: the full response case.

Efficiency Measures	n=30	n=40	n=50
Pulmonary Flux 081	0,894	0,896	0,904
Mitral Flux-Gram 86	0,938	0,943	0,943
Maximal Cardiac Frequency 144	0,948	0,948	0,948
Myocardiac Efficiency 1,7.	0,936	0,941	0,948

Table 1 suggests that the normal approximation of test statistic is not acceptable for the pulmonary flux but in the rest of the measures performed adequately close to $1 - \alpha$. The sample size has not a role in the approximation.

3.3. The non-response problem.

The non-response problem used the samples generated previously. For the second visit non—responses were generated for each selected patient. Then, the subsamples were obtained and the non-response based estimator calculated. Missing data were generated using Bernoulli random variables with parameters 0,444, for the control group, and 0,622 for the treatment one. The subsamples rules used the values K=H=2, 5, 10- The accuracy of the estimator was evaluated computing

$$\vartheta_{nr|\theta} = \frac{1}{1000}\sum_{h=1}^{1000} \left|\widehat{\Delta}_{nr|\theta} - \Delta\right|_h ; \theta = K, H, \frac{n_2}{n}; \; n = 30, 40, 50$$

Table 3: The non-response case: Accuracy of $\widehat{\Delta}_{nr|\theta}$ in the efficiency measures.

Efficiency Measures	n=30	n=40	n=50
Pulmonary Flux			
K=2	0,0781	0,0714	0,0712
K=5	0,0833	0,0821	0,0813
K=10	0,1023	0,1011	0,1001
H=2	0,1061	0,1052	0,0975
H=5	0,1873	0,1832	0,1819
H=10	0,2404	0,2400	0,2397
n_2/n	0,0594	0,0594	0,0571
Mitral Flux-Gram			
K=2	0,1383	0,1338	0,1285
K=5	0,1494	0,1330	0,1323
K=10	0,1524	0,1513	0,1513
H=2	0,1928	0,1922	0,1922
H=5	0,2172	0,1972	0,1959
H=10	0,2289	0,1904	0,1896
n_2/n	0,1127	0,1227	0,1105
Maximal Cardiac Frequency			
K=2	0,0899	0,0883	0,0848
K=5	0,0947	0,0926	0,0922
K=10	0,1054	0,1010	0,0971
H=2	0,1735	0,1713	0,1611
H=5	0,2578	0,1889	0,1689
H=10	0,2945	0,2486	0,2282
n_2/n	0,0757	0,0723	0.0704
Myocardiac Efficiency			
K=2	0,0178	0,0176	0,0169
K=5	0,0468	0,0318	0,0236
K=10	0,0596	0,0583	0,0401
H=2	0,0398	0,0367	0,0355
H=5	0,0717	0,0669	0,0604
H=10	0,4879	0,4872	0,4765
n_2/n	0,0165	0,0162	0,0157

As deducible from the formula of the error the use of Srinath (1971) rule is more inaccurate and the rule of Bouza (1981) is the most accurate. Larger values of K and H increase t$\vartheta_{nr|\theta}$. From the literature we have that Hansen-Hurwitz rule (1946) has a smaller expected cost than the other rules. Therefore it is commonly preferred as its accuracy is similar to Bouza`s.

For performing tests $H_0: E(\widehat{\Delta}_{nr|\theta}) = \Delta \; vs \; H_1: \widehat{\Delta}_{nr|\theta} \neq \Delta$ is needed to use a non-parametric method, as normality is not a natural approximation. Resampling methods provide tools for testing hypothesis when dealing with complex sampling

Resampling methods are commonly used for inference in complex survey sampling. See Booth et al. (1994), Antal-Tillé (2011, 2014). They allow solving the difficulties of estimating the sampling errors. Resampling identifies a set of inferential techniques as randomization-based tests, cross-validation, Jacknife and Bootstrap. Their principles are very similar. Efron (1979) in his seminal paper conceived Bootstrap to be used for inferential purposes, see also Efron-Tsibirani (1993).The initial sample s is treated as the population and pseudo-populations are randomly generated. A large number of resamples s_b , b=1,..,B, of size n are selected randomly, from the original sample of size n, with replacement. It performs

better than some other asymptotic statistical methods. Bootstrap method also provides consistent estimates of the distribution of the estimator. The complexity of the finite population sampling design poses a challenge for finding a valid bootstrap procedure. A good Bootstrap procedure should support that the Bootstrap-bias estimate be 0 . See Antal-Tillé (2011) and Booth et al. (1994). The model developed previously determines a smooth function of finite population means. In such cases, in practical situations, Bootstrap enables to implement adequate tests of hypothesis.
The Bootstrap implemented involves the following steps:

1. Select randomly and independently a sample s_b using SRSWR from s.
2. For b=1,…,B calculate $\widehat{\Delta}_{nr|\theta(b)}$.
3. Compute $\bar{\bar{\Delta}}_{nr|\theta}=\frac{1}{B}\sum_{b=1}^{B} \widehat{\Delta}_{nr|\theta(b)}, v_{nr|\theta}=\frac{1}{B-1}\sum_{h=1}^{B} (\widehat{\Delta}_{nr|\theta(b)}-\bar{\bar{\Delta}}_{nr|\theta})^2$.

For B sufficiently large the output allows using the percentile method or the T-Student Bootstrap for testing hypothesis. See for example the confidence intervals formula proposed in Rao et al.(1992). See also Tillé (2006).

$$T_{nr|\theta}=\frac{\bar{\bar{\Delta}}_{nr|\theta}-\Delta}{\sqrt{\frac{1}{B(B-1)}\sum_{h=1}^{B} (\widehat{\Delta}_{nr|\theta H(h)}-\bar{\bar{\Delta}}_{nr|\theta})^2}}\sim N(0,1)$$

It works well for any smooth statistic.
A Monte Carlo experiment was developed with the data and H=1000 samples were selected from the population of patients. $T_{nr|\theta}$ was computed in each generated sample and was tested
$H_0: E(\widehat{\Delta}_{nr|\theta})=\Delta\ vs\ H_1: \widehat{\Delta}_{nr|\theta}\neq\Delta$,
$1-\alpha=0,95$ was estimated using

$$\hat{\gamma}_{nr|\theta}=\frac{1}{1000}\sum_{h=1}^{1000} L_{nr|\theta(h)};\ L_{nr|\theta(h)}=\begin{cases}1\ if\ H_0\ is\ accepted\\ 0\ otherwise\end{cases};$$

$n=30,40,50;\ \theta=K,H,\frac{n_2}{n}$.The Monte Carlo experiment generated the results in table 4 below. They sustain that the parametric Bootstrap tests developed using of the rules of Hansen-Hurwitz and Bouza generated are good alternatives as the estimates of $1-\alpha$, as they are close to 0,95.
The increase in the parameters of the rules of Hansen-Hurwitz and Srinath are more important in the convergence of the test statistics then having larger sample sizes.

Table 4: Estimated probability of accepting that Δ was the population DiD for $1-\alpha=0,95$: the non- response case.

Efficiency Measures	n=30	n=40	n=50
Pulmonary Flux			
K=2	0,9180	0,9271	0,9388
K=5	0,8566	0,8590	0,8876
K=10	0,8810	0,8887	0,8907
H=2	0,8611	0,8610	0,8821
H=5	0,8590	0,8591	0,8604
H=10	0,8529	0,8557	0,8593
n_2/n	0,9196	0,9196	0,9297
Mitral Flux-Gram			
K=2	0,9358	0,9379	0,9388
K=5	0,9211	0,9276	0,9365
K=10	0,9086	0,9123	0,9198
H=2	0,9054	0,9074	0,9077
H=5	0,8977	0,9021	0,9039
H=10	0,8941	0,8966	0,8953
n_2/n	0,9300	0,9352	0,9375
Maximal Cardiac Frequency			
K=2	0,9334	0,9387	0,9421

K=5	0,9322	0,9342	0,9342
K=10	0,9310	0,9327	0,9327
H=2	0,9210	0,9289	0,9305
H=5	0,9053	0,9071	0,9088
H=10	0,8996	0,9009	0,9039
n_2/n	0,9305	0,9312	0,9409
Myocardiac Efficiency			
K=2	0,9033	0,9084	0,9189
K=5	0,8976	0,8976	0,8985
K=10	0,8663	0,8701	0,8785
H=2	0,8965	0,8995	0,9005
H=5	0,8884	0,8902	0,8974
H=10	0,8752	0,8789	0,8896
n_2/n	0,9008	0,9027	0,9085

RECEIVED: JULY, 2023.
REVISED: JANUARY, 2024.

REFERENCES

[1] ABADIE, A., and CATTANEO, M. D. (2018): Econometric methods for program evaluation. **Annual Review of Economics,** 10, 465–503.

[2] AGGARWAL, V. A., and HSU, D. H. (2014): Entrepreneurial exits and innovation. **Management Scienc**e, 60, 867–887.

[3] ANTAL E, TILLÉ Y (2014): A new resampling method for sampling designs without replacement: the doubled half bootstrap. **Computational Statistics** 29, 1345-1363.

[4] ASHENFELTER, O. (1978): Estimating the effect of training programs on earnings. **The Review of Economics and Statistics,** 60, 47–57.

[5] ATHEY, S., and IMBENS, G. W. (2017): The state of applied econometrics: Causality and policy EVALUATION. **JOURNAL OF ECONOMIC PERSPECTIVES**, 31, 3–32.

[6] BERGER, A. N., KICK, T., and SCHAECK, K. (2014): Executive board composition and bank risk taking. **Journal of Corporate Finance**, 28, 48–65.

[7] BERTRAND, M., DUFLO, E., and MULLAINATHAN, S. (2004): How much should we trust differences-in-differences estimates? **The Quarterly Journal of Economics**, 119, 249–275.

[8] BOOTH JG, BUTLER RW, HALL P (1994): Bootstrap methods for finite populations. **J Am Stat Assoc** 89, 1282–1289

[9] BOUZA C. N.; PRHABU-AJGAONKAR, S. G.: (1993): Estimation of the difference of population means when observations are missing. **Biom. J**.,.35, 245-252

[10] CASTELLANOS VÁZQUEZ, J., A. SANTIAGO MORENO, C. BOUZA HERRERA AND J. M. SAUTTO VALLEJO (2019): Valoración de riesgo cardiovascular mediante modelos de clasificación**. Revista Investigacion Operacional** 40 , 81-88.

[11] CONYON, M. J., HASS, L. H., PECK, S. I., SADLER, G. V., and ZHANG, Z. (2019): Do compensation consultants drive up CEO pay? Evidence from UK public firms. **British Journal of Management**, 30, 10–29.

[12] DISTELHORST, G., HAINMUELLER, J., and LOCKE, R. M. (2016): Does lean improve labor standards? Management and social performance in the Nike supply chain. **Management Science**, 63, 707–728.

[13] EFRON, B. (1979): Bootstrap methods: Another look at the Jackknife. **Ann. Statist**. 7, 1-26.

[14] EFRON, B, and TIBSHIRANI, R.J. (1993): **An Introduction to the Bootstrap**. Chapman and Hall, London.

[15] HANSEN, M.H.. and HURWITZ, W.N. (1946): The problem of non-responses in sample surveys. **J. American. Statistical Association.** 41, 517-529

[16] HE, P. and ZHANG, B. (2018): Environmental tax, polluting plants' strategies and effectiveness: Evidence from China. **J. of Policy Analysis and Management**, 37, 493–520.

[17] HOLM, J. M. (2018): Successful problem solvers? Managerial performance information use to improve low organizational performance. **J. of Public Administration Research and Theory,** 28, 303–320.

[18] KUMAR, A., BEZAWADA, R., RISHIKA, R., JANAKIRAMAN, R., and KANNAN, P. K. (2016): From social to sale: The effects of firm-generated content in social media on customer behavior. **J. of Marketing,** 80, 7–25.
[19] SINGH, S. (2003): **Advanced Sampling Theory with Applications,** Kluwer Academic Publishers, Dordrecht, Amsterdam
[20] SRINATH, K. P. (1971): Multi-Phase Sampling In Non-Response Problems. **J. American Stat. Ass**. 66, 583-589.
[21] RAO J. N. K, WU C. F. J., YUE K. (1992): Some recent work on resampling methods for complex surveys. **Surv. Methodol** 18, 209–217.
[22] SAIGO H., SHAO J., SITTER R. R. (2001): A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data. **Surv. Methodol** 27, 189–196
[23] TILLÉ Y. (2006): **Sampling algorithms**. Springer, New York.
[24] WING, C., SIMON, K., and BELLO-GOMEZ, R. A. (2018): Designing difference in difference studies: Best practices for public health policy research. **Annual Review of Public Health**, 39, 453–469.