# A METHOD FOR FORECASTING POPULATION CHANGES IN ALPINE, SEMI-ALPINE AND LOWLAND COMMUNITIES OF EPIRUS REGION IN GREECE

**Georgios XANTHOS**
Technological Education Institute of Crete, Greece
xanthosg@ret.forthnet.gr.

**Christos Ap. LADIAS**
University of Central Greece
caladias@otenet.gr

**Christos GENITSAROPOULOS**
University of Central Greece and
Technological Education Institute of Lamia, Greece.
cgenits@ucg.gr

**Abstract**
*The prediction to the extent possible of the future size of the population of a region and its distribution among the settlements of the region is evidently significant for regional planning. In this paper, deriving data from population censuses, we attempted to predict the population of alpine, semi-alpine and lowland settlements of the Epirus Region. To perform the prediction, we applied a modified version of the shift and share method. In addition to applying the method, we evaluated its predictive ability in relation to a) the fixed share method, b) the fixed magnifications share method and c) the classical shift and share method, suggesting the relevant statistical tests.*

**Keywords:** prediction of population shift-share method.
**JEL classification:** R10, R58.

## 1. Introduction

Assessing the future total population of a region and its distribution among the settlements is necessary for regional planning (Bendavid 1991, Klosterman 1990, Isard 1970). To begin with, the difficulties of this attempt derive from the difficulty to find regional data in both time and space. Moreover, the difficulty is increased due to the fact that the prediction does not only pertain to the total population size of a region, but it also pertains to its distribution among the settlements. These difficulties have led to the use of methods which are based on “poor data” (Isard 1970, Wang 2007) and do not have any special computational requirements. One of these methods is the shift share method (Helman 1976, Isard 1970, Kurre 1989). Although this method was invented for the purpose of analyzing mainly employment changes at a regional level (Dunn 1960), it was modified (Esteban-Marquillas, 1972) and its applications were expanded. Moreover, modifications (Bergeg 1984, Knudsen1991, 2000, Nissan 1994) pertaining to the method’s weakness to accept statistical tests, were suggested. In this paper, we will use its modified version (Nissan 1994) to perform population predictions, but we will differentiate the stochastic hypotheses that have been set forth. The region on which we will apply the method is the Epirus Region. We will compare the results of the method with the results of other simple and frequently used methods, which do not have any special computational requirements and are also based on poor data.

## 2. The model

Let the random variables $P_{ij0}, P_{ij1}$ concerning the population size of the settlement $i = 1,2,3,..,n$ which belongs to the category $j$, be at times zero (0) and one (1) respectively. We then expect that the population size $P_{ij0}$ is connected with the population size $P_{ij1}$ via a linear relationship with a positive slope. The population difference between the two times is represented as:

$$P_{ij1} - P_{ij0} = (\hat{P}_{ij1} - P_{ij0}) + (P_{ij1} - \hat{P}_{ij1}) \quad (1)$$

Where $\hat{P}_{ij1}$ is the expected population of the settlement, based on the following relationship

$$\hat{P}_{ij1} = a + b * P_{ij0} \quad (2)$$

With $(P_{ij1} - \hat{P}_{ij1})$ being the random term which expresses the difference between the observed population size $P_{ij1}$ and the expected one according to (2). The estimation of the parameters of (2) result from $\hat{b} = r * (S_{Pj1} / S_{Pj0})$ and $\hat{a} = \mu_{j1} - \hat{b}\mu_{j0}$, where $r$ is the Pearson's linear correlation coefficient, $\mu_{j1}, \mu_{j0}$ are the numerical average populations of the category $j$ settlements at times one and zero, and $S_{pj1}, S_{pj0}$ are the corresponding standard deviations. By substituting (2) in (1) and after adding and subtracting the mean value $\mu_{j0}$ we have:

$$(P_{i1} - P_{i0}) \equiv [(\mu_{j1} - \mu_{j0}) + (\hat{b} - 1)(P_{ij0} - \mu_{j0})] + (P_{ij1} - \hat{P}_{ij1}) \quad (3)$$

Identity (3) is a different expression of the shift-share method[1] and it breaks the difference $(P_{ij1} - P_{ij0})$ into a sum of three components. The first component $(\mu_{j1} - \mu_{j0})$ pertains to the part of the change which can be attributed to the effect exerted on the settlement by the change in the total population of the category to which the settlement belongs[2]. The second component pertains to the part of the change which can be attributed to the net effect exerted on the formation of the difference by the initial population size of the settlement. The third component, despite the fact that it is essentially treated as a random variable, pertains to the part of the change which is formed under the influence of random and non-random factors. If $(P_{ij1} - \hat{P}_{ij1}) > 0$, the settlement has changed its population size faster than what (2) predicts and reversely if $(P_{ij1} - \hat{P}_{ij1}) < 0$. We applied model (3) to analyze into components the change of population data of alpine, semi-alpine and lowland settlements of the Epirus Region during the period 1971-2001. We used the components of the change to make predictions. Finally, we assessed the predictive ability of (3) in relation to the following models. (a) Fixed share model, (b) Fixed magnification share model and (c) Shift share model. These three models belong to the general category of "ratio" models and its central hypothesis is that the population of a small area or a settlement is proportionate to the population of a broader metropolitan area. If $p^S_{i,-1}, p^S_{i,0}, p^S_{i1}$ are the populations of the small area and $P^L_{-1}, P^L_0, P^L_1$ are the populations of the larger area at times minus one (-1), zero (0) and one (1) respectively, then the algebraic expression of the three models a, b, c is (United Nations 1974 , Wang 2007,Smith 2002 ):

[1]The hypotheses from which model (3) is derived in the current study are not exactly the same as the ones set by Nissan and Carter. The current study maintains the main hypothesis that variables $P_{ij0}, P_{ij1}$ are random, but it does not adopt the hypothesis of the joint normal distribution.

[2]j=1,2,3 where 1= Alpine- 2= Semi-alpine- 3= Lowland communities in the current study.

(a): $p^{S}{}_{i,1} = P_1^L(\frac{p_{i,0}^S}{P_0^L})$

(b): $p^{S}{}_{i,1} = p_{i,0}^S + [(P_1^L - P_0^L)(\frac{p_{i,0}^S - p_{i,-1}^S}{P_0^L - P_{-1}^L})]$

(c): $p^{S}{}_{i,1} = P_1^L[\frac{p_{i0}^S}{P_0^L} + (\frac{p_{i,0}^S}{P_0^L} - \frac{p_{i,-1}^S}{P_{-1}^L})]$

## 3. Areas of Application

We applied the model to lowland, semi-alpine and alpine communities of the Epirus Region. Epirus has an area of 9203 square kilometers, with just 10% of its land being lowland, thus rendering it the most alpine region of Greece[3]. According to the population censuses of the years 1971-1981-1991-2001 the picture of the development of the real population of Epirus, per category of alpine, semi-alpine and lowland communities is displayed on table -1-:

**Table-1-** Development of the Real Population of Epirus

| | Census year 1971 | Census year 1981 | Census year 1991 | Census year 2001 |
|---|---|---|---|---|
| Population of Alpine Communities | 119.531 | 119.222 | 109.115 | 108.714 |
| Population of Semi-alpine Communities | 52.933 | 57.168 | 62.012 | 63.785 |
| Population of Lowland Communities | 137.873 | 148.151 | 168.599 | 181.321 |
| Total Population | 310.337 | 324.541 | 339.726 | 353.820 |

The reduction in the population of alpine communities is evident. During the period 1971-2001, the Alpine communities had an average annual rhythm of population loss that equaled 0.31%. In contrast, Semi-alpine and Lowland Communities increased their population with an average annual rhythm of 0.62% and 0.92% respectively, for the same period. In its total, the Epirus Region increased its population with an average annual rhythm of 0.438%. This rhythm is significantly slower compared to the corresponding rhythm of the country, which is 0.748% for the period 1971-2001. For the application of model (3), we used the data of the real population of the settlements based on the censuses.

Table -2- displays: The value of the difference between the average values $\mu_{j1} - \mu_{j0}$, as well as the estimation of the difference $\hat{b} - 1$ .

[3]Greek Geographical Encyclopedia volume A, Editions TEGOPOULOS MANIATEAS

**Table-2-**

| Alpine Settlements | $\mu_{j1}-\mu_{j0}$ | $\hat{b}-1$ | Number[4] of Settlements |
|---|---|---|---|
| 1971-1981 | -3.80 | -0.077 | 701 |
| 1981-1991 | -13.30 | -0.023 | 733 |
| 1991-2001 | -1.10 | 0.043 | 730 |
| **Semi-alpine Settlements** | $\mu_{j1}-\mu_{j0}$ | $\hat{b}-1$ | **Number of Settlements** |
| 1971-1981 | 21.40 | 0.193 | 171 |
| 1981-1991 | 18.20 | 0.099 | 182 |
| 1991-2001 | 6.70 | 0.193 | 189 |
| **Lowland Settlements** | $\mu_{j1}-\mu_{j0}$ | $\hat{b}-1$ | **Number of Settlements** |
| 1971-1981 | 63.60 | 0.085 | 134 |
| 1981-1991 | 140.70 | 0.224 | 105 |
| 1991-2001 | 88.10 | 0.088 | 146 |

## 4. Tests for the application of the model

In this study, the limiting hypothesis of the joint normal distribution of random variables $P_{ij0}, P_{ij1}$ has not been adopted. For this reason, the tests that will follow are based on Spearman and Kendall's rank correlation coefficients (Cooper 1983, Gopal, 2006). The first test concerns the independence of random variables $P_{ij0}, P_{ij1}$. The second test concerns the acceptance or not of the linear relationship between the variables (2). This test is necessary because, even though the likely rejection of $H_0$ at the first test will lead us to conclude that there is a monotonic association between the variables, the existence of monotony alone does not entail a linear relationship between the variables. Finally, the third test concerns the components of (3). Due to the fact that the limiting hypothesis of the joint normal distribution of random variables $P_{ij0}, P_{ij1}$ has not been adopted, the significance test of the components' contribution to the difference $(P_{ij1}-P_{ij0})$ will be performed with the non-parametric analysis of variance, which pertains to correlated samples (Friedman 1937, Siegel 1956). For the first two tests, the relevant statistics in large samples for Spearman and Kendall's correlation coefficients are $TS = r_s\sqrt{n-1}$, and $TS = \tau / \sqrt{2(2n+5)/9n(n-1)}$ respectively. For the third test, we use the statistic $\chi_r^2 = \frac{12}{nk(k+1)}\sum_{j=1}^{k}(R_j)^2 - 3n(k+1)$ (Gopal 2006, Siegel 1956). At the level of significance $a = 0.05$, the test of the $H_0$ was applied: Variables $P_{i0}, P_{i1}$ are mutually independent, with $H_1$ as an alternative hypothesis: Variables $P_{i0}, P_{i1}$ are not mutually independent. On the basis of our data, the $H_0$ hypothesis was not accepted. Then, again, for $a = 0.05$ the second test was performed for the acceptance or not of a linear relationship between the variables (2). The test pertaining to the $H_0$ hypothesis: that no linear relationship exists between the variables will not be applied by testing the Spearman correlation coefficient, because it presupposes normal distribution. The test was performed with the Kendall coefficient (Gopal, 2006) and the $H_0$ hypothesis is rejected. Therefore, the linear relationship is accepted. To reinforce the relevant linearity hypothesis, Pearson's linear

[4] The number of the settlements is different due to differences in the volumes of the censuses of the Hellenic Statistical Authority.

correlation coefficient, as well as the elementary correlation coefficient (eta) were calculated, and their values do not differ significantly, pleading for the linearity hypothesis. Table 3 exhibits the values of the Spearman, Kendall, and Pearson correlation coefficients, as well as the value of the elementary correlation coefficient.

**Table-3-**

| **Period 1971-1981** | **Alpine Communities** | **Semi-alpine Communities** | **Lowland Communities** |
|---|---|---|---|
| Kendall $\tau$ | 0.700 | 0.803 | 0.851 |
| Spearman $r_S$ | 0.862 | 0.935 | 0.935 |
| Pearson $r$ | 0.941 | 0.970 | 0.997 |
| $\sqrt{\eta} - eta$ | 0.999 | 0.998 | 0.997 |
| **Period 1981-1991** | **Alpine Communities** | **Semi-alpine Communities** | **Lowland Communities** |
| Kendall $\tau$ | 0.771 | 0.830 | 0.839 |
| Spearman $r_S$ | 0.918 | 0.940 | 0.934 |
| Pearson $r$ | 0.965 | 0.983 | 0.997 |
| $\sqrt{\eta} - eta$ | 0.999 | 0.998 | 0.997 |
| **Period 1991-2001** | **Alpine Communities** | **Semi-alpine Communities** | **Lowland Communities** |
| Kendall $\tau$ | 0.779 | 0.869 | 0.879 |
| Spearman $r_S$ | 0.931 | 0.971 | 0.964 |
| Pearson $r$ | 0.973 | 0.983 | 0.999 |
| $\sqrt{\eta} - eta$ | 0.999 | 0.998 | 0.997 |

The third test pertains to the components of (3). At the level of significance $a = 0.05$, the zero hypothesis, which has been tested, is: $H_0$: the three components have no significantly different effect on the formation of the difference $(P_{ij1} - P_{ij0})$. The alternative hypothesis is $H_1$: the three components have significantly different effect on the formation of the difference $(P_{ij1} - P_{ij0})$. The relevant test led to the rejection of the hypothesis. Consequently, the individual components, into which (3) has been divided, contribute statistically significantly to the difference $(P_{ij1} - P_{ij0})$.

## 5. Predictions and conclusions

The assessment of the predictive ability of model (3) was based on the comparison of results from the three models we have already mentioned. The central hypothesis for the application of the three models above is that each time we have a good estimation of the future total population of Epirus. Thus, given that the total future population of Epirus has been estimated with great accuracy, the prediction process was performed as follows: with the population data of censuses 1971-1981, we attempt a prediction for year 1991; with the data of censuses1981-1991, we attempt a prediction for year 2001; and finally, with the data of censuses 1991-2001, we attempt a prediction for year 2011. Similarly, for model (3) estimations of $\mu_{j1} - \mu_{j0}, \hat{b} - 1$ were used for the prediction of both the total population of each category of settlements, and the distribution of the population predicted in the settlements of each category. The evaluation of the methods with regard to the prediction was performed on the basis of three criteria. The first one is the mean absolute percentage error

($MAPE$). The second one is concerned with which method gives the smallest absolute percentage difference in the total population of each category. The third criterion uses Kendall's correlation coefficient, so that we can establish the degree of agreement in the ranking order between real values of population per settlement and predicted values per settlement. The following tables 4 and 5 exhibit for each settlement category the MAPE values, the values involving the absolute percentage error (absPE) for the total population of each category to which the settlement belongs, and finally, the values of the Kendall coefficient.

**Table-4- Alpine Settlements**

| | Prediction for 1991 | | | Prediction for 2001 | | | Prediction for 2011 | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAPE | absPE | Kendall | MAPE | absPE | Kendall | MAPE | AbsPE | Kendall |
| Fixed share method | 42.86 | 14.4 | 0.758 | 30.97 | 4.5 | 0.794 | 78.84 | 30.8 | 0.743 |
| Fixed magnification share method | 74.48 | 9.0 | 0.618 | 49.89 | 8.3 | 0.665 | 110.06 | 38.1 | 0.677 |
| Shift Share Method | 75.75 | 8.8 | 0.614 | 55.15 | 10.5 | 0.635 | 88.48 | 24.9 | 0.622 |
| Suggested method | 43.86 | 7.3 | 0.758 | 25.14 | 8.9 | 0.795 | 95.80 | 36.9 | 0.743 |

**Table-5- Semi-alpine Settlements**

| | Prediction for 1991 | | | Prediction for 2001 | | | Prediction for 2011 | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAPE | absPE | Kendall | MAPE | absPE | Kendall | MAPE | AbsPE | Kendall |
| Fixed share method | 28.88 | 3.5 | 0.797 | 33.29 | 1.3 | 0.861 | 89.09 | 2.0 | 0.597 |
| Fixed magnification share method | 43.39 | 0.5 | 0.717 | 51.05 | 4.3 | 0.806 | 99.94 | 0.55 | 0.561 |
| Shift Share Method | 25.78 | 12.6 | 0.797 | 28.58 | 7.9 | 0.806 | 90.15 | 1.1 | 0.597 |
| Suggested method | 38.03 | 2.2 | 0.797 | 33.81 | 2.8 | 0.861 | 148.41 | 4.4 | 0.597 |

**Table-6- Lowland Settlements**

| | Prediction for 1991 | | | Prediction for 2001 | | | Prediction for 2011 | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAPE | absPE | Kendall | MAPE | absPE | Kendall | MAPE | AbsPE | Kendall |
| Fixed share method | 22.58 | 8.0 | 0.70 | 23.87 | 3.2 | 0.786 | 58.0 | 11.8 | 0.835 |
| Fixed magnification share method | 35.82 | 6.0 | 0.659 | 41.93 | 4.3 | 0.736 | 90.61 | 0.9 | 0.792 |
| Shift Share Method | 37.82 | 5.6 | 0.653 | 43.95 | 4.6 | 0.730 | 85.27 | 9.0 | 0.787 |
| Suggested method | 22.61 | 6.9 | 0.70 | 115.52 | 5 | 0.786 | 63.87 | 0.3 | 0.835 |

Taking all the aforementioned into consideration, we will underline the following. As for the alpine communities, all $\mu_{j1} - \mu_{j0}$ differences are negative. They therefore represent the trend towards the reduction of the total population of the alpine communities. This trend seems to be generalized for decades 1971-81 and 1981-1991, given that the component of the population size of the settlement also has a negative sign. In contrast, while for the decade 1991-2001 the total population of the alpine communities continues to decrease, this decrease is decelerated. For lowland and semi-alpine communities, all $\mu_{j1} - \mu_{j0}$ differences are positive, thus indicating the upward trend, even marginally, in the total population of semi-alpine and evidently in lowland communities. This trend is reinforced by the fact that the component of the population size of the community contributes positively, reinforcing the trend. As for the predictive ability of the model, in all three community categories the results of the model are similar to the results of the fixed share model, as long as the prediction for the total population of Epirus is accurate enough. If the population of Epirus is entered with a 5% error, the suggested method prevails, with the exception of the prediction involving year 2011. However, its lagging behind may be attributed to the fact that the population data of 2011 concern the permanent population, as opposed to the real population of the censuses. In closing, we can conclude that the suggested model provides useful information with the division of population changes into components on the one hand, and it provides satisfactory prediction for both the total population of the region and the distribution of the population among the settlements of the area, on the other hand. Moreover, it is not demanding in terms of data; neither does it require laborious calculations, providing useful information for regional planning with simple statistical tests.

## References

Berzeg Korhan., "A note on statistical approaches to shift share analysis", Journal of Regional Science, Vol. 24, No 2, 1984.

Bendavid val Avrom., "Regional and Local Economic Analysis for practitioners", Praeger Publishers, 1991.

Dunn Jr ES., "A statistical and analytical technique for Regional analysis", Papers of the regional Science Association ,6,1960.

Esteban-Marquillas JM., "Shift and share analysis revisited", Regional and Urban Economics, 2,1972.

Friedman M., "The use of Ranks to avoid the assumption of Normality implicit in the analysis of variance", Journal of American Statistical Association, Vol. 32,1937.

Gopal K. K., "100 Statistical Tests" Sage Publications London, 2006.

Helman D., "Shift Share Model as predictive Tool", Growth and Change 1976, Vol. 8, Issue 3,1976.

Isard W., "Methods of Regional analysis : an Introduction to Regional science" The MIT Press, 1970.

Kurre J., Weller B., "Forecasting the local Economy using time series and shift share techniques", Environment and Planning A., Vol., 21,1989.

Klosterman R., "Community Analysis and Planning Techniques", Rowman & Littlefield Publishers co, 1990.

Knudsen D., Barf R., "Shift share analysis as a linear model", Environment and planning A, 23,1991.

Knudsen D., "Shift share analysis: further examination of models for the description of economic change" , Socio economic Planning Sciences ,34,2000.

Nissan E., Carter G., "An Extension to the shift share techniques for predicting and evaluating changes in Employment Growth", Regional Science Perspectives, Vol. 24, No 2,1994.

Smith S., Tayman J., Swanson D., "State and Local Population projections Methodology and Analysis", Kluwer Academic Publishers,2002.

Siegel S., "Non Parametric Statistics for the behavioral science", McGraw Hill New York,1956.

United Nations "Methods for projection of Urban and Rural Population, manual on methods of estimating Population (manual VIII)", New York,1974.

Wang X., Rainer vom Hofe., "Research Methods in Urban and Regional Planning", Springer, 2007.