Instituto Nacional de Estadística

Working Papers

03/2011

Study of variance estimation methods in the Spanish Labour Force Survey (EPA)

Gerardo Azor Martínez

Juan V. Jiménez Llorente

Carlos Pérez Arriero

Juana Porras Puga

First draft:

This draft:

Study of variance estimation methods in the Spanish Labour Force Survey (EPA)

Abstract

The aim of this paper is to compare different methods for calculating sampling errors in the Spanish Labour Force Survey (EPA). Half sample replication (HSR) is the method currently employed to this end. We compare its results with those obtained with two other more recent techniques, standard delete-one jackknife and Rao-Wu-Yue bootstrap. The paper begins with a brief description of the EPA methodology, and goes on with a theoretical presentation of the above mentioned methods, followed by the coefficient of variation (CV) calculated for the estimates of the most important EPA variables in 2009.

Finally, we present a more detailed study for the autonomous community of Galicia. In this NUTS2 the sample has been enlarged in the third quarter of 2009, and this fact allows us to study the changes in the estimates of the variance, in relation to the change in sample size.

Keywords

Sampling errors, half sample replication, jackknife, bootstrap

Authors and Affiliations

Gerardo Azor Martínez, Juan V. Jiménez Llorente, Carlos Pérez Arriero y Juana Porras Puga

S. G. of Sampling and Data Collection

I. INTRODUCTION

I.1. Background

The Spanish Labour Force Survey (EPA) is a continuous quarterly survey addressed to the population living in family dwellings. The main objective of EPA is to reveal information on economic activities regarding to the human component. It has been carried out by the INE since 1964.

I.2 Sample Design

The sampling plan is a two stages stratified sampling:
In the first stage a stratified sample of geographical areas (PSUs- Census Sections-) is selected with probability proportional to size, measured by the number of dwellings. In the second stage, within selected PSUs, a sample of dwellings (SSUs) is selected with equal probability.
The sample size is about 65.000 dwellings, in each of them all people aged 16 and over are interviewed.
The sample weights required to obtain estimates are calculated through the following steps:

Weights based in sample design probabilities (Horvitz-Thompson).
Nonresponse adjustment.
Calibration to external sources.

I.3 Variance Estimation

Variance estimation is an essential indicator of the quality of estimates. Complex sample designs present particular challenges to estimate variance, due to the difficulties of an analytical approach. Increase in computing efficiency has made the use of resampling techniques feasible for large surveys. These methods are easy to implement because they always use the same estimation process repeated many times. The main idea consists of getting subsamples from the sample, compute the estimate for the full sample and for each subsample with identical procedures, and then combine these results in a simple variance expression. The variation among the subsamples estimates is used to calculate the variance for the full sample.
EPA utilizes a replication method based on methodology used for estimating variances for the Current Population Survey labour force estimates conducted by the U.S. Census Bureau and the U.S. Bureau of the Labour Statistics.

I.4 Countries Comparison: Different Ways of Variance Estimation

There is not an international agreement about the best methodology for estimating variance in household surveys, except the use of indirect techniques. The variance estimation procedures among countries is varied, different methods are in use (see Table below).

Country	*Method of estimation*
Canada	Jackknife variance estimation procedures
Germany	Taylor linearization
France	Direct estimation of the theoretical formulas
Italy	Variance of estimator GREG
Portugal	Jackknife technique
Poland	Bootstrap method with calibration of bootstrap weights
United Kingdom	Linearization methods
USA	Replication methods

(Sources: Eurostat, Statistics Canada, U.S. Census Bureau)

II. VARIANCE ESTIMATION METHODS

This section presents the methods for estimating the variance to be studied.

II.1 Half-Sample Replication (HSR) methods

The original idea was introduced by Mc Carthy in 1966 for the case of stratified sampling with two sampling units per stratum (Balanced Half Sample method - BHS). Different extensions of this method were developed for cases with more than two sampling units in any stratum (see Shao - Chen 1999).
In EPA, sampling errors are estimated with one of these extensions. The method consists of dividing in pairs at random the sample of PSUs in each stratum. Then a half sample is built taking at random one unit of each pair. After that, another half sample is built taking from each pair the unit that has not been chosen in the first step. So that, all the strata are represented in each replicate, and these first two subsamples contain the information from the whole sample.
The estimate of the variance in EPA is obtained using forty halfsamples obtained in such a way. In each of them the EPA general estimation has been employed.

The variance estimation for a estimator of the parameter θ is $\hat{V}(\hat{\theta}) = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\theta}_i - \hat{\theta}\right)^2$

where $\hat{\theta}$ is the estimate based on the full sample, $\hat{\theta}_i$ is the estimate based on the i-th replicate half-sample and r is the number of replicates.

II.2. Jackknife method

The jackknife method is first introduced by Quenouille (1949) as a method to estimate, and consequently to reduce, the bias of an estimator. It has become a more valuable tool since Tukey (1958) demonstrated that the jackknife can also be used to construct variance estimators.

For the use of the jackknife method in a stratified multi-stage sampling design, we form $n = \sum_{n} n_h$ replicates. Each replicate is formed by deleting one PSU.

The jackknife estimator of $V(\hat{\theta})$ is given by

$$\hat{V}_J(\hat{\theta}) = \sum_h \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{\theta}_{hj} - \hat{\theta})^2$$

where:

$\hat{\theta}_{hj}$ is the estimation of the total, using the replicate *hj*, what is formed deleting the j-th PSU within stratum h.

$\hat{\theta}$ is the estimation of the total using the whole sample.

The formula above is known as the **standard delete-one jackknife**.

Valliant (1993) showed theoretically and empirically that poststratification weights must be recomputed for every replicate in order to get a consistent jackknife estimation in two-stage sampling. Yung and Rao (1996,2000) obtained similar results for the jackknife method in stratified, multistage sampling. For this reason, and taking into account that calibrated EPA estimates are very close to poststratification method, in this paper the jackknife estimates has been calculated using the overall EPA estimation process in each replicate.

II.3. Bootstrap method

In 1979 B. Efron introduced bootstrap methods. On the one hand he tries to clarify the theoretical basis of the jackknife, "*an intriguing nonparametric method*". On the other he proposes a new technique for estimation problems, "*more widely applicable than the jackknife and also more dependable*".

Since then, original or modified bootstrap, has been used as the variance estimation method for many surveys, because it works properly and it is relatively easy to implement.

The simplest version of bootstrap is:

a) Draw a simple random sample with replacement from the original sample with the same size: $(x^*_1, \ldots x^*_n)$.
b) Compute the estimator from this resample in the same way that the one of the survey.
c) Repeat this routine M times, and so, obtain M estimates $\{\hat{\theta}_k\}_{k=1,\ldots,M}$
d) The estimation of the variance of θ is: $\hat{V}_{BT}(\hat{\theta}) = \frac{1}{M-1}\sum_{k=1}^{M}(\hat{\theta}_k - \hat{\theta})^2$

Rao and Wu (1988) proposed a bootstrap method for stratified multi-stage designs and with-replacement sampling of PSU's. The method applied a scale adjustment directly to the survey data values.

Rao, Wu and Yue (1992) presented a modification of the 1988 method, where the scale adjustment is applied to the survey weights rather than to the data values. Below is presented a description of this method.

Let θ denote a finite population parameter and $\hat{\theta}$ its estimator based on the full survey sample. To estimate the variance of $\hat{\theta}$, we repeatedly select bootstrap samples from the full survey sample and apply the procedure given below.

Let $\hat{\theta}^*$ denote a bootstrap replicate of $\hat{\theta}$.

For b=1,…,B, where B is large (typically, B=500 for Statistics Canada Surveys), repeat independently steps (i) to (iv):

(i) Independently in each stratum h, h=1,…,H, select a bootstrap sample by drawing a simple random sample of n_h^* PSUs with replacement from the sample of n_h PSUs. Let $t_{hi,b}^*$ be the number of times that PSU *hi* is selected in the bootstrap sample b $\left(\sum_{i=1}^{n_h} t_{hi,b}^* = n_h^*\right)$.

(ii) For each secondary sampling unit (SSU) *k* in PSU *hi*, calculate the initial bootstrap weight $d_{hik,b}^* = d_{hik}\left\{\left(1-\sqrt{\frac{n_h^*}{n_h-1}}\right)+\sqrt{\frac{n_h^*}{n_h-1}}\cdot\frac{n_h}{n_h^*}\cdot t_{hi,b}^*\right\}$, where d_{hik} is the initial sampling weight of the SSU *hik*, equal to the inverse of its selection probability, i.e. $d_{hik} = 1/\pi_{hik}$.

(iii) For each SSU, the final bootstrap weight $w_{hik,b}^*$ is calculated by applying, to the initial bootstrap weight $d_{hik,b}^*$, the same adjustment procedures (re-weighting for non-response and calibration) that were applied to the initial sampling weight $d_{hik,b}$ to obtain the final survey weight w_{hik} .

(iv) Calculate $\hat{\theta}_b^*$, the b-th bootstrap replicate of $\hat{\theta}$, by replacing w_{hik} with $w_{hik,b}^*$ in the formula for $\hat{\theta}$.

The bootstrap variance estimator of $\hat{\theta}$ is given by $\hat{V}_{BS}(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}\right)^2$. If $n_h^* \le n_h - 1$ then the bootstrap weights are never negative. Usually $n_h^* = n_h - 1$, which simplifies bootstrap weights given above to $d_{hik,b}^* = d_{hik}\left\{\frac{n_h}{n_h^*}\cdot t_{hi,b}^*\right\}$.

III. RESULTS

To compare the above methods, variance estimates were obtained for estimates of major EPA target variables in each of the four quarters of 2009.

Tables A1, A2 y A3, report variance estimation for unemployment, employment and inactive population, respectively, at national and NUTS2 level.

In this group of tables it can be seen a general similarity between the three types of estimates. Appart from that, the following considerations may be taken into account:

- In a few cases there are some non negligible differences. As expected, the largest absolute differences are presented with the smallest proportion, the unemployed people in Table A1.
- At national level all estimates are very close in every quarter.
- Jackknife and bootstrap are generally closer to each other than those obtained by half sample replication, particularly in Tables A2 and A3.
- Ceuta and Melilla are geographic areas much smaller than the rest of NUTS2. For this reason its results should be viewed with some caution.

TABLE A1. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF UNEMPLOYED POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE AND BOOTSTRAP METHODS, AT NATIONAL AND NUTS2 LEVEL

UNEMPLOYED

	QUARTER											
	1/2009			2/2009			3/2009			4/2009		
REGION	H-S	J	B	H-S	J	B	H-S	J	B	H-S	J	B
National	1,25	1,27	1,26	1,17	1,21	1,21	1,12	1,22	1,23	1,17	1,23	1,21
Andalucía	2,36	2,27	2,26	1,83	2,16	2,16	1,83	2,33	2,48	2,23	2,39	2,49
Aragón	6,24	6,23	5,99	7,62	6,72	6,75	8,06	6,97	7,08	6,77	6,33	5,99
Asturias, Princ. de	8,72	6,83	6,78	7,74	6,64	6,41	7,13	5,90	5,94	6,60	6,05	6,10
Balears, Illes	5,30	6,11	5,98	7,80	5,81	5,65	6,51	6,66	6,53	6,79	5,49	5,40
Canarias	3,85	4,22	4,36	5,45	4,71	4,55	4,70	4,67	4,53	5,40	4,53	4,48
Cantabria	6,63	7,39	7,45	8,64	7,02	6,95	8,98	8,98	9,00	8,83	7,93	7,41
Castilla y León	2,95	3,78	3,81	4,64	4,45	4,33	3,70	3,77	3,76	3,70	3,47	3,54
Castilla - La Mancha	4,06	4,05	3,94	4,37	3,75	3,64	3,36	3,88	3,71	4,17	4,04	3,93
Cataluña	3,20	3,64	3,62	2,89	3,58	3,59	3,01	3,67	3,65	2,99	3,12	3,16
Comunitat Valenciana	4,19	4,35	4,24	3,76	4,03	3,96	3,02	3,66	3,73	3,57	4,15	4,01
Extremadura	5,66	5,32	5,35	5,57	5,34	5,09	6,23	5,16	4,90	4,67	5,31	5,23
Galicia	4,56	4,57	4,40	4,87	4,48	4,34	4,03	3,02	3,16	3,71	3,25	3,03
Madrid, Comun. de	6,90	5,45	5,44	6,59	5,00	4,99	5,28	4,92	4,92	5,67	5,39	5,40
Murcia, Región de	6,37	5,87	5,69	4,55	5,22	5,39	5,08	5,87	5,54	4,86	5,09	4,79
Navarra, Comun. F. de	9,80	8,49	8,30	8,81	7,89	7,76	9,25	8,07	7,77	10,16	8,32	8,59
País Vasco	5,47	5,95	5,85	5,22	5,70	5,81	5,56	5,21	5,30	5,98	5,35	5,40
Rioja, La	9,01	10,20	9,37	9,27	9,95	9,41	8,29	7,95	8,11	9,39	9,36	9,06
Ceuta	16,53	23,04	21,56	18,83	21,48	20,96	15,13	18,02	17,22	17,62	16,24	16,67
Melilla	23,18	28,33	24,91	17,62	16,87	16,83	14,25	19,02	18,86	16,53	15,28	15,79

TABLE A2. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF EMPLOYED POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE AND BOOTSTRAP METHODS, AT NATIONAL AND NUTS2 LEVEL

EMPLOYED

	QUARTER											
	1/2009			2/2009			3/2009			4/2009		
REGION	H-S	J	B	H-S	J	B	H-S	J	B	H-S	J	B
National	0,38	0,34	0,33	0,40	0,33	0,34	0,32	0,34	0,33	0,36	0,35	0,35
Andalucía	1,00	0,90	0,90	0,85	0,90	0,86	0,84	0,96	0,98	0,81	0,97	0,95
Aragón	1,45	1,34	1,28	1,42	1,41	1,44	1,22	1,30	1,32	1,46	1,37	1,31
Asturias, Princ. de	1,46	1,60	1,60	1,54	1,76	1,73	1,53	1,57	1,57	1,64	1,68	1,67
Balears, Illes	1,52	1,92	1,75	1,55	1,81	1,80	1,52	1,62	1,65	2,24	1,80	1,71
Canarias	1,36	1,74	1,70	1,94	1,79	1,65	1,60	1,69	1,66	1,68	1,71	1,68
Cantabria	1,47	1,60	1,61	1,61	1,46	1,38	2,15	1,83	1,81	1,64	1,60	1,55
Castilla y León	1,16	0,97	0,95	1,17	0,98	0,95	0,97	0,91	0,93	1,20	0,99	0,99
Castilla - La Mancha	1,19	1,09	1,10	1,26	1,15	1,12	1,04	1,09	1,08	1,09	1,11	1,08
Cataluña	0,96	0,87	0,87	0,91	0,88	0,88	0,88	0,91	0,97	0,84	0,90	0,93
Comunitat Valenciana	1,00	1,08	1,10	1,29	1,24	1,30	1,15	1,33	1,32	1,13	1,38	1,31
Extremadura	2,14	1,84	1,78	1,56	1,72	1,66	2,09	1,90	1,82	2,03	1,67	1,58
Galicia	0,87	1,07	1,06	0,95	1,08	1,07	0,80	0,76	0,76	0,85	0,79	0,76
Madrid, Comun. de	1,45	1,15	1,15	1,43	1,02	1,05	1,07	0,99	1,00	1,20	1,12	1,14
Murcia, Región de	1,82	1,73	1,70	1,54	1,47	1,48	1,86	1,83	1,85	1,94	1,85	1,79
Navarra, Comun. F. de	2,14	1,85	1,76	1,92	1,62	1,59	1,63	1,29	1,33	1,80	1,36	1,46
País Vasco	0,99	1,09	1,08	1,04	1,02	1,02	1,18	1,08	1,10	1,29	1,07	1,06
Rioja, La	1,49	1,89	1,84	1,50	1,58	1,58	1,62	1,86	1,84	1,92	1,96	1,95
Ceuta	7,24	6,19	6,00	6,57	5,74	5,60	8,15	7,04	6,38	9,02	8,55	8,42
Melilla	9,20	10,21	9,62	7,30	8,36	8,75	8,27	8,28	8,58	10,87	9,65	9,34

TABLE A3. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF INACTIVE POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE AND BOOTSTRAP METHODS, AT NATIONAL AND NUTS2 LEVEL

INACTIVE POPULATION

	QUARTER											
	1/2009			2/2009			3/2009			4/2009		
REGION	H-S	J	B	H-S	J	B	H-S	J	B	H-S	J	B
National	0,38	0,32	0,33	0,37	0,32	0,33	0,35	0,31	0,31	0,35	0,32	0,32
Andalucía	0,93	0,79	0,76	0,87	0,76	0,73	0,70	0,72	0,70	0,70	0,72	0,70
Aragón	1,57	1,16	1,17	1,56	1,44	1,39	1,17	1,10	1,13	1,10	1,18	1,20
Asturias, Princ. de	1,13	1,42	1,39	1,02	1,39	1,42	1,11	1,38	1,40	1,36	1,36	1,31
Balears, Illes	1,99	2,08	1,95	2,33	2,17	2,19	2,50	2,32	2,29	1,92	1,76	1,71
Canarias	1,18	1,53	1,53	1,40	1,48	1,45	1,53	1,61	1,58	1,45	1,53	1,51
Cantabria	1,46	1,48	1,49	1,28	1,42	1,45	1,83	1,55	1,51	1,06	1,34	1,28
Castilla y León	0,91	0,82	0,81	0,78	0,78	0,79	0,85	0,81	0,81	0,79	0,85	0,84
Castilla - La Mancha	1,25	0,93	0,94	1,00	0,96	0,96	1,08	0,89	0,90	1,13	0,94	0,90
Cataluña	1,16	0,94	0,91	0,99	0,96	0,99	1,08	1,00	1,01	1,02	0,96	0,99
Comunitat Valenciana	1,09	1,25	1,27	1,06	1,21	1,20	1,31	1,19	1,20	0,95	1,07	1,06
Extremadura	1,38	1,27	1,26	1,19	1,23	1,21	1,49	1,34	1,28	1,31	1,39	1,34
Galicia	0,90	1,04	1,03	0,89	1,00	0,97	0,61	0,73	0,70	0,66	0,73	0,72
Madrid, Comun. de	1,43	1,22	1,21	1,48	1,28	1,28	0,98	1,17	1,15	1,39	1,37	1,35
Murcia, Región de	1,34	1,63	1,61	1,10	1,57	1,55	1,52	1,64	1,62	1,34	1,58	1,54
Navarra, Comun. F. de	1,80	1,82	1,72	1,50	1,72	1,68	1,57	1,60	1,68	1,53	1,45	1,62
País Vasco	0,84	1,11	1,10	0,85	1,04	1,04	1,03	1,12	1,10	1,30	1,12	1,15
Rioja, La	1,78	2,12	2,12	1,90	1,92	1,92	1,54	1,86	1,84	2,25	1,97	1,93
Ceuta	7,26	6,18	6,17	6,54	5,12	4,94	6,83	5,58	5,18	6,53	6,24	6,17
Melilla	6,41	5,48	5,55	5,41	5,19	5,72	8,60	7,87	7,69	9,56	7,91	7,50

Tables B1, B2 y B3, report variance estimation for unemployment, employment and inactive population in the autonomous community of Galicia (NUTS2), and in its provinces (NUTS3) for the four quarters of 2009.

The same general considerations as in the Tables-A can be done with Tables-B:

- The three methods have similar overall results
- Estimates obtained by jackknife and bootstrap are similar
- Major differences in estimates of small proportions

TABLE B1. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF UNEMPLOYED POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE, BOOTSTRAP AND SIMPLE RANDOM SAMPLING (SRS) METHODS, IN GALICIA (NUTS2) AND ITS PROVINCES (NUTS3)

UNEMPLOYED

PROVINCE	QUARTER 1/2009				2/2009				3/2009				4/2009			
	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS
Galicia	4,56	4,57	4,40	3,76	4,87	4,48	4,34	3,67	4,03	3,02	3,16	2,70	3,71	3,25	3,03	2,65
A Coruña	9,42	8,28	8,09	6,64	8,25	7,45	7,41	6,41	7,37	5,47	5,65	4,78	6,66	5,76	5,98	4,73
Lugo	14,64	15,19	14,36	11,75	22,86	18,70	17,71	11,36	9,97	10,55	10,84	8,56	10,43	10,71	10,62	7,68
Ourense	21,73	17,42	18,08	12,73	18,88	18,74	18,48	12,62	10,38	10,37	10,14	8,07	9,25	10,13	9,90	7,57
Pontevedra	6,82	6,78	6,53	5,42	6,64	6,41	6,24	5,34	4,45	4,24	4,12	4,00	4,64	4,66	4,82	4,01

TABLE B2. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF EMPLOYED POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE, BOOTSTRAP AND SIMPLE RANDOM SAMPLING (SRS) METHODS, IN GALICIA (NUTS2) AND ITS PROVINCES (NUTS3)

EMPLOYED

PROVINCE	QUARTER 1/2009				2/2009				3/2009				4/2009			
	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS
Galicia	0,87	1,07	1,06	1,04	0,95	1,08	1,07	1,06	0,80	0,76	0,76	0,76	0,85	0,79	0,76	0,76
A Coruña	2,11	2,50	2,49	1,72	2,29	2,48	2,29	1,73	1,97	1,71	1,62	1,24	1,81	1,80	1,81	1,25
Lugo	4,62	4,24	4,08	2,78	4,22	3,79	3,79	2,87	2,69	2,66	2,53	1,92	3,15	2,67	2,54	1,99
Ourense	5,93	5,91	5,36	2,87	5,75	5,31	4,86	3,05	3,72	3,50	3,35	2,30	3,73	3,14	2,98	2,27
Pontevedra	1,81	2,29	2,27	1,74	1,76	2,15	2,10	1,76	1,17	1,57	1,59	1,27	1,38	1,61	1,64	1,29

TABLE B3. COMPARISON OF THE COEFFICIENTS OF VARIATION (IN %) OF INACTIVE POPULATION, ACCORDING TO EPA 2009, CALCULATED BY HALF-SAMPLE REPLICATION, JACKKNIFE, BOOTSTRAP AND SIMPLE RANDOM SAMPLING (SRS) METHODS, IN GALICIA (NUTS2) AND ITS PROVINCES (NUTS3)

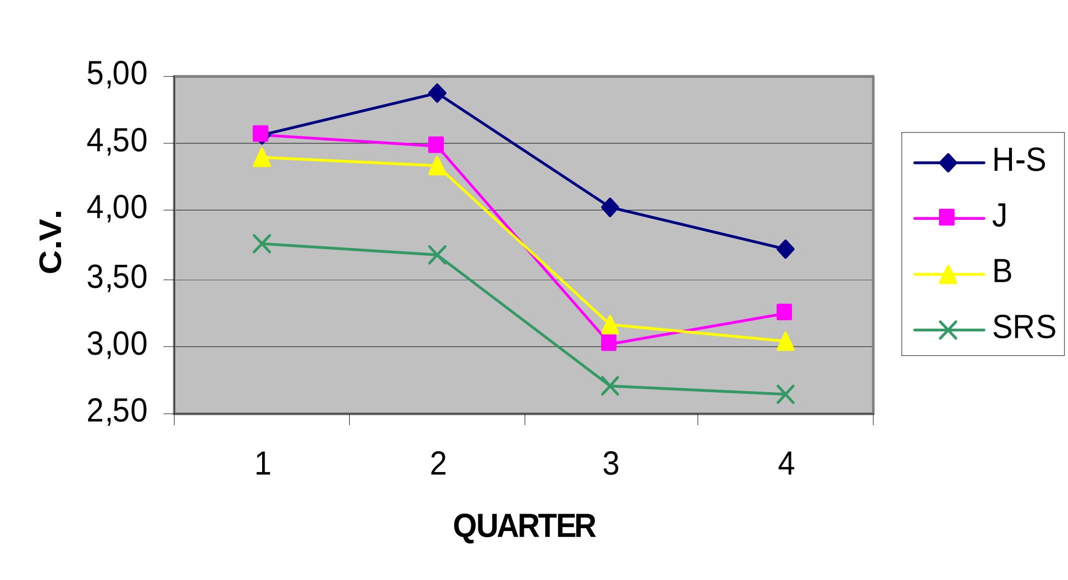
INACTIVE POPULATION

PROVINCE	QUARTER 1/2009				2/2009				3/2009				4/2009			
	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS	H-S	J	B	SRS
Galicia	0,90	1,04	1,03	1,14	0,89	1,00	0,97	1,13	0,61	0,73	0,70	0,81	0,66	0,73	0,72	0,80
A Coruña	2,92	3,19	3,16	2,08	2,81	2,91	2,73	2,07	1,89	1,96	1,90	1,45	1,78	1,97	2,00	1,46
Lugo	4,66	4,34	4,11	2,50	4,15	3,71	3,63	2,45	2,81	2,48	2,33	1,79	2,89	2,55	2,45	1,76
Ourense	5,76	5,53	5,03	2,59	5,18	4,81	4,44	2,51	2,77	2,63	2,44	1,74	2,99	2,52	2,44	1,78
Pontevedra	2,32	2,64	2,67	1,98	2,18	2,51	2,50	1,98	1,26	1,83	1,83	1,45	1,38	1,80	1,85	1,42

As stated earlier, the sample of EPA in Galicia has doubled its size in the third quarter of 2009. It is therefore of particular interest to study the evolution of the estimates of sampling errors this year. To this end, figures F1, F2 y F3 have been made with data from this group of tables. In these figures we have added the coefficient of variation calculated as if a simple random sampling (SRS) had been used, just to have a reference.

The first consideration about these figures is, again, the proximity between jackknife and bootstrap estimates, what are represented by pink and yellow lines respectively.
A second point to take into account is the evolution on time marked by the green line of reference (SRS). Jackknife and bootstrap lines (pink and yellow) seem to follow closely the SRS evolution. Although it must be kept in mind that, this behavior varies depending on the variable being analyzed.

FIGURE F1. UNEMPLOYED GALICIA EPA 2009



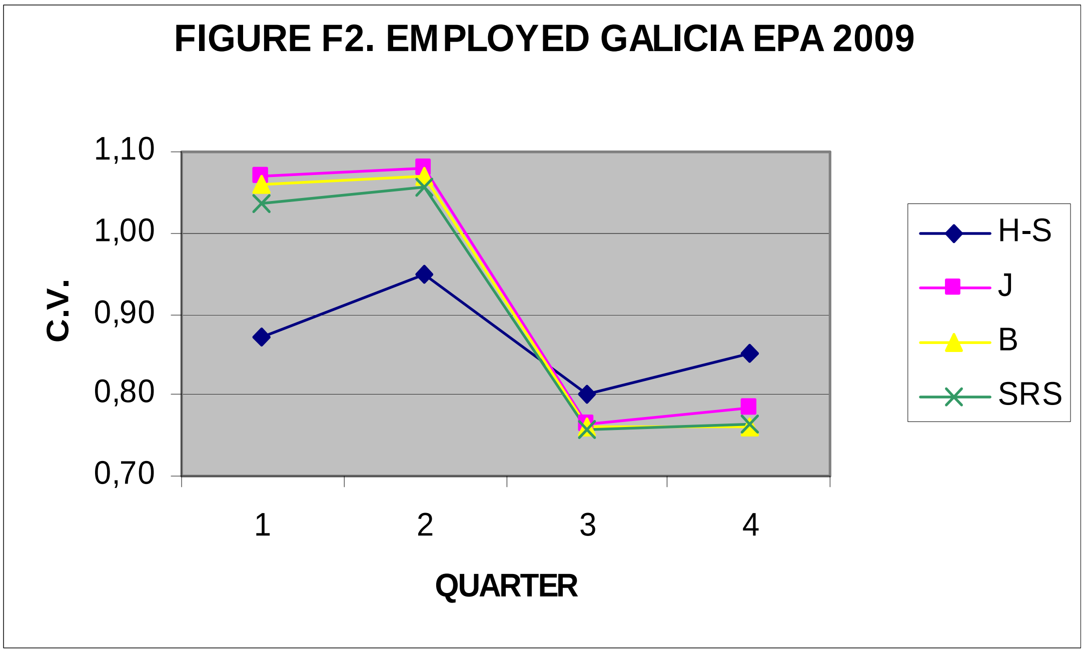


FIGURE F2. EMPLOYED GALICIA EPA 2009
C.V.
1,10
1,00
0,90
0,80
0,70
1
2
3
4
QUARTER
H-S
J
B
SRS

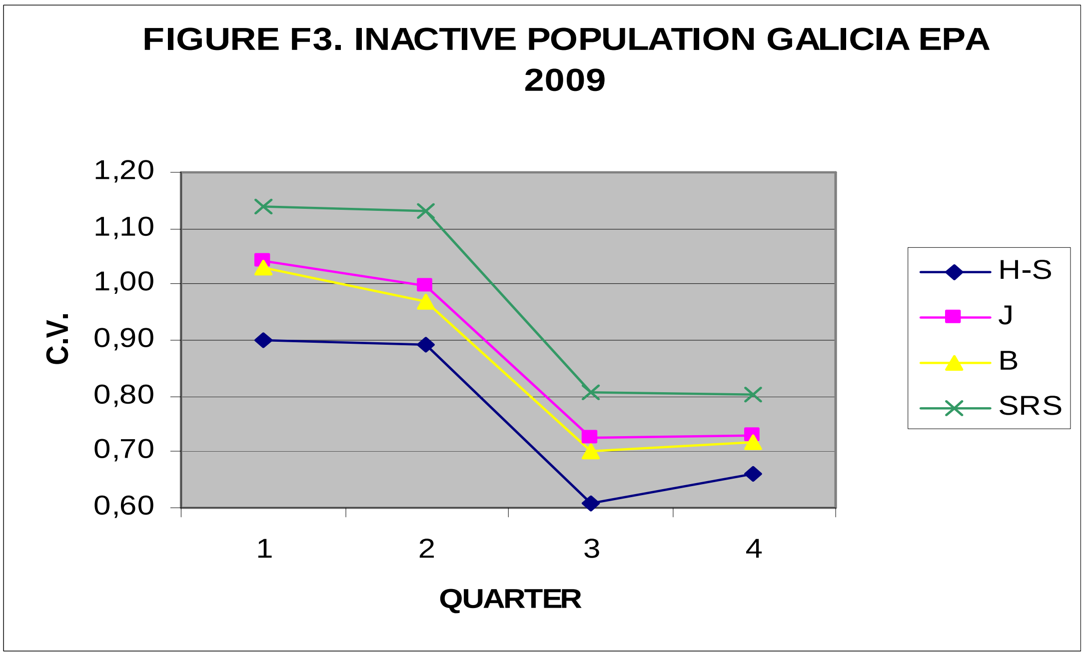


FIGURE F3. INACTIVE POPULATION GALICIA EPA 2009
C.V.
1,20
1,10
1,00
0,90
0,80
0,70
0,60
1
2
3
4
QUARTER
H-S
J
B
SRS

IV. SUMMARY

We have conducted a comparative study of sampling errors estimates in EPA, using half sample replication, jackknife and bootstrap methods.
Half sample replication, what is the current method for estimating variance in EPA, provides consistent results with the breakdown of estimates, and close to those provided by the other two methods.
Jackknife and bootstrap could be a good alternative to calculate sampling errors in EPA, because their estimates seem to be also consistent and they describe a little better changes in the sample size.
More in-depth studies are needed before incorporating changes in methodology of EPA sampling errors estimation, in order to confirm the quality of possible new CV estimates. These studies should also include a proper way to apply these new techniques in the calculation of sampling error of annual estimates and estimates of change.

V. References

Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *The Annals of Statistics*, 7, 1-26.

Mach, L., Dumais, J. and Robidou, L. (2005). A study of the properties of a bootstrap variance estimator under sampling without replacement. *Proceedings of the FCSM Research Conference.*

McCarthy, P. J. (1966). Replication: An approach to the analysis of data from complex surveys. *Vital and health statistics,* Series 2 No.14.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Rao, J. N. K., Wu, C.F.J., and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology*, 18, 209-217.

Särndal C. E., Swensson, B. and Wretman. J. (1992). *Model assisted survey sampling*. Springer-Verlang. New York, Inc.

Shao, J. and Chen Y. (1999). Approximate Balanced Half Sample and related Replication Methods for imputed survey data. *The Indian Journal of Statistics*, 61, 187-201.

Shao, J. and Tu, D. (1996). *The Jackknife and Bootstrap*. Springer series in statistics. Springer-Verlag, New York.

Valliant, R., Brick, J. M., and Dever, J. A. (2008). Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Official Statistics*, 24, 469-488.