HELLENIC REPUBLIC

HELLENIC STATISTICAL AUTHORITY

Piraeus, 18 December 2025

HELLENIC STATISTICAL AUTHORITY (ELSTAT)

Population and Migration Statistics Section

Short Methodological Note

Population Revision

for the Intercensal years 2012 – 2021

and

Estimation of Post-census Migration Models

Contents

Chapter A: Population Revision for the Intercensal years 2012 – 2021

A.1: Introduction

Aim of the methodological note

This methodological note provides documentation on the methodology, sources, relevant reviews, and the final decisions that were qualified as most appropriate for completing the revision of population data for the years 2012 – 2021 between the two Population Censuses of 2011 and 2021. The Task Force set up at ELSTAT for population revision was assisted by an external expert specialized in demography.

Necessity of revision for the intercensal population estimates

The revision of population estimates between censuses is an essential statistical process that aims to improve the depiction of the demographic profile and the evolution of the population structure. Furthermore, as population is used as a denominator in the compilation of various statistical measures, population revision is considered essential for the update and the quantitative and qualitative improvement of demographic indicators (fertility, mortality), but also for other domains of official statistics, such as household, employment and labour force surveys, health, national accounts (per capita income), for estimating the number and structure of private households and their members, and for compiling socioeconomic and other indicators and measures covering various activities of the population. Therefore, the task of revising the population estimates for the intermediate years between Population and Housing Censuses of 2011 and 2021 is considered particularly important.

The need to revise both the population on 1st of January of each year and the annual migration flows to and from Greece was more apparent after the announcement of the final data from the 2021 Population and Housing Census. Based on these data, it was verified that the existing estimation methodology had led to an overestimation of the population by approximately 140,000 people.

Therefore, it seemed necessary to analyze and review the existing model used to estimate population on 1st January of each year, and to apply a new methodology for estimating population data, for the revision of both the intercensal years 2012–2021 and post-censal years.

A.2 Assessment of available population data

Available data sources

The following data are required to produce intercensal population estimates:

- Usual resident population of the Country by age and sex at national and regional level (former prefecture) according to Population and Housing Censuses of 2011 and 2021, respectively.

- Number of births by sex and usual residence of the mother and number of deaths by age, sex and usual residence of the deceased at national and regional level (former prefecture) for the entire reference period (2011-2021).

- Number of migrants (immigrants and emigrants) by sex, age and place of settlement or previous residence respectively (former prefecture) for the entire reference period (2011-2021).

The available data sources are as follows:

- Population and Housing Census data for 2011 and 2021
- Vital statistics data (births and deaths)
- Administrative sources on migration (Residence permits for third-country nationals (Ministry of the Interior) – Acquisitions of Greek citizenship (Ministry of the Interior) – Data on refugees (Ministry of Migration and Asylum)

Assessment of data

The analysis and evaluation of the data revealed the following:

- **Population and Housing Censuses for 2011 and 2021**. Data from the Population and Housing Censuses of 2011 and 2021 are considered reliable and particularly useful, as in addition to quantitative data, they also provide a complete database of qualitative characteristics of the population and immigrants (sex, age, nationality, country of origin, and year of settlement for immigrants, etc.).
- **Vital Statistics data (births - deaths).** Civil registry records of births and deaths are considered reliable. They cover the entire population, population groups and regions of the country for the entire reference period (2011-2021).
- **Administrative migration data.** In Greece, as in almost all countries, migration data are inadequate and of poor quality and reliability. Specifically, there are gaps in the recording of immigrants and incorrect or incomplete declarations of their characteristics (age, education, etc.), while data on emigration is essentially non-existent, due to the fact that there is no obligation to declare one's departure from the Country. The relevance of data on residence permits for third-country nationals held by the Ministry of the Interior, although an important source of information on the characteristics of foreigners residing in the country, is subject to restrictions, as their number is influenced by any incentives offered by the Country and the benefit of a residence permit does not necessarily means usual residence in the country.

The conclusion is that there is:

- High quality of census and registry data
- Low quality and significant gaps in migration data, especially for outflows
- Necessity to use indirect calculations and regression models.

A.3 Methodological Approach to Population Data Revision

This chapter presents the technical method for calculating and adjusting population data for the intercensal years. Initially, regression models were used to estimate migration flows using various sources, such as residence permits, changes in gross domestic product (GDP) at constant and current prices, migration data from the destination countries of emigrants from Greece. However, these models were considered unreliable because they did not demonstrate satisfactory fit, statistical significance, or sufficient stability.

The option adopted for the revision of population and migration data for the intercensal period 2011-2021 is the application of a technical method of population adjustment, maintaining the method of estimating immigration and estimating emigration by subtraction according to the demographic balance principle.

To revise the population data for the intercensal period, alternative technical estimation methods were studied, designed, and applied.

The method that was finally selected distributes evenly the difference between the estimated population on 1st January 2022, as it would have been calculated based on the existing methodology (without taking into account the results of the 2021 Population Census) and the corresponding published population on 01.01.2022, as a fixed reference point, as calculated on the basis of the results of the 2021 Population-Housing Census, throughout the intercensal period, i.e. in all years from 2012 to 2021, creating a consistent and smoothly changing population time series that maintains the evolution of the population estimates of the intercensal period.

Specifically, a process of smoothing and reallocating the population difference that occurs at the end of a period is applied when the estimated population is compared with the actual census population. The method is applied at a low level, for each sex, age (at birth year level) and Regional Unit. The population estimates for each year for total Country and at regional level, by sex and age, are obtained by adding up the population estimates of Regional Units. A detailed description of the method is provided in Chapter A.4 of this methodological note.

With regard to the post-census period, it was decided to maintain the calculation procedure for immigration and emigration, with appropriate adjustment of the estimated coefficients.

A.4 Population Revision Process

A.4.1 Calculation of annual population estimates 01.01.2012 – 01.01.2021

The algorithm describing the population redistribution process from 2011 to 2022 can be summarized in the following steps:

Step 1 — Calculation of coefficient D

$$D = \frac{\text{Population on 01.01.2022 based on the 2021 Population Census}}{\text{Estimated population on 01.01.2022}}$$

The denominator (*Estimated Population 01.01.2022*) was produced exclusively for the purposes of this revision process, applying the estimation method used for all intercensal years.

Step 2 — Calculation of n

The variable n equals to:

$$n = \frac{x}{11}$$

where 11 is the number of intervals between 2011 and 2022 and x is each year between 2011–2022 (x = 0 for year 2011, …, x = 11 for year 2022)

Step 3 — Calculation of coefficient a

This coefficient equals to:

$$a = \left(\frac{1}{D}\right)^{n}$$

It constitutes the 'annual rate of adjustment' with respect to the final deviation.

Step 4 — Calculation of coefficient c

Coefficient c is defined as:

$$c = \frac{1}{a}$$

The coefficient c is applied to the annual population in order to correct it proportionally.

Step 5 — Adjustment of the population using coefficient *c*

For each year x:

$$\text{Revised Population} = c \times Existing\ Population$$

Step 6 — Repeat for all years

Steps 2–5 are repeated for each year 2011–2022.

A.4.2 Distribution of population estimates by sex, age, and geographical level (Regions & Regional Units)

The procedure for calculating the annual population estimates for the period 01.01.2012–01.01.2021 was applied at the level of Regional Unit for each sex and age (approximate year of birth). The population estimates for each year at regional level and for total Country, by sex and age, are obtained by summing up the population estimates of the Regional Units.

A.4.3 Distribution of population estimates by country of citizenship and country of birth

The distribution of the population by groups of countries of citizenship and countries of birth for the intercensal period is based on a weighted linear approach that uses data from the two consecutive Population-Housing Censuses of 2011 and 2021 as fixed reference points. For each group of countries, the population of the 2011 and 2021 Population Censuses are combined using

weighting coefficients that vary linearly over time, so that the contribution of the first census decreases and that of the second increases as the time index (which expresses the distance of each intercensal year from 2011) diverges from 2011 and approaches 2021. Specifically, for each intercensal year, the respective weights of the two censuses are calculated, the sum of which remains constant. The population for each year is calculated as the weighted average of the census figures and then rounded up to the nearest integer to ensure that the results are expressed in integer population units. The process is applied repeatedly for all intercensal years and for each group of countries separately, ensuring a smooth and consistent transition between the two census reference points.

A.4.4 Estimation of migration flows by sex and age

The evaluation of alternative regression models for estimating immigration and emigration using various explanatory variables, as well as the study of available sources on migration flows, highlighted the following with regard to the estimation methodology currently applied:

1. The method used to estimate immigration is considered adequate and reliable, as it combines various statistical and administrative sources and includes as many categories of immigrants as possible (countries within and outside the EU, refugees, asylum seekers, illegal immigrants).
2. The method for estimating emigration could be improved, for example by re-evaluating the coefficients of the regression model parameters. For the intercensal period, emigration was calculated by subtraction, applying the demographic equation to the data for each year.
3. The proportional distribution of migration figures by age and sex is necessarily based on the available official census data, due to the lack of other sources, as is the case in most countries.

Based on the above findings, the data and distributions for immigration were retained, while for the recalculated emigration, the existing distributions by sex and age were applied.

A.4.5 Statistical adjustment procedure

In the final stage of the revision process of population estimates and given that certain inconsistencies were identified in the regional population estimates for older age groups, particularly those aged 85 and over, an adaptive statistical approach was applied. Unrealistic fluctuations, especially at very old ages both during the intercensal period and geographically between Regional Units stressed the need for this statistical adjustment.

The application of statistical adjustment aims to smooth the estimates using appropriate techniques and to ensure the internal consistency of the results, without distorting the overall population distribution at national level. Through this process, the produced population estimates are compatible with both the temporal evolution of the population in the intercensal period and the geographical distribution of the population, enhancing the statistical reliability and demographic interpretability of the final results.

A.5 Quality and Reliability Checks

Checks on sex distribution and age consistency

The assessment of the quality of population estimates is an integral part of the methodological process and focuses on checking the internal demographic consistency of the estimates, with key indicators being the distribution of the population by sex and age, in accordance with international statistical practices.

The table below shows the evolution of the sex ratio index for the period 2012–2021, by broad age groups (0–14, 15–64, 65+) and for the total population, comparing the existing with the revised population.

The results show that the adopted revision method maintains the stability of the sex ratio over time and the expected demographic behavior, confirming the internal consistency of the estimates and its suitability for estimating the population in the intercensal years.

(males per 100 females)		2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
EXISTING SEX RATIO	**0-14**	104.6	104.8	104.9	105.1	105.3	105.4	105.5	105.7	105.7	105.6
	15-64	99.4	98.5	97.8	97.2	96.8	97.0	97.1	97.5	98.2	98.3
	65+	79.2	79.2	79.2	79.3	79.5	79.7	79.8	79.8	79.9	80.0
	Total	95.8	95.2	94.6	94.3	94.0	94.1	94.2	94.4	94.8	94.8
REVISED SEX RATION	**0-14**	104.6	104.7	104.9	105.0	105.2	105.3	105.5	105.7	105.9	105.8
	15-64	99.7	99.1	98.5	98.1	97.8	98.2	98.4	98.9	99.6	99.8
	65+	78.5	78.0	77.6	77.5	77.6	77.8	78.1	78.3	78.8	79.2
	Total	95.8	95.2	94.7	94.4	94.2	94.4	94.6	94.9	95.4	95.5

Checks based on key demographic indicators

In addition, comparisons of the main demographic indicators were made for the existing and revised population of each intercensal year, which are presented in the tables below.

Ageing Ratio

TOTAL GREECE	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Existing	134.1	136.7	139.9	143.8	147.3	149.2	151.3	153.8	156.2	159.4	168.6
Revised	134.0	136.8	140.4	144.4	148.2	150.8	153.5	156.9	160.3	164.3	168.6

Dependency Ratio

TOTAL GREECE	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Existing	52.3	53.3	54.2	54.9	55.5	56.2	56.7	57.2	57.5	58.0	56.9
Revised	52.0	52.8	53.6	54.2	54.8	55.4	55.8	56.2	56.4	56.6	56.9

Old-Age Dependency Ratio

TOTAL GREECE	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Existing	30.0	30.8	31.6	32.4	33.1	33.7	34.1	34.6	35.1	35.6	35.7
Revised	29.8	30.5	31.3	32.0	32.7	33.3	33.8	34.3	34.7	35.2	35.7

Total Fertility Rate

TOTAL GREECE	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Existing	1.36	1.31	1.31	1.34	1.39	1.36	1.36	1.35	1.40	1.44
Revised	1.36	1.31	1.31	1.34	1.39	1.36	1.37	1.36	1.41	1.45

A.6 Final Results

The two tables below demonstrate a comparison of the results of the population and migration flow estimates during the intercensal period before and after the revision.

Year	Existing Population on 1st January	Revised Population on 1st January	Difference
2012	11,086,406	11,072,725	-13,681
2013	11,003,615	10,980,006	-23,609
2014	10,926,807	10,894,294	-32,513
2015	10,858,018	10,818,041	-39,977
2016	10,783,748	10,735,415	-48,333
2017	10,768,193	10,707,527	-60,666
2018	10,741,165	10,667,336	-73,829
2019	10,724,599	10,635,213	-89,386
2020	10,718,565	10,612,591	-105,974
2021	10,678,632	10,555,344	-123,288
2022	10,461,627	10,461,627	0

Year	Immigrants	Existing emigrants	Revised emigrants
2011	60,089	92,404	106,398
2012	58,200	124,194	135,073
2013	57,946	117,094	126,464
2014	59,014	106,804	114,276
2015	64,446	109,351	118,218
2016	116,867	106,535	119,409
2017	112,247	103,327	117,123
2018	119,489	103,049	118,545
2019	129,459	95,020	111,603
2020	84,221	77,837	95,439
2021	57,120	79,596	92,775

CHAPTER B: Post-Census Migration Models

This analysis is part of an ongoing effort to monitor and record migration flows to and from Greece, using models that have already been used in previous methodological approaches by ELSTAT. The selection of these specific models is based on their ability to capture both the dynamic behavior of migration flows and the impact of key economic factors.

The aim of the analysis is not to introduce new models, but to review and validate the existing coefficients of the immigration and emigration models, based on updated data resulting from the revision of population data. Maintaining the same structure of the equations allows for the comparability of results over time and ensures consistency with the previous methodological framework.

In this context, the model coefficients were re-estimated using the Ordinary Least Squares (OLS) method, while the prescribed residual diagnostics (i.e., simultaneous satisfaction of non-autocorrelation, homoscedasticity, and normality) were tested to confirm that the basic assumptions of the method were not violated. This ensures that the updated estimates are theoretically consistent, statistically reliable, and suitable for use in updated analyses and forecasts of migration flows.

B.1 Modeling Immigration

The immigration model's dependent variable is immigration (I_t) and the explanatory variables are the immigration of the previous year (I_{t-1}) and the percentage change in GDP (Source: ELSTAT) of the previous year (GDP_{t-1}). The reason for choosing these explanatory variables is that they reflect both the dynamic behavior of migration flows and the way in which economic conditions influence migration decisions. Immigration of the previous year (I_{t-1}) is included to capture the time dependence, as the volume of migration in the previous year serves as an indicator of existing conditions and trends that directly influence how many immigrants are likely to enter the country in the current year. At the same time, the percentage change in GDP of the previous year is incorporated into the model to take into account the impact of economic conditions in the host country.

The above functional form allows for the capture of both the economic and dynamic dimensions of immigration flows. The analysis is performed on logarithmic values in order to limit the phenomenon of heteroscedasticity.

Using the OLS method, the following results are obtained from the estimation of the immigration model, where the exact significance levels of the model variables (p-values) are shown in parentheses:

$$\ln(\widehat{I_t}) = 2.621435 + 0.765123 \cdot \ln \hat{I}_{t-1} + 0.007012 \cdot GDP_{t-1}$$
$$(0.0114) \quad (0.0000) \qquad (0.0001)$$

Where: I_t = total annual immigration

I_{t-1} = total annual immigration of the previous year

GDP_{t-1} = percentage change in GDP in current prices with 2020 as the base year and 2021 as the reference (of the previous year)

The signs of the explanatory variables remain unchanged and are consistent with the theoretically expected effects, as lagged immigration positively affects current immigration flows, confirming the positive time dependence of the phenomenon, while the percentage change in GDP also shows a positive relationship with immigration, which suggests that economic growth and migration flows move in the same direction, in line with the relevant economic theory.

The necessary residual diagnostics of the regression model were tested, and it was found that none of the basic assumptions were violated, since the residuals do not show autocorrelation, are characterized by homoscedasticity, and follow a normal distribution, which allows for reliable conclusions. The results show that both the constant term and the explanatory variables of lagged immigration and the percentage change in GDP of the previous year are statistically significant at a 5% significance level, while the overall regression model is statistically significant. Meanwhile, the high coefficient of determination (R^2=0.988043) demonstrates the very good fit of the model to the data, as almost all of the variability in immigration is explained by the variables included in the model.

B.2 Modeling Emigration

The emigration model's dependent variable is emigration (E_t) and the explanatory variable is the immigration to Germany from Greece in the same year, based on data from the Statistical Office of Germany (DESTATIS), which is symbolized as D_t.

The reason why this explanatory variable was chosen is that Germany is consistently the main destination for Greek emigrants.

Using the OLS method, the following results are obtained from the estimation of the emigration model, where the exact significance levels of the model variables (p-values) are shown in parentheses:

$$\widehat{\mathrm{E_t}} = 13{,}804.30 + 3.319701 \cdot \mathrm{D_t}$$

$$(0.2314) \qquad (0.0000)$$

Where: E_t = total emigration in year t

D_t = immigration to Germany from Greece in the same year

The signs of the estimated coefficients are consistent with theory, as emigration to Germany has a positive and amplifying effect on total emigration, suggesting a strong positive relationship between the two variables. The necessary residual diagnostics of the regression model were tested and no violation of the basic assumptions of the model was found, since the residuals do not show autocorrelation, are characterized by homoscedasticity, and follow a normal distribution. The estimation results show that the explanatory variable of immigration to Germany from Greece is statistically significant and adequate to explain emigration, while the model fits the data very well, as reflected in the high value of the coefficient of determination (R^2=0.904750).