

ESBSD: AN ESSAY ON THE NEW EXPONENTIAL SMOOTHING METHODOLOGY APPLIED TO THE PROJECTION OF THE POPULATION OF BELO HORIZONTE

Daniel Henrique Pereira¹

¹ Student, Business Administration, Pontifical Catholic University of Minas Gerais (PUC Minas)

E-mail: researchdh.pereira@gmail.com

Orcid: <https://orcid.org/0000-0003-4750-9659>

Abstract: Population projection is essential for the public and private sectors, because it is possible to estimate the number of people that will be served by infrastructure works and services. In this paper, a mathematical model based on alternative exponential smoothing was presented, tested and approved in thousands of real case studies in different fields of study. Using this method and after obtaining data from previous years, projections were raised for the years 2022 to 2034 for the city of Belo Horizonte, Minas Gerais, Brazil.

Keywords: Population Projection; Exponential Smoothing; Time Series; Mathematics; Belo Horizonte

I INTRODUCTION

According to Abubakar and Usaini (2019) population projection is essential for public and private policy planning, as it helps to make decisions about the resources needed to meet the demand of a specific geographical location. When it comes to the public sector, population projection, when done well, will meet the needs of the population through infrastructure works, avoiding unnecessary costs with projections overestimating actual values and projections underestimating actual population values that imply the fact a certain majority of the population will not have their needs fully met. Therefore, we must find a "balance" a method that minimizes errors.

Generally, when we discuss population projections, we also turn to a concept called Time Series. As Adhikari and Agrawal (2013) say, time series is a dynamic field of study in which its main goal, is to collect data from past observations and through them identify patterns that can be useful for making future predictions.

According to Gawatre et al. (2016) we have different methodologies and each of them can be useful at certain stages of population dynamics. In this sense, this paper aims to present an alternative mathematical model for projecting population growth in the city of Belo Horizonte, Brazil. Through data obtained by PopulationStat, projections for the years 2022 to 2034 will be raised.

II MATERIALS AND METHODS

The author collected data from 2017 to 2021 from the Belo Horizonte population present in the data provided by the PopulationStat platform and also data on the future population of Belo Horizonte obtained through projections registered by the same platform. As a quantitative measure of the prediction error, the Absolute and Relative Error was considered.

III ESBSD METHODOLOGY

3.1 Population of the city of Belo Horizonte between 2017 to 2021

Historical series of the population of Belo Horizonte for the years 2017 to 2021 was according to the data collected through the PopulationStat platform that collects its data from official sources of each country and other data such as World Bank and United Nations.

Table 1

Population of the city of Belo Horizonte between 2017 and 2021, data by PopulationStat

Time	Years	Population	Growth Rate ($\pm\Delta\%$)
1	2017	5,899,000	-
2	2018	5,972,000	+1.24%
3	2019	6,028,000	+.94%
4	2020	6,084,000	+.93%
5	2021	6,140,000	+.92%

3.2 Understanding ESBSD Methodology

In 2018, I have been writing for the first time a draft about the methodology will be presented ahead due to some content I had as a business student. Only now has it been possible to present it. The ESBSD model, which stands for "*Exponential Smoothing Based on Standard Deviation*", is a forecasting model applied to environments in which the observed variable has a constant growth or decay and with similar behavior as Delmas (2004) states about the so-called *saturation population* (Ps) concept was introduced by Pierre François Verhulst through his proposal with Logistic Regression.

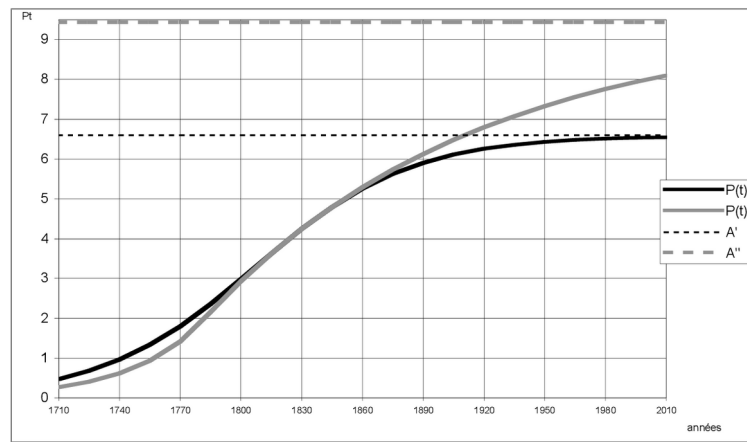


Figure 1. Logistic Regression and the concept of Saturation Population by Pierre François Verhulst (DELMAS, 2004)

The motivation behind the presentation of this model is the fact in certain sets of data present in a time series with characteristics of constant growth or decay trend, even choosing the lowest or the highest value of the constants, usually lies between 0 and 1, it is still not possible for the future projections F_{t+n} to follow real values. In this sense, ESBSD presents itself as a way to minimize these issues by "widening" a range of possible future outcomes.

As an introduction, this model can be represented by the equation below:

$$F_{t+1} = \gamma(A_{t-1}) + (1 - \gamma) \sqrt{\frac{A_{t-1}}{\frac{\sum x_i}{n}}} A_{t-1}$$

Where,

F_{t+1} = Future value

γ = Weighting factor between 0 and 1

A_{t-1} = Last analyzed period

$n = "n"$ past periods

Using this method, we should note that on one side of the formula there is the last period $t-1$ multiplied by an "arbitrary" weighting factor between 0 and 1, and on the other side, there is the multiplication of this same weighting factor by the square root of the "basic growth/decline rate" which is given by the ratio between the last period analyzed with the average of " n " periods of past observations and then multiplying again by the last period we find a relative value of what may occur in the future.

Besides what was previously mentioned, in the ESBSD there is also a "Complementary Factor \pm " (CF_{\pm}) in which the result found will be added or subtracted from the results obtained through the calculation performed by the CF_{\pm} .

Through this CF_{\pm} present in the formula is derived its name (ESBSD), since, in a historical series if the values of the n periods observed are dispersed, we will have a high standard deviation, consequently when we subtract by the difference between the mean and each of its component values, we will also have a relatively high CF_{\pm} value. Similarly, if we take into account a historical series in which data are less scattered around the mean, then the CF_{\pm} will also have a relatively lower value. Clear and understandable the behavior of the standard deviation of an analyzed data set together with the weighting factor, strongly influences the future results we want to project.

In summary, if by historical series we have an environment with constant growth, it will be considered CF_{+} (a positive value) in a similar way if we have an environment with constant decay we will soon consider a CF_{-} (a negative value).

Below is the CF_{\pm} formula:

$$CF_{\pm} = \left[\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} - ABS \left(ABS \frac{\sum_{i=1}^n x_i - A_{t-n}}{n} \right) \right]$$

This CF_{\pm} consists of subtracting the same standard deviation from n time periods by the difference between the mean and each of its component values from this same n periods considered. Absolute Value Function (ABS) in this case is intended to balance the data in the

series, i.e. preventing a possible outlier in the growth/decay rate in an n time period from significantly distorting the data, making a good future projection impossible.

Therefore, an average of each of these "remaining" values must be taken until the value of the CF_{\pm} is determined.

It is interesting to share if the data of a given time series presents a high level of volatility and/or regardless of the smallest or largest constant value is not able to approximate the actual values in the forecast simulations, it is interesting to disregard the use of the square root over the ratio between the last analyzed period with the average of n considered periods, as we can see below:

$$F_{t+1} = \gamma(A_{t-1}) + (1 - \gamma) \frac{A_{t-1}}{\frac{\sum x_i}{n}} A_{t-1}$$

In this way, it will be possible to expand the range of possibilities so that we have more freedom with the choice of constants and that they can be useful in the medium and long term and above all, they can approximate actual values in the simulations and in the future to support decision making.

Through thousands of case studies analyzed from business administration (sales forecasting) to demography and biology, the use of the square root has demonstrated recurrent practical applicability. However, you will have some cases where without its use it can be more interesting as will be ahead presented. In this sense, as in any other similar method, it is recommended that the user previously perform tests to identify which will be the best value of the constant between 0 and 1 that approximates the actual values.

After emphasizing some particularities that should be considered, below we can see the general formula that contains CF_{\pm} :

$$F_{t+1} = \gamma(A_{t-1}) + (1 - \gamma) \sqrt{\frac{A_{t-1}}{\frac{\sum x_i}{n}}} A_{t-1} \pm \left[\sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} - ABS \left(ABS \frac{\sum x_i - A_{t-n}}{n} \right) \right]$$

3.3 ESBSD Calculation Process

The steps to reach the final result of the ESBSD projection consist of the following steps:

1st: Define the number of n periods to be considered by the "basic rate of growth/decay" with respect to the average considered (3 n periods are recommended in this study).

2nd: Define the number of n periods to be considered in the "complementary factor \pm " (4 n periods are recommended in this study).

3rd: Define a value between 0 and 1 for the "weighting factor".

Taking into account the stages of the ESBSD, the "projection" for the year 2021 and a comparison of the actual result obtained by the PopulationStat of the same year will be carried out, in order to evaluate its practical applicability.

3.3.1 Population projection for 2021

1°: Number of n periods to be considered by the "basic growth/decay rate" will be $n = 3$;

2°: Number of n periods to be considered by the "Complementary Factor \pm " will be $n=4$;

3°: "Weighting Factor" to be considered will be $\gamma = 0.538$

The same specifications will apply for the projection of the following years.

Considering the data presented in table 1:

$$\begin{aligned}\bar{X}_4 &= 5,995,750.00 & S_4 &= 79,062.74 \\ [79,062.74 - (5,995,750.00 - 5,684,000)] &= - 9,187.26 \\ [79,062.74 - (5,995,750.00 - 5,347,000)] &= 46,812.74 \\ [79,062.74 - (5,995,750.00 - 5,044,000)] &= 55,312.74 \\ [79,062.74 - (5,995,750.00 - 4,668,000)] &= - 17,687.26\end{aligned}$$

We can notice in the process of calculating CF_{\pm} we can find negative values both in the subtraction between the average and each of the component values of the set of n past periods, as in the "remaining" values, but in this case, we will consider the same values found, but with a positive sign, because we are considering the Absolute Value Function ABS that, according to Tao (2011) among its properties is non-negativity $|f| : x \rightarrow |f(x)|$ **and** $|x| = 0$ **if and only if** $x = 0$.

Another point to note in this case study is that since \forall n number $\in \mathbb{R} \neq 0$ of this time series is shown to grow continuously despite the growth rate tending to $\cong 0$ (zero) over time \therefore we can consider a positive sign for CF_{\pm} .

$$CF_{\pm 2021} = 9,187.26 + 46,812.74 + 55,312.74 + 17,687.26 / 4$$

$$CF_{\pm 2021} = + 32,250.00$$

After performing the calculation of CF_{\pm} which will be used in the formula, the calculation will be performed on the second side of the model where an average of the last 3 periods was considered:

$$\bar{X}_3 = 6,028,000.00$$

We can note in this case study, the particularity of the fact occurred in our first test (exclusive for the choice of the value of the constant considered ideal) that regardless of using the smallest or largest value of the constant between 0 and 1 it was still not possible to reach a range in which it approximates the actual values (2021). For this reason, it was used in this second part of the formula, the ratio between the last analyzed period with the average of n time periods without extracting the square root. As the same tends to happen with other similar exponential smoothing models, this example reinforces the main reason for proposing this model.

Using the second side of the model:

$$\frac{A_{t-1}}{\frac{\sum_{i=1}^n x_i}{n}} A_{t-1}$$

$$\frac{6,084,000}{6,028,000.00} 6,084,000 = 6,140,520.24$$

Therefore, we get as value = 6,140,520.24.

Final calculation of the model is then reached:

$$F_{t_{2021}} = 0,538 (6,084,000) + 0,462 (6,140,520.24) + 32,250.00$$

$$F_{t_{2021}} = 3,273,192.00 + 2,836,920.35 + 32,250.00$$

$$F_{t_{2021}} = \mathbf{6,142,362}$$

Therefore, 6,142,362 would be the number of inhabitants for the city of Belo Horizonte according to the ESBSD model.

As Makridakis et al. (1983) states, the Absolute and Relative (%) Error analysis is very useful in time series analysis. In this sense, below - in table 2 - is the absolute and relative error for the model taking into account the comparison with the actual data obtained by the PopulationStat in 2021:

Table 2

Comparison of the results of the ESBSD model with actual. Data obtained by the PopulationStat

Population by PopulationStat (2021)	Population by ESBSD (2021)	Absolute Error	Relative Error (%)
6,140,000	6,142,362	2,362	.038%

In the same sense, the population will be projected for the subsequent years from 2022 to 2034.

3.3.2 Population projection for 2022

$$\bar{X}_4 = 6,056,000.00 \quad S_4 = 72,295.69$$

$$[72,295.69 - (6,056,000.00 - 6,140,000)] = - 11,704.31$$

$$[72,295.69 - (6,056,000.00 - 6,084,000)] = 44,295.69$$

$$[72,295.69 - (6,056,000.00 - 6,028,000)] = 44,295.69$$

$$[72,295.69 - (6,056,000.00 - 5,972,000)] = - 11,704.31$$

It should be noted that for the 2022 projection the last actual value (6,140,000) was considered part of the n=4 elements, carried out by the 2021 census and not the ESBSD projection in the same year, because we still have actual data. Below, the calculations continue:

$$CF_{\pm 2022} = 11,704.31 + 44,295.69 + 44,295.69 + 11,704.31 / 4$$

$$CF_{\pm 2022} = + 28,000.00$$

$$\bar{X}_3 = 6,084,000.00$$

$$\frac{6,140,000}{6,084,000.00} 6,140,000 = 6,196,515.45$$

$$F_{t 2022} = 0,538 (6,140,000) + 0,462 (6,196,515.45) + 28,000.00$$

$$F_{t 2022} = 3,303,320.00 + 2,862,790.14 + 28,000.00$$

$$F_{t 2022} = \mathbf{6,194,110}$$

3.3.3 Population projection for 2023

$$\bar{X}_4 = 6,111,527,50 \quad S_4 = 71,566.19$$

$$[71,566.19 - (6,111,527,50 - 6,194,110)] = - 11,016.31$$

$$[71,566.19 - (6,111,527,50 - 6,140,000)] = 43,093.69$$

$$[71,566.19 - (6,111,527,50 - 6,084,000)] = 44,038.69$$

$$[71,566.19 - (6,111,527,50 - 6,028,000)] = - 11,961.31$$

In the 2023 projection, the last value found (2022) will be considered part of the n=4 elements, but as we do not have actual values available, the projection given by the ESBSD in the year 2022 will be considered as "actual value". This same reasoning will be considered for the other years.

$$CF_{\pm 2023} = 11,016.31 + 43,093.69 + 44,038.69 + 11,961.31 / 4$$

$$CF_{\pm 2023} = + 27,527.50$$

$$\bar{X}_3 = 6,139,370.00$$

$$\frac{6,194,110}{6,139,370.00} 6,194,110 = 6,249,338.07$$

$$F_{t2023} = 0,538 (6,194,110) + 0,462 (6,249,338.07) + 27,527.50$$

$$F_{t2023} = 3,332,431.18 + 2,887,194.19 + 27,527.50$$

$$F_{t2023} = \mathbf{6,247,153}$$

3.3.4 Population projection for 2024

$$\bar{X}_4 = 6,166,315.75 \quad S_4 = 70,179.73$$

$$[70,179.73 - (6,166,315.75 - 6,703,311)] = - 10,657.52$$

$$[70,179.73 - (6,166,315.75 - 6,363,604)] = 42,385.48$$

$$[70,179.73 - (6,166,315.75 - 6,028,000)] = 43,863.48$$

$$[70,179.73 - (6,166,315.75 - 5,684,000)] = - 12,136.52$$

$$CF_{\pm 2024} = 10,657.52 + 42,385.48 + 43,863.48 + 12,136.52 / 4$$

$$CF_{\pm 2024} = + 27,260.75$$

$$\bar{X}_3 = 6,193,754.33$$

$$\frac{6,247,153}{6,193,754.33} 6,247,153 = 6,301,012.04$$

$$F_{t2024} = 0,538 (6,247,153) + 0,462 (6,301,012.04) + 27,260.75$$

$$F_{t2024} = 3,360,968.31 + 2,911,067.56 + 27,260.75$$

$$F_{t2024} = \mathbf{6,299,297}$$

3.3.5 Population projection for 2025

$$\bar{X}_4 = 6,220,140.00 \quad S_4 = 68,545.64$$

$$[68,545.64 - (6,220,140.00 - 7,045,273)] = - 10,611.36$$

$$[68,545.64 - (6,220,140.00 - 6,703,311)] = 41,532.64$$

$$[68,545.64 - (6,220,140.00 - 6,363,604)] = 42,515.64$$

$$[68,545.64 - (6,220,140.00 - 6,028,000)] = - 11,594.36$$

$$CF_{\pm 2025} = 10,611.36 + 41,532.64 + 42,515.64 + 11,594.36 / 4$$

$$CF_{\pm 2025} = + 26,563.50$$

$$\bar{X}_3 = 6,246,853.33$$

$$\frac{6,299,297}{6,246,853.33} 6,299,297 = 6,352,180.94$$

$$F_{t 2025} = 0,538 (6,299,297) + 0,462 (6,352,180.94) + 26,563.50$$

$$F_{t2025} = 3,389,021.79 + 2,934,707.59 + 26,563.50$$

$$\mathbf{F_{t2025} = 6,350,293}$$

Using the same reasoning, we proceed with the calculation of the population projection until 2034.

IV RESULTS

The results found through the ESBSD method for the years 2024 to 2039 were as follows according to the table below:

Table 3

Belo Horizonte's population for the years 2022 to 2034 according to the ESBSD methodology

Years	Population by ESBSD Methodology	Growth Rate (%)±
2022	6,194,110	+ .88%
2023	6,247,153	+ .86%
2024	6,299,297	+ .83%
2025	6,350,293	+ .81%
2026	6,400,233	+ .79%
2027	6,449,173	+ .76%
2028	6,497,097	+ .74%
2029	6,544,027	+ .72%
2030	6,589,988	+ .70%
2031	6,634,994	+ .68%
2032	6,679,062	+ .66%
2033	6,722,210	+ .65%
2034	6,764,455	+ .63%

Below, in Table 4, the proposed methodology will be compared with the projections for the following years presented by the PopulationStat platform, which takes into consideration too many other variables in its calculation, such as national wealth, quality of life index, birth rate, fecundity and mortality, for example.

Table 4

A comparison between the projection made by the ESBSD methodology and that made by the PopulationStat

Years	Population by PopulationStat	Population by ESBSD Methodology	Difference (±)
2022	6,194,000	6,194,110	+ 110
2023	6,248,000	6,247,153	- 847
2024	6,300,000	6,299,297	- 703
2025	6,352,000	6,350,293	- 1,707
2026	6,402,000	6,400,233	- 1,767
2027	6,450,000	6,449,173	- 827
2028	6,496,000	6,497,097	+ 1,097
2029	6,541,000	6,544,027	+ 3,027
2030	6,583,000	6,589,988	+ 6,988
2031	6,624,000	6,634,994	+ 10,994
2032	6,663,000	6,679,062	+ 16,062
2033	6,699,000	6,722,210	+ 23,210
2034	6,734,000	6,764,455	+ 30,455

As a complementary information and of importance for the following projections, the Instituto Brasileiro de Geografia e Estatística (*Brazilian Institute of Geography and Statistics, IBGE*) through its studies with multiple variables strongly considers the possibility the population of Belo Horizonte, as well as the whole of Brazil, will decline after 2040, in this sense, it is sensible and realistic to make projections with mathematical models related to the growth until the given period.

V DISCUSSIONS

It is important to note that, as in other exponential smoothing models, in ESBSD, the higher the value of the chosen constant, means that the user will be giving more importance to past data, likewise, the lower the value of the constant, the more importance will be given to more recent data. Another issue to highlight is that the model is adaptable with the input of new data, thus adapting to new changes in the growth/decay dynamics.

Through this case study applied to population projection, ESBSD has again demonstrated interesting applicability. Despite being a mathematical model, it was able to adapt to the trends expected by official agencies such as IBGE (Instituto Brasileiro de Geografia e Estatística) and the United Nations (whose data is presented by the PopulationStat platform) as to the future number of inhabitants expected for the population of the city of Belo Horizonte by the year 2034.

Through table 4, for example, we can see the sum of the values between the years 2022 to 2034, when compared with the values presented by PopulationStat, returns a total difference of 86,092.00. With this result we can affirm its potential high level of accuracy for the following years if the projections also raised by official agencies such as IBGE and the United Nations are on the same path of the actual values in which we will know in the future.

VI CONCLUSION

Population projection, when properly carried out may contribute to the planning of public policies and the private sector and consequently contribute to the economic and social development of a given geographic location under analysis, since resources will tend to be applied in a more rational manner while seeking to meet the needs of the population and avoiding the lack and misuse of resources.

ESBSD, in its many experiments considering real case studies (from sales forecasting to areas such as demographics and biology, for example) and a continuously growing/decreasing trending environment (and, additionally, also trending towards a "saturation limit" as Pierre François Verhulst raised in his study), can present a practical applicability in the short, medium and long term if the value of the constant between 0 and 1 is well established by the user in his testing and validation.

Considering the case study with the survey of the population growth projection of the city of Belo Horizonte, it was possible to observe its practical applicability only in the short term, because official agencies of Statistics of Brazil (IBGE) points to a strong trend of population decay of the Brazilian population from 2040, when taking into account too many other variables intrinsic to the field of study of demography as birth rate; fertility, mortality, in addition to other important indicators such as national wealth, migration; emigration; quality of life index, unemployment rate, access to education, among other examples directly influence the population dynamics.

Mathematical models, when applied to population projections, can be useful in aiding decision-making when a given geographic location does not have a complete Demographic Census and has too many resource constraints. With regard to ESBSD, whether in population projection analysis or in other fields of study, it has proven to be an interesting model for its proposal. In this sense, it remains as one more model for the portfolio to successfully address real problems.

REFERENCES

- Abubakar, U. M., & Usaini, S. Note on Mathematical Modeling for Population Projection and Management: A Case Study of Niger State.
- Adhikari, R., and R. K. Agrawal. 2013. "An Introductory Study on Time Series Modelling and Forecasting." arXiv:1302.6613.
- Delmas, B. Pierre-Francois Verhulst et la loi logistique de la population. Math Sci Hum / Mathematical Social Sciences, v42, n 167, p. 51-81, 2004.
- Gawatre, D.W., Kandgule, M.H., Kharat, S.D., Comparative Study of Population Forecasting Methods. IOSR Journal of Mechanical and Civil Engineering (IOSRJMCE):,13, 16-19. (2016).
- IBGE. (2022). Projeção da população do Brasil e das Unidades da Federação [*Population Projection of the population of Brazil and the Units of the Federation*]. Retrieved August 2, 2022, from https://www.ibge.gov.br/apps/populacao/projecao/index.html?utm_source=portal&utm_medium=popclock&utm_campaign=novo_popclock
- Makridakis, Spyros, Wheelwright, Steven C. and McGee, Victor E. Forecasting: Methods and Applications. 2a ed. New York: John Wiley & Sons, 1983.
- Tao, T. An Introduction to Measure Theory (American Mathematical Society, Providence, 2011), Vol. 126.
- World Population Statistics. (2022). Belo Horizonte, Brazil Population. Retrieved August 4, 2022, from <https://populationstat.com/brazil/belo-horizonte>.
- Yahaya, A.A., Kandgule, M.H., Audu, P.M., Aisha, H.S., Mathematical Modeling for Population Projection and Management: A Case Study Of Niger State . IOSR Journal of Mathematics (IOSR-JM):,13, 51-57. (2017).