Dual Frame Implementation for the Colombian National Agricultural Survey

Trujillo, Leonardo(1st author)
Department of Statistics
National University of Colombia
Bogota, Colombia
Email: ltrujilloo@bt.unal.edu.co

Merchan, Oscar (2nd author)
Department of Statistics
National University of Colombia
Bogota, Colombia
Email: ofmerchanl@unal.edu.co

Ferraz, Cristiano (3rd author)
Department of Statistics
Federal University of Pernambuco
Pernambuco, Brasil
E-mail: cferraz.stat@gmail.com

Abstract

Multiple frame surveys are very useful when it is not possible to guarantee a complete coverage of the target population and may result in considerable cost savings over a single frame design with comparable precision. However, this technique is not very often applied from national statistical offices (NSO) due to its complexity and also because in most surveys only a single sample frame is available. Regarding the Colombian case, the implementation of multiple frame surveys is very rare and therefore, this paper presents a proposal for a dual frame implementation in the National Agricultural Survey at DANE (National Statistical Administrative Department - Colombian NSO) considering the three most well known available estimators for the population total in the recent statistical literature.

Introduction

This paper considers a first approximation for the application of the multiple frame surveys theory in one of the most important surveys for the Colombian government such as the Colombian National Agricultural Survey (NAS). Multiple frame surveys are very useful when it is not possible to guarantee a complete coverage of the target population and may result in considerable cost savings over a single frame design with comparable precision. However, this technique is not very often applied from national statistical offices (NSO) due to its complexity and also because in most surveys only a single sample frame is available. The statistical literature about multiple frame surveys started around 1960 and its development has evolved very quickly. Hartley and Rao have been two importants names of researchers working in this subject and for the Colombian particular case there are not many references about its application. A possible reason for this absence of applications on this topic could be the practical problems (costs, data availability, among others) to build several sampling frames but the recent advances on information systems and communications could help on the dissemination of these techniques over the country.

There are several available estimators in the statistical literature and the first question to take into account is how to choose which one of them would be more suitable for this application. The decision was taken according to the evaluation of the accuracy and precision of the more relevant estimators with a Monte Carlo simulation. Some simulation analysis have been done before from different authors but the usual sampling designs only range from simple random sampling without replacement (SI) to two stage sampling with a SI design in each stage (SI-SI design). In this paper, we will consider pps, stratified and multistage sampling designs and some of their combinations. The comparison is done using the Colombian National Agricultural Survey which is implemented every single year.

In the case of the NAS, some indicators are estimated in terms of soil use, production and yield for both transitory and permanent cultivations, pasture area, milk production and animal inventory. In parallel, some specialized agricultural surveys are implemented separately to the NAS in order to estimate some variables of interest for legumes, rice and cattle. Then, several indicators are generated for the same agricultural indicators. However, it does not exist so far a methodological proposal in order to integrate several surveys in order to get a single estimation and in order to reduce costs and variance of the estimations.

Multiple Frame Surveys

In survey sampling, a sampling frame is any device or instrument to identify and locate all the elements in a population. However, in the classic theory of sampling, a single sampling frame is obtained in order to get a sample. The sampling unit could be either an element or a subset of element and in this sense, one could define two types of sampling frames: element sampling frames or subset sampling frames as in the case of area sampling frames.

If it is possible to use two sampling frames in order to get a complete coverage of the population we are facing a dual frame estimation problem (Hartley, 1962). In this case, we have to work under two main assumptions: firstly, every unit in the population belongs to at least one of the two sampling frames considered and secondly, for every unit in the sample of one frame will be possible to determine if this unit belongs or not to the other sampling frame.

A graphical representation of the samples under a dual sampling frame approach is shown in the figure 1:

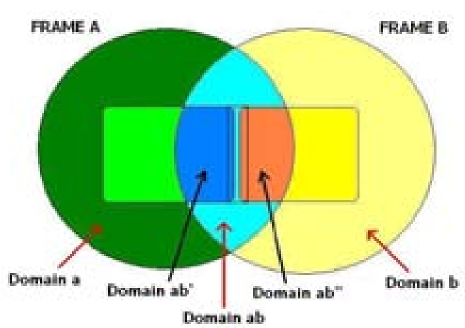

Figure 1: General scheme of samples under a dual sampling frame approach

We will consider the notation in the Table 1 (Fernandes, 2007). There are four possible scenarios according to the knowledge of the total number of units in each sampling frame, the domains and the possibility to assign the number of elements to be selected in each domain (Hartley,1962). The first scenario is when all the domain sizes are known and then is possible to assign sample sizes to them (Domain = Strata). The second one is when all the domain sizes are known, but the sampling sizes can be assigned only to the sampling frames (Domains = Post-strata). The third one is when the domain sizes are unknown but the sampling frame sizes are known. In this case, sampling sizes can be assigned only to the sampling frames (Domains = Proper Domains). Finally, both the domain sizes

Table 1: Notation

Object	SAMPLING FRAME		DOMAIN			
	A	B	a	b	ab	
Population	U_A	U_B	U_a	U_b	U_{ab}	
Population Size	N_A	N_B	N_a	N_b	N_{ab}	
Population Total	t_{yA}	t_{yB}	t_{ya}	t_{yb}	t_{yab}	
Population Mean	μ_{yA}	μ_{yB}	μ_{ya}	μ_{yb}	μ_{yab}	
Sample	S_A	S_B	S_a	S_b	S_{ab}	
Sample Size	n_A	n_B	n_a	n_b	n'_{ab}	n''_{ab}
Sample Total	$\hat{t}_{yA}$	$\hat{t}_{yB}$	$\hat{t}_{ya}$	$\hat{t}_{yb}$	$\hat{t}'_{yab}$	$\hat{t}''_{yab}$
Sample Mean	$\overline{y}_A$	$\overline{y}_B$	$\overline{y}_a$	$\overline{y}_b$	$\overline{y}'_{ab}$	$\overline{y}''_{ab}$

and the sampling frame sizes are unknown but the relative size of the sampling frames are known. In this case, sampling sizes can be only assigned to the sampling frames (Domains = Domains in populations of unknown size). In this paper, we will only consider the scenarios 2 and 3 as these are the particular cases for our application.

Estimation in Dual Frame Surveys

Hartley Estimator

We will consider that our interest is to estimate a population total t_y, using information from two independent samples taken over each sampling frame. Under this dual sampling frame approach, the total population t_y can be expressed as $t_y = t_{ya} + t_{yab} + t_{yb}$, where t_{ya} is the domain total in a, t_{yab} is the domain total in the intersection ab, and t_{yb} is the domain total in b. Notice then a, ab and b are mutually exclusive sets. Hartley (1962) proposed an estimator under this dual frame sampling approach. The estimator corresponds to a weighted average in order to estimate the domain total in the population by $\hat{t}_{yH}(p) = \hat{t}_{ya} + p\hat{t}'_{yab} + (1-p)\hat{t}''_{yab} + \hat{t}_{yb}$, where $\hat{t}_{ya}$ is the estimated population total for the domain a, $\hat{t}'_{yab}$ is the estimated total population for the domain ab using the sample in the sampling frame A, $\hat{t}''_{yab}$ is the estimated population total for the domain ab using the sample in the sampling frame B, $\hat{t}_{yb}$ is the estimated population total for the domain b. If the estimators $\hat{t}_{ya}$, $\hat{t}_{yb}$, $\hat{t}'_{yab}$ and $\hat{t}''_{yab}$ are unbiased for t_{ya}, t_{yb}, t_{yab} and t_{yab} respectively, $\hat{t}_{yH}(p)$ will be unbiased for t_y. Hartley (1962) proposed to calculate p in order to minimize the variance of $\hat{t}_{yH}(p)$, considering the assumption that both samples in both sampling frames are independent, this variance can be expressed by $V[\hat{t}_{yH}(p)] = V[\hat{t}_{ya} + p\hat{t}'_{yab}] + V[(1-p)\hat{t}''_{yab} + \hat{t}_{yb}]$ and the optimum value of p minimizing this variance is given by $p_{opt} = \frac{V(\hat{t}''_{yab}) + Cov(\hat{t}''_{yab}, \hat{t}_{yb}) - Cov(\hat{t}_{ya}, \hat{t}'_{yab})}{V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})}$

Fuller and Burmeister Estimator

Fuller and Burmeister (1972) considered a modification to Hartley's estimator incorporating the estimated size of the domain ab according to the samples in both sampling frames $\hat{t}_{yFB}(p_1, p_2) = \hat{t}_{ya} + \hat{t}_{yb} + p_1\hat{t}'_{yab} + (1-p_1)\hat{t}''_{yab} + p_2(\hat{N}'_{ab} - \hat{N}''_{ab})$, where $\hat{N}'_{ab}$ and $\hat{N}''_{ab}$ are the respective estimators of the size of domain ab under the samples from A and B respectively. If the estimators $\hat{t}_{ya}$, $\hat{t}_{yb}$, $\hat{t}'_{yab}$, $\hat{t}''_{yab}$, $\hat{N}'_{ab}$ and $\hat{N}''_{ab}$ are unbiased for t_{ya}, t_{yb}, t_{yab}, t_{yab}, N_{ab} and N_{ab} respectively; $\hat{t}_{yFB}(p_1, p_2)$ will be unbiased for t_y. Analogously to the Harvey's estimator, values for p_1 y p_2 are obtained in order to minimize the variance of $\hat{t}_{yFB}$, to finally get

$$\begin{bmatrix} p_{1opt} \\ p_{2opt} \end{bmatrix} = -\left(Cov \begin{bmatrix} \hat{t}'_{yab} - \hat{t}''_{yab} \\ \hat{N}'_{ab} - \hat{N}''_{ab} \end{bmatrix} \right)^{-1} \begin{bmatrix} Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}''_{yab}, \hat{t}'_{yab} - \hat{t}''_{yab}) \\ Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}''_{yab}, \hat{N}'_{ab} - \hat{N}''_{ab}) \end{bmatrix}$$

The optimal values of p_1 y p_2 are function of some population covariances that could be estimated from the information in the selected samples.

Pseudo Maximum Likelihood Estimator

Skinner and Rao (1996) proposed the Pseudo Maximum Likelihood (PML) estimator. This estimator permits to calibrate the domain total estimations to the general total under complex sampling designs like stratified or cluster sampling designs. The estimator is given by:

$$\hat{t}_{yPMV}(p) = \frac{N_A - \hat{N}_{ab}^{PMV}(p)}{\hat{N}_a}\hat{t}_{ya} + \frac{\hat{N}_{ab}^{PMV}(p)}{p\hat{N}'_{ab} + (1-p)\hat{N}''_{ab}}[p\hat{t}'_{yab} + (1-p)\hat{t}''_{yab}] + \frac{N_B - \hat{N}_{ab}^{PMV}(p)}{\hat{N}_b}\hat{t}_{yb}$$

where $\hat{N}_{ab}^{PMV}(p)$ is the smaller root of the quadratic equation:

$$(n_A + n_B)\hat{N}_{ab}^{2PMV}(p) - (n_A N_B + n_B N_A + n_A \hat{N}'_{ab} + n_B \hat{N}''_{ab})\hat{N}_{ab}^{PMV}(p) + n_A \hat{N}'_{ab} N_B + n_B \hat{N}''_{ab} N_A = 0$$

The expected value and the variance of the PML estimator is obtained under asymptotic theory and it is necessary to find an optimal value of p minimizing the asymptotic variance of $\hat{N}_{ab}^{PMV}(p)$. The minimization with respect to the asymptotic variance of $\hat{t}_{yPMV}(p)$, have the problem that the sampling weights would depend on the values of y.

Skinner and Rao (1996) show that the optimal value of p minimizing the asymptotic variance of $\hat{N}_{ab}^{PMV}(p)$ is given by $p_{opt} = \frac{N_a N_B V(\hat{N}''_{ab})}{N_a N_B V(\hat{N}''_{ab}) + N_b N_A V(\hat{N}'_{ab})}$

Methodology

In order to compare and to evaluate the three estimators above to be implemented in the Colombian ANS under a dual frame approach, we will consider an area sampling frame for the ANS itself and an available list sampling frame from the Legumes Survey. Both studies have been designed and implemented by DANE (Colombian National Statistical Office) and we will consider the current sampling designs that have been used for these surveys. In this section, we will present the sampling designs for the Colombian National Agricultural Survey (NAS) and the Legumes Survey (LS).

Colombian National Agricultural Survey - Sampling Design

The sampling frame for the Colombian National Agricultural Survey (NAS) corresponds to an area sampling frame with a coverage of 37,900,546 has (169,587 SSUs in 31,588 PSUs). The survey uses a stratified two stage sampling design: the first stage corresponds to a probability proportional to size sampling without replacement (pips) and a simple random sampling without replacement sampling design (SI) in the second stage. The auxiliary variable for the first stage is the PSU's planimetric area. The stratification variables are elevation and vegetable coverage. The NAS is applied twice per year and for 2010, the sampling size during the first semester was 2,537 SSUs corrsponiding to 293,252 has and for the second semester was 5,894 SSUs corresponding to 959,767 has

Colombian Legumes Survey (LS) - Sampling Design

Together with the NAS, another specific surveys are done over the country to estimate variables of interest for some specific products such as corn, fish and legumes, among others. Currently, DANE takes what is in these specialized frames from the area sampling frame and considered them as strata (scenario 1). Then, the idea is to take information from the LS to make some simulation studies and to evaluate some accuracy and precision indicators of some estimators under scenarios 2 and 3.

The sampling frame of legumes has a coverage of 10,874 has (588 sampling units). The sampling design is a srswor of these sampling units, considering the same stratification than in the ANS and the sampling size for LS in 2010 was 95 sampling units corresponding to 1,735 has. There are four possible combinations between srswor and πps for each design in the NAS and the LS.

Simulation

The simulation was done making an artificial study population, taking as reference sampling data from the Agricultural National Survey (NAS in frame A) and the Legumes Survey (LS in frame B) and replicating this information until reaching the population sizes in both frames. In both surveys, three variables were measured: cultivated area, harvested area and production. However in terms of the simulation, we chose cultivated area only. Then, from each sampling frame we used different sample sizes. In the case of the ANS (frame A), there were two sampling sizes, each one for every selection stage. On the first stage, we took 1,000 (2.5%) and 1,750(5%) PSUs. On the second stage, we decided to take all the sampling elements in the PSU (cluster sampling design). In the case of the LS (frame B), we took three different sample sizes corresponding to 59(10%), 160(20%) and 117(30%). Finally, different random samples were obtained from each sample frame under the eight sampling designs considered. The simulation was done using the libraries survey and TeachingSampling in R, repeating this procedure a thousand of times and calculating the respective values of relative bias, mean square error and the coverage probability. In order to evaluate the efficiency and precision of the estimators considers, some precision measures such as the relative bias, the mean square error and the coverage probability were calculated. The relative bias is calculated by $SR = \frac{1}{M}\sum_{m=1}^{M}\frac{\hat{t}_{y,m}-t_y}{t_y}$, where $\hat{t}_{y,m}$ is the estimated value $\hat{t}_y$ for the m-th repetition. The mean square error is calculated by $ECME = \frac{1}{M}\sum_{m=1}^{M}(\hat{t}_{y,m}-t_y)^2$. Finally, if we define the confidence interval for the total of the variable of interest y under the normal approximation like $IC(m) : \hat{t}_{y,m} \pm z_{1-\alpha/2}\sqrt{Var(\hat{t}_{y,m})}$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution and if we denote the indicator variable $\gamma(m)$ as 1 if t_y belongs to $IC(m)$ and 0 otherwise, the coverage probability is calculated by $PC = 100\% M^{-1}\sum_{m=1}^{M}\gamma(m)$

Main Results

The most relevant results of the simulation study are summarized in Tables 2 - 4:

Table 2: Relative bias (RB x 10^{-3}) for the three estimators and the four sampling designs considered (scenarios 2 and 3).

Sampling Design	Sample size		FB	H	PML	FB	H	PML
Design A - Design B	n_A	n_B		Scenario 2			Scenario 3	
$srswor - srswor$								
	1000	117	-1.0	-1.0	-1.1	-5.1	-5.0	-5.0
	1000	160	-3.9	-3.9	-4.0	5.3	5.4	5.4
	1000	59	-3.9	-3.9	-4.2	-0.3	-0.2	-0.4
	1750	117	4.3	4.3	4.5	1.4	1.5	1.4
	1750	160	-4.0	-4.0	-3.9	2.6	2.7	2.7
	1750	59	0.6	0.6	0.7	-2.3	-2.2	-2.4
$srswor - \pi ps$								
	1000	117	-1.5	-1.5	-1.6	4.0	4.1	4.0
	1000	160	0.0	0.0	-0.2	2.0	2.1	2.0
	1000	59	0.8	0.8	0.5	12.6	12.7	12.3
	1750	117	-1.2	-1.2	-1.2	0.4	0.5	0.3
	1750	160	0.4	0.4	0.5	0.3	0.3	0.2
	1750	59	-2.6	-2.6	-2.5	-2.7	-2.6	-2.7
$\pi ps - srswor$								
	1000	117	-1.4	-1.4	-1.5	-2.5	-2.4	-2.5
	1000	160	-0.3	-0.3	-0.3	5.1	5.2	5.0
	1000	59	5.6	5.6	5.4	3.9	4.0	3.8
	1750	117	4.2	4.2	4.4	1.5	1.6	1.5
	1750	160	1.0	1.0	1.2	3.3	3.3	3.3
	1750	59	2.7	2.7	2.9	-0.8	-0.7	-0.9
$\pi ps - \pi ps$								
	1000	117	-0.9	-0.9	-0.9	-2.6	-2.4	-2.6
	1000	160	0.4	0.4	0.3	10.2	10.3	10.2
	1000	59	5.5	5.5	5.3	-4.8	-4.7	-4.9
	1750	117	3.4	3.4	3.5	6.5	6.6	6.3
	1750	160	3.5	3.5	3.6	-0.3	-0.3	-0.5
	1750	59	4.9	4.9	5.0	1.9	2.0	1.8

Table 3: Mean square error (MSE x 10^8) with the respective variance (s^2 x 10^8) for the three estimators and the four sampling designs considered (scenarios 2 and 3).

Sampling Design	**Sample size**		**FB**	**H**	**PML**	**FB**	**H**	**PML**
Design A - Design B	n_A	n_B		Scenario 2			Scenario 3	
$srswor - srswor$								
	1000	117	1.52(1.51)	1.52(1.51)	1.52(1.51)	1.51(1.5)	1.51(1.5)	1.51(1.5)
	1000	160	1.44(1.43)	1.44(1.43)	1.44(1.43)	1.57(1.56)	1.57(1.56)	1.57(1.56)
	1000	59	1.52(1.51)	1.52(1.51)	1.52(1.51)	1.45(1.44)	1.45(1.44)	1.45(1.44)
	1750	117	0.83(0.82)	0.83(0.82)	0.83(0.82)	0.79(0.78)	0.8(0.79)	0.8(0.79)
	1750	160	0.85(0.84)	0.85(0.84)	0.85(0.84)	0.89(0.88)	0.89(0.88)	0.89(0.88)
	1750	59	0.77(0.76)	0.77(0.76)	0.77(0.76)	0.83(0.82)	0.83(0.82)	0.84(0.83)
$srswor - \pi ps$								
	1000	117	1.52(1.51)	1.52(1.51)	1.52(1.51)	1.48(1.47)	1.49(1.48)	1.48(1.47)
	1000	160	1.4(1.39)	1.4(1.39)	1.4(1.39)	1.39(1.38)	1.39(1.38)	1.38(1.37)
	1000	59	1.41(1.4)	1.41(1.4)	1.41(1.4)	1.55(1.54)	1.55(1.54)	1.55(1.54)
	1750	117	0.85(0.84)	0.85(0.84)	0.85(0.84)	0.81(0.8)	0.81(0.8)	0.81(0.8)
	1750	160	0.8(0.79)	0.8(0.79)	0.8(0.79)	0.83(0.82)	0.83(0.82)	0.83(0.82)
	1750	59	0.86(0.85)	0.86(0.85)	0.86(0.85)	0.82(0.81)	0.82(0.81)	0.82(0.81)
$\pi ps - srswor$								
	1000	117	2.15(2.14)	2.15(2.14)	2.15(2.14)	2.28(2.27)	2.28(2.27)	2.28(2.27)
	1000	160	2.27(2.26)	2.27(2.26)	2.27(2.26)	2.32(2.31)	2.33(2.32)	2.32(2.31)
	1000	59	2.26(2.25)	2.26(2.25)	2.26(2.25)	2.1(2.09)	2.1(2.09)	2.1(2.09)
	1750	117	1.17(1.16)	1.17(1.16)	1.17(1.16)	1.23(1.22)	1.24(1.23)	1.23(1.22)
	1750	160	1.22(1.21)	1.22(1.21)	1.22(1.21)	1.23(1.22)	1.23(1.22)	1.23(1.22)
	1750	59	1.12(1.11)	1.12(1.11)	1.12(1.11)	1.32(1.31)	1.32(1.31)	1.32(1.31)
$\pi ps - \pi ps$								
	1000	117	2.34(2.33)	2.34(2.33)	2.34(2.33)	2.21(2.2)	2.22(2.21)	2.21(2.2)
	1000	160	2.07(2.06)	2.07(2.06)	2.06(2.05)	2.18(2.17)	2.18(2.17)	2.18(2.17)
	1000	59	2.38(2.37)	2.38(2.37)	2.38(2.37)	2.21(2.2)	2.21(2.2)	2.21(2.2)
	1750	117	1.16(1.15)	1.16(1.15)	1.16(1.15)	1.38(1.37)	1.38(1.37)	1.38(1.37)
	1750	160	1.25(1.24)	1.25(1.24)	1.25(1.24)	1.28(1.27)	1.28(1.27)	1.28(1.27)
	1750	59	1.17(1.16)	1.17(1.16)	1.17(1.16)	1.23(1.22)	1.23(1.22)	1.23(1.22)

Table 4: Relative bias of the variance (RBV x 10^2) and Coverage Probability (CP) for the three estimators and the four sampling designs considered (Scenarios 2 and 3).

Sampling Design	**Sample size**		**FB**	**H**	**PML**	**FB**	**H**	**PML**
Design A - Design B	n_A	n_B		Scenario 2			Scenario 3	
$srswor - srswor$								
	1000	117	-0.06 (0.93)	-0.06 (0.93)	0.15 (0.94)	1.13 (0.93)	1.11 (0.93)	0.57 (0.93)
	1000	160	4.47 (0.94)	4.48 (0.94)	4.34 (0.94)	-0.90 (0.94)	-0.87 (0.94)	-1.48 (0.93)
	1000	59	-0.35 (0.94)	-0.35 (0.94)	-0.18 (0.95)	5.89 (0.94)	5.88 (0.94)	4.98 (0.94)
	1750	117	5.67 (0.95)	5.67 (0.95)	5.61 (0.95)	10.62 (0.96)	10.56 (0.96)	9.85 (0.96)
	1750	160	1.60 (0.94)	1.60 (0.94)	1.61 (0.94)	-1.22 (0.94)	-1.23 (0.94)	-1.66 (0.94)
	1750	59	12.92 (0.95)	12.93 (0.95)	12.88 (0.95)	4.31 (0.94)	4.29 (0.94)	3.82 (0.94)
$srswor - \pi ps$								
	1000	117	-0.15 (0.93)	-0.15 (0.93)	-0.12 (0.93)	3.38 (0.95)	3.32 (0.95)	3.01 (0.95)
	1000	160	8.45 (0.95)	8.45 (0.95)	8.39 (0.95)	11.30 (0.96)	11.32 (0.96)	10.81 (0.96)
	1000	59	8.79 (0.95)	8.80 (0.95)	9.05 (0.95)	1.73 (0.96)	1.76 (0.96)	1.11 (0.96)
	1750	117	1.76 (0.95)	1.76 (0.95)	1.78 (0.95)	7.55 (0.94)	7.56 (0.94)	7.13 (0.95)
	1750	160	8.95 (0.95)	8.95 (0.95)	8.85 (0.95)	5.57 (0.95)	5.56 (0.95)	4.82 (0.95)
	1750	59	0.58 (0.94)	0.58 (0.94)	0.51 (0.94)	6.16 (0.95)	6.15 (0.95)	5.42 (0.95)
$\pi ps - srswor$								
	1000	117	2.40 (0.94)	2.40 (0.94)	2.62 (0.94)	-0.36 (0.93)	-0.33 (0.93)	-3.53 (0.93)
	1000	160	-2.84 (0.93)	-2.84 (0.93)	-2.59 (0.93)	-1.42 (0.93)	-1.46 (0.93)	-4.25 (0.93)
	1000	59	-2.49 (0.94)	-2.48 (0.94)	-2.10 (0.93)	7.90 (0.94)	7.97 (0.94)	4.55 (0.94)
	1750	117	9.37 (0.96)	9.37 (0.96)	9.45 (0.96)	5.41 (0.95)	5.39 (0.95)	2.22 (0.94)
	1750	160	5.71 (0.94)	5.71 (0.94)	5.60 (0.94)	5.63 (0.94)	5.61 (0.94)	2.53 (0.94)
	1750	59	13.09 (0.95)	13.09 (0.95)	13.27 (0.95)	-0.86 (0.93)	-0.88 (0.93)	-3.76 (0.93)
$\pi ps - \pi ps$								
	1000	117	-5.05 (0.93)	-5.05 (0.93)	-4.81 (0.93)	2.33 (0.93)	2.29 (0.93)	-0.68 (0.93)
	1000	160	7.94 (0.95)	7.95 (0.95)	8.32 (0.94)	8.29 (0.95)	8.27 (0.95)	4.50 (0.94)
	1000	59	-4.24 (0.92)	-4.24 (0.92)	-4.31 (0.92)	2.17 (0.91)	2.16 (0.91)	-1.02 (0.91)
	1750	117	9.70 (0.95)	9.70 (0.95)	9.65 (0.95)	-3.76 (0.92)	-3.75 (0.92)	-6.45 (0.92)
	1750	160	1.40 (0.95)	1.40 (0.95)	1.17 (0.95)	2.62 (0.93)	2.61 (0.93)	-0.52 (0.93)
	1750	59	9.81 (0.95)	9.82 (0.95)	9.90 (0.95)	6.17 (0.94)	6.13 (0.94)	2.98 (0.94)

Conclusions

The property of unbiasedness of the Hartley (1962) and the Fuller-Burmeister (1972) estimators was observed empirically. Also, for the Skinner and Rao (1996) estimator the relative bias for all the sampling designs considered were negligible confirming the asymptotic unbiasedness of the Pseudo Maximum Likelihood Estimator. Regarding the empirical Mean Square Error, this one decreases when the sample size of the first stage in the frame A increases, but not when the sample size in the frame B increases. Comparing scenarios 2 and 3, the empirical Mean Square Errors for the scenario 2 are smaller than the empirical Mean Square Errors for the scenario 3. The values for the relative bias of the variance are small, with the three estimators behaving similarly. The relative bias of the variance is smaller in the scenario 2 than in the scenario 3. Finally, the coverage probability is nearly the same for most of the considered sampling designs in both scenarios but these probabilities vary from 92% to 97%.

REFERENCES (RÉFERENCES)

Fuller, W. A., and Burmeister, L. F. (1972), "Estimators for samples selected from two overlapping frames", in ASA Proceedings of the Social Statistics Section, 245-249.

Hartley, H. O. (1962), "Multiple frame surveys", Proceedings of the Social Statistics Section, American Statistical Association, 203-206.

Lohr S. L., and Rao J. N. K. (2000) " Inference from dual frame surveys", Journal of the American Statistical Association, 95, 271-280.

Lohr S. L., and Rao J. N. K. (2006) " Estimation in multiple-frame surveys", Journal of the American Statistical Association, 101, 1019-1030.

Skinner, C. J., and Rao, J. N. K. (1996), "Estimation in dual frame surveys with complex designs", Journal of the American Statistical Association, 91, 349-356.

Fernandes, Hemilio. (2007). "A abordagem de cadastro duplo (dual frame); estimacao assistida por modelos lineares com aplicacao em pesquisas agropecuarias". (M. Sc. Thesis - Federal University of Pernambuco)