İstatistik Araştırma Dergisi
Journal of Statistical Research
ISSN: 2791-7614

(2021), 11 (2)
Başvuru/Received: 25/10/2021
Kabul/Accepted: 23/12/2021

https://journal.tuik.gov.tr/

ARAŞTIRMA MAKALESİ RESEARCH ARTICLE

Econometric Analysis of Population Increase and Population Projections in Turkey

Cemil İSKENDER
Retired Research Economist
iskenderc@ttmail.com
Orcid No: 0000-0003-2841-5964

Abstract

Turkish Statistical Institute (TurkStat), which has been carrying out the official censuses in Turkey since 1927, published backward mid-year population estimates for all years from the first census year of 1927 to 2006, address-based census results and mid-year estimates for 2007-2017 and population projections for 2018-2080 period. According to the main scenario of the projections, population will reach 107.7 million in 2069, and from this point, it will decrease to 107.1 million in 2080. TurkStat population projections method covers the dimensions which is based on the institution's long term experience and data, the principles of the discipline of demography e, and international data compiled. My aim in this econometric study is to investigate the representation of the main scenario projections and the results of past censuses with five and six parameter Gompertz population growth model which was developed during this study and also sinusoidal function applications for the residuals. In addition, I also applied Verhulst five and six parameter functions and reached positive statistical and linearity test results made with all functions. The results of the econometric study confirmed that the main scenario projections of the Statistical Institute are valid and realistic.

Keywords: Gompertz Function, Nonlinear Functions, SAS MODEL Procedure, SAS NLIN Procedure, Turkish Population Study, Verhulst Function.

Corresponding Author / Sorumlu Yazar: 1-Cemil İSKENDER, Retired Research Economist.

Citation / Atıf: İSKENDER C. Econometric Analysis of Population Increase and Population Projections in Turkey. İstatistik Araştırma Dergisi, 2021, 11 (2), 30-55.

Citation / Atıf: İSKENDER C. Econometric Analysis of Population Increase and Population Projections in Turkey. İstatistik Araştırma Dergisi, 2021, 11 (2), 30-55.

Türkiye'deki Nüfus Artışı ve Nüfus Projeksiyonlarının Ekonometrik Analizi

Özet

Türkiye'de 1927 yılından bu yana resmi nüfus sayımını yapan Türkiye İstatistik Kurumu (TÜİK), 1927 yılının birinci nüfus sayımı yılından başlayarak 2006 yılına kadar tüm yıllar için geriye dönük yıl ortası nüfus tahminleri, 2007-2017 yıllarına ait adrese dayalı nüfus sayımı sonuçları ve yıl ortası tahminleri ile 2018-2080 dönemine ilişkin nüfus projeksiyonlarını yayınlamıştır. Tahminlerin ana senaryosuna göre, nüfus 2069'da 107,7 milyona ulaşacak ve bu noktadan sonra 2080'de 107,1 milyona inecektir. TÜİK nüfus projeksiyonları yöntemi çalışması, kurumun uzun vadeli deneyim ve verilerine, demografi disiplini ilkelerine ve derlenen uluslararası verilere dayanan boyutları kapsamaktadır. Bu ekonometrik çalışmadaki amacım, çalışma sırasında geliştirdiğim beş ve altı parametreli Gompertz nüfus artış modeli ve ayrıca artıklar için sinüzoidal fonksiyon uygulamaları ile ana senaryo projeksiyonlarının ve geçmiş sayımların sonuçlarının temsilini araştırmaktır. İlaveten Verhulst beş ve altı parametre fonksiyonları için de uygulama yaptım. Tüm fonksiyonlarla yapılan pozitif istatistiksel ve doğrusallık test sonuçlarına ulaştım. Ekonometrik çalışma sonuçları, İstatistik Kurumu'nun ana senaryo projeksiyonlarının geçerli ve gerçekçi olduğunu doğrulamaktadır.

Anahtar Kelimeler: Doğrusal Olmayan Fonksiyonlar, Gompertz Fonksiyonu, SAS MODEL Yöntemi, SAS NLIN Yöntemi, Türkiye Nüfus Çalışması, Verhulst Fonksiyonu.

1. Introduction and Scope of Research

The aim of this econometric study is to investigate the representation of Turkey's population projections published by TurkStat (Turkish Statistical Institute) for the 2018-2080 period and the past census results (1927-2017) together by applying growth functions of Gompertz and Verhulst. As based on economic theories, explained by mathematical models, applied and supported by data and includes advanced statistical applications, we define our research as econometric analysis.

TurkStat has prepared population projections of Turkey for 2018-2080 period and published them on its website in 2018. Projections are an important source of data when used together with the past censuses. With an advanced statistical study of this data, it is possible to obtain concrete results for the population growth trend of Turkey. However, this long-term data need to be handled within a certain economic and statistical framework. Otherwise, our statistical study remains at a simple level of regression work: If we had done a short-to-medium-term work for 10-15 years, inter-census interpolation study or a two to three-year forecast, we could have used linear, polynomial, etc. regression methods. But that is not an affair and purpose of this study and the things are quite different than that of ordinary regression analysis. At this point, the theory of population growth, advanced mathematics and statistical analysis is included into the phases of the study.

Theorists who studied population growth for this purpose also established the theory of population development. There must be a consistent approach behind the statistical study. The first and most important examples of this approach are theorists Malthus, Gompertz and Verhulst, whose theories survived to the present day. While population growth in human communities in a certain geography increases according to biological and economic principles of scarce resources, growth rate is limited and may even fall to zero with the increase of factors against population growth after a certain stage. Verhulst wrote[1]:

> There is a limit to the growth of that population to the extent permitted by the land that meets the need for a population's housing…Therefore, all formulas that will try to represent the laws of the population must accept the requirement that *maximum* can be reached in an extremely distant period. This *maximum* will be a population that has become stable. (1838, pp. 113–115)

The mathematical population growth theories developed by Verhulst and Gompertz will be the basis for our study of population growth in Turkey. We used Verhulst logistic and Gompertz functions as they were specifically developed for the analysis of human population. In addition to the five and six-parameter versions of Verhulst

[1] For a very detailed explanation of economic fundamentals and mathematics of population theories see (İskender, 2021)

logistic function, we also used five and six parameter versions of Gompertz functions including sinusoidal ones which we developed in this study.

TurkStat population projections methods are very different from the mathematical and statistical applications used in this research and in our previous "Turkish Population Growth Study" (İskender, 2018). In that study which was based on population counting and forecasting data for the 1927-2015 period, we had made projections with mathematical growth functions for the 2016-2080 period and predicted the final levels of the population. In this study, we will use same methodology of the previous manuscript and newly developed Gompertz function for the evaluation of TurkStat population projections.

In the last section, the final evaluation of the study and its conclusions are presented.

2. Turkish Population Data, Projections and Methodology of Turkstat

Population projections of 2018-2080 period published in three series, namely; *the main scenario, high scenario, and low scenario,* all are extending until 2080. According to the *main scenario,* the population of 81.9 million[2] people in 2018 will reach 107.7 million in 2069, and from this point, it will decrease to 107.1 million in 2080. At the *low scenario*, the population will reach a maximum of 99.7 million in 2055, then with a decrease to 94.1 million in the year 2080. As to the *high scenario,* the population will be 121 million people in 2080, and although not too fast, it will continue to increase as far as the Figure 1 indicates.

TurkStat has explained the assumptions and calculation methods on which the population projections are based, as follows:

> (i) Analytical Framework, Concepts, Definitions, and Classifications:
>
> The aim of this study is to produce national and provincial population projections by using the current data sources. Cohort-component method, which reflects the effects of various demographic indicators on age-sex structure of the population, was used. Cohort-component method is based on lifelong monitoring of cohorts of the same age group according to their fertility, mortality and migration patterns. The cohorts are annual birth cohorts (age-cohorts). The components are births, deaths and migration… (TurkStat, 2021)
>
> (iii) Accounting Conventions:
>
> During the design and production stages of the projections, a committee that consists of representatives from academic and official institutions/organizations and sub-working groups for fertility, mortality and migration were constituted under chairmanship of TurkStat. Within the context of activities carried out by sub-working groups, evaluations about the current and future demographic circumstances of Turkey were made; backward analysis of historical data, problematic points and limitations were discussed. Provincial age and sex distribution of demographic data and their trends were analyzed. Thus, projections and assumptions regarding future dynamics of fertility, mortality and migration have been determined...(TurkStat, 2021)

After almost a century of successful activities and achievements of important statistical publications, TurkStat has changed its mode of operation and vision to the next level by publishing the population projections for the2018-2080 period. In addition to the population figures of the past years, the projections published on the TurkStat website are also vital reference source for researchers.

Before explaining the data we use, it is useful to look at the population curve of Turkey indicated in Figure 1.

The major event that disrupted the population trend of the 1925 -2080 period was the Second World War. Even though Turkey was not in the war, the Second World War had a negative effect on the population growth due to the high number of soldiers and the longer than normal military service period. In 1935-40, five-year population growth was 10.3%, compared to 5.4% in 1940-45. Despite the negative effects of the Second World War, the statistical representation of the curves remained intact.

TurkStat has carried out its first census in 1927 with a daily curfew and by visiting households[3], starting from the establishment of the Republic of Turkey in 1923. Following the negative effects of the world-wide economic crisis in 1930, the second census was held in 1935, two years after the first five-year industrial development plan which

[2] By following the SAS Studio statistical software methodology, we use dots to separate the decimal places in our manuscript.

[3] For a detailed explanation of census methodologies in Turkey see Canpolat & Gök, (2020).

began in 1933 and was the first major economic initiative of the Republic. In the subsequent years, the censuses were maintained for every five-year, until 1990 (including). Last census with daily curfew carried out in 2000.

Based on the data of 12 censuses for 1927-1985 period, TurkStat has prepared and included census mid-year estimates, as well as the census years figures in its published tables (59 observations). In determining the mid-year population figures for the 1986-1999 period, estimates produced from the Turkey Demographic and Health Surveys and 2008 address-based population registration system were used (14 observations). Address-based population registration system results of 2007-2020 and population projections of 2018-2080 period have been used for the population estimates of 2000-2006 period (7 observations). Mid-year census results for the 2007-2017 period were produced from annual results of the address-based registration system for 2007-2017 period (11 observations). Thus, we have 91 mid-year-based data for each year of 1927-2017 period.

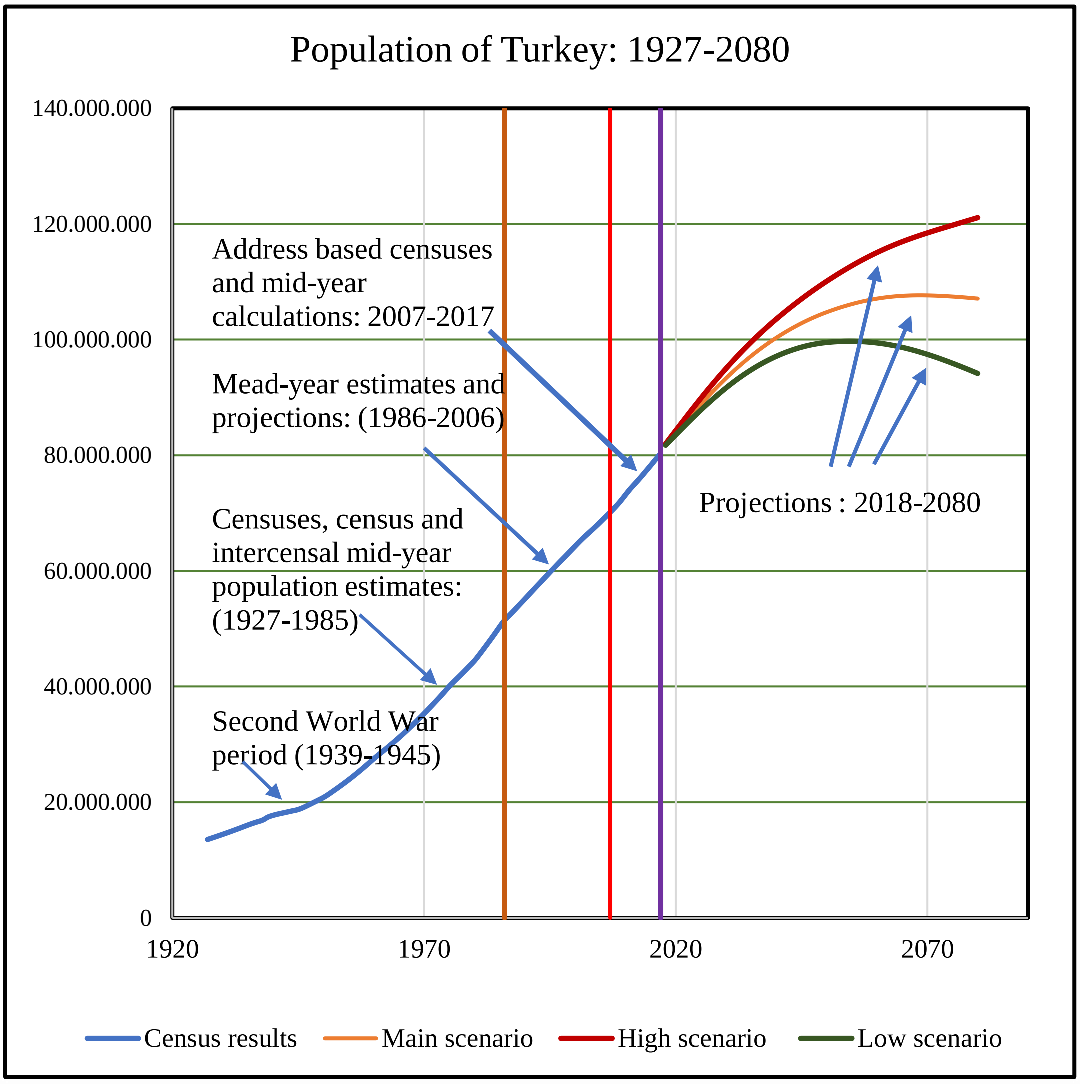

Figure 1: Turkish Population Curves of TurkStat Data

When we combine the estimates of the sixty-three years of the main scenario projections with historical data, we reach a total of 154 observations. 154 figures will represent all sections of the mathematical growth curves, including the lower and upper asymptote regions which are critical to statistical application. Our objective is to statistically analyze the consistency of *main scenario projections* with the census results.

Population statistics compiled and used are summarized in Table 1.

Table 1: Summary Table of Population Statistics and Their Sources

Index	Years	Number of observations	Census results (summary)	Source
1-59	1927-1985	59	13,538,000…50,307,000	TurkStat counts and estimates
60-80	1986-2006	21	51,480,000…69,295,000	TurkStat estimates and projections
81-91	2007-2017	11	70,158,000-80,313,000	TurkStat censuses of population registration system based on resident addresses
92-154	2018-2080	63	81,867,223...107,100,904	TurkStat *main scenario* projections

3. Mathematics of Particular Integral or Trend Functions

In accordance with our purpose of this study, we have selected five and six-parameter Gompertz and Verhulst growth functions. We developed and implemented five and six-parameter Gompertz functions in this study for the first time. Gompertz function will be discussed and presented with the addition of fifth and sixth parameters of a new approach and will be demonstrated with the Turkish population data. In addition, the application of the complementary sinusoidal function for residual series of growth functions will be discussed.

In our previous study, we already applied the six-parameter Verhulst growth function to Turkish population data. Here, the reason we examined the population projections of Turkey with the newly developed Gompertz growth function is to show that the successful statistical results obtained in all cases are not accidental or coincidental. Verhulst and Gompertz growth functions come from completely different differential equation definitions. Verhulst's logistic growth function is based on the dilemma of population growth and forces opposing to growth, while the Gompertz function is based on a geometric decrease in survivors and increase in deaths. The upper asymptotes obtained with each group of functions are different and in general Gompertz asymptote figures in Turkey's population data are higher than Verhulst's. On the other hand, this curve has less curvature flexibility when coming to the upper asymptotic zone. It takes starting level of population as lower asymptote. However, the logistic curve is based on a lower asymptote which is less than the initial level of the population and this gives flexibility to Verhulst curve.

3.1. Gompertz Function

Although five and six-parameters forms of Gompertz function come from the roots of the main function (three-parameter), by definition they are quite different: Two more parameters; lower asymptote (L) and allometric coefficient (v), and two new definitions; separated growth base coefficients (Q_1 and Q_2) and weight's function ($w_{(t-T)}$) have been added to growth formula. These changes make formulas completely different. Despite these mathematical differences, statistical test results obtained from the application of Gompertz's five and six-parameter functions are very close to each other and are more preferable relative to four-parameter forms when applied with Turkish population data.

Symbols used for Gompertz functions are:

Y: Population (dependent variable),

t: Time (explanatory variable),

T: Base year[4],

K: Upper asymptote,

L: Lower asymptote,

N: Starting level;

Q: Coefficient of exponential growth base (e).

Q_1 and Q_2: Coefficients of exponential growth base of (e),

Y_0: Initial population,

r: Intrinsic growth rate,

v: Allometric coefficient which indicates the direction and magnitude of the structure of the curve,

$w_{(t-T)}$: Weight's distribution function as being the key to distributing weights in the function and,

$\overline{H}_w$: As harmonic mean of Q_1 and Q_2, is the coefficient of *e (exponential base)* in the weight's function.

Based on approach of Ricketts and Head (1999), we separated Q - multiplier of the naturel logarithm (e) base of the growth function- into two multipliers as $Q_1$and Q_2. Also, the weight's distribution function as weighted with the harmonic, arithmetic or geometric mean of Q_1 and Q_2 were included in the equation defining $w_{(t-T)}$ for Q_1 and $(1 - w_{(t-T)})$ for Q_2. Thus, asymmetry of the Gompertz function happened to be more asymmetrical. The $Qe^{-r(t-T)}$ term of the Gompertz equation is as follows in the five-parameter form:

$$Qe^{-r(t-T)} \equiv w_{(t-T)}Q_1e^{(-r(t-T))} + \left(1 - w_{(t-T)}\right)Q_2e^{(-r(t-T))} \tag{1}$$

When Q_1becomes equal to Q_2, we obtain four-parameter form. We also applied the weight's function of (2) to the coefficients in the denominator of Gompertz function. It defines a Gompertz weighting function varying between 0 and 1 similar to main curve.

The generalized five-parameter Gompertz growth function is a mathematical function with two variables and six parameters (K, L, Q_1, Q_2, r) and has a nonlinear structure.

The weight's function and separation of Q into two multipliers as $Q_1$and Q_2 overcome the constraints imposed by the symmetry event and to better reflect the asymmetrical nature of the data. The flexibility of this function is as much an advantage as it is, one and most important drawback is that the number of parameters to be estimated increases by one. Difficulties are felt in estimating an increased number of parameters during the computer software application. Same argument is valid in a wider scale for six-parameter case. More parameters mean more conformity with the data on the one hand, and calculation difficulties on the other hand. To overcome this situation, a dataset which is refined, high quality, smooth (without acute changes) and having large number of observations (including sufficient data for the lower and upper asymptotic regions) needs to be implemented.

Consequently, we defined the generalized five - parameter population growth function of Gompertz (G5) based on the changes and additions of the components given below;

(i) the addition of lower asymptote (L),
(ii) the separation of the base coefficient (Q) of (e) into two (as Q_1 and Q_2),
(iii) the inclusion of the weight's function $w_{(t-T)}$ in the denominator of equation.

Constituents:

Harmonic mean:

$$\overline{H}_w = \frac{2Q_1Q_2}{|Q_1 + Q_2|} \tag{2}$$

[4] *T* and *t* are different.

Weight's distribution function:

$$w_{(t-T)} = \frac{1}{e^{\overline{H}_w e^{(-r(t-T))}}} \tag{3}$$

Starting values and *e*-bases:

$$A = Q_1 e^{(-r(t-T))} \tag{4}$$

$$B = Q_2 e^{(-r(t-T))} \tag{5}$$

Five-parameter Gompertz function (G5):

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{e^{[w_{(t-T)}A + (1 - w_{(t-T)})B]}} \tag{6}$$

or:

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{e^{\left[\left(\frac{1}{e^{\left[\frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}\right]}}\right)e^{(-r(t-T))} + \left(1 - \frac{1}{e^{\left[\frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}\right]}}\right)Q_2 e^{(-r(t-T))}\right]}} \tag{7}$$

We also included allometric coefficient of Richards (1959) to denominator of equation. The variable *v* provides flexibility to the predicted dependent variable of the function and increases the compliance of the curve to the actual data. *v* is specified as the "allometric coefficient." Thus, we reach six-parameter form of Gompertz function. When *v*=1, (8) becomes (7).

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{\left[e^{\left[\left(\frac{1}{e^{\left[\frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}\right]}}\right)e^{(-r(t-T))} + \left(1 - \frac{1}{e^{\left[\frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}\right]}}\right)Q_2 e^{(-r(t-T))}\right]}\right]^{(1/v)}} \tag{8}$$

We defined the generalized six - parameter population growth function of Gompertz as the addition of the allometric coefficient (1/*v*) to five-parameter form.

3.2. Verhulst Logistic Function

Six-parameter Verhulst logistic growth function is a mathematical function with two-variables as time and population, six-parameters (K, L, Q_1, Q_2, r, v) and has a nonlinear structure.

Constituents of Function:

Harmonic mean:

$$\overline{H}_w = \frac{2Q_1Q_2}{|Q_1 + Q_2|} \tag{9}$$

Weight's distribution function:

$$w_{(t-T)} = \frac{1}{1 + \overline{H}_w e^{(-r(t-T))}} \tag{10}$$

Starting values and e bases:

$$A = Q_1 e^{(-r(t-T))} \tag{11}$$

$$B = Q_2 e^{(-r(t-T))} \tag{12}$$

Six-parameter short form of Verhulst function (V6):

$$Y_{(t-T)} = L + \frac{K-L}{\left(1 + w_{(t-T)}A + \left(1 - w_{(t-T)}\right)B\right)^{(1/v)}} \tag{13}$$

Six-parameter long form of Verhulst function (V6):

$$Y_{(t-T)} = L + \frac{K-L}{\left(1 + \frac{1}{1 + \frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}} Q_1 e^{(-r(t-T))} + \left(1 - \frac{1}{1 + \frac{2Q_1Q_2}{|Q_1+Q_2|}e^{(-r(t-T))}}\right) Q_2 e^{(-r(t-T))}\right)^{(1/v)}} \tag{14}$$

Five-parameter Verhulst logistic growth function (V5) is a growth function without the allometric coefficient (1/v) as exponent of denominator of six-parameter form of Verhulst logistic growth function.

4. Mathematical Model of Analysis Applied in Development of Population

Population developments must be addressed under two main headings: (i) The main development trend of the population according to the natural birth rate and mortality rates and (ii) increases and decreases in population caused by external events and data errors. The mathematical growth models of Gompertz and Verhulst (particular solution equations) explain the first heading. For the second title, which describes autonomous changes outside the trend in the population, it will be required to add a new constituent to the growth function. These additional functions will explain most of the residuals at statistical application phase of the study.

The general solution of the mathematical model (*complete primitive*)[5] is equation (18). "$\bar{Y}(x)$ is called the *particular integral* and the reminder makes up the *complementary function…*" (Allen, 1956, p. 137). In our case particular integral -*trend, equilibrium path*- is the Gompertz population growth function. For the path to follow in statistical application Allen wrote:

> …Some particular solution $Y = \bar{Y}(x)$ is first sought. Any particular solution will serve but the most relevant, from the point of view of interpreting the result, is a solution which represents the *trend* or equilibrium variation of Y. Interest is usually concentrated on the path of Y as x increases indefinitely, e.g. the time path of Y if x is time. The particular solution $Y = \bar{Y}(x)$ sought then describes that path which is consistent with the given equation (15) and which is in some sense a norm or equilibrium. The next step is to write $y = Y - \bar{Y}$ as the deviation of Y from the trend or equilibrium value. It follows that y must satisfy the homogeneous form (16), as is seen by substituting Y and $\bar{Y}$ in (15)—both being solutions—and subtracting. The last step is to obtain the general solution of the homogeneous form (16) in the way indicated in (17). The solution is now both complete and ready for interpretation. The deviations y of Y from $\bar{Y}$ are given by (17); translating back into the original variables gives solution (18). Of the two terms in this solution, one is the trend or equilibrium path (the particular integral) and the other is deviation from trend or equilibrium (the complementary function)…(Allen, 1956, pp. 137–138)

While a little long, the quote explains the state of affairs very well, so we kept it as it was. As a complement to the population growth functions, contributions of mathematical economics we have taken from the 20th century economics and statistics lecturer and author Allen (1956, pp. 91-173; 1967, pp.342-363),, are as valuable and

[5] (Piaggio, 1920, p. 4)

explanatory as the population theory models that are taken from the early 19th century authors Gompertz and Verhulst[6].

Differential equation with constant coefficients of non-homogenous form of order n:

$$a_0 \frac{d^nY}{dx^n} + a_1 \frac{d^{n-1}Y}{dx^{n-1}} + \cdots + a_{n-1} \frac{dY}{dx} + a_nY = f(x) \tag{15}$$

Homogenous form of corresponding equation:

$$a_0 \frac{d^ny}{dx^n} + a_1 \frac{d^{n-1}y}{dx^{n-1}} + \cdots + a_{n-1} \frac{dy}{dx} + a_ny = 0 \tag{16}$$

Solution of (16):

$$y = A_1y_1(x) + A_2y_2(x) + \cdots + A_ny_n(x) \tag{17}$$

General solution (Complete Primitive):

$$Y = \bar{Y}(x) + A_1y_1(x) + A_2y_2(x) + \cdots + A_ny_n(x) \tag{18}$$

Since we give trend functions in section 3 as mathematics of population growth of Gompertz and Verhulst which were based on population growth theories by authors (İskender, 2021), now we will explain complementary functions.

Second order differential equation (19) and its homogeneous form (20) can be written as follows:

$$\frac{d^2Y}{dx^2} + a\frac{dY}{dx} + bY = f(x) \tag{19}$$

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = 0 \tag{20}$$

With two equations, values of *a* and *b* can be found. However, a particular solution $Y = \bar{Y}$(x) is found first. Then complementary functions are obtained.

In case of $a^2 < 4b$, solution equation of four-parameter sinusoidal function as complementary of growth function:

$$y_{(t-T)} = Ae^{\alpha(t-T)}cos(\omega(t-T) - \varepsilon \tag{21}$$

or eight-parameter form[7] is;

$$y_{(t-T)} = Ae^{\alpha(t-T)}cos(\omega(t-T) - \varepsilon) + Be^{a1(t-T)}cos(\theta(t-T) - \varphi) \tag{22}$$

(21) and (22) are complementary sinusoidal functions with two variables (as *y* and *t*), have four and eight parameters respectively and a non-linear structure. Since it has a nonlinear structure, data applied with this function has to provide statistical and linearity tests of the NLIN procedure[8].

y is the autonomous population which will be added to main trend function of $F_{(t-T)}$.

± *A* and *B*: Initial amplitude of oscillation or peak and trough coefficient,

[6] There are also quite wide and useful explanations on sinusoidal functions in Tenenbaum and Pollard (1963, pp. 211-220, 313-392)

[7] See Allen (1956, pp. 123–124, 147–148) for proof from differential equation to general solution.

[8] All of the auxiliary sinusoidal functions we applied have had satisfactory statistical tests through this study.

$\omega, \varepsilon, \theta, \varphi$ are constants of cosines functions.

a and a1: Damping or anti-damping multipliers, growth rates

If α and/or $\alpha 1 > 0$ *anti-damped* or *explosive* oscillations occur, when α and/or $\alpha 1 < 0$ amplitude *damped* and when α and/or $\alpha 1 = 0$ amplitude has *regular* oscillations.

In sinusoidal study, we found α very close to zero in conformity with theory. This means that functions of residuals we studied have a *regular* amplitude structure, neither dumped nor oscillates.

Consecutive application of sinusoidal function in the five-parameter form of Gompertz function is considered similar to Fourier series whose parameters change at each stage, not only once as in (21), since the next stage parameters are determined, calculated and included in the function through the residuals formed at previous stage. Statistics and linearity tests are performed as in the growth function in determining the coefficients of each of the sinusoidal functions. Thus, if it satisfies statistical tests of NLIN procedure, the residual series might be included in the main function via auxiliary sinusoidal functions.

Allen wrote:

> "The interpretation of Fourier series is that any type of oscillation with given period $\left(\frac{2\pi}{\omega}\right)$ can be approximated by the sum of *sinusoidal* oscillations by taking a sufficient number of terms in the series… The first of the sinusoidal oscillations has the same period or frequency as *F(t),* the fundamental frequency. The latter sinusoidal oscillations in the sum have shorter periods, or higher frequencies in multiples of the fundamental frequency. Hence, in expressing *F(t)* as a sum of sinusoidal functions, there is a range of decreasing periods $\left(\frac{2\pi}{n\omega}\right)$ …Where the series is stopped is a matter of convenience or closeness of approximation. It is to be noticed that terms can be added without affecting those already written."(1956, p. 121)

5. SAS® Studio Software

In our statistical studies, we used the SAS/STAT NLIN procedure, SAS/ETS ARIMA, AUTOREG and MODEL procedures of "SAS® OnDemand for Academics (SAS® Studio)" which is an advanced web-based SAS development environment for statistical and econometric analysis.

NLIN procedure which is developed for the analysis of nonlinear mathematical functions and provides very wide and flexible results in the statistical applications. Partial derivatives are taken according to the parameters in the function, the linearity of the function is checked with the results found and the relevant linearity criteria are included in the output in addition to known statistical tests: Hougaard's skewness, Box's bias measure and Bates and Watts' global nonlinearity measures are among them[9]. Linearity tests provide close-to linear behavior of functions.(Ratkowsky, 1993)

ARIMA uses autoregressive moving-average errors to perform multiple regression analysis.

AUTOREG implements regression models that use time series data in which the errors are autocorrelated.

MODEL handles nonlinear simultaneous systems of equations, such as econometric models.

Procedure analyzes models in which the relationships among the variables form a system of one or more nonlinear equations. Primary uses of the MODEL procedure are estimation, simulation, and forecasting of nonlinear simultaneous equation models.

PROC MODEL features include the following:

_ SAS programming statements to define simultaneous systems of nonlinear equations

_ tools to analyze the structure of the simultaneous equation system

_ ARIMA, PDL, and other dynamic modeling capabilities

_ tools to specify and estimate the error covariance structure

_ tools to estimate and solve ordinary differential equations

_ the following methods of parameter estimation:

[9] For details of nonlinearity see (Gebremariam, 2014; SAS Institute Inc., 2018b).

– ordinary least squares (OLS)

– two-stage least squares (2SLS)

– seemingly unrelated regression (SUR) and iterative SUR (ITSUR)

– three-stage least squares (3SLS) and iterative 3SLS (IT3SLS)

– generalized method of moments (GMM)

– simulated method of moments (SMM)

– full information maximum likelihood (FIML)

– general log-likelihood maximization

_ simulation and forecasting capabilities

_ Monte Carlo simulation

_ goal-seeking solutions. (SAS Institute Inc., 2018a, pp. 1423–1424)

6. Statistical Application Results of Turkish Population Data

Aiming to research the statistical representability of Turkish population growth through the entire series (1927-2080) with the mathematical curves, now we shall analyze the Turkish population in 2081, as if the population growth has already been completed in the country. Reliable long-term population projections published by TurkStat in 2018 facilitated the work of researchers. With the long-term projections which include sufficient data for the upper asymptotic region of the curves, we had the opportunity to focus on determining parameters, statistical test applications and obtaining statistical test values rather than deal with problems of census data.

6.1. Gompertz Growth Function (G5)

At the first stage of statistical works of the Gompertz growth function (G5), we determined the values of the six parameters ($K,\ L,\ Q_1,\ Q_2,\ r,\ v$) simultaneously together with statistical test results of these parameters. Results are included in the Table 2a, 2b and 2c.

Convergence criterion/iteration phase is successful and the error sum of squares found as 1.883E14, which is quite reasonable amount for this study. Error sum of squares/total sum of squares (1.779E17) is at a level that is at a negligible level as 1058 per million. The time variable is successful in the explanation of population growth and the function/data compatibility is complete. Standard errors of the parameters, F-value, and 95% confidence limits are acceptable.

Hougaard's skewness figures of parameters are below 0.25%, which is the level of acceptance as the absolute value and Box's bias figures are below the acceptable level of 1% and almost zero. Nonlinear growth function of Gompertz is close to linear in case of Turkish population data (Ratkowsky, 1993). Global nonlinearity measures – first four row values of Table 2c are below the critical value of 0.6804 (5th row in the table 2c). Significant statistical values confirm the linearity of the nonlinear generalized function.

According to the Gompertz growth function, the lower asymptote is 15.4 million, and the upper asymptote is 112.7 million people. The figure to be focused on is the upper asymptote. Upper asymptote figure (112.7 million) is above the maximum figure of TurkStat's *main projections* which is 107.7 million (2069 max.). However, the Gompertz curve captured almost all of the TurkStat estimates for 2040-2080 period with zero deviation, except for 0.1 million deviation in 2050 (see table-11 in the "Results and Comments" below). We can say that Gompertz's five- parameter function is successful in identifying the details in the development of the population.

Table 2: SAS NLIN Procedure Statistical Test Results of Gompertz Curve (G5y1y2)

2a) Model Estimates

	Source	DF	Sum of Squares	Mean Square	F Value	Approx. Pr.>F
1	Model	4	1.777E17	4.442E16	35142.6	<.0001
2	Error	149	1.883E14	1.264E12		
3	Corrected Total	153	1.779E17			

2b) Parameter Estimates

	Parameter	Estimate	Approx. Std Error	Approximate 95% Confidence Limits		Hougaard's Skewness	Box's Bias	Percent Bias
1	K	1.1265E8	373008	1.1192E8	1.1339E8	0.1363	13330.1	0.012
2	N	15474709	335398	14811958	16137460	-0.0639	-7386.9	-0.05
3	Q1	7.9392	0.2561	7.4333	8.4452	0.00227	-0.00739	-0.09
4	Q2	2.8475	0.0770	2.6953	2.9997	0.000859	-0.00059	-0.02
5	r	0.0433	0.000644	0.0420	0.0446	-0.1261	-0.00002	-0.05

2c) Global Nonlinearity Measures

1	Max Intrinsic Curvature	0.1068
2	RMS Intrinsic Curvature	0.0562
3	Max Parameter-Effects Curvature	0.2124
4	RMS Parameter-Effects Curvature	0.1188
5	Curvature Critical Value	0.6630
6	Raw Residual Variance	1264E9
7	Projected Residual Variance	249E8

The first stage of the statistical study has provided the results we wanted until now: ESS/TSS figures are 104 and 1058 per million (0.0104% and 0.1058%) for Verhulst V6 and Gompertz G6 respectively. These figures are more than reasonable to work with. There's nothing to worry about. On the other hand, we cannot be satisfied that the analysis is sufficient and results are perfect by saying that time explains 99.9% of the population growth. Although the ESS's figures are low and relatively insignificant as a percentage, the residual figures for some years are large enough to affect the estimates as an absolute figure. Therefore, investigating possibilities of further lowering the absolute levels of ESSs' should be the next important research subject. On the other hand, we will explain the theoretical framework indicated above.

When the residuals curve of G5 is examined (blue line of figure 2), it is seen that the curve shows a structure of sinusoidal trend. The sinusoidal trend moves around the timeline, it does not disappear during the 155-year population growth period, on the contrary its period grows. Last period duration has grown significantly. The upper and lower boundaries of the residuals curve have an interval of -3 and +2 million. Figure 2 shows sinusoidal periods of residuals approximately:

1) 1935-1975: 40 years,
2) 1976-2026: 50 years,
3) 2027-2080: 53 years (three-quarter of period).

First and second periods run forty-fifty years successively, but projections for 2027-2080 period differ with three-quarters being 53 years and the 2060-2080 period is still exploding towards downside. By doing an additional mathematical study, we have to investigate and explain this autonomous behavior which is second constituent of the main function.

When we include the sinusoidal functions of the residual elements which are two in our case, the main growth function becomes as follows:

$$Y_{(t-T)} = y1_{(t-T)} + y2_{(t-T)} + Y_0 + \frac{K - Y_0}{\left[e^{\left[w_{(t-T)}A + (1-w_{(t-T)})B\right]}\right]} \quad (23)$$

Note: y1 applied as eight-parameter, then y2 as four-parameter respectively.

y's are sinusoidal functions that are independent of trend of growth function. They have their own growth coefficients. The last part of the equation is the Gompertz growth function driven by intrinsic rate of growth. During the application of sinusoidal function, linearity of the main function was almost not affected by changes, making it easier to work.

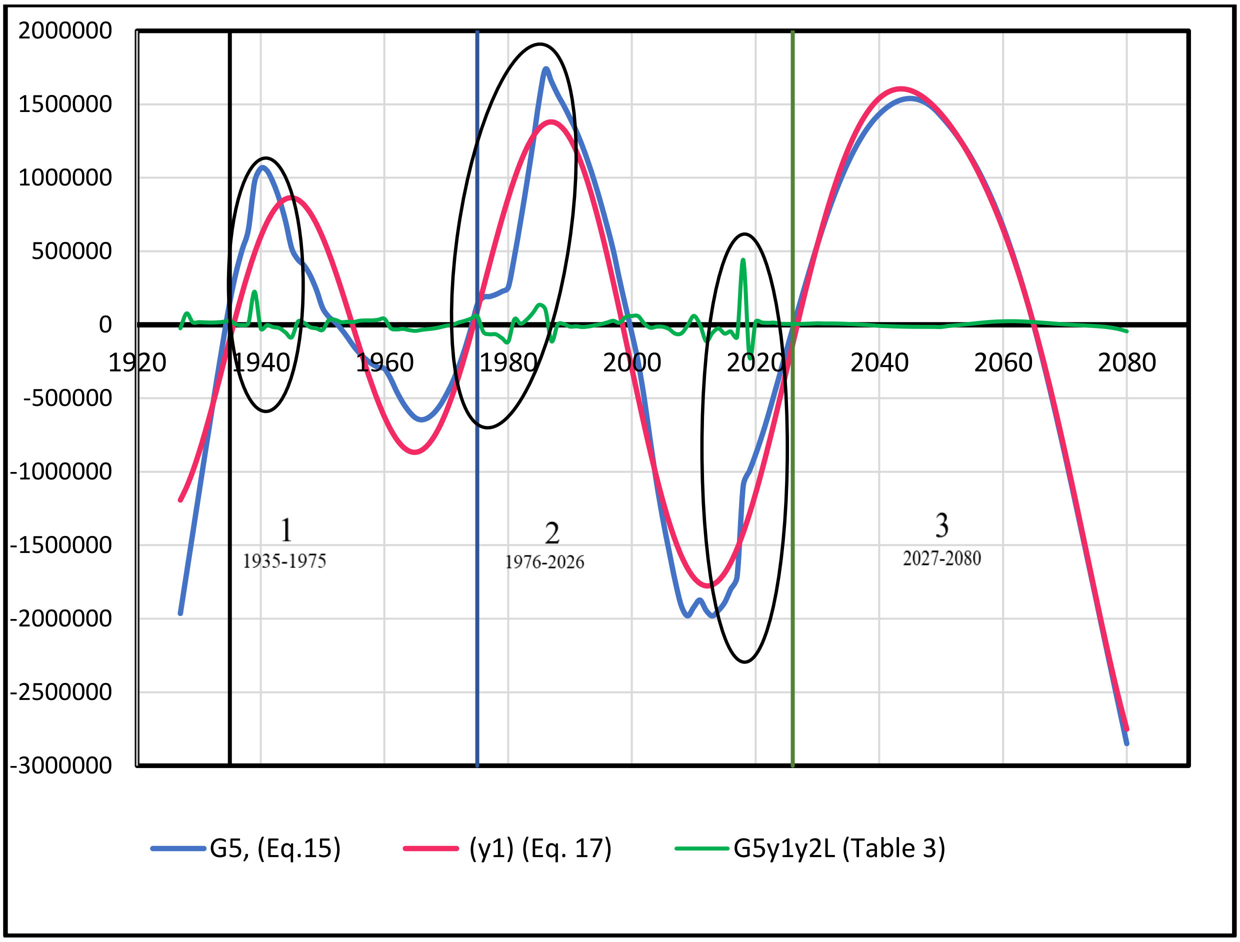

Figure 2: Residuals Curves of Five-parameter Gompertz Curve Family

Note to Figure 2: Black circles show data inconsistencies.

At second stage, curve fitting was made for the residuals series of the main equation obtained at first stage and y1- equation of the residuals- was obtained as a complementary function[10].

[10] We found that with eight parameters (equation 7) in the first stage then with four parameters in second stage (equation 8) gave more advantageous results in statistical application.

$$y1_{(t-1951)} = -366560e^{(0.0167(t-1950))}\cos(0.0914(t-1951)-0.4816) + 1163610e^{(-0.00678(t-1950))}\cos(0.1443(t-1950)+0.6589) \quad (24)$$

At third stage, trend function (G5) plus y1 were combined to obtain the statistical results of the G5y1 function.

At fourth stage, y2 function was obtained through the residuals of the G5y1 function.

$$y2_{(t-1950)} = -414476e^{(2.6762(t-1950))}\cos(0.2225(t-1950)-0.3572) \quad (25)$$

At fifth stage, G5y1y2 was run together to achieve its statistical results.

In five stages, positive and satisfactory results were obtained from the statistical and linearity tests of both the main equation and complementary equations.

The application was terminated when the study on the residuals of the G5y1y2 equation did not achieve sufficient linearity results. However, an important step has already been taken to reduce residuals with $y1$ and $y2$ in these two-stage applications (as main and complementary functions).

After obtaining the main function together with complementary functions, the results of general statistical validity were investigated at the sixth stage.

At the statistical studies of G5y1y2, we applied the following methods of parameter estimation: Nonlinear Ordinary Least Squares (NOLS), Nonlinear Seemingly Unrelated Regression (NSUR) and Nonlinear Full Information Maximum Likelihood (NFIML). Since the results of the other two methods are very close to each other and to NFIML, we have only given the results of NFIML below to save some space in manuscript.

On the relationship between population development and time yielding autoregressive results, the model was redefined and the residuals of population's series -lag1 and lag2 -were added to the function as explanatory lagged variables and then desired results were reached with three explanatory variables and seven parameters. The final state of the function is as follows. The results were collected in thirteen subheadings in five tables and two figures.

$$Y_{(t-T)} = y1_{(t-T)} + y2_{(t-T)} + Y_0 + \frac{K - Y_0}{\left[e^{[w_{(t-T)}A + (1-w_{(t-T)})B]}\right]} + X1 * lag1(\hat{Y} - Y) + X2 * lag2(\hat{Y} - Y) \quad (26)$$

Table 3: Main Statistical Test Results of SAS Software Proc MODEL Procedure

3a) Nonlinear FIML Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Durbin Watson
Population	7	147	4.914E11	3.3425E9	57814.6	1.0000	1.0000	-
Residual Population	-	147	4.898E11	3.3317E9	57720.5	-	-	2.1747

3b) Nonlinear FIML Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|	Label
K	1.1263E8	231050	487.46	<.0001	Upper asymptote, at most level of population
N	15196918	108975	139.45	<.0001	Starting value
Q1	7.848847	0.0784	100.16	<.0001	First coefficient of exp growth base
Q2	2.794677	0.0314	89.01	<.0001	Second coefficient of exp growth base
r	0.043219	0.000232	186.03	<.0001	Intrinsic growth rate of population
Population_l1	1.228335	0.0441	27.84	<.0001	AR(Population) Population lag1 parameter
Population_l2	-0.46892	0.0722	-6.50	<.0001	AR(Population) Population lag2 parameter

3c) FIML Log Likelihood: -1903

Note: Population_l1: lag 1 variable. Population_l2: lag 2 variable.

The t-values of parameters are significantly above the t-table values and estimates are valid. The probabilities of invalidity of estimates are less than one in ten thousand.

SSE/TSS (4.898E11/1.779E+17) (2.75 per million) is at the acceptable level. The Durbin Watson coefficient indicates that there is no autocorrelation.

SAS Institute writes:

> The option FIML (full information maximum likelihood) requests full information maximum likelihood estimation. If the errors are distributed normally, FIML produces efficient estimators of the parameters. If instrumental variables are not provided, the starting values for the estimation are obtained from a SUR (seemingly unrelated regression) estimation. If instrumental variables are provided, then the starting values are obtained from a 3SLS (three-stage least squares) estimation. The log-likelihood value and the l2 (lag 2) norm of the gradient of the negative log-likelihood function are shown in the estimation summary. (SAS Institute Inc., 2018a, p. 1500)

Table 4: Parameter Wald and Likelihood Ratio 95% Confidence Intervals

		a) Parameter Wald 95 % Confidence Intervals		b) Parameter Likelihood Ratio 95% Confidence Intervals	
Parameter	Value	Lower	Upper	Lower	Upper
K	1.1263E8	1.1218E8	1.1308E8	1.1257E8	1.1269E8
N	15196918	14983331	15410506	15170220	15223617
Q1	7.8488	7.6953	8.0024	7.8296	7.8680
Q2	2.7947	2.7331	2.8562	2.7870	2.8024
r	0.0432	0.0428	0.0437	0.0432	0.0433
Population_l1	1.2283	1.1419	1.3148	1.1419	1.3148
Population_l2	-0.4689	-0.6104	-0.3274	-0.4866	-0.4512

Note: Population_l1: lag 1 variable. Population_l2: lag 2 variable.

Table 5: Collinearity Diagnostics and Heteroscedasticity Test

5a) Collinearity Diagnostics

			Proportion of Variation						
Number	Eigenvalue	Condition Number	K	N	Q1	Q2	r	Pop_l1	Pop_l2
1	3.103608	1.0000	0.0102	0.0116	0.0033	0.0053	0.0021	0.0000	0.0001
2	1.830841	1.3020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0842	0.0841
3	1.452319	1.4618	0.0075	0.0437	0.0009	0.1158	0.0012	0.0000	0.0000
4	0.299204	3.2207	0.2499	0.0682	0.0299	0.1239	0.0007	0.0020	0.0002
5	0.168533	4.2913	0.0010	0.0000	0.0002	0.0000	0.0000	0.9129	0.9151
6	0.131736	4.8538	0.0918	0.7735	0.0272	0.7365	0.0097	0.0007	0.0000
7	0.013760	15.0184	0.6396	0.1031	0.9385	0.0185	0.9863	0.0001	0.0004

5b) Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Pop	White's Test	49.32	29	0.0107	Cross of all vars
	Breusch-Pagan	2.28	2	0.3192	1, s, Population

Condition numbers of Table 5a are well below dangerous level of thresholds. Authors consider levels 10, 30 and 100 as thresholds. In general researchers agree that no single number can handle all situations. There is no collinearity in our study according to the thresholds 30 and 100. By threshold 10, the K, Q1, and r parameter values exceed 0.5 in 7th row (conditional number: 15.02 >10). But, values of these parameters in 1-6 rows are very close to zero. Generally, we can say that there is no collinearity in the population study.

One assumption of regression analysis is heteroscedasticity. Accordingly, the variance of errors in observations are constant. Although the variance of errors is not distributed exactly normally in our study, our tests appear to be valid enough.

Table 6: Godfrey's Serial Correlation Test, Durbin-Watson Statistics and Normality Test Tables

6a) Godfrey's Serial Correlation Test

Equation	Alternative	LM	Pr>LM
Pop	1	5.74	0.0165
	2	6.97	0.0306
	3	7.53	0.0567

6b) Durbin-Watson Statistics

Equation	Order	DW	Pr<DW	Pr>DW
Pop	1	2.17	0.7430	0.2570
	2	1.81	0.0648	0.9352
	3	1.95	0.3209	0.6791
	4	1.93	0.3034	0.6966

6c) Normality Test

Equation	Test Statistic	Value	Prob
Pop	Shapiro-Wilk W	0.71	<.0001
System	Mardia Skewness	236.5	<.0001
	Mardia Kurtosis	63.67	<.0001
	Henze-Zirkler T	12.16	<.0001

SAS Institute Inc. writes:

> The GODFREY= option in the FIT statement produces the Godfrey Lagrange multiplier test for serially correlated residuals for each equation…The three variations of the test reported by the GODFREY=3 options are designed to have power against different alternative hypothesis. Thus, if the residuals in fact have only first-order autocorrelation, the lag 1 test has the most power for rejecting the null hypothesis of uncorrelated residuals. If the residuals have second- but not higher-order autocorrelation, the lag 2 test might be more likely to reject; the same is true for third-order autocorrelation and the lag 3 test. (SAS Institute Inc., MODEL Procedure, 2018a, p. 1531)

The generalized Durbin-Watson Statistics are above p-values in all four levels.

The null hypothesis of normality test of the Shapiro-Wilks assumes that residuals distributed normally. In the preparatory phase of our study, although we supported main function with complementary functions, inconsistencies from the data on the residuals have not been lost. In fact, the outliers have become even more obvious in the end. It is clear that the final series of residuals in Figure 2 is not normally distributed, and the outliers in the black circles specifically indicate this situation. The first two is due to the figures of the 1940 and 1986 censuses, and the last one is due to an outlier emerging from difference between 2017 census figure and projection figure of 2018. We chose not to do any corrective work with the data. However, arbitrariness did not disappear completely after complementary function studies. Although inconsistencies are still valid, the probability of the hypothesis that the normality test is invalid is less than one in ten thousand.

The covariance and correlation values of the parameters in Table 5a and b respectively are also at reasonable levels: only K and N highly interact with other parameters in the covariance table. There are three values in the correlation table that exceed 0.5: A high correlation of r with K and Q1 on the one hand and Q2 and N appear on the other. The figures are not sufficient high to adversely affect the validity of the model.

SAS Institute Inc. proposes: “Hausman’s m-statistic can also be used, in principle, to test the null hypothesis of normality when comparing 3SLS to FIML. Because of the poor performance of this form of the test, it is not offered in the MODEL procedure. For a discussion of why Hausman’s test fails for common econometric models, see Fair (1984, pp. 246–247)” (SAS Institute Inc., 2018a, p. 1556).

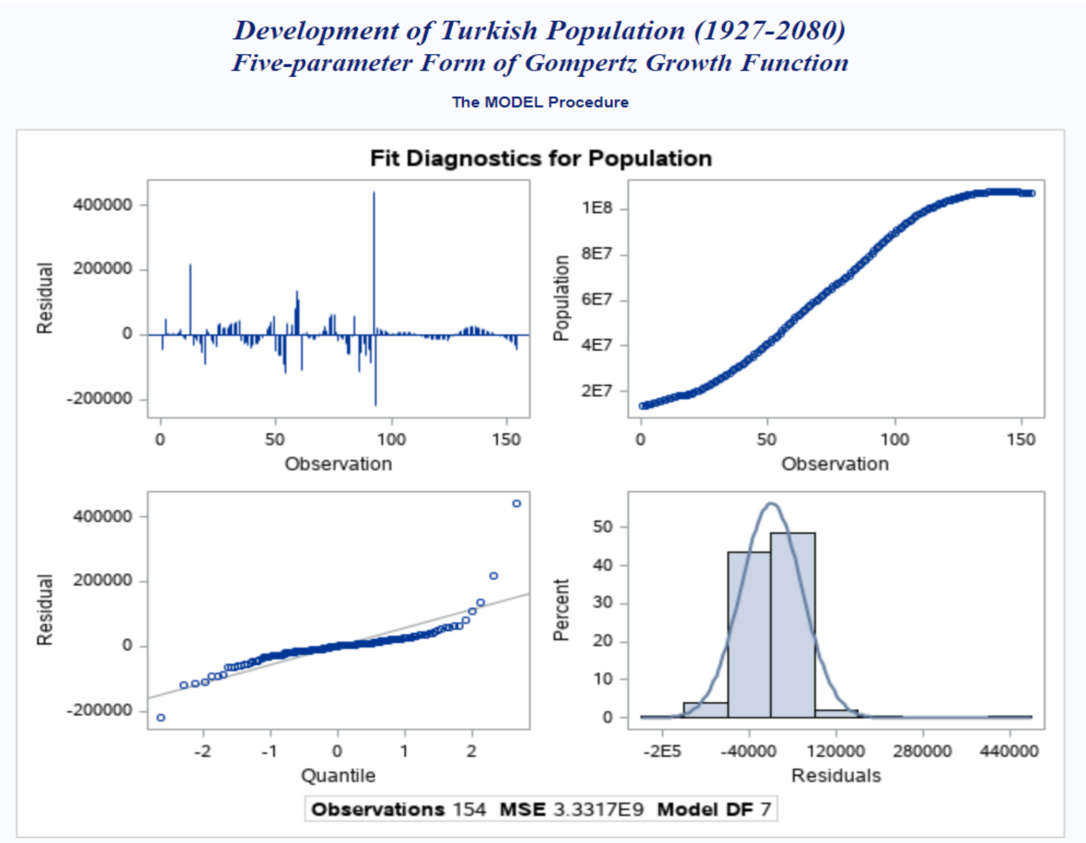

Figure 3: Fit Diagnostic Figures of Gompertz Function

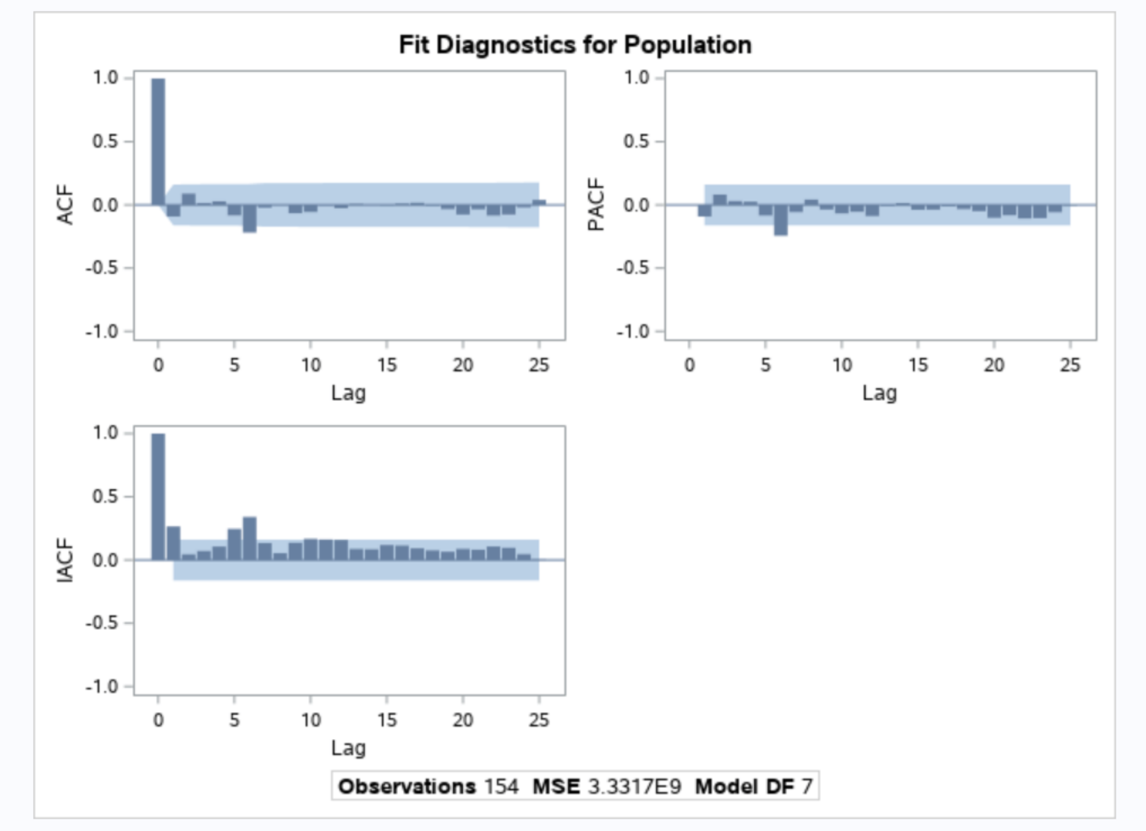

Figure 4: Fit Diagnostic Figures of Gompertz Function

Table 7: Covariances and Correlations of Parameter Estimates

7a) Covariances of Parameter Estimates

	K	N	Q1	Q2	r	Pop_l1	Pop_l2
K	5.3384E10	-2.4356E9	-6737.432	421.966	-39.75519	-183.4953	-1917.902
N	-2.4356E9	1.1876E10	3666.762	2839.149	5.79392	-420.9370	710.099
Q1	-6737.432	3666.7617	0.006	0.000	0.00002	0.0002	0.002
Q2	421.96617	2839.1488	0.000	0.001	0.00000	-0.0001	0.000
r	-39.75519	5.7939243	0.000	0.000	0.00000	0.0000	0.000
Pop_l1	-183.4953	-420.937	0.000	-0.000	0.00000	0.0019	-0.001
Pop_l2	-1917.902	710.09938	0.002	0.000	0.00001	-0.0013	0.005

7b) Correlations of Parameter Estimates

	K	N	Q1	Q2	r	Pop_l1	Pop_l2
K	1.0000	-0.0967	-0.3721	0.0582	-0.7406	-0.0180	-0.1150
N	-0.0967	1.0000	0.4294	0.8298	0.2288	-0.0876	0.0903
Q1	-0.3721	0.4294	1.0000	0.1836	0.8720	0.0455	0.4336
Q2	0.0582	0.8298	0.1836	1.0000	0.0064	-0.1010	0.0755
r	-0.7406	0.2288	0.8720	0.0064	1.0000	0.0253	0.4750
Pop_l1	-0.0180	-0.0876	0.0455	-0.1010	0.0253	1.0000	-0.4001
Pop_l2	-0.1150	0.0903	0.4336	0.0755	0.4750	-0.4001	1.0000

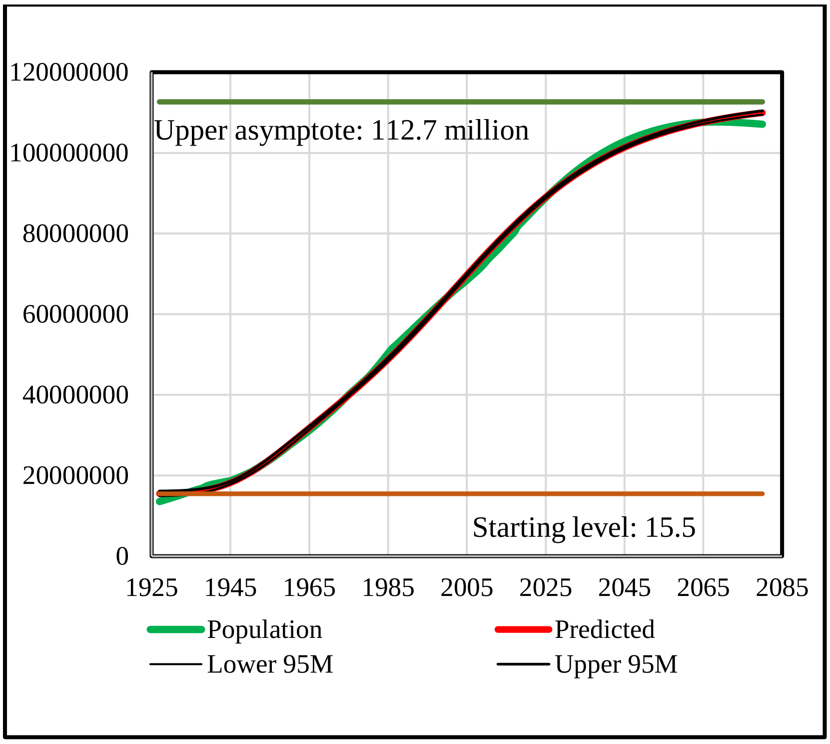

Figure 5: Population Curve of Five-parameter Gompertz Growth Function (G5)

After two complementary applications, ESS of main function (G5y1y2) which is 2.053E+12 reduced to 0.011% of the main function (G5) ESS value of 1.883E+14 (table 6). Gain is 1.862E+14. In this way, a further curve representation was realized within the narrower limits of confidence. We see that population growth curve, which has not yet reached the upper asymptote (112.6 million) from the path, is leveled off with the TurkStat projections (107 million) in the upper asymptote region. After application of sinusoidal equations, when we examine the Figure 3, we will see that the range of residuals has reduced from an interval of -3 and +2 million to almost interval of less than plus-minus two hundred thousand, and the regular trend of oscillations before 2000 have almost been disappearing after 2018 till 2080 (projections period of TurkStat) except for the tail which indicates population decrease estimate of TurkStat.

Table 8: Four-parameter (G4) Summary and Complementary Function Application Results of Five-parameter Gompertz Function (G5)

Equation name	G4	G5	G5y1	G5y1y2	G5y1y2L
Equation	-	-	-	Table 2	Table 3
Parameters	(K, L, Q, r)	(K, L, Q_1, Q_2, r)			
Upper asymptote (million)	119.1	112.7	112.6	112.6	112.6
Starting value (million)	15.3	15.5	15.4	15.2	15.2
Error sum of squares	4.111E+14	1.883E+14	7.671E+12	2.053E+12	4.898E11
Total sum of squares	1.779E+17	1.779E+17	1.779E+17	1.779E+17	1.779E+17
ESS/TSS (per million)	2311	1,058	43	12	2.76
ESS as % of four-parameter ESS	100	45.80	1.87	0.50	0.001191

Although we have obtained almost same results in G6 works- the 6-parameter Gompertz function (eq. 14)- we did not include results in our manuscript to save space because the ESS figure of G6 is already very close to the G5 ESS figure.

6.2. Verhulst Logistic Growth Function (V5 and V6)

In this section, we have decided to provide summary information about the Verhulst V6 function. Positive results were obtained in other test tables that we didn't provide. All comments on the Gompertz function also apply here. As can be seen from the pivot tables, the Verhulst function is more successful in capturing 108 million, the highest level of TurkStat estimates.
Complementary functions of Verhulst V6:

$$\hat{y}1_{(t-1960)} = 393510\mathrm{e}^{(-0.0130(t-1960))}\cos(-0.2071(t-1960)+0.9723) \quad (27)$$

$$\hat{y}2_{(t-1960)} = -221926\mathrm{e}^{(0.00515(t-1960))}\cos(-0.1529(t-1960)-0.0483) \quad (28)$$

$$\hat{y}3_{(t-1960)} = -0.00066(\mathrm{t}-1960)^5 + 0.1611(\mathrm{t}-1960)^4 - 11.6469(\mathrm{t}-1960)^3 + 141.1(\mathrm{t}-1960)^2 + 8153.2(\mathrm{t}-1960) - 93450.2 \quad (29)$$

Statistical tables of V6y1y2y3L:

Table 9: Statistical Test Results of Verhulst (V6) Growth Function

9a) Nonlinear FIML Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Durbin Watson
Population	8	146	5.008E11	3.4303E9	58568.9	1.0000	1.0000	
Residual Population		146	5.014E11	3.4343E9	58603.1			2.1857

9b) Nonlinear FIML Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|	Label
Q1e	6.417653	0.2307	27.82	<.0001	Log of first coefficient of exp growth base
re	-2.33339	0.0300	-77.74	<.0001	Log of intrinsic growth rate of population
Q2e	4.615833	0.1828	25.25	<.0001	Log of second coefficient of exp growth base
v	3.048595	0.1170	26.05	<.0001	Allometric coefficient
L	5158064	277384	18.60	<.0001	Lower asymptote, at least level of population
K	1.0797E8	186188	579.91	<.0001	Upper asymptote, at most level of population
Population_l1	1.260139	0.0465	27.09	<.0001	AR(Population) Population lag1 parameter
Population_l2	-0.48668	0.0767	-6.35	<.0001	AR(Population) Population lag2 parameter

a) FIML Log Likelihood: -1905

Equivalences of log parameters:

$$Q_1 = e^{(\text{Q1e})} = e^{6.417653} = 612.5637$$
$$Q_2 = e^{(\text{Q2e})} = e^{4.615833} = 101.072$$
$$r = e^{(\text{re})} = e^{-2.33339} = 0.096976$$

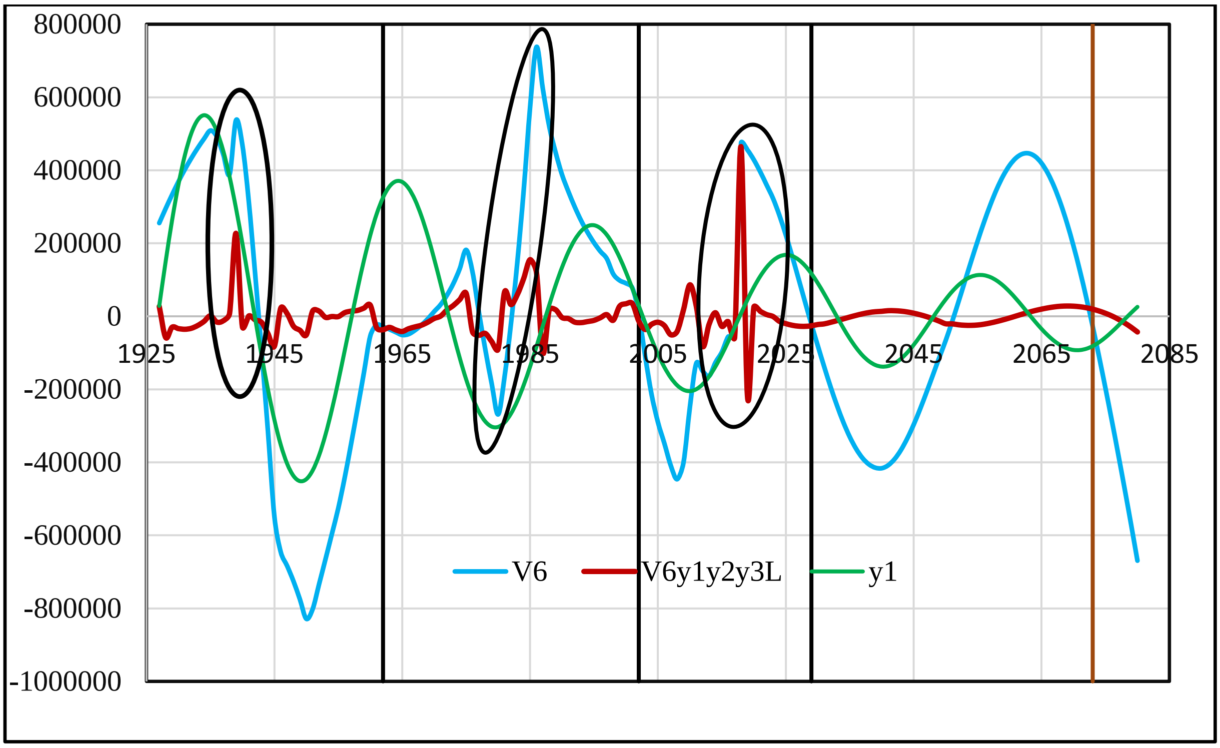

Figure 6: Residual Curves of Verhulst Function Family

Figure 6 note: Data inconsistencies of Figure 2 appear here as well.

Table 10: Summary Statistics of Verhulst Logistic Growth Functions

Equation name	5-prm V5	6-prm				
		V6	V6y1	V6y1y2	V6y1y2y3	V6y1y2y3L
Parameters	(K, L, Q_1, Q_2, r)	(K, L, Q_1, Q_2, r, v)				
Upper asymptote(million)	110.1	108.0	108.0	108.0	108.0	108
Lower asymptote (million)	9.7	5.8	5.4	5.1	5.1	5.2
Error sum of squares	1.095E+14	1.858E+13	1.109E13	3.540E+12	2.303E+12	5.014E+11
Total sum of squares	1.779E+17	1.779E+17	1.779E+17	1.779E+17	1.789E+17	1.789E+17
ESS/TSS (per million)	616	104	62	20	13	2.8
ESS as % of five-prm ESS	100	17	10	3.2	2.1	0.46
ESS as % of six-prm ESS	-	-	59.7	19.1	12.4	2.7

Prm: parameter.

7. Results and Comments

Even though we did not have data for the upper asymptotic region (*K*) of the mathematical curves in our previous population study (İskender, 2018), we made population estimates with ten mathematical functions that we defined and statistically implemented. As a result of this, we were not able to determine the statistical test values for the parameters of the lower and upper asymptotic regions (*L* and *K*), since we followed the minimization of the sum of squares by taking the lower and upper asymptotic parameters externally. In other words, we set our expectations for the mathematical functions we applied somewhat strongly high. However, we see that the results obtained[11] in this study are confirmed by the results obtained here in this study of TurkStat projections. The results of the study met our expectations. Turkey's population is expected to reach 108-114 million by the end of the twenty first century.

Both the known statistical and linearity test results obtained in this study demonstrate that it is possible to represent the *main scenario*. The projection methods used by Institute overlap with the mathematical modeling applied here. In the summary, Table 11 predictions for the years selected from the *main scenario* projections are given together with the estimates obtained from the six-parameter Verhulst and Gompertz functions. Although the TurkStat projections have foreseen the decline in the last decade of forecasts, the six-parameter logistic and Gompertz growth functions almost catch up with the annual estimates of the Institute.

Table 11: Population Figures Comparison for Selected Years (million)

Selected Years	Main Scenario Projections of TurkStat	Verhulst (V6y1y2y3L)	Difference of Verhulst	Gompertz (G5y1y2L)	Difference of Gompertz
2020	83.9	83.8	0.1	83.9	0.0
2030	93.3	93.4	-0.1	93.3	0.0
2040	100.3	100.3	0.0	100.3	0.0
2050	104.7	104.8	-0.1	104.8	-0.1
2060	107.1	107.1	0.0	107.1	0.0
2070	107.6	107.6	0.0	107.6	0.0
2080	107.1	107.1	0.0	107.1	0.0

We may summarize results of the econometric study as follows:

Our econometric analysis is fully integrated with population theory, economic modeling and statistical applications. These three aspects of the article were especially taken into account during this study. We think that it is the first study in our knowledge of mathematical economics and applied statistical methods. The development of Gompertz function as a population growth function with five and six parameters and its application in long-term population statistics data is the first study with successful results. Preferred Verhulst and Gompertz growth functions and Allen's approach of mathematical economics analysis have produced successful results in the Turkish population data. In this statistical study, reducing residuals with sinusoidal approach played an important role in achieving successful results from statistical criteria. Our analysis of TurkStat's 1927-2018 population data and 2018-2080 population estimates are a new approach to address in econometric study with the methods described.

Statistical applications and solutions that cannot be done manually, in a short time and accurately, have been obtained and achieved with the opportunities offered by SAS Studio software which is an advanced econometrics and statistics online environment. There was no way to carry out our work with ordinary softwares. On the other hand, only a small number of the obtained tables and results were included to the article for the sake of brevity. The software we use has much more possibilities. New statistical tests developed in the last fifty years namely;

[11] Upper asymptote values of 2018 study are: Four-parameter logistic 94000000, six-parameter logistic 115000000, four-parameter Gompertz 113800000.

Hougaard, Box and Bates, Ratkovsky etc. linearity tests, Godfrey serial correlation test, normality tests, White and Breusch-Pagan and others were extensively used in our study. Our software allows this. The linearity tests we mentioned are not available in any software other than SAS Studio. And it is necessary for a healthy application of mathematically nonlinear growth functions. Advanced statistical applications have been made possible with this software. All contributions of this software have developed this manuscript in the direction of both the purpose and results of our article.

TurkStat's 2018-2080 population projections are a valuable and important source to be considered by policy makers involved in population policies of Turkey. With its approach, the institution has achieved a significant success not only in preparation and publishing the statistics of the past years, but also with its future works: we can perceive this approach as a first in the *back to the future* way of understanding.

Five and six-parameter forms (inc. sinusoidal) of all growth curves applied in this study can be taken as a reference for population analysis of Turkey.

Econometric study has confirmed that Turkstat's *main scenario* projection is *consistent* in terms of statistical evaluation. All statistical test results of functions and curves successfully represent the Turkish population growth and its projections.

At the statistical analysis of economic categories such as income, demand, population, etc., the explanation of the residuals should be included in the study as well as the explanation of the main trend. This approach makes the results more valuable. Application of sinusoidal functions has led to significant decreases in ESS values and increased representation of curves, while it has highlighted features that were not appear in the trend of the population. My first excerpt from Allen (subdivision 7.1) explains the theoretical basis of state of affairs from mathematical economics to statistics. This is the method to be followed in econometrics.

Although sinusoidal approach reduced absolute level of residuals at a great extent, linearity measures are left unchanged which is an advantage of applied methodology.

It is possible to represent Turkish population development not with simple regression analyses, but with advanced econometric models, statistical applications and software in accordance with the growth theory taken from Allen, Verhulst and Gompertz.

From 2017 to 2080 population growth figure of 26.3 million is very close to Turkey's population level of 27.7 million people in 1960. In other words, in the next half century Turkey's population figure close to that of 1960 will be added to the 2017 population.

Considering Turkey's current political, social, economic and geographical problems, it is necessary to consider the planning for the next fifty years from today onwards.

It is clear that Turkey will need significant capital accumulation and investment when the economic demands of the future population to live by at least today's standards, not to mention of standards of nineteen sixties.

It is difficult to plan what kind of geography the future population of 26.3 million will live in. It is quite obvious that Turkey will face significant settlement problems when considering the earthquake structure, water resources, the settlement difficulties brought by the mountainous structure etc. It will be an important task of the public due to the fact that horizontal or vertical urbanization approaches are widely discussed.

It is inevitable that the population's settlement area demands will conflict with the needs for agricultural land.

Turkey is a state that has successfully protected its borders and geography throughout the history of the Republic. It is realistic to continue this assumption for future. Inevitably, population growth, problems and solutions will be shaped and located within these borders and framework.

References

Allen, R. G. D. (1956). *Mathematical Economics*. London: Macmillan & Co Ltd.

Allen, R. G. D. (1969). *Mathematical Analysis for Economists*. London, UK: Macmillan and Co. Ltd. (Original work published 1938).

Canpolat, S. B. & Gök, M. G. (2020). Census Methodology in Turkey: Transformation from Traditional to Register-based Administrative Data. *Statistical Journal of the IAOS*, *36*(1), 177–185. https://doi.org/10.3233/SJI-190561

Gebremariam, B. (2014). *Is Nonlinear Regression Throwing You a Curve? New Diagnostic and Inference Tools in the NLIN Procedure* (No. Paper SAS384-2014). Retrieved from https://support.sas.com/resources/papers/proceedings14/SAS384-2014.pdf

İskender, C. (2018). Türkiye Nüfus Büyümesi ve Tahminleri: Matematiksel Büyüme Modelleri ve İstatistiksel Analiz ile Kuramsal ve Uygulamalı bir Yaklaşım. *Istanbul University Econometrics and Statistics E-Journal*, *14*(28), 75–141. https://doi.org/10.26650/ekoist.2018.14.28.0004

İskender, C. (2021). Mathematical Study of the Verhulst and Gompertz Growth Functions and Their Contemporary Applications. *Ekoist : Journal of Econometrics and Statistics*, (34). Retrieved from https://dergipark.org.tr/tr/pub/ekoist/issue/64260/876749

Piaggio, H. T. (1920). *An Elementary Treatise on Differential Equations and Their Applications* (W. P. Milne, ed.). London: G. Bell and Sons, Ltd.

Ratkowsky, D. A. (1993). Principles of Nonlinear Regression Modeling. *Journal of Industrial Microbiology*, *12*(3–5), 195–199. https://doi.org/10.1007/BF01584190

Richards F. J. (1959). A Flexible Growth Function for Emprical Use. *Journal of Experimental Botany*, *10*(2), 290–300.

Ricketts, J. H. & Head, G. A. (1999). A Five-Parameter Logistic Equation for Investigating Asymmetry of Curvature in Baroreflex Studies. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, *277*(2), R441–R454. https://doi.org/https://doi.org/10.1152/ajpregu.1999.277.2.R441

SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide The NLIN Procedure (Chapter). In *SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.* Retrieved from https://support.sas.com/documentation/onlinedoc/stat/930/whatsnew.pdf%0A

SAS Institute Inc. (2013). SAS/STAT ® 13.1 User's Guide The NLIN Procedure. In *SAS Institute Inc. 2013. SAS/STAT® 13.1 User's Guide. Cary, NC: SAS Institute Inc*. Retrieved from https://support.sas.com/documentation/onlinedoc/stat/131/nlin.pdf

SAS Institute Inc. (2018a). SAS/ETS® 15.1 User's Guide The MODEL Procedure. In *SAS Institute Inc. 2018. SAS/ETS® 15.1 User's Guide. Cary, NC: SAS Institute Inc. SAS/ETS®* (2018th ed.). Retrieved from https://support.sas.com/documentation/onlinedoc/ets/151/model.pdf

SAS Institute Inc. (2018b). SAS/STAT® 15.1 User ' s Guide The NLIN Procedure. In *SAS Institute Inc. 2018. SAS/STAT® 15.1 User's Guide. Cary, NC: SAS Institute Inc.* Retrieved from https://support.sas.com/documentation/onlinedoc/stat/143/nlin.pdf

Tenenbaum, M., & Pollard, H. (1963). *Ordinary Differential Equations*. New York: Dover Puplications.

TurkStat. (2021). Population and Dempgraphy Statistics. Retrieved from https://data.tuik.gov.tr/Kategori/GetKategori?p=Population-and-Demography-109

Verhulst, P.-F. (1838). Notice Sur La Loi Que La Population Suit Dans Son Accroissement. In A. Quetelet & J. G. Garnier (Eds.), *Correspondance Mathématique et Physique* (Vol. 4, pp. 113–121). Retrieved from https://archive.org/details/correspondancem02belggoog/page/n122/mode/2up