# Boletim SPE

Sociedade Portuguesa de Estatística

*Publicação semestral* *primavera de 2020*

## INE - 85 anos de estatísticas a servir o país



**Informação Editorial**
**Endereço:** Sociedade Portuguesa de Estatística.
Campo Grande. Bloco C6. Piso 4.
1749-016 Lisboa. Portugal.
**Telefone:** +351.217500120
**e-mail:** *spe*@spestatistica.pt
**URL:** https://www.spestatistica.pt
**ISSN:** 1646-5903
**Depósito Legal:** 249102/06
**Tiragem:** Edição digital
**Execução Gráfica e Impressão:** Gráfica Sobreirense
**Editor:** Fernando Rosado, *fernando.rosado@fc.ul.pt*

**Sociedade Portuguesa de Estatística** desde 1980

# Short-Term Regional Demographic Forecasts with Time Series Methods and Machine Learning Algorithms

Jorge M. Bravo[1], *jbravo@novaims.unl.pt*
*Universidade Nova de Lisboa (NOVA IMS) & MagIC & CEFAGE-EU & Université Paris-Dauphine PSL*

Edviges Coelho, *edvides.coelho@ine.pt*
*Departamento de Estatísticas Demográficas e Sociais, Instituto Nacional de Estatística* & *Universidade Lusófona (ECEO-UHLT)*

## 1 Introduction

Forecasts of births and deaths are a critical input in the computation of resident population estimates since they determine, together with net migration, the dynamics of both the population size and its age distribution in the territory. In this study we evaluate the short-term forecasting accuracy of alternative traditional linear and non-linear seasonal time series methods (SARIMA, Holt-Winters, State Space models) and new advanced machine learning algorithms (Artificial Neural Networks, Bagging) to birth and death monthly forecasting at the sub-national level using a backtesting cross-validation approach. We use a time series of monthly data from 2000 to 2018 disaggregated by sex for the 25 Portuguese NUTS3 regions. Our results provide valuable tools for policymakers in assessing how changes in population size and age composition affect economic outcomes and their geographical patterns and in developing, implementing and coordinating active regional policy instruments.

Empirical evidence shows that there is a territorial dimension of demographic change. In recent years there is an increasing acknowledgment that the local and regional levels provide a more suitable ground for designing and implementing policy responses to the complex interaction of factors that dictate the wide-ranging patterns of demographic change (EU, 2016). Population forecasts are widely used for analytical, planning and policy purposes (e.g., education, health, housing, pensions, labour market, security, spatial planning, transportation, public infrastructure and social policy planning) at national, regional and local levels (Bravo, 2016, 2019; Bravo et al., 2018, 2020; Ayuso, Bravo & Holzmann, 2020; Bravo & Herce, 2020). Concerns about the possible long-term effects of demographic change on population size, dynamics and structure increased the importance of producing accurate population projections at the subnational level. The population of a given territorial area and its age distribution changes over time through the interaction of three possibly correlated factors: fertility, mortality, and (international and regional) migration. To project the population size and age structure at a future date, economists and demographers typically using the cohort-component method and stochastic time series methods to project the dynamics of the three components of demographic change. Forecasts of monthly births and deaths are a critical input in the computation of monthly estimates of resident population (MERP).[2]

Birth and death forecasts can in principle be produced using, among others, statistical time series methods (univariate or multivariate), structural models (e.g., VAR models) or machine learning methods (e.g., Artificial Neural Network (ANN), Support Vector Machines (SVM)). Births and deaths time series are typically non-stationary, contain a trend and exhibit strong seasonality patterns at both national and

[1] This communication brief is an abridged and updated version of a research paper presented at the ASMDA, 26th APDR and CAPSI 2019 Conferences.

[2] To produce MERP, for each subpopulation and gender it is necessary to: (i) obtain monthly forecasts of the total number of births and deaths, (ii) estimate age-specific mortality rates considering period/cohort life tables derived from stochastic mortality models, eventually considering for heterogeneity in longevity (Ayuso, Bravo & Holzmann, 2017a,b), (iii) estimate the level and age pattern of net international migration, and (iv) consider a number of assumptions such as the distribution of age-specific fertility rates or the sex ratio a birth (Bravo

regional levels. For vital events computed for small populations on monthly time intervals, the need to uncover complex structures of temporal interdependence in time series data is critically challenged in the presence of seasonal variability. In recent decades a substantial amount of research has focused on the development and application of traditional time series models in population forecasts, focusing either on total population growth or on individual components of growth (see, e.g., Pflaumer, 1992; Lee 1992; Lee and Tuljapurkar 1994; Keilman, Pham & Hetland, 2002; Booth, 2006; Tayman, Smith, and Lin 2007; Alho, Bravo and Palmer, 2012; Abel et al. 2013; Wiśniowski et al., 2015; Bravo and El Mekkaoui de Freitas, 2018; Li et al., 2018). The main focus of these studies is largely on the identification and measurement of uncertainty in population forecasts, with little interest in the assessment of the models forecasting accuracy or the out-of-sample validity of the prediction intervals. Much of the research concerning the evaluation of time series models for birth and death forecasting has been focused on univariate time series ARIMA models at the national level, with little research on the predictive accuracy of these models at the sub-national level, particularly in small population areas. Fewer still have explored the use of the Holt-Winters exponential smoothing (HW) and State Space (SS) time series models in small population exercises. Up to our knowledge, no attempt has been made to use machine learning and deep learning methods to forecasts monthly demographic data.

In this study, we address this gap and investigate and compare the predictive accuracy of alternative linear and non-linear traditional time series models (seasonal ARIMA, HW and SS) and more advanced machine learning time series methods (ANN, Bootstrapp Aggregating or Bagging) to birth and death monthly forecasting at the sub-national level using up-to-date demographic data. Using a series of monthly data from 2000 to 2018 disaggregated by sex for the 25 Portuguese NUTS3 regions, we compare the short-term (one year) method's forecasting accuracy. We adopt a backtesting time series cross-validation approach, i.e., we consider a multi-step forecasting approach with re-estimation in which the training data or base period is extended before re-selecting and re-estimating the model at each iteration and computing forecasts. Our main contributions are the following. First, we summarise and analyse the out-of-sample error performance of commonly used Seasonal ARIMA, HW and SS forecasting models together with new powerful machine learning algorithms, using a rich and large set of subpopulations and two different demographic events with different dynamics over time. Second, we evaluate the out-of-sample performance of the prediction intervals produced by these models. Third, we assess the consistency of the predictive performance of these methods in populations of different size and nature. Fourth, we evaluate the existence of significant differences in the model's forecasting accuracy between subpopulations of different sex. Fifth, we investigate how well the models perform in terms of predicting the uncertainty of future monthly birth and death counts.

The selection of the appropriate forecasting method depends on several factors, including the past behaviour pattern of the time series, previous knowledge about the nature of the phenomenon being studied, the availability of statistical data and the predictive capacity of the model. Our results show that these simulations provide valuable insights regarding the forecasting performance of alternative time series models in small population forecasting exercises and on the validity of using such models as predictors of population forecast uncertainty and, thus, have significant practical implications in territorialising public policies to address demographic change. The remaining part of the study is organised as follows. Section 2 describes the materials and methods used in this study. Section 3 details the research design used to produce forecasts and assess model’s performance. Section 4 presents and discusses the results. Section 5 concludes.

## 2 Materials and Methods

### 2.1. Seasonal ARIMA Model

The seasonal ARIMA model is an extension to the classical ARIMA model that supports the direct modelling of both the trend and seasonal components of a time series and it is widely used for forecasting. The model includes new parameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality (Hyndman and Athanasopoulos, 2018). In this study, we combine the seasonal and non-seasonal components into a multiplicative seasonal autoregressive moving average model, or SARIMA model, given by

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \delta + \Theta_Q(B^s)\theta(B)w_t \quad (1)$$

where $w_t$ denotes the Gaussian white noise process. The general model can be expressed as $ARIMA(p,d,q) \times (P,D,Q)_s$, where the ordinary autoregressive (AR) and moving average (MA) components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders $p$ and $q$, respectively, the seasonal AR and MA components are denoted by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders $P$ and $Q$, respectively. The non-seasonal and seasonal difference components are represented by $\nabla^d = (1-B)^d$ and $\nabla^D = (1-B^s)^D$, respectively. The seasonal period $s$ defines the number of observations that make up a seasonal cycle (e.g., $s = 12$ for monthly observations).

The estimation process for the parameters in (1) for each of the 100 time series follows the standard Box-Jenkins methodology in an iterative 3-step procedure comprising the identification, estimation and evaluation and diagnostic analysis stages, testing for unit roots and white noise errors and optimizing a stepwise algorithm for the AIC Criterion. When the data suggest the inexistence of seasonal unit roots in the series and the seasonality is deterministic, we can express it as a function of seasonal dummy variables (and time eventually). In this case, an ARIMA model if fitted to the residuals of the equation:

$$Y_t = \alpha + \sum_{i=1}^{s-1} \gamma_{i,t} D_{i,t} + \beta t + \epsilon_t \quad (2)$$

where $Y_t$ is the variable of interest, $D_{i,t}$ are seasonal dummies, $t$ denotes time and $\epsilon_t$ is a white-noise error term. Additionally, we examined the residuals of the selected model and formally examined the null hypothesis of independence of the residuals using the Box-Pierce/Ljung-Box test. We also tested the normality of the residuals using the Jarque-Bera Test. After examining different models, the best SARIMA model was selected, parameters were estimated using the nonlinear least squares method, and the model was used for forecasting monthly births and deaths.

## 2.2. Holt-Winters' Seasonal Method

The Holt-Winters method is a univariate automatic forecasting method that uses simple exponential smoothing (Holt 1957; Winters 1960). The forecast is obtained as a weighted average of past observed values in which the weight function declines exponentially with time, i.e., recent observations contribute more to the forecast than earlier observations. Forecasted values are dependent on the level, slope and seasonal components of the series being forecast. The model-specific formulation depends on whether seasonality is modelled in an additive or multiplicative way. The additive method is selected when the seasonal variations are approximately constant through the series, whereas the multiplicative method is preferred when the seasonal variations change proportionally to the level of the series (Hyndman and Athanasopoulos, 2018). The additive method is specified as:

$$\begin{aligned} l_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} - b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\ y_{t+h|t} &= l_t + hb_t + s_{t-m+h} \end{aligned} \quad (3)$$

where $l_t$, $b_t$ and $s_t$ denote the level, trend and seasonal components, respectively, with corresponding smoothing parameters $\alpha$, $\beta$ and $\gamma$; $y_{t+h|t}$ is the forecast for $h$ periods ahead at time $t$. The Holt-Winters' **multiplicative method** is defined as:

$$\begin{aligned} l_t &= \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} - b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma\frac{y_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m} \\ y_{t+h|t} &= (l_t + hb_t)s_{t-m+h} \end{aligned} \quad (4)$$

We initialize the model's hyperparameters using the decomposition approach suggested by Hyndman et al. (2008) and implemented in the forecast package in R. After examining each time series for both the additive and multiplicative versions of the Holt-Winters' seasonal method, we finally selected the model showing lower residual sum of squares to produce forecasts of monthly births and deaths.

### 2.3. Exponential smoothing state space model

State Space models consist of a measurement equation that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonal) change over time. The general Gaussian state space model involves a measurement equation relating the observed data to an unobserved state vector $x_t = (b_t, s_t, s_{t-1}, \dots, s_{t-(m-1)})$, an initial state distribution and a Markovian transition equation that describes the evolution of the state vector over time state. In this study, we investigate both the additive and multiplicative error versions of SS models that underlie the exponential smoothing methods of the form (Hyndman et al., 2002):

$$Y_t = \mu_t + k(x_{t-1})\varepsilon_t \tag{5}$$

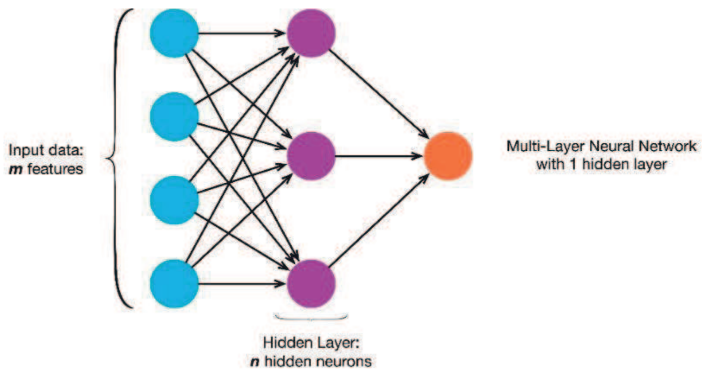
$$x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t \tag{6}$$

where $\varepsilon_t \sim N(0, \sigma^2)$, $\mu_t = Y_{t-1}$ and where, for additive error models $k(x_{t-1}) = 1$, such that $Y_t = \mu_t + \varepsilon_t$, whereas for multiplicative error models $k(x_{t-1}) = \mu_t$ such that $Y_t = \mu_t(1 + \varepsilon_t)$. Model estimation involves measuring the unobservable state (prediction, filtering and smoothing) and estimating the unknown parameters using MLE methods.

### 2.4. Artificial Neural Network Algorithms

Artificial neural networks are forecasting methods (algorithms) that are based on individual, interconnected units called neurons that allow complex nonlinear relationships between the response variable and its predictors. The typical neural network architecture consists of a network of "neurons" organised in layers in which the predictors form the bottom layer, the forecasts form the top layer and there may be intermediate layers containing hidden neurons (Figure 1).

Figure 1 – A hypothetical example of Multilayer Perceptron Network.




Source: Author's preparation

In this latter case, each layer of nodes receives inputs from the previous layers, the neural network becomes non-linear and is known as a multilayer feed-forward network (MFFN). Forecasts are obtained by a linear combination of the inputs (or features) through an activation (or transfer) function, with weights automatically selected using a learning algorithm that minimises a cost function (Hyndman and Athanasopoulos, 2018). The following equation summarises the forecasted output.

$$f(\boldsymbol{x}, \boldsymbol{w}) = \varphi(\boldsymbol{x} \cdot \boldsymbol{w}) = \varphi\left(\sum_{j=1}^{P} (x_j w_j) + \theta\right) \tag{7}$$

where $\boldsymbol{x}$ and $\boldsymbol{w}$ represent, respectively, the input vector and weight vector of the neuron when there are $P$ inputs into the neuron, $\theta$ is a bias term and $\varphi$ denotes an activation function (e.g., sigmoid function). The process results in a single output from a neuron. With time series data, lagged values of the time series are used as inputs to a neural network in what is called a neural network autoregression. In this study we only consider MFFN with one hidden layer and add the last observed values from the same month as inputs.

### 2.5. Bootstrapp Aggregating (Bagging)

Bootstrap aggregating or bagging is a machine learning ensemble meta-algorithm that involves generating $m$ new training sets by sampling from the original time series $D = \{(t_1, y_1), \ldots, (t_k, y_k)\}$ uniformly and with replacement. Then, for each of the $m$ bootstrap samples, time series methods are fitted and combined using ensemble techniques. More formally,

Figure 2: The Bagging algorithm

**Bagging Algorithm**

**Input:** Training set $D = \{(t_1, y_1), \ldots, (t_k, y_k)\}$
$B$ number of iterations
Forecasting methods / Base learners $M$

**Process:**

1. **for** $t = 1, \ldots, M$:
   $t_t = \varphi(D, D_{bs})$ % $D_{bs}$ is the bootstrap distribution
   $Apply\ method\ M$ to the dataset $D_{bs}$ and obtain $f^{bs}(x)$
2. $\boldsymbol{Return}$: $ensemble\ \{f^1(x), \ldots, f^B(x)\}$
3. **Prediction** with ensemble

$$f(\boldsymbol{x}) = \frac{1}{B}\sum_{h=1}^{B} f^h(x)$$

Source: Author's preparation

The idea is that a set of individual potentially weak learners (models) can be combined (averaged) to create a strong learner that outperforms individual models by reducing variance, and bias. First, the time series is Box-Cox-transformed, and then decomposed into trend, seasonal and residual components. We then bootstrapp the residual series, add them back to the trend and seasonal components, and reverse the Box-Cox transformation to obtain variations on the original time series. In this study we use 1000 bootstrapped series in combination with an SS model.

## 3 Research Methodology

### 3.1. Research Design

We set out a backtesting framework applicable to single-period ahead forecasts with steps:[3] (i) Selection the metric of interest (monthly births or deaths by sex and subpopulation); (ii) For each time series and period, selection of the historical "lookback window" for model calibration. We adopt a time series cross-validation approach, i.e., we consider a multi-step forecasting approach with re-estimation in which the training data or base period is extended before re-selecting and re-estimating the model at each iteration and computing forecasts. We adopt an expanding lookback window approach; (iii) Selection of the forecasting horizon ("lookforward window") over which to make forecasts. We focus on short-term horizon forecasts (1-year ahead of monthly births and deaths forecasts, i.e.,12 observations) since our interest is to use them as an input for computing MERP; (iv) Select a rolling fixed-length horizon backtesting approach in which we consider the accuracy of forecasts over fixed-length horizons as the jump-off date moves sequentially forward through time; (v) Select the evaluation criteria. We computed several criteria but, due to space constrains, we report the results for the Mean Absolute Percent Error (MAPE). For a given lookback and lookforward window, the MAPE for model $j$ is defined as

$$MAPE_j = \frac{1}{n}\sum_{t=1}^{n} \frac{\left|\hat{y}_{t,j} - y_t\right|}{y_t} \times 100 \qquad (8)$$

where $n$ is the number of forecasted values, $\hat{y}_t$ is the number of monthly births/deaths predicted by the model for time point $t$, and $y_t$ is the corresponding value observed at time point $t$.

[3] For a similar approach used in evaluating the forecasting performance of stochastic mortality models and interest rate and credit risk models see, e.g., Dowd et al. (2010), Bravo & Silva (2006) and Chamboko & Bravo (2016, 2019a,b).

Each of the different time series models constructed (using a different lookback window and jump-off year) implies a different set of prediction intervals for the forecast horizon. To better understand the performance of the models analysed in terms of predicting the uncertainty of future births and deaths we computed the number of birth and death counts falling outside the 95% prediction intervals associated with each set of forecasts. Parameter estimation and model forecasting assessment were carried out using a computer routine written in R-script (R Development Core Team 2019).

### 3.2. Data

In this study, we use demographic data for Portugal comprising monthly data on live births and deaths broken down by sex and 25 different NUTS 3 regions from January 2000 to December 2018 provided by Statistics Portugal. The demographic dataset consists of 228 monthly observations for each one of the 100 different subpopulations of different size, the smallest with 38,753 resident individuals in December 2017 (Beira Baixa, male), the largest with 1,505,435 individuals (Lisbon Metropolitan Area, female). Of the 100 subpopulations tested, four (Lisbon and Oporto metropolitan areas male and female populations) correspond to highly populated areas with, in the case of Lisbon, more than one million residents. In contrast, the dataset tested includes several small population areas with less than 50,000 residents (e.g., Beira Baixa, Alto Tâmega, Alentejo Litoral). This archive is a challenging dataset in which to assess the monthly forecasting performance of time series methods since the data exhibits significant trend and seasonal components and high volatility in some cases, particularly in small population areas. Figure 2 represents the time series plot of monthly deaths of one representative small NUTS3 subpopulation.

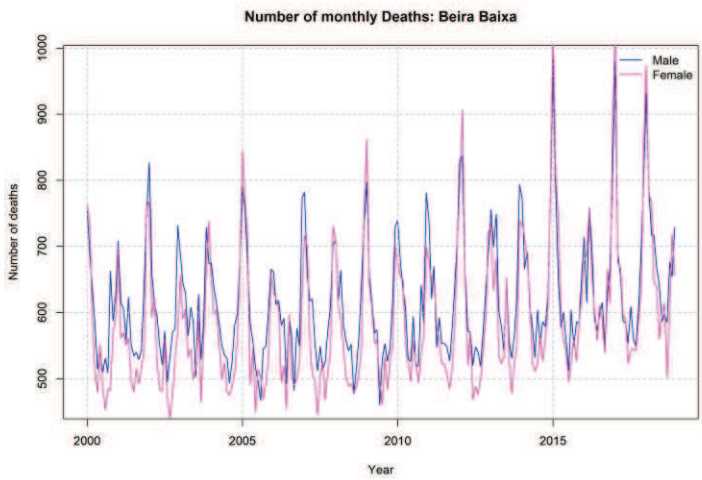



Figure 2 – Number of monthly deaths: Beira Baixa NUTS3 Region

## 4 Empirical results

The five time series methods are used as predictive models for making forecasts for future values of live births and deaths by sex and NUTS3 regions in Portugal. The MAPE results of 1-year ahead forecasts of monthly births and deaths by sex and NUTS3 regions for the period 2014-2018 averaged over all jump-off years with the different models are given in Tables 1 and 2, respectively. The results averaged (simple and weighted averages) over all 25 regions and five launch years are shown in the Tables. Additionally, Tables 1 and 2 include data on the population size of each NUTS3 region in December 2017 to ascertain whether the model's relative forecasting performance is a function of population size.

We first discuss the results related to monthly births forecasting. The all regions and launch years simple and weighted average forecasting performance for the five methods tested are similar for both male and female subpopulations showing relatively small average MAPE results, with the exception of ANN with one hidden layer that clearly underperforms. The simple average results show that the precision of the SARIMA forecasts is better than that of HW, SS and Bagging methods for the female subpopulations whereas, for the male counterparts, the Bagging method exhibits slightly lower forecasting errors.

Table 1 – Births Forecasting - Average MAPE by Model, Sex and NUTS3

| Births | Females | | | | | | Males | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NUTS 3 | Popul. | AR | HW | SS | ANN | Bag | Popul. | AR | HW | SS | ANN | Bag |
| Alto Minho | 124583 | 13.20 | 13.38 | 14.13 | 15.69 | 13.60 | 107595 | 12.46 | 13.07 | 12.87 | 12.55 | 12.61 |
| Cávado | 211950 | 10.23 | 10.39 | 9.63 | 12.70 | 9.65 | 192003 | 11.61 | 10.44 | 9.62 | 10.83 | 10.06 |
| Ave | 215975 | 10.50 | 9.57 | 8.67 | 11.71 | 8.80 | 197879 | 9.52 | 6.90 | 8.13 | 8.19 | 7.08 |
| Área Metrop. Porto | 910200 | 6.14 | 5.35 | 5.45 | 7.75 | 5.48 | 809502 | 4.74 | 5.40 | 5.04 | 5.58 | 5.16 |
| Alto Tâmega | 46044 | 18.49 | 18.86 | 21.03 | 19.17 | 20.38 | 41113 | 23.66 | 21.11 | 21.26 | 26.71 | 21.95 |
| Tâmega e Sousa | 216999 | 8.96 | 8.39 | 9.18 | 9.86 | 8.93 | 201769 | 8.62 | 7.61 | 8.55 | 8.17 | 8.51 |
| Douro | 101142 | 13.49 | 13.76 | 13.97 | 14.54 | 12.95 | 90904 | 14.27 | 14.86 | 13.88 | 15.12 | 13.31 |
| Terras Trás-os-Montes | 56870 | 15.33 | 15.02 | 16.95 | 18.77 | 15.75 | 51677 | 16.38 | 16.44 | 15.70 | 19.94 | 16.30 |
| Oeste | 186405 | 9.91 | 10.18 | 9.53 | 10.52 | 9.52 | 171301 | 7.84 | 8.82 | 8.07 | 11.46 | 7.95 |
| Região de Aveiro | 190926 | 8.84 | 8.75 | 8.22 | 10.11 | 8.19 | 172169 | 8.47 | 7.56 | 6.95 | 7.73 | 7.33 |
| Região de Coimbra | 231654 | 8.15 | 8.21 | 7.84 | 10.48 | 8.00 | 205294 | 7.86 | 7.70 | 7.34 | 9.88 | 7.16 |
| Região de Leiria | 149784 | 9.20 | 8.85 | 7.84 | 9.80 | 8.06 | 136525 | 9.86 | 10.79 | 9.75 | 10.78 | 9.83 |
| Viseu Dão Lafões | 134679 | 12.15 | 11.62 | 12.21 | 12.91 | 12.31 | 119952 | 12.64 | 12.53 | 12.12 | 13.38 | 11.85 |
| Beira Baixa | 43061 | 21.60 | 21.31 | 24.19 | 26.73 | 24.07 | 38753 | 16.40 | 17.92 | 17.23 | 18.77 | 16.70 |
| Médio Tejo | 123699 | 10.53 | 11.54 | 12.01 | 10.81 | 11.70 | 110956 | 9.99 | 10.63 | 10.21 | 14.13 | 10.59 |
| Beiras, Serra da Estrela | 114163 | 12.72 | 12.43 | 12.97 | 13.17 | 12.14 | 102025 | 11.11 | 10.75 | 12.37 | 11.08 | 11.73 |
| Área Metrop. Lisboa | 1505435 | 3.38 | 3.60 | 3.11 | 5.07 | 3.09 | 1328244 | 3.59 | 4.18 | 3.29 | 6.89 | 3.41 |
| Alentejo Litoral | 47551 | 16.99 | 17.56 | 17.78 | 18.75 | 17.68 | 46223 | 17.32 | 18.77 | 19.11 | 18.76 | 18.13 |
| Baixo Alentejo | 60669 | 11.59 | 12.45 | 12.57 | 12.41 | 12.33 | 57199 | 14.33 | 14.17 | 14.59 | 16.94 | 14.84 |
| Lezíria do Tejo | 124049 | 8.00 | 9.50 | 8.66 | 10.01 | 8.73 | 114666 | 11.56 | 10.94 | 11.48 | 12.16 | 11.08 |
| Alto Alentejo | 56092 | 18.80 | 19.01 | 18.54 | 18.80 | 18.39 | 50965 | 16.32 | 16.33 | 17.15 | 19.30 | 16.70 |
| Alentejo Central | 80677 | 12.81 | 13.87 | 13.01 | 14.05 | 13.05 | 73859 | 12.01 | 13.61 | 12.80 | 11.87 | 12.88 |
| Algarve | 229719 | 7.33 | 8.30 | 7.02 | 9.21 | 7.14 | 209898 | 7.40 | 7.46 | 7.24 | 8.45 | 7.13 |
| RA Açores | 125052 | 10.77 | 10.49 | 10.59 | 11.10 | 10.54 | 118810 | 9.78 | 9.78 | 9.52 | 9.53 | 9.27 |
| RA Madeira | 135957 | 12.00 | 12.30 | 14.02 | 13.22 | 12.97 | 118411 | 10.39 | 11.51 | 11.93 | 14.20 | 10.88 |
| All regi.Simple Average | 216933 | 11.64 | 11.79 | 11.96 | 13.09 | 11.74 | 194708 | 11.53 | 11.57 | 11.45 | 12.90 | 11.30 |
| Weighted Average | | 8.00 | 7.99 | 7.83 | 9.49 | 7.74 | | 7.70 | 7.87 | 7.52 | 9.30 | 7.48 |
| Max | 1505435 | 21.60 | 21.31 | 24.19 | 26.73 | 24.07 | 1328244 | 23.66 | 21.11 | 21.26 | 26.71 | 21.95 |
| Min | 43061 | 3.38 | 3.60 | 3.11 | 5.07 | 3.09 | 38753 | 3.59 | 4.18 | 3.29 | 5.58 | 3.41 |

**Source:** Authors preparation; **Notes:** Average Mean Absolute Percent Error (MAPE) by model (AR=ARIMA; HW; SS, ANN, Bagging) Sex and NUTS3 Region for the period 2014-2018. Weighted Average computed using the proportion of region's male or female population in the corresponding (sex) total population.

Note, however, that when considering the weighted average results (with weights given by the proportion of the region's subpopulation in the total resident population) the bagging method in combination with an exponential smoothing state space model exhibit higher forecasting accuracy due to their superior performance in highly populated regions. Using this later metric, the Bagging model advantages the SARIMA, HW, SS and ANN models by 0.26 (0.22), 0.25 (0.39), 0.09 (0.04) and 1.76 (1.82) percentage points in the female (male) subpopulations, respectively. On average for all models and for 59.2% of the subpopulations the forecasting errors are smaller for the male subpopulations when compared to their female counterparts. As expected, the average MAPE results over the five launch years are larger, the smaller the region's population size. The largest average forecasting error (26.73%) is found in the Beira Baixa female subpopulation using the ANN model whereas the highest accuracy (having 3.09% MAPE) is attained in the Lisbon metropolitan area ("Área Metropolitana de Lisboa") using the Bagging model. The forecasting error is less than 10% in 52.8% of the subpopulations considered.

Moving now to the results related to 1-year ahead monthly deaths forecasting, Table 2 shows once again that, with the exception of ANN that underperforms, the all regions and launch years simple and weighted average forecasting performance for the different models was relatively similar for both the male and female subpopulations, although the differences between the worst and the best performing model is higher in the male subset. Compared to births results, the average (weighted) forecasting accuracy of the alternative univariate time series methods continue to be lower in the male subpopulations and higher in the female group. The weighted average results show that the precision of SARIMA forecasts is consistently better than that of the HW, SS, ANN and Bagging models although the differences towards the predictive performance of SS and Bagging methods is small. The SARIMA model advantages the HW, SS, ANN and Bagging models by 0.59 (0.31), 0.19 (0.09), 1.51 (1.35) and 0.10 (0.11) percentage points in the female (male) subpopulations, respectively. On average for all models and for 71.2% of the subpopulations the forecasting errors are notably smaller for the male subpopulations when compared to their female counterparts.

Table 2 – Deaths Forecasting - MAPE by Model, Sex and NUTS3

| Births | Females | | | | | | Males | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NUTS 3 | Popul. | AR | HW | SS | ANN | Bag | Popul. | AR | HW | SS | ANN | Bag |
| Alto Minho | 124583 | 10.36 | 11.27 | 10.64 | 10.74 | 10.59 | 107595 | 9.20 | 9.29 | 9.44 | 12.48 | 9.41 |
| Cávado | 211950 | 11.45 | 11.21 | 11.07 | 12.28 | 11.05 | 192003 | 8.88 | 9.14 | 8.99 | 9.03 | 8.84 |
| Ave | 215975 | 7.73 | 9.27 | 8.55 | 8.57 | 8.69 | 197879 | 8.89 | 9.31 | 9.05 | 10.79 | 8.92 |
| Área Metrop. Porto | 910200 | 7.24 | 8.07 | 7.85 | 8.87 | 7.79 | 809502 | 5.76 | 6.33 | 6.12 | 7.99 | 6.10 |
| Alto Tâmega | 46044 | 11.71 | 12.00 | 11.35 | 13.68 | 11.43 | 41113 | 12.69 | 15.07 | 14.94 | 15.14 | 14.73 |
| Tâmega e Sousa | 216999 | 9.02 | 11.21 | 10.33 | 13.00 | 10.05 | 201769 | 8.94 | 9.61 | 8.99 | 9.96 | 9.02 |
| Douro | 101142 | 11.25 | 13.74 | 12.21 | 13.57 | 12.27 | 90904 | 9.88 | 10.27 | 9.73 | 11.51 | 9.29 |
| Terras Trás-os-Montes | 56870 | 11.46 | 12.35 | 11.35 | 12.32 | 11.37 | 51677 | 10.53 | 11.07 | 10.71 | 12.72 | 10.59 |
| Oeste | 186405 | 7.30 | 8.11 | 7.65 | 9.07 | 7.73 | 171301 | 8.09 | 7.99 | 7.75 | 9.22 | 7.75 |
| Região de Aveiro | 190926 | 10.29 | 10.47 | 10.04 | 9.98 | 9.99 | 172169 | 7.94 | 9.20 | 8.20 | 9.33 | 8.34 |
| Região de Coimbra | 231654 | 7.54 | 7.56 | 7.57 | 7.44 | 7.25 | 205294 | 7.29 | 7.16 | 7.38 | 8.03 | 7.45 |
| Região de Leiria | 149784 | 9.57 | 9.98 | 9.63 | 10.92 | 9.68 | 136525 | 9.74 | 9.90 | 9.62 | 10.28 | 9.65 |
| Viseu Dão Lafões | 134679 | 9.91 | 10.35 | 9.80 | 12.78 | 9.88 | 119952 | 8.28 | 8.79 | 7.80 | 10.53 | 7.80 |
| Beira Baixa | 43061 | 14.26 | 14.13 | 14.96 | 14.22 | 14.95 | 38753 | 12.75 | 11.73 | 13.95 | 13.97 | 13.06 |
| Médio Tejo | 123699 | 8.10 | 7.95 | 7.74 | 9.56 | 7.57 | 110956 | 8.87 | 9.12 | 9.11 | 9.77 | 8.87 |
| Beiras, Serra da Estrela | 114163 | 10.29 | 11.48 | 10.34 | 11.36 | 10.42 | 102025 | 8.46 | 8.10 | 8.07 | 9.44 | 8.07 |
| Área Metrop. Lisboa | 1505435 | 6.01 | 6.07 | 5.89 | 7.70 | 5.71 | 1328244 | 5.07 | 5.01 | 4.99 | 5.59 | 5.16 |
| Alentejo Litoral | 47551 | 11.97 | 13.04 | 11.46 | 12.76 | 11.39 | 46223 | 13.24 | 16.11 | 15.24 | 15.97 | 14.86 |
| Baixo Alentejo | 60669 | 11.80 | 13.06 | 12.48 | 13.91 | 12.38 | 57199 | 10.09 | 10.29 | 10.00 | 12.26 | 9.99 |
| Lezíria do Tejo | 124049 | 9.48 | 10.33 | 9.97 | 11.11 | 10.05 | 114666 | 9.07 | 9.85 | 8.87 | 11.62 | 8.96 |
| Alto Alentejo | 56092 | 10.65 | 11.56 | 10.57 | 13.54 | 10.52 | 50965 | 11.29 | 11.71 | 11.48 | 13.08 | 11.50 |
| Alentejo Central | 80677 | 9.74 | 10.49 | 10.93 | 11.35 | 10.39 | 73859 | 9.22 | 9.52 | 8.98 | 11.82 | 9.25 |
| Algarve | 229719 | 9.26 | 9.65 | 8.94 | 10.88 | 8.96 | 209898 | 7.57 | 7.50 | 7.33 | 8.84 | 7.32 |
| RA Açores | 125052 | 10.90 | 11.72 | 11.33 | 11.94 | 11.07 | 118810 | 9.67 | 11.31 | 10.52 | 12.52 | 10.75 |
| RA Madeira | 135957 | 9.78 | 10.86 | 9.82 | 11.45 | 9.77 | 118411 | 9.53 | 10.05 | 9.50 | 11.17 | 9.54 |
| All regi..Simple Average | 216933 | 9.88 | 10.64 | 10.10 | 11.32 | 10.04 | 194708 | 9.24 | 9.74 | 9.47 | 10.92 | 9.41 |
| Weighted Average | | 8.25 | 8.83 | 8.44 | 9.76 | 8.35 | | 7.35 | 7.66 | 7.43 | 8.70 | 7.46 |
| Max | 1505435 | 14.26 | 14.13 | 14.96 | 14.22 | 14.95 | 1328244 | 13.24 | 16.11 | 15.24 | 15.97 | 14.86 |
| Min | 43061 | 6.01 | 6.07 | 5.89 | 7.44 | 5.71 | 38753 | 5.07 | 5.01 | 4.99 | 5.59 | 5.16 |

**Source:** Authors preparation; **Notes:** Average Mean Absolute Percent Error (MAPE) by model (AR=ARIMA; HW; SS, ANN, Bagg) Sex and NUTS3 Region for the period 2014-2018. Weighted Average computed using the proportion of region's male or female population in the corresponding (sex) total population.

Similar to the births results, the average MAPE results over the five launch years are smaller, the more populated the region is. The largest average forecasting error (16.11%) is found in the Alentejo Litoral male subpopulation using the HW model whereas the highest accuracy (4.99%) is attained in the Lisbon metropolitan area ("Área Metropolitana de Lisboa") male subpopulation using the SS model. The forecasting error is less than 10% in 37.6% of the subpopulations considered. To measure how well models perform in terms of predicting the uncertainty of future monthly birth/death counts over 1-year forecasting horizons, we computed the percentage of monthly birth and death counts falling outside the 95% prediction interval estimated for each model, sex and NUTS3 Region.[4] The results show that for the birth and death count forecasting exercises the prediction intervals for the SARIMA, SS and Bagging models consistently provide appropriate measures of uncertainty for short-term forecasting horizons. The SARIMA, SS and Bagging models perform equally well in terms of predicting the uncertainty of future monthly death counts, with SS and Bagging models slightly overperforming in births forecasting. On the contrary, the HW and ANN model consistently fail in predicting the uncertainty of future monthly birth and deaths with, in some regions, up to 17% of observed death counts falling out of the 95% prediction interval (Alto Minho, ANN model).

[4] Due to space constraints, the full set of results is not displayed in the study but can be obtained from the authors upon request.

## 5 Conclusion

Monthly time series of live births and deaths exhibit significant and persistent seasonality patterns, requiring the adoption of appropriate forecasting methods to increase the accuracy of population forecasts. In this study we empirically evaluated the forecasting performance of traditional time series methods (seasonal ARIMA, HW and SS) and new machine learning approaches applied to birth and death monthly forecasting by sex and NUTS 3 regions for Portugal using a backtesting framework and monthly data for the period 2000-2018. With the exception of the ANN structure with a single hidden layer that clearly underperformed, the all regions and launch years simple and weighted average forecasting performance for the three models was relatively similar for both male and female subpopulations births and deaths. However, our results show that the Bagging method combined with an exponential smoothing state space model exhibit higher forecasting accuracy for births whereas for deaths forecasting the seasonal ARIMA slightly overperformed both alternative univariate time series methods and machine learning algorithms. As expected, the weighted average precision is higher, the more populated the region is. The prediction intervals for the SARIMA, SS and Bagging models consistently provide appropriate measures of uncertainty for short-term forecasting horizons. Further research should check for the robustness of these results against alternative forecasting horizons and fixed lookback windows using rolling fixed-length horizon backtests. Future research will also investigate the robustness of these results against alternative primary, extended, composite, and hybrid performance metrics used in machine learning regression, forecasting and prognostics, considering for competing distance measures and normalization and aggregation procedures.

## References


Abel, G., Bijak, J., Forster, J., Raymer, J., Smith, P. & Wong, J., (2013) Integrating uncertainty in time series population forecasts: An illustration using a simple projection model. *Demographic Research*, 29 (43), 1187-1226.

Alho, J., Bravo, J. M. & Palmer, E. (2013). *Annuities and Life Expectancy in NDC*. In Holzmann, R. E. Palmer and D. Robalino (Eds.), Nonfinancial defined contribution Pension Schemes in a Changing Pension World, Vol. 2 Gender, Politics, and Financial Stability, 395 - 436.

Ayuso, M., Bravo, J. M. & Holzmann, R. (2020). Getting Life Expectancy Estimates Right for Pension Policy: Period versus Cohort Approach. Journal of Pension Economics and Finance, 1-20. doi: doi:10.1017/S1474747220000050;

Ayuso, M., Bravo, J. M., & Holzmann, R. (2017a). Addressing Longevity' Heterogeneity in Pension Scheme Design. *Journal of Finance and Economics*, 6(1), 1–21.

Ayuso, M., Bravo, J. M., & Holzmann, R. (2017b). On the Heterogeneity in Longevity among Socioeconomic Groups: Scope, Trends, and Implications for Earnings-Related Pension Schemes. *Global Journal of Human Social Sciences - Economics*, 17(1), 31–57.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting, 22 (3), 547-581.

Bravo, J. M. (2016). Taxation of Pensions in Portugal: A Semi-Dual Income Tax System. *CESifo DICE Report - Journal for Institutional Comparisons*. 14 (1), 14-23.

Bravo, J. M. (2019). Funding for Longer Lives: Retirement Wallet and Risk-Sharing Annuities. EKONOMIAZ *Basque Economic Review*, Nº 96 (II-2019), 268–291.

Bravo, J. M. & Coelho, E. (2019). Forecasting Subnational Demographic Data using Seasonal Time Series Methods. In *Proceedings of the 19th Portuguese Association of Information Systems Conference: digital disruption: living between data science, IoT and ... people* (pp. 40). Associação Portuguesa de Sistemas de Informação. ISSN 2183-489X

Bravo, J. M. & Coelho, E. (2020). Forecasting small population monthly fertility and mortality data with seasonal time series methods. In: Linhares, W. (Ed.). As Ciências Sociais Aplicadas e a Interface com vários Saberes 2, Atena Editora, 158-176.

Bravo, J. M. & Coelho, E. (2020). Modelling monthly birth and deaths using Seasonal Forecasting Methods as an input for population estimates. In: *Demography of Population Health, Aging and Health Expenditures. The Springer Series on Demographic Methods and Population Analysis*. In Press.

Bravo, J. M. & Silva, C. (2006). Immunization Using a Stochastic Process Independent Multifactor Model: The Portuguese Experience. *Journal of Banking and Finance*, 30 (1), 133-156.

Bravo, J. M., & El Mekkaoui de Freitas, N. (2018). Valuation of longevity-linked life annuities. *Insurance: Mathematics and Economics*, 78, 212–229.

Bravo, J. M., & Herce, J. A. (2020). Career breaks, Broken pensions? Long-run effects of early and late-career unemployment spells on pension entitlements. Preprint submitted to Journal of Pension Economics and Finance.

Bravo, J. M., Ayuso, M., Holzmann, R. & Palmer, E. (2020). Addressing Life Expectancy Gap in Pension Policy. Preprint submitted to Insurance: Mathematics and Economics.

Bravo, J. M., Coelho, E., & Magalhães, M. G. (2010). *Mortality projections in Portugal*. In EUROSTAT - European Commission (eds.), Work session on demographic projections, EUROSTAT-EC Collection: Methodologies and working papers, Theme: Population and Social Conditions, 241–252.

Bravo, J. M., Rodrigues, T., Ribeiro, S. & Inácio, A. (2018). *Portugal. Projeções de População Residente 2011-2040*. In Teresa Rodrigues & Marco Painho (Coord.). Modelos Preditivos e Segurança Pública. Fronteira do Caos Editores, 168-214.

Chamboko, R., & Bravo, J. M. (2016). On the modelling of prognosis from delinquency to normal performance on retail consumer loans. *Risk Management*, 18(4), 264–287.

Chamboko, R. & Bravo, J. M. (2019a). Modelling and forecasting recurrent recovery events on consumer loans. *International Journal of Applied Decision Sciences*, 12(3), 271-287.

Chamboko, R. & Bravo, J. M. (2019b). Frailty correlated default on retail consumer loans in developing markets. *International Journal of Applied Decision Sciences*, 12(3), 257–270.

Dowd, K., Cairns, A., Blake, D., Coughlan, G. Epstein, D. & Khalaf-Allah, M. (2010) Backtesting Stochastic Mortality Models, North American Actuarial Journal, 14:3, 281-298

EU (2016). The impact of demographic change on European regions. European Union - Committee of the Regions (doi:10.2863/26932).

Holt, C.C. (1957). Forecasting seasonals and trends by exponentially weighted averages. O.N.R. Memorandum 52/1957, Carnegie Institute of Technology. Reprinted with discussion in 2004, *International Journal of Forecasting*, 20, 5–13.

Hyndman, R. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. 2nd edition, OTexts.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18, 439–454.

Hyndman, R. J., Koehler, A.B., Ord, J. K., & Snyder, R.D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer.

Keilman, N., Pham, D., & Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research*, 6, 409–453.

Lee, R. (1992). Stochastic demographic forecasting. *International Journal of Forecasting*, 8, 315–327.

Lee, R., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89, 1175–1189

Li S., Yang, Z., Li, H. & Shu, G. (2018). Projection of population structure in China using least squares support vector machine in conjunction with a Leslie matrix model. Journal of Forecasting, 37(2), 225–34.

Pflaumer, P. (1992). Forecasting U.S. population totals with the Box–Jenkins approach. *International Journal of Forecasting*, 8, 329–338.

R Development Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org.

Tayman, J., Smith, S. K., and Lin, J. (2007). Precision, bias, and uncertainty for state population forecasts: an exploratory analysis of time series models. *Population Research and Policy Review,* 26(3), 347–369.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342

Wiśniowski A., Smith P., Bijak, J., Raymer, J. & Forster, J. (2015). Bayesian Population Forecasting: Extending the Lee-Carter Method. Demography. Jun; 52(3), 1035-59.