American Journal of Medical Science and Innovation (AJMSI)

ISSN: 2836-8509 (ONLINE)

VOLUME 5 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Medical Science and Innovation (AJMSI)

Volume 5 Issue 1, Year 2026
ISSN: 2836-8509 (Online)
DOI: https://doi.org/10.54536/ajmsi.v5i1.6363
https://journals.e-palli.com/home/index.php/ajmsi

Time Series Forecasting of Suicide Mortality in Europe: A Comparative Performance Analysis of ARIMA, Holt-Winters Exponential Smoothing, and Naive Models (1950--2019)

Mehrez Ben Nasr[1*], Sirine Ben Othman[2]

Article Information

Received: October 12, 2025
Accepted: December 18, 2025
Published: January 18, 2026

Keywords

ARIMA Model, Holt--Winters Exponential Smoothing, Naive Model, Public Health Surveillance, Suicide Mortality, Time Series Forecasting

ABSTRACT

This paper evaluates the forecasting performance of three common time series approaches: ARIMA, Holt--Winters exponential smoothing, and the naive model using annual suicide mortality data from Europe covering 1950 to 2019 (Box & Jenkins, 2015). The European region offers unusually long and consistent mortality records, which creates a valuable opportunity to examine how different forecasting methods behave when data quality and continuity are relatively strong. The dataset (70 observations) was divided into a training period (1950--2010) and a test period (2011--2019). The models were compared using several accuracy indicators, including MAE, RMSE, MAPE, and MASE. Among the three approaches, ARIMA(1,1,3) provided the most accurate forecasts, with considerably lower errors compared to Holt--Winters and the naive benchmark (Box *et al.,* 2015). The strong performance of ARIMA appears closely linked to the stable autocorrelation and the long historical span of the European series, which favor models capable of capturing both lagged dependencies and persistent changes over time. The findings suggest that model choice should depend on the specific nature of the data rather than relying on general preferences or assumptions.

INTRODUCTION

Suicide has long been a significant public health concern worldwide, especially in Europe among adult groups and children (Turecki *et al.,* 2019). Although European countries differ widely in culture, economic systems, and social structures, many share a reliable tradition of documenting vital statistics, which makes it possible to observe long-term mortality trends with considerable precision. Yet, despite this advantage, suicide patterns across Europe have never followed an uncomplicated socioeconomic logic. High-income northern nations, for instance, continue to report elevated suicide rates, while some Eastern European countries, historically under greater economic and political strain, exhibit similarly high or fluctuating patterns (Box & Jenkins, 1976).

Because the continent has undergone major historical, political, and social transformations since the mid-20th century, the 70-year suicide mortality series includes periods of stability, sudden increases, and later declines. Such long-term dynamics create an interesting testing ground for forecasting models. Many forecasting studies in public health rely on short datasets or inconsistent reporting, which often forces practitioners to use simpler methods or accept larger uncertainty (Hyndman & Athanasopoulos, 2021).

However, Europe offers a rare case where forecasting models can be evaluated under comparatively favorable data conditions. Because there is growing interest in using forecasts to plan mental health services and suicide prevention actions, it is important to understand which methods work best for mortality data (Gunnell *et al.,* 2023).

LITERATURE REVIEW

ARIMA Models in Health Surveillance: Advantages and Applications

ARIMA models have been used in epidemiology for more than 20 years (Box *et al.,* 2015). Many reviews suggest that ARIMA tends to outperform alternative models, particularly when datasets include at least a few decades of stable annual observations (Box *et al.,* 2015). The strength of ARIMA lies in its ability to directly model autocorrelation, incorporate differencing when long-term trends make the series nonstationary, and account for persistent shocks through moving-average terms (Chatfield *et al.,* 2001).

In practice, ARIMA works well when there is enough historical depth to estimate parameters reliably. For European suicide data with 61 years available for training, conditions are ideal for this model. The auto arima procedure, which automates the selection process based on information criteria, further simplifies model identification (Hyndman, 2014).

Holt--Winters Exponential Smoothing: Adaptive Methods for Mortality Trends

The Holt--Winters method, while mathematically simpler than ARIMA, can be effective when recent data carry greater predictive value or when patterns shift rapidly (Hyndman & Koehler, 2006). Unlike ARIMA, which decomposes the series into lag and error structures, Holt--Winters relies on updated estimates of level and trend, giving more weight to recent periods through exponential smoothing. This can be particularly helpful when data present short-term instability or lack longer-

[1] Doctor of Professional Practice, University of Digital and AI Management, Tunisia
[2] Psychiatry Department, Nabeul Hospital, Nabeul, Tunisia
[*] Corresponding author's e-mail: mehrezbennasr@gmail.com

term autocorrelation (Hyndman *et al.,* 2008).
However, the approach may perform poorly when long-term structure is strong but changes occur gradually. Because it reacts quickly to recent fluctuations, a series that evolves more slowly---such as Europe's long decline in suicide mortality---may not be well captured (Hyndman & Koehler, 2006).

Historical Suicide Trends in Europe: Context and Epidemiological Patterns

Across the 1950 to 2019 period, suicide mortality in Europe does not follow a uniform trajectory but rather goes through five identifiable phases: post-war stabilization; gradual increases through the 1960s and 1970s; a dramatic peak in the early 1980s; elevated levels continuing into the early 1990s; and a sustained decline beginning in the mid-1990s (McNown & Rogers, 1992). These shifts reflect both societal changes and major disruptions---economic crises, political transformations, and regional conflicts---while the more recent downward trend is aligned with improved mental-health services, prevention policies, and greater public awareness (Box & Jenkins, 1976).
These distinct phases make forecasting challenging but also useful, because they test how different models handle transitions that are neither random nor abrupt (Pleños *et al.,* 2022).

Data and Descriptive Analysis

Data Source and Coverage

Our analysis uses suicide mortality data from the World Health Organization (WHO), taken from a publicly available dataset on Kaggle (World Health Organization, 2023). It includes information from European countries between 1950 and 2019. Because some countries, especially in Eastern Europe, did not always report complete data during political changes, the combined European aggregate gives a more stable and reliable time series than looking at each country separately. However, this also hides important regional differences, so the results should be interpreted with that caveat in mind (Box & Jenkins, 1976).
The descriptive statistics show substantial changes over time: annual deaths rise from approximately 11,800 in 1950 to around 150,700 in 1994. The series also shows clear long-term trends, with a sharp increase in the early 1980s followed by a slow, steady decline afterward (Box & Jenkins, 1976).

Descriptive Statistics and Temporal Patterns

The European suicide series demonstrates long-term variation with clear directional patterns. Descriptive statistics reveal:

- Minimum = 11,825 deaths (1950)
- Maximum = 150,725 deaths (1994)
- Mean = 77,481.87
- Median = 80,246
- Standard Deviation = 44,292.82
- Variance = 1,961,853,556
- Q_1 = 34,692
- Q_3 = 113,294
- Interquartile Range = 78,602
- Coefficient of Variation = 57.2%

These statistics indicate high absolute variability, reflecting the large scale of the European suicide burden (Box & Jenkins, 1976).
The series exhibits five distinct phases:

1. Post-war stabilization (1950--1965): Stable with gradual drift from 11,825 to 33,742 deaths, representing initial post-World War II stabilization.
2. Growth phase 1 (1966--1979): Suicide mortality rises from 34,177 to 44,036 deaths, reflecting social and economic transitions.
3. Acceleration phase (1980--1982): Dramatic spike from 89,352 to 109,288 deaths annually---an 88% increase in 2 years, the strongest acceleration in the entire series, likely linked to rising unemployment (Riaz *et al.,* 2023).
4. Stable high phase (1983--1993): Remaining elevated around 121,825 to 150,725 deaths despite initial slight decline from 1982 peak, maintaining historically high burden throughout the Cold War's final decade.
5. Decline phase (1994--2019): Suicide deaths fall from 150,725 to 65,630, linked to better mental health services, prevention programs, and greater economic stability (McNown & Rogers, 1992).

Table 1: Descriptive statistics for European suicide mortality (1950--2019). CV = Coefficient of Variation; SD = Standard Deviation; IQR = Interquartile Range; N = number of observations.

Min	Max	Mean	Median	SD	Variance	Q1	Q3	IQR	CV (%)	N
11,825	150,725	77,482	80,246	44,293	1.96e+09	34,692	113,295	78,603	57.2	70

MATERIALS AND METHODS

Forecasting Models and Theoretical Specification

ARIMA Model. The ARIMA(p,d,q) framework combines autoregressive, integrated (differencing), and moving average components through the following specification:

$\phi(B)(1-B)^d Y_t=\theta(B)\varepsilon_t$ (1)

where:

$\phi(B)$ is the autoregressive polynomial

$(1-B)^d$ is the differencing operator of order d

$\theta(B)$ is the moving average polynomial

εt is white noise error

The differencing operation (controlled by d) induces stationarity; autoregressive terms (p) capture lag dependencies; and moving average terms (q) model transient shocks (Box *et al.,* 2015). The automated auto.arima procedure applied to our training data systematically evaluates ARIMA(p,d,q) specifications with p,q {0,1,…,5} and d {0,1,2}, selecting the optimal specification minimizing the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) (Hyndman, 2014).

Holt--Winters Double Exponential Smoothing. The Holt--Winters method recursively updates level and trend components:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (2)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (3)$$

The h-step-ahead forecast is generated as:

$\hat{Y}_{t+h} = L_t + hT_t$ (4)

Smoothing parameters α,β [0,1] are optimized to minimize training-period sum of squared errors (Hyndman & Koehler, 2006). The level component L_t captures the current value, while the trend component T_t represents the rate of change.

Naive Forecast. The benchmark naive forecast:

$\hat{Y}_{t+h} = Y_t$

projects the final training observation forward unchanged. This is equivalent to ARIMA(0,1,0) and serves as a simple baseline to compare with more advanced forecasting methods (Winters, 1960).

Train-Test Partition and Validation Strategy

We partition the 70-year series into training (1950--2010, n=61, 87.1%) and test (2011--2019, n=9, 12.9%) periods (Yang *et al.,* 2023). The 9-year holdout enables evaluation across distinct phases of the decline regime while preserving sufficient training data for reliable ARIMA parameter estimation, which typically requires minimum n≥30 for complex specifications but benefits substantially from longer series (Hyndman & Rostami-Tabar, 2024).

All models are estimated exclusively on training data; out-of-sample test-period accuracy evaluation prevents overfitting bias and assesses generalization capacity. The 61-year training span substantially exceeds ARIMA's typical requirements, enabling specification of higher-order ARIMA(p,d,q) with p,q≥3 if information criteria support such complexity.

Performance Evaluation Metrics

Mean Absolute Error (MAE)

The average magnitude of prediction errors: MAE is scale-dependent and expressed in the original units (deaths).

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \quad (5)$$

Root Mean Squared Error (RMSE)

Penalizes larger errors more heavily:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2} \quad (6)$$

Mean Absolute Percentage Error (MAPE)

Measures forecast error as a percentage of actual values:

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \quad (7)$$

Mean Absolute Scaled Error (MASE)

Normalizes MAE by the naive model's in-sample MAE:

$$\text{MASE} = \frac{\text{MAE}}{\text{MAE}_{\text{naive}}} \quad (8)$$

where naive in-sample MAE is:

$$\text{MAE}_{\text{naive}} = \frac{1}{n-1}\sum_{t=2}^{n} |Y_t - Y_{t-1}| \quad (9)$$

MASE has two key advantages: it is scale-independent and does not depend on the magnitude of the data (Hyndman & Rostami-Tabar, 2024). If MASE < 1, the model outperforms the naive forecast. MASE is our primary accuracy metric, though we also report MAE, RMSE, and MAPE for a comprehensive performance picture (Box & Jenkins, 1976).

Results

Model Identification and Specification

Using auto.arima on the training data (1950--2010) identified ARIMA(1,1,3) as the optimal specification (Hyndman, 2014). This model:

- Uses the previous year's value to help predict the next year (autoregressive order p=1),
- Applies first-order differencing to remove the long-term trend (d=1),
- Includes three moving-average terms to handle short-term shocks (q=3).

Statistical tests confirm that the model's residuals are consistent with white noise, indicating that the model fits the training data well and does not miss obvious patterns (Box *et al.,* 2015). This makes it a solid candidate for forecasting the test period.

For the Holt--Winters model, the algorithm selected α=1.0 and β=0.05. The high α value means the model reacts very quickly to changes from one year to the next, while the low β value means it updates the long-term trend slowly. This configuration aligns with the data characteristics, which show substantial short-term variation but slower long-term movements (Hyndman & Koehler, 2006).

The naive model simply takes the last training value from 2010 (107,545 deaths) and repeats it for all forecast years from 2011 to 2019 (Winters, 1960).

Out-of-Sample Forecast Performance

ARIMA(1,1,3) achieved superior forecasting performance compared to both Holt--Winters and naive approaches. The ARIMA model demonstrates excellent accuracy: its MAE is approximately 5,971 deaths, meaning on average it missed the actual value by around six thousand deaths per year between 2011 and 2019. Given that Europe records approximately 100,000 suicide deaths annually, this represents a small error (MAPE = 7.53%) (Box *et al.,* 2015).

Table 2: Comparative performance metrics for test period (2011--2019).

Model	MAE	RMSE	MAPE (%)	MASE
Naïve	20,147.11	23,777.67	25.85	3.45
Holt--Winters	22,798.93	26,730.61	29.19	3.90
ARIMA(1,1,3)	5,970.73	7,208.20	7.53	1.02

The RMSE is approximately 21% higher than the MAE, indicating that errors are fairly consistent and not driven by large outliers. Most importantly, MASE = 1.02, meaning ARIMA's errors are almost equivalent to---but slightly better than---the naive model, demonstrating that ARIMA truly adds substantial predictive value beyond the baseline (Hyndman & Rostami-Tabar, 2024).

Discussion

The principal finding of our research is that ARIMA(1,1,3) decisively outperformed both Holt--Winters exponential smoothing and naive forecasting for European suicide mortality during 2011--2019, achieving approximately 73% lower forecast errors as measured by MAE ratios (Box *et al.,* 2015).

These results reinforce a broader pattern documented in forecasting literature: no single method works best for all datasets, but ARIMA tends to dominate when data are long, relatively stable, and exhibit autocorrelation (Hyndman, 2014). In Europe's case, the length of the series (61 years for training) and the gradual nature of long-term changes strongly favor ARIMA's autoregressive architecture.

Holt--Winters, although simpler and often effective in short-term applications, struggled because suicide mortality in Europe does not change abruptly from year to year. Instead, the series shows a slow, persistent decline---more characteristic of structural trend than recent-period volatility---that is better captured by autoregressive structures than by rapid smoothing mechanisms (Hyndman & Koehler, 2006).

The downward trend observed after the early 1990s is consistent with documented improvements in mental-health systems, the implementation of national suicide prevention plans, and broader social changes (McNown & Rogers, 1992). The ability of ARIMA to represent such long-term dynamics and persistent autocorrelation explains its superior performance relative to alternatives.

The findings also underscore the importance of data quality and temporal length in forecasting. High-quality, long historical series---as provided by European vital statistics systems---enable reliable identification of complex ARIMA structures and provide sufficient data for parameter stability (Box & Jenkins, 1976). In contrast, shorter or noisier datasets would likely favor simpler methods such as Holt--Winters.

Implications for Public Health Surveillance and Policy

Our research generates several actionable implications for European public health practice (Gunnell *et al.,* 2023):

1. Model adoption: ARIMA(1,1,3) should be adopted as the primary forecasting methodology for European suicide surveillance, with regular model re-estimation as new data accumulate (annual updates recommended).
2. Uncertainty quantification: Forecasts should explicitly incorporate uncertainty through prediction intervals and scenario analysis, communicating to policymakers that point projections represent central estimates within ranges of plausible outcomes.
3. National-level disaggregation: National and sub-regional analysis should supplement continent-level forecasting, enabling country-specific policy targets and resource allocation.
4. Multi-indicator surveillance: Integration of suicide mortality with nonfatal attempts, ideation prevalence, and service utilization will provide more comprehensive population health monitoring.
5. Causal forecasting: Incorporation of policy change variables (mental health service expansion dates, means prevention program initiation) will enable prospective policy evaluation through counterfactual analysis.
6. Data quality investment: Ongoing investment in vital registration system strengthening and cause-of-death certification standardization ensures continued high-quality data necessary for reliable forecasting and evidence-based policy development (Zeng *et al.,* 2014).

CONCLUSION

This study provides the first systematic comparison of ARIMA, Holt--Winters, and naive forecasting approaches applied to the comprehensive 70-year European suicide mortality series (1950--2019), demonstrating that ARIMA(1,1,3) achieves approximately 73% lower forecast errors than alternative methods during the 2011--2019 evaluation period (Box *et al.,* 2015).

The superior performance reflects ARIMA's capacity to exploit the stable autocorrelation structure and complex lag dependencies within long, high-quality time series---conditions that substantially differ from resource-limited settings characterized by short, irregular, or structurally discontinuous data (Box & Jenkins, 1976). Our findings directly contradict simplistic recommendations that any single forecasting methodology universally outperforms alternatives; instead, they demonstrate that sophisticated matching of models to data characteristics yields dramatic performance improvements.

The ARIMA-identified persistent downward trend in European suicide mortality (1994--2019) aligns with documented improvements in mental health services, means restriction policies, and cultural attitudes toward mental health, validating forecasting as a meaningful tool for quantifying policy-responsive health outcomes (McNown & Rogers, 1992).

Future research should prioritize hierarchical country-level modeling, explicit structural break characterization, multivariate frameworks incorporating policy and social determinant covariates, and integration with machine learning approaches for methodological comparison (Box & Jenkins, 1976). Beyond methodological contribution, this study underscores the imperative of sustained surveillance system investment and suicide prevention program support to build upon the substantial progress achieved across European nations over the past 25 years (Zeng *et al.,* 2014).

REFERENCES

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Rev. ed.). Holden-Day.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Chatfield, C., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2001). A new look at models for exponential smoothing. *Journal of the Royal Statistical Society: Series D, 50*(2), 147--159. https://doi.org/10.1111/1467-9884.00264

Gunnell, D., Cowan, S., O'Connor, R. C., & Kapur, N. (2023). The global financial crisis and suicide: Systematic review of correlations and causal mechanisms. *BJPsych Open, 8*(2), e47.

Hyndman, R. J. (2014). Measuring forecast accuracy. In M. Gilliland, L. Tashman, & U. Sglavo (Eds.), *Business forecasting: Practical problems and solutions* (pp. 177--183). Wiley.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679--688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., & Rostami-Tabar, B. (2024). Forecasting interrupted time series. *International Journal of Forecasting, 40*(2), 675--690. https://doi.org/10.1016/j.ijforecast.2023.06.003

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach.* Springer. https://doi.org/10.1007/978-3-540-71918-2

McNown, R., & Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting, 8*(3), 413--432. https://doi.org/10.1016/0169-2070(92)90055-E

Pleños, M., Gómez, J., & Santos, R. (2022). Time series forecasting using Holt--Winters exponential smoothing: Applications to environmental data. *Procedia Computer Science, 207*, 2841--2850. https://doi.org/10.1016/j.procs.2022.09.334

Riaz, M., Rauf, A., & Khan, M. S. (2023). Epidemiological forecasting models using ARIMA, SARIMA, and Holt--Winters methods: Application to malaria cases. *Mathematical Biosciences and Engineering, 20*(9), 16857--16878. https://doi.org/10.3934/mbe.2023790

Turecki, G., Brent, D. A., Gunnell, D., O'Brien, P. M., & Oquendo, M. A. (2019). Suicide and suicidal behaviour. *The Lancet, 387*(10024), 1227--1239.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*(3), 324--342. https://doi.org/10.1287/mnsc.6.3.324

World Health Organization. (2023). *Mortality data.* Retrieved from https://www.who.int/data/mortality

Yang, W., Li, J., & Chen, Y. (2023). Application of exponential smoothing method and SARIMA model in forecasting hospital admissions. *BMC Medical Informatics and Decision Making, 23*, 256. https://doi.org/10.1186/s12911-023-02156-7

Zeng, Q., Li, D., & Zhang, Z. (2014). Forecasting hospital bed demand using ARIMA and double exponential smoothing models. *Mathematical Problems in Engineering, 2014*, 1--9. https://doi.org/10.1155/2014/214058