

USING SAMPLING TECHNIQUES TO ESTIMATE THE DE JURE POPULATION OF GREECE AND APPORTION PARLIAMENT SEATS AMONG ELECTORAL DISTRICTS

Vasilis Michalacopoulos and Evdokia Kekalaki

Abstract—During the last ten years there has been much heated debate in academic journals, the popular press, and governments about the use of sampling methods to estimate the de jure population for each prefecture or district of a country, on the basis of which the allocation of parliament seats is made or/and national funds are distributed. The allocation of parliament seats was the original purpose for which decennial population censuses were established for most of the democratic countries. However, given that population censuses are carried out every ten years, it is obvious that many of the national election results are based on data that do not represent the real status of the population distribution at the time of the election. The aim of this paper is to explore the possibility of utilizing estimates of the size of the de jure population of a country for such purposes. An estimation procedure is proposed and studied. The implications of using the resulting estimates on the allocation of parliament seats are examined in the case of two national elections that took place in Greece a little while before or following the 1991 census in the years 1989, and 1993. In particular, using stratified proportional allocation, several samples of size 10000 are taken from the 1991 Population Census Data Base of the National Statistical Service of Greece. The use of ARIMA modelling is then considered in estimating the size of the de facto population for each electoral district for a specific year between two successive decennial censuses and determining the respective size of the de jure population. The apportionment of the Hellenic Parliament seats corresponding to the de jure population of each electoral district, as estimated by the sample for the year 1991, is compared to the official apportionments that the above elections resulted in on the basis of the complete census results considered in each case.

Keywords—population census; sampling from population census data; de jure population; de facto population; elections.

1. Introduction

The number of persons legally registered in the Rolls of each electoral district of a country (legal or de jure population), whether or not they are physically present on the census day, is the measure according to which the Parliament seats are apportioned among all the electoral districts. This measure is also the main criterion for the funding of districts by the central administration. As it is well known, population counts are taken every ten years. So, worldwide, the official apportionment of Parliament seats resulting from any national elections in the decennial interval between two successive censuses is made on the basis of the latest census. In Greece, for example, by law, national

elections take place every four years, but of course, the times between two consecutive elections are often shorter (about three years). A question, therefore arises as to the degree to which the results of the census considered are representative of the actual size of the de jure population, particularly when the time of the election is quite distant from the census year. The 1993 elections in Greece constitute a rather extreme such case as the apportionment of Parliament seats was based on the results of the 1981 census due to processing problems of the 1991 census data. Therefore, exploring the possibility of utilizing estimates of the de jure population for such purposes would be interesting and has been the topic of much heated debate in the last decade, the prime concern being on the magnitude of sampling errors in small geographical regions and on the possibility of data manipulation for political gains.

Of course, even in the case of censuses, the degree of accuracy is of equal concern. As [3] and [7] have pointed out, "*censuses, if done correctly, have the potential advantage of providing precise, detailed and credible information on all population units. On the other hand, samples have the advantage of providing richer, more complex, accurate, inexpensive and timely information for a smaller portion of the population units. By making use of sampling methods in a census context, it is hoped that we can realize the advantages of both methodologies*". Obtaining data on census items, however, has only been used to a limited extent in cases where the facilities required for a complete enumeration do not exist or as an auxiliary means or quality check to improve the final count, e.g. in estimating non-response and for increasing the overall coverage by reducing the differential net undercount (i.e., the difference between the counted population and the best estimate of the actual population). The use of a sampling method for obtaining up-to-date estimates for census items for their use in connection with various governmental decisions (e.g. allocation of national funds) or in the allocation of Parliament seats following an election appears not to have been considered. It is a rather sensitive issue that is only reluctantly being thought of. It would be interesting to note at this point that, although recognizing the advantages of sampling, the US Congress, expressed concern about the constitutionality of sampling, the possibility that the use of statistical methods would allow the data to be manipulated for political advantage and, finally, the magnitude of

Department of Statistics, Athens University of Economics and Business. E-mail for correspondence: basmix@infocae.gr, exek@aub.gr

The same procedure was implemented in the case of the 1981 Population Census Data Base, from which eight (8) samples were extracted yielding estimates that were also very close to the actual values with a mean absolute difference of about 1%.

The implementation of the sampling technique described above presupposes knowledge of the *de facto* population for each district for the particular year. To recover lack of this type of information, an electronic data base of the mid-year *de facto* population maintained by the Greek Statistical Bureau on a yearly basis was used. This data base is yearly updated and contains information on the mid-year real population of Greece, from the year 1981 through to the year immediately preceding the year of the update. (So, currently, the data base covers the years 1981 through to 2000). This one year delay in the updating of the data base by the Greek Statistical Bureau, is due to the fact that the mid-year real population is calculated, for any particular year, as the average of the population sizes on the first and the last day of the year. So, for the year 1999, for example, the corresponding population size is calculated as the average of the sizes of the real population on 1/1/1999 and on 31/12/1999. To overcome this shortcoming the size of the *de facto* population for each department was estimated using ARIMA modeling [?]. The Box-Pierce test for excessive autocorrelations was used for testing the residuals. For each district, the criterion for the selection of the appropriate model was the smallest mean square error (MSE) during the estimation period.

The estimates thus obtained turn out to be very close to those provided by the mid-year real population data base. As an illustration, consider Table 2 where estimates of the size of the *de facto* population obtained by appropriately chosen ARIMA models for the census year 1991 (column labeled F1991) were compared to the corresponding values provided by the data base (column labeled R1991) for some of the 55 electoral districts of Greece. The column labeled C1991 represents the percentages of the *de facto* population as taken from the official census and the columns labeled P2000 and P2001 represent the projection values for the year 2000 and the census year 2001, respectively.

The formula used for calculating the sample size for each of the 55 districts, according to the stratified proportional allocation method was $n_i = n \times N_i/N$, where

n = sample size (= 10000 persons from the *de facto* population),

n_i = size of the subsample associated with the i^{th} district,

N = *de facto* population of Greece in the year on which estimation is required,

N_i = estimated size of the *de facto* population of the i^{th} district. $i = 1, 2, \dots, 55$.

The percentage N_i/N for each district was estimated through the use of an appropriate ARIMA models.

After the estimation of the sample sizes n_i , $i = 1, 2, \dots, 55$ for all the districts, each of n_i individuals (persons) from district i , located by the enumerators on the basis of a

Code	District	F1991	R1991	C1991	P2000	P2001
13	Ahaia	0.00000	0.00000	0.00000	0.00000	0.00000
94	Hania	0.00000	0.00000	0.00000	0.00000	0.00000
85	Hios	0.00000	0.00000	0.00000	0.00000	0.00000
71	Evros	0.00000	0.00000	0.00000	0.00000	0.00000
63	Florina	0.00000	0.00000	0.00000	0.00000	0.00000
14	Ilia	0.00000	0.00000	0.00000	0.00000	0.00000
91	Iraklio	0.00000	0.00000	0.00000	0.00000	0.00000
41	Karditsa	0.00000	0.00000	0.00000	0.00000	0.00000
23	Kefalonia	0.00000	0.00000	0.00000	0.00000	0.00000
22	Kerkira	0.00000	0.00000	0.00000	0.00000	0.00000
15	Korinthia	0.00000	0.00000	0.00000	0.00000	0.00000
59	Pella	0.00000	0.00000	0.00000	0.00000	0.00000
93	Rethymno	0.00000	0.00000	0.00000	0.00000	0.00000
03	Voiotia	0.00000	0.00000	0.00000	0.00000	0.00000

TABLE II
PREDICTED SIZES OF THE DE FACTO POPULATION WITH THE USE OF
ARIMA MODELS

simple random sample is asked to specify the district of the country in which they are legally registered in. The possible different answers correspond to the 55 different departments of Greece. On the basis of all the above results, the sample data obtained can then be tabulated in the form of Table 1 and the percentages and the estimated values of the *de jure* population of each district can be estimated. The estimates obtained can then be extrapolated to provide the *de jure* population of each district for the country as a whole. These values are subsequently used for the apportionment of Parliament seats.

Note that using the estimates of the *de facto* population derived from the ARIMA models considered for the year 1991 to determine the population percentages per district relative to the total sample size, leads to results strikingly close to those derived from the official values of *de facto* population for the census year 1991.

III. Apportioning Parliament Seats on the Basis of the *de Jure* Population Estimates: The Case of the 1989 and 1993 Greek Elections

The Greek Parliament consists of 300 seats of which 288 are directly allocated according to the *de jure* population sizes of the 55 districts. There is a first phase of apportioning of the seats by dividing the size of each district by the "electoral measure" which is the quotient derived from the total population of Greece divided by the fixed number 288. In the second phase, the remaining seats are allocated on the basis of priority scores (biggest remainder) until all seats have been apportioned.

Years 1993 and 1989 were both election years for the Greek Parliament. Due to problems related to the processing of the 1991 census data, the apportionment of Parliament seats of the election year 1993 was based on the results of the previous census year 1981. Table 3 shows the official apportionment of the Hellenic Parliament seats on the basis of the 1981 census (column 1) as well as the offi-

cial apportionment of the Hellenic Parliament seats on the basis of the 1991 census announced by the Greek Ministry of Interior and Public Administration (column 2). Between these two official apportionments there is a displacement of $20/2 = 10$ seats in a total of 288 (Column 3).

This finding underlines the necessity for considering the use of appropriate estimates of the district population sizes based on more up-to-date estimates. Such estimates can only be provided through samples of the population taken at or near the time of the election. With an appropriately designed sampling scheme that would contain the probability of sampling errors, these estimates would more accurately reflect the actual population distribution. To illustrate this, the 1991 census is regarded in the sequel, as being an accurate representation of the 1993 population and treated as the "true 1993 population" of Greece from which an appropriate sample will provide up-to-date estimates for the apportionment of Parliament seats.

Indeed, using the sampling scheme described in section 2, a sample of size 10000 was taken from the 1991 census data. The population estimates obtained were subsequently used for the apportionment of Parliament seats for the 1993 elections. The apportionment obtained is presented in Table 4. As can be verified from this table, a displacement of 8 seats results in on comparing the sample based apportionment of Parliaments seats to the 1991 census based official one.

This is caused mainly by the difference in the ways in which the Ministry of Interior and the National Statistical Service classify the population. The largest displacement appears in the area of Attica which, for the purposes of the Ministry is divided into five regions, but for the purposes of the Greek Statistical Service, into only two districts. In order to smoothen the consequences of this problem and considering the fact that the existent data set cannot be adjusted to counter-balance that difference, two new samples of size 50000 were formed by merging firstly the samples 1 to 5 (Grouped Sample 1-5) and secondly the samples 6 to 10 (Group Sampled 6-10). Actually, the result was a displacement of only one (1) seat for Grouped Sample 1-5 and of two (2) seats for Grouped Sample 6-10. (see Tables 5 and 6).

A comparison of the estimates taken from Grouped Sample 1-5 to the official apportionment of census year 1981 shows a displacement of 8 seats. This difference is large compared to the corresponding difference of 1 seat in the apportionment based on the 1991 census. This is precisely in support of the usefulness of up-to-date sample based population estimates.

To demonstrate the implications that the use of census data has on the apportionment of Parliament seats for elections that take place towards the end of the decade and quite close to the next census, the sample taken from the 1991 census data was also considered for the apportionment of the Parliament seats for the 1989 election.

Comparisons similar to those made for the 1993 election based on Grouped Sample 6-10 yield the same difference

District	Off.1981	Off.1991	Dis/ment
ATTIKH I	21	19	2
ATTIKH II	32	38	6
ATTIKH III	7	9	2
PIREAS I	8	7	1
PIREAS II	8	8	0
THES/NIKI I	13	14	1
THES/NIKI II	7	7	0
XALKIDIKI	3	3	0
KERKIRA	3	3	0
LAKONIA	3	3	0
KILKIS	3	3	0
LESVOS	4	3	1
SERRES	8	7	1
ARTA	3	3	0
KIKLADES	3	3	0
RODOPI	3	3	0
IRAKLEIO	7	8	1
EVROS	5	4	1
IMATHIA	4	4	0
HANIA	4	4	0
KEFALONIA	1	1	0
DODEKANISA	4	4	0
ZAKINTHOS	1	1	0
KORINTHIA	4	4	0
IOANNINA	6	5	1
KAVALA	4	4	0
LASITHI	2	2	0
PREVEZA	2	2	0
RETHIMNO	2	2	0
ILIA	6	6	0
ETOL/NIA	8	8	0
GREVENA	2	1	1
KASTORIA	2	2	0
SAMOS	1	1	0
MAGNHSIA	6	6	0
PELLA	5	5	0
FOKIDA	2	2	0
ARKADIA	4	4	0
DRAMA	4	3	1
AHAIA	9	9	0
FLORINA	2	2	0
PIERIA	4	4	0
ARGOLIDA	3	3	0
THESPROTIA	2	2	0
HIOS	2	2	0
XANTHI	3	3	0
LEYKADA	1	1	0
EYRYTANIA	1	1	0
KARDITSA	5	5	0
KOZANI	5	5	0
TRIKALA	5	5	0
VIOTIA	4	4	0
EVIA	6	6	0
MESSINIA	7	6	1
LARISA	8	8	0
TOTAL	288	288	20

TABLE III

OFFICIAL APPORTIONMENT OF PARLIAMENT SEATS FOR THE CENSUS YEARS 1981 AND 1991

District	Sample '91	Off.1991	Dis/ment
ATTIKH I	84	19	3
ATTIKH II	-	38	-
ATTIKH III	-	9	-
PIREAS I	-	7	-
PIREAS II	-	8	-
THES/NIKI I	21	14	0
THES/NIKI II	-	7	-
XALKIDIKI	3	3	0
KERKYRA	3	3	0
LAKONIA	3	3	0
KILKIS	3	3	0
LESVOS	3	3	0
SERRES	7	7	0
ARTA	3	3	0
KYKLADES	3	3	0
RODOPI	3	3	0
IRAKLEIO	7	8	1
EVROS	4	4	0
IMATHIA	4	4	0
HANIA	4	4	0
KEFALONIA	1	1	0
DODEKANISA	4	4	0
ZAKINTHOS	1	1	0
KORINTHIA	4	4	0
IOANNINA	5	5	0
KAVALA	4	4	0
LASITHI	2	2	0
PREVEZA	2	2	0
RETHIMNO	2	2	0
ILIA	6	6	0
ETOL/NIA	8	8	0
GREVENA	1	1	0
KASTORIA	1	2	1
SAMOS	1	1	0
MAGNHSIA	5	6	1
PELLA	4	5	1
FOKIDA	2	2	0
ARKADIA	4	4	0
DRAMA	4	3	1
AHAIA	9	9	0
FLORINA	2	2	0
PIERIA	4	4	0
ARGOLIDA	3	3	0
THESPROTIA	2	2	0
HIOS	2	2	0
XANTHI	3	3	0
LEYKADA	1	1	0
EYRYTANIA	1	1	0
KARDITSA	5	5	0
KOZANI	5	5	0
TRIKALA	5	5	0
VIOTIA	4	4	0
EVIA	6	6	0
MESSINIA	6	6	0
LARISA	8	8	0
TOTAL	288	288	8

TABLE IV
APPORTIONMENT OF PARLIAMENT SEATS FOR THE 1993 ELECTION
BASED ON A SAMPLE OF SIZE 10000

District	Sample '91	Off.1991	Dis/ment
ATTIKH I	82	19	1
ATTIKH II	-	38	-
ATTIKH III	-	9	-
PIREAS I	-	7	-
PIREAS II	-	8	-
THES/NIKI I	21	14	0
THES/NIKI II	-	7	-
XALKIDIKI	3	3	0
KERKYRA	3	3	0
LAKONIA	3	3	0
KILKIS	3	3	0
LESVOS	3	3	0
SERRES	7	7	0
ARTA	3	3	0
KYKLADES	3	3	0
RODOPI	3	3	0
IRAKLIO	8	8	0
EVROS	4	4	0
IMATHIA	4	4	0
HANIA	4	4	0
KEFALONIA	1	1	0
DODEKANISA	4	4	0
ZAKINTHOS	1	1	0
KORINTHIA	4	4	0
IOANNINA	5	5	0
KAVALA	4	4	0
LASITHI	2	2	0
PREVEZA	2	2	0
RETHIMNO	2	2	0
ILIA	6	6	0
ETOL/NIA	8	8	0
GREVENA	1	1	0
KASTORIA	2	2	0
SAMOS	1	1	0
MAGNHSIA	6	6	0
PELLA	5	5	0
FOKIDA	1	2	1
ARKADIA	4	4	0
DRAMA	3	3	0
AHAIA	9	9	0
FLORINA	2	2	0
PIERIA	4	4	0
ARGOLIDA	3	3	0
THESPROTIA	2	2	0
HIOS	2	2	0
XANTHI	3	3	0
LEYKADA	1	1	0
EYRYTANIA	1	1	0
KARDITSA	5	5	0
KOZANI	5	5	0
TRIKALA	5	5	0
VIOTIA	4	4	0
EVIA	6	6	0
MESSINIA	6	6	0
LARISA	8	8	0
TOTAL	288	288	2

TABLE V
APPORTIONMENT OF PARLIAMENT SEATS FOR THE 1993 ELECTION
BASED ON GROUPED SAMPLE 1-5

District	Sample '91	Off.1991	Dis/ment
ATTIKH I	83	19	2
ATTIKH II	-	38	-
ATTIKH III	-	9	-
PIREAS I	-	7	-
PIREAS II	-	8	-
THES/NIKI I	21	14	0
THES/NIKI II	-	7	-
XALKIDIKI	3	3	0
KERKYRA	3	3	0
LAKONIA	3	3	0
KILKIS	3	3	0
LESVOS	3	3	0
SERRES	7	7	0
ARTA	3	3	0
KYKLADES	3	3	0
RODOPI	3	3	0
IRAKLIO	7	8	1
EVROS	4	4	0
IMATHIA	4	4	0
HANIA	4	4	0
KEFALONIA	1	1	0
DODEKANISA	4	4	0
ZAKINTHOS	1	1	0
KORINTHIA	4	4	0
IOANNINA	5	5	0
KAVALA	4	4	0
LASITHI	2	2	0
PREVEZA	2	2	0
RETHIMNO	2	2	0
ILIA	6	6	0
ETOL/NIA	8	8	0
GREVENA	1	1	0
KASTORIA	2	2	0
SAMOS	1	1	0
MAGNHSIA	6	6	0
PELLA	4	5	1
FOKIDA	2	2	0
ARKADIA	4	4	0
DRAMA	3	3	0
AHAIA	9	9	0
FLORINA	2	2	0
PIERIA	4	4	0
ARGOLIDA	3	3	0
THESPROTIA	2	2	0
HIOS	2	2	0
XANTHI	3	3	0
LEYKADA	1	1	0
EYRYTANIA	1	1	0
KARDITSA	5	5	0
KOZANI	5	5	0
TRIKALA	5	5	0
VIOTIA	4	4	0
EVIA	6	6	0
MESSINIA	6	6	0
LARISA	8	8	0
TOTAL	288	288	4

TABLE VI
APPORTIONMENT OF PARLIAMENT SEATS FOR THE 1993 ELECTION
BASED ON GROUPED SAMPLE 6-10

of 8 seats. This is again too large compared to the corresponding difference of 2 seats in the apportionment based on the 1991 census. This result casts some doubts on the appropriateness of the use of censuses taken as far back as almost ten years before an election. On the contrary, it is indicative of the necessity to consider using sample estimates of the population at or near the time of an election using estimates of the real population for the same point in time obtained from appropriate ARIMA models.

IV. Discussion - Conclusions

Sample censuses have mostly been taken where difficulties over funds, personnel, transport and the like have precluded a complete count. They are also considered a logical alternative if circumstances are such that data with acceptable accuracy cannot be obtained unless special precautionary measures are taken during the field work. However, as demonstrated in this paper, sample censuses can offer a valuable tool in obtaining sample estimates of the population size that can be accurate and up-to-date.

The use of appropriate sampling methods for estimating the de jure population for each district of the Greek country, according to which the allocation of parliament seats, or/and the national funds was demonstrated to be quite reliable. The effect of sampling errors could probably be the displacement of one to two Parliament seats across the districts. As demonstrated in section 3, if the sample size is large enough and if the selected stratum is not the district but the municipality area, the effect of sampling error on the apportioning of Parliament seats may be further contained. This change in the level of stratum enables one to counter-balance the implications of differentiation between electoral and statistical districts.

It would be interesting to examine possible modifications of the sampling method proposed so as to take account of the difference in the definition of a statistical district and that of an electoral district.

The question of which districts can be more sensitive to a reallocation effect of the sample based apportionment of Parliament seats or funds is also worth exploring.

Of course, applying the proposed methodology on the field for the national election that took place in Greece in 2000, a year before the 2001 population census, and comparing the 1991 census based official apportionment of the Parliament seats to the apportionment that would result in on the basis of an up-to-date sample estimate of the legal population is undoubtedly a stimulating issue.

The sampling technique proposed in this paper can obviously be used for the estimation of various other variables pertaining to the design of the population census.

REFERENCES

- [1] Erickson, E., and J. Kadane. (1985). Estimating the Population of a Census Year. *Journal of the American Statistical Association*, 80, 98-116.
- [2] Anderson, M. and Fienberg, S. (2000). Census 2000 Controversies. *Chance*, 03, 22-30.
- [3] Kish, L. (1979). Samples and Censuses. *International Statistical Review*, 17, 99-109.

- [4] Michalopoulos, V. and Kekelaki, E. (2000). Using Sampling Techniques to Estimate the De Jure Population of Greece and Apportion Parliament Seats Among Electoral Districts. Technical Report no 148, Department of Statistics, Athens, University of Economics and Business, July 2001.
- [5] National Statistical Service of Greece. (1996). Census of Population.
- [6] Schirm, A. (1991). The Effects of Census Undercount Adjustment on Congressional Apportionment. *Journal of the American Statistical Association*, 86, 526-540.
- [7] Wright, T. (1998). Sampling and Census 2000: The Concepts. *American Scientist*, 86, 245-253.