Revealing the Nature of Italian Life Expectancy: A Comparative Study of ARIMA Models Using COVID-19 Shock

Girolamo Franchetti and Massimiliamo Politano(✉)

Economics and Statistics Department, University of Naples Federico II, Naples, Italy
{girolamo.franchetti,politano}@unina.it

Abstract. This study investigates how alternative ARIMA model specifications can be used to infer the underlying trend structure—deterministic or stochastic—of the life expectancy at birth time series for the Italian population over the period 19742024. By comparing two ARIMA(1,d,0) models and two ARIMA(1,d,1) models, each estimated with and without a deterministic trend component, we aim to assess not only forecast accuracy but also the capacity of each model to capture the structural dynamics of the series, particularly in the presence of exogenous shocks. The COVID-19 outbreak in 2020 is treated as a structural shock and serves as a testing ground for evaluating model adaptability and long-run behavior. Our analysis employs stationarity tests, residual diagnostics, impulse response functions (IRFs), model fit statistics, and forecast error measures. Results indicate that while trend-based ARIMA models tend to provide better in-sample statistical fit, they often fail to capture the persistent deviations induced by structural breaks. In contrast, the ARIMA(1,d,1) model without a deterministic trend offers greater flexibility and superior post-shock forecasting performance. The paper concludes by proposing a structured approach to model selection under structural uncertainty, highlighting how comparative model analysis can inform our understanding of time series behavior over a given historical period.

Keywords: Life expectancy · ARIMA models · COVID.19

1 Introduction

Forecasting demographic time series, such as life expectancy, requires explicit assumptions regarding the nature of the underlying trend—in particular, whether it follows a deterministic trajectory (characterized by smooth and predictable growth) or a stochastic one (where shocks can have persistent, long-term effects) [1–5]. These assumptions are not merely technical, but fundamentally shape how we interpret temporal dynamics and respond to structural changes in

A. D'Ambrosio et al. (Eds.): CLADAG-VOC 2025, SCDAKO, pp. 156–171, 2025.
https://doi.org/10.1007/978-3-032-03042-9_14

the data [8,9,11,13]. This issue becomes especially salient when a series is subjected to significant exogenous shocks [7], which may either temporarily deviate from or permanently alter its trajectory.

Life expectancy reflects a multifactorial process influenced by health interventions, socioeconomic development, public policy, environmental factors, and, crucially, epidemiological events [12,16]. The COVID-19 pandemic in 2020 represents a unique, global, and exogenous disruption, offering an empirical setting to investigate the resilience and adaptability of different time series models [17,18].

Research Question: *Does the time series of Italian life expectancy over the period 1974–2024 behave more like a deterministic-trend process subject to temporary fluctuations, or a difference-stationary process characterized by permanent innovations? Furthermore, how do different ARIMA model specifications—with and without deterministic trends—perform in the presence of the COVID-19 shock in terms of model fit, forecasting accuracy, and dynamic adjustment?*

To address these questions, we undertake a structured econometric analysis using ARIMA models [6,13]. Our methodological strategy comprises five analytical steps:

1. **Step 1:** Examine the stationarity properties of the life expectancy series using visual diagnostics and formal unit root tests;
2. **Step 2:** Identify appropriate ARIMA model structures through ACF/PACF diagnostics;
3. **Step 3:** Estimate and compare model performance using in-sample fit statistics and residual diagnostics;
4. **Step 4:** Evaluate out-of-sample forecast accuracy, especially over the post-pandemic period (2020–2034);
5. **Step 5:** Analyze impulse response functions (IRFs) to assess each model's dynamic behavior in response to exogenous shocks.

Through this comparative framework, we aim to shed light on both empirical model performance and the deeper structural properties of the underlying demographic process. Ultimately, our findings contribute to the ongoing methodological debate on time series specification under structural uncertainty, particularly in the context of public health and demographic forecasting.

2 Data and Initial Diagnostics

The empirical object of analysis is the annual life expectancy at birth for the Italian population over the period 1974–2024.[1] The series exhibits a pronounced upward trend, consistent with improvements in medical technology, healthcare coverage, and living standards. However, a sudden and substantial decline is visible in 2020, coinciding with the onset of the COVID-19 pandemic (Fig. 1).

[1] Data were obtained from the ISTAT database: https://esploradati.istat.it/databrowser/#/en.

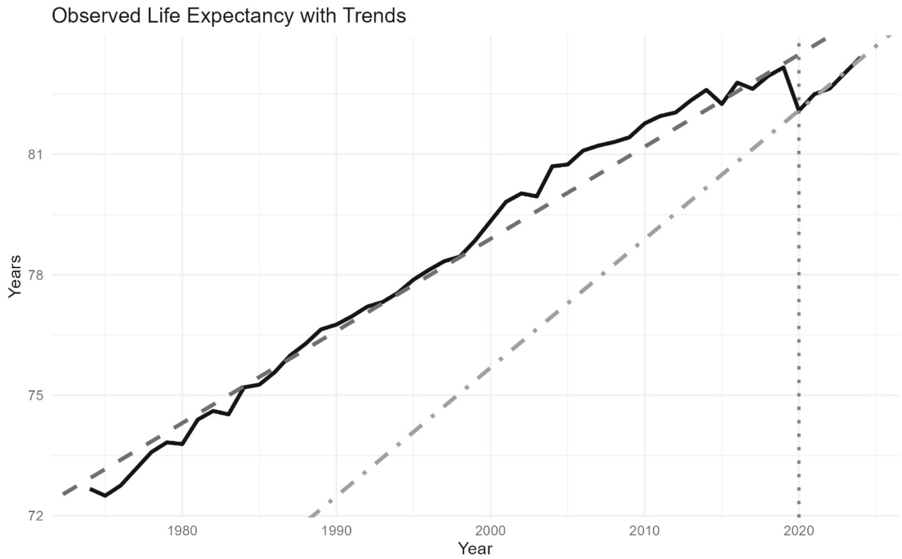

Fig. 1. Original Series with Fitted Linear Trend (1974–2024). 2020 reflects the COVID-19-induced drop.

Prior to estimating any model, it is essential to determine whether the series is stationary, and if not, whether the non-stationarity arises from a deterministic trend or from stochastic integration. These considerations inform whether the appropriate transformation is differencing, detrending, or both.

We apply two complementary statistical tests:

- The **Augmented Dickey-Fuller (ADF)** test, which evaluates the null hypothesis that the series contains a unit root (i.e., it is non-stationary);
- The **Kwiatkowski-Phillips-Schmidt-Shin (KPSS)** test, which evaluates the null hypothesis that the series is stationary around a deterministic trend (Table 1).

The ADF test fails to reject the null hypothesis of a unit root, indicating that the series is non-stationary. Concurrently, the KPSS test rejects the null of trend stationarity, implying that stationarity cannot be achieved by removing a deterministic trend. The joint interpretation of these results—in line with econometric best practices—strongly supports the presence of a stochastic trend. This finding

Table 1. Stationarity Tests: ADF and KPSS on the Original Series

Test	Test Statistic	5% Critical Value
ADF (with drift)	−1.832	−2.930
KPSS (with trend)	0.208	0.146

justifies modeling the series as integrated of order one, i.e., I(1), and applying first-order differencing as a necessary step toward achieving stationarity.

Trend Nature Derivation. Taken together, the ADF and KPSS results point to the presence of a difference-stationary process in the data, wherein shocks may induce permanent shifts in the level of the series. However, both tests have well-documented limitations: they are sensitive to sample size, lag structure, and the presence of structural breaks—a crucial factor in our case, given the abrupt decline in 2020.

As such, while these preliminary diagnostics provide a foundation for transformation strategy, they must be complemented by model-based validation. This motivates the transition to the next analytical step, where ARIMA models with different structural assumptions are estimated and evaluated based on their empirical adequacy, explanatory power, and robustness to structural change.

3 Model Estimation and Specification

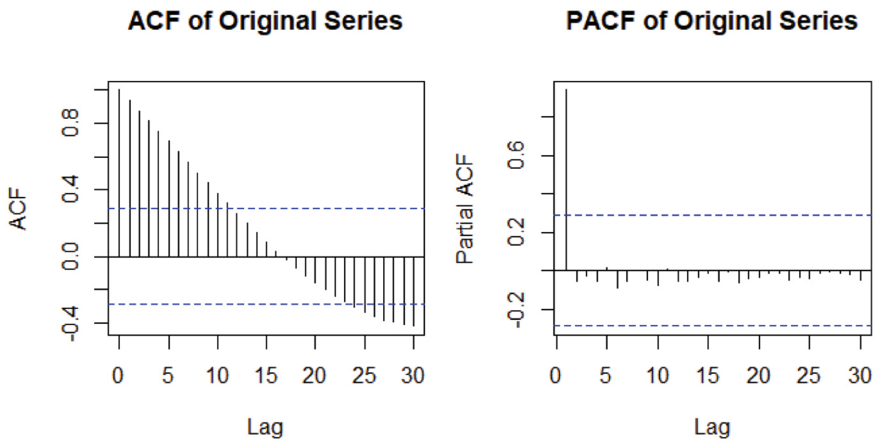

Fig. 2. ACF and PACF of the original (non-differenced) series.

Preliminary Identification via ACF and PACF. The initial model selection process begins with the analysis of autocorrelation patterns. The autocorrelation function (ACF) of the original life expectancy series displays a slowly decaying structure across the first several lags. This long memory behavior is consistent with a non-stationary series possessing a unit root. For instance, the ACF coefficients remain relatively high at lags 1 through 10 (e.g., 0.941, 0.878, 0.817), indicating strong temporal dependence (Fig. 2).

Conversely, the partial autocorrelation function (PACF) demonstrates a sharp cut-off after the first lag. The PACF at lag 1 is large and significant, while subsequent lags are statistically insignificant and fluctuate around zero. This signature is typical of an AR(1) process in the presence of differencing.

These observations are consistent with the stationarity tests discussed in Sect. 2, and they support an initial specification of an ARIMA(1,1,0) model. However, to capture richer short-term dynamics and potential measurement errors, we also consider models with a moving average (MA) component, such as ARIMA(1,1,1).

Trend Treatment and Differencing Strategy. An essential component of ARIMA modeling involves deciding how to transform a non-stationary series into a stationary one. This process depends on whether the trend is assumed to be deterministic or stochastic. Accordingly, we apply two distinct preprocessing strategies to the life expectancy data, aligned with the conceptual structure of the models under comparison.

- **Models without a deterministic trend:** In these specifications, we assume that the trend component is stochastic and best addressed through differencing. The original series is differenced once, based on the results of the ADF and KPSS tests and supported by the slow decay pattern in the ACF. First-order differencing is thus employed to remove the stochastic trend and achieve mean stationarity.
- **Models with a deterministic trend:** Here, we posit the existence of a linear deterministic trend in the data, which is removed prior to differencing. This is achieved via ordinary least squares (OLS) regression on a time index. The residuals from this regression—interpreted as deviations from the trend—are subsequently differenced as needed. This approach is suitable when the trend is viewed as exogenous or policy-driven and not subject to random shocks.

This dual approach allows us to explore whether the life expectancy series is better modeled by an exogenous, smooth trend or an endogenous, evolving one. In both cases, the resulting transformed series are subjected to further unit root testing to verify that stationarity has been successfully achieved (Table 2).

Table 2. Number of Differences Applied for Stationarity

Model	Differencing Order
ARIMA(1,d,0) without trend	1
ARIMA(1,d,1) without trend	1
ARIMA(1,d,0) with trend	1
ARIMA(1,d,1) with trend	1

The table above confirms that all model variants ultimately require a single difference to induce stationarity. However, the interpretation of the differencing operation differs across model classes. In models without a trend, differencing captures the full stochastic evolution of the process. In trend-adjusted models, it captures fluctuations around a pre-estimated deterministic path. As we will show, these conceptual differences have significant implications for model performance, especially in the face of structural shocks.

Model Candidates and Rationale. To systematically explore the structure of the time series, we estimate and compare four ARIMA model variants. Each model differs in the presence or absence of autoregressive (AR) and moving average (MA) components, and all are specified with a first-order difference ($d = 1$) based on the unit root tests and autocorrelation diagnostics.

The general ARIMA(p, d, q) model can be written as:

$$\Phi(L)(1 - L)^d y_t = \Theta(L)\varepsilon_t,$$

where:

- y_t is the observed time series (life expectancy);
- L is the lag operator, such that $Ly_t = y_{t-1}$;
- $(1 - L)^d$ denotes differencing of order d;
- $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is the autoregressive polynomial of order p;
- $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$ is the moving average polynomial of order q;
- ε_t is a white noise error term, with mean zero and constant variance.

Model Candidates and Rationale. Based on the preliminary diagnostics—specifically, the slow exponential decay of the autocorrelation function (ACF) and the sharp cut-off in the partial autocorrelation function (PACF) after lag 1—we focus our analysis on autoregressive integrated moving average (ARIMA) models with first-order differencing ($d = 1$), an autoregressive term of order one (AR(1)), and optionally a moving average term of order one (MA(1)).

This leads to two principal model specifications:

- **ARIMA(1,1,0):**
  $$(1 - \phi_1 L)(1 - L)y_t = \varepsilon_t$$
  This model includes a first-order autoregressive component ϕ_1 applied to the differenced series. It captures persistence in the deviations from the stochastic trend and is directly motivated by the PACF structure observed in the data.
- **ARIMA(1,1,1):**
  $$(1 - \phi_1 L)(1 - L)y_t = (1 + \theta_1 L)\varepsilon_t$$
  This more flexible specification combines both autoregressive and moving average terms, allowing the model to capture richer short-term dynamics and potentially serially correlated shocks or measurement noise.

Each of these specifications is estimated in two versions:

1. **Without deterministic trend:** The original series is differenced directly, under the assumption that the trend is stochastic (i.e., driven by a unit root process).
2. **With deterministic trend:** A linear trend is first removed via ordinary least squares (OLS) regression on time, and the residuals are subsequently differenced to isolate the stochastic component.

This results in a total of four candidate models, defined by the presence or absence of a moving average component (MA(1)) and a deterministic trend. All models assume first-order differencing ($d = 1$), as supported by unit root tests and autocorrelation diagnostics.

– **ARIMA(1,1,0) without trend:**

$$(1 - \phi_1 L)(1 - L)y_t = \varepsilon_t$$

This is a difference-stationary model with a single autoregressive term. It models stochastic trends without any deterministic component, allowing for permanent level shifts in response to shocks.

– **ARIMA(1,1,1) without trend:**

$$(1 - \phi_1 L)(1 - L)y_t = (1 + \theta_1 L)\varepsilon_t$$

This more flexible specification adds a moving average term to capture serial correlation in the innovations of the differenced series. It improves short-run fit without imposing a deterministic trend.

– **ARIMA(1,1,0) with trend:**

$$(1 - \phi_1 L)(1 - L)(y_t - \alpha - \beta t) = \varepsilon_t$$

This model assumes a deterministic linear trend $\alpha + \beta t$, which is removed before differencing. The stochastic component is then modeled as an AR(1) process.

– **ARIMA(1,1,1) with trend:**

$$(1 - \phi_1 L)(1 - L)(y_t - \alpha - \beta t) = (1 + \theta_1 L)\varepsilon_t$$

This specification combines a deterministic trend with both autoregressive and moving average components. It provides the greatest flexibility for capturing both structural trend and transitory deviations.

These models are selected not only for their theoretical consistency with the observed autocorrelation patterns, but also for their parsimony and interpretability in the presence of structural shocks. Their adequacy is evaluated using standard information criteria—namely, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)—alongside residual diagnostics and out-of-sample forecasting performance. The next section further examines how these specifications differ in the treatment of trend components and in the transformations required to ensure stationarity.

4 Impulse Response and Dynamic Behavior

To analyze how each ARIMA model propagates the effect of an exogenous shock over time, we compute the impulse response function (IRF) associated with a one-time innovation. The IRF represents the expected dynamic adjustment of the process following a unit shock to the error term at time t, i.e., $\varepsilon_t = 1$, while all other future shocks are assumed to be zero.

We focus on the ARIMA(1,1,1) model, specified as:

$$(1 - \phi_1 L)(1 - L)y_t = (1 + \theta_1 L)\varepsilon_t,$$

which can be rearranged into its state-space equivalent for the differenced series $\Delta y_t = y_t - y_{t-1}$:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Now, consider a one-time shock $\varepsilon_t = 1$ at time t, with $\varepsilon_{t+1} = \varepsilon_{t+2} = \ldots = 0$. The impulse response function (IRF) then traces the effect on Δy_{t+h} and, cumulatively, on y_{t+h}.

The dynamic path is obtained recursively:

$$\Delta y_t = 1 + \theta_1 \cdot 0 = 1,$$

$$\Delta y_{t+1} = \phi_1 \cdot \Delta y_t + 0 + \theta_1 \cdot 1 = \phi_1 + \theta_1,$$

$$\Delta y_{t+2} = \phi_1 \cdot \Delta y_{t+1} + 0 + \theta_1 \cdot 0 = \phi_1(\phi_1 + \theta_1),$$

$$\Delta y_{t+3} = \phi_1 \cdot \Delta y_{t+2} = \phi_1^2(\phi_1 + \theta_1),$$

$$\vdots$$

$$\Delta y_{t+h} = \phi_1^{h-1}(\phi_1 + \theta_1), \quad h \geq 2.$$

Thus, the cumulative effect on the level y_{t+h} is given by:

$$IRF(h) = \sum_{j=0}^{h} \Delta y_{t+j},$$

which shows that the autoregressive parameter ϕ_1 governs the rate at which the effect of the initial shock decays over time. When $|\phi_1| < 1$, the impact of the shock gradually diminishes but may persist for many periods, depending on the value of ϕ_1. When ϕ_1 is close to 1, the decay is slow, and the shock has near-permanent effects—characteristic of highly persistent stochastic trends.

By contrast, in models with a deterministic trend, such as:

$$(1 - \phi_1 L)(1 - L)(y_t - \alpha - \beta t) = \varepsilon_t,$$

the deterministic trend $\alpha + \beta t$ acts as an anchor, forcing the series to eventually return to a fixed growth path. Consequently, the IRF in such models tends to converge back to zero, even after large shocks, implying that structural breaks are treated as temporary.

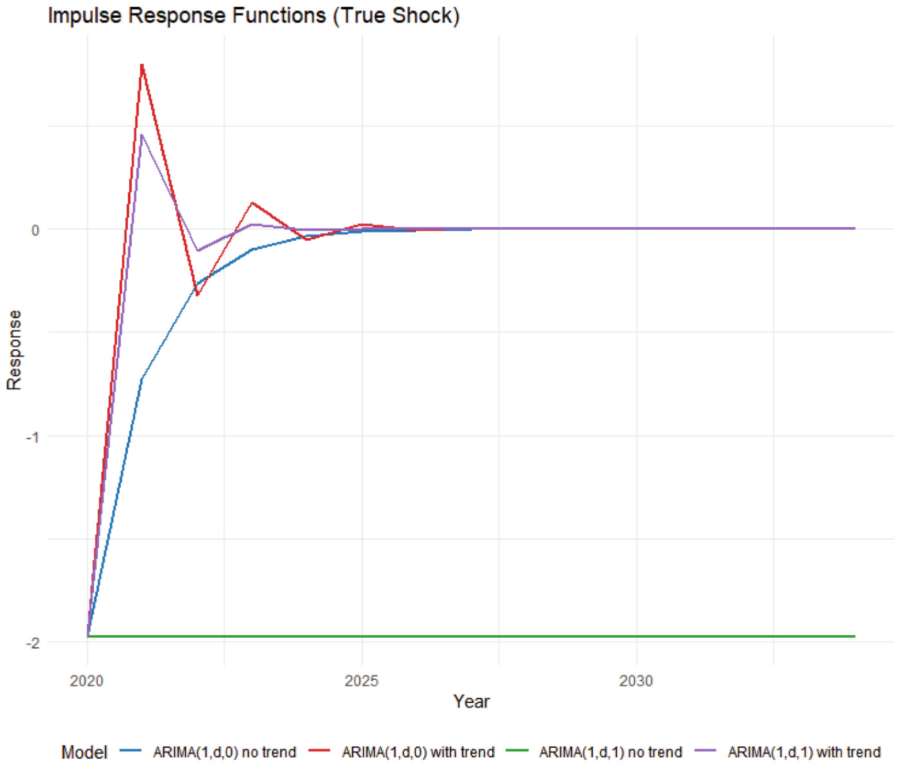

Fig. 3. Impulse Response Functions to a Negative Shock in 2020

As shown in Fig. 3, the ARIMA(1,1,1) model without trend produces a persistent yet gradually decaying IRF, consistent with the partial recovery in life expectancy observed after the 2020 pandemic shock. Conversely, models with deterministic trends exhibit rapid reversion, underestimating the long-term impact of the disruption.

In summary, the autoregressive coefficient ϕ_1 *plays a central role in shaping the post-shock trajectory of the series. The IRF provides a structural lens through which the plausibility and realism of different ARIMA models can be assessed in the presence of rare but impactful events.*

It is important to emphasize that the IRF represents the deviation from the model's expected path in the absence of the shock. In practical terms:

– For models **without a deterministic trend**, the IRF is added to the last observed level y_{t-1}, since the model assumes a stochastic trend. The post-shock trajectory is therefore:

$$y_{t+h} = y_{t-1} + \sum_{j=0}^{h} IRF(j).$$

– For models **with a deterministic trend** (e.g., linear $\alpha + \beta t$), the IRF is added to the extrapolated trend component:

$$y_{t+h} = \hat{\alpha} + \hat{\beta}(t + h) + IRF(h).$$

In this way, the impulse response function reflects the *transitory or permanent displacement* relative to the model's baseline expectation—whether that baseline is a constant (in difference-stationary models) or a time-dependent trend (in trend-stationary models).

This distinction is critical when interpreting the effect of structural shocks: in difference-stationary processes, the entire path is shifted, while in trend-stationary processes, the shock represents only a temporary deviation from the pre-defined trajectory.

5 Forecasting Performance and Model Evaluation

Beyond capturing historical dynamics, a robust time series model must deliver accurate and interpretable forecasts—especially under conditions of structural change. To this end, we evaluate the forecasting performance of each ARIMA specification over the period 20202034, with a particular focus on the response to the COVID-19 shock.

Forecast accuracy is assessed using standard performance metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and pseudo-R^2 values. The pseudo-R^2 is calculated as:

$$\text{pseudo-}R^2 = 1 - \frac{\text{MSE}_{\text{model}}}{\text{Var}(y_{\text{obs}})},$$

which compares the mean squared forecast error to the variance of the observed series, offering an intuitive measure of explanatory power over the forecast window.

To ensure that the forecasts reflect structural dynamics and not just trend extrapolation, we explicitly incorporate impulse response functions (IRFs) into the forecast generation. The 2020 shock is introduced as a one-time innovation in the error term, and each model's internal dynamics are allowed to propagate this disturbance over time. This approach embeds structural realism into the projections by simulating how each model would have reacted ex-ante to the COVID-19 shock.

Figure 4 illustrates the forecast trajectories implied by each model. Deterministic trend models project a relatively fast return to pre-shock trajectories,

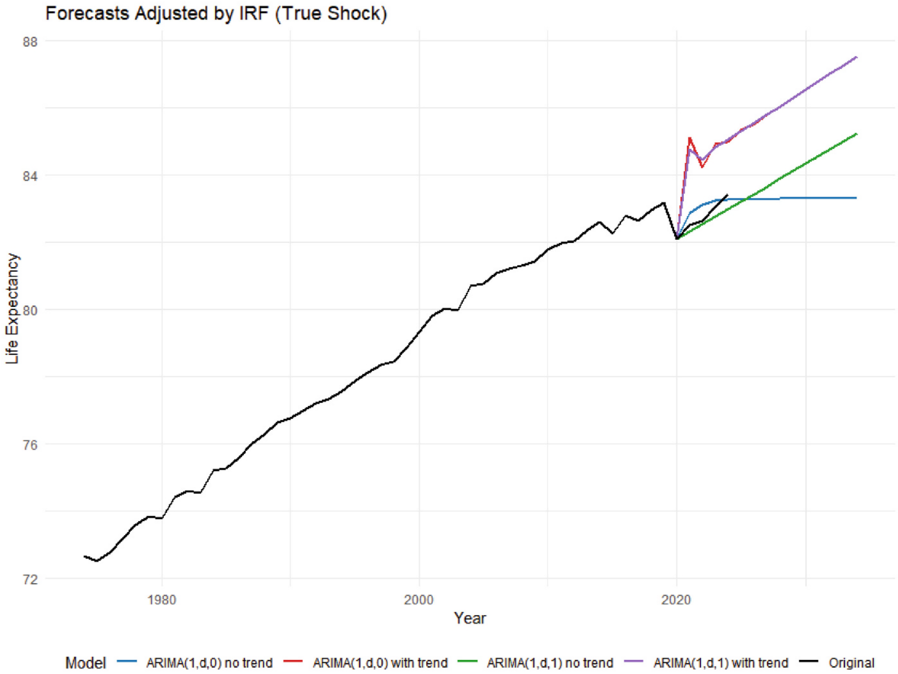

Fig. 4. Forecasting Comparison Across Model Specifications (19742034)

while stochastic-trend models—particularly **ARIMA(1,1,1) without trend**—forecast more persistent effects and a gradual recovery. These visual differences reflect fundamentally distinct assumptions about shock persistence and system memory.

The accuracy of the forecasts is summarized in Table 3, which compares predicted and observed life expectancy values over 20202024.

The **ARIMA(1,1,1) without trend** model demonstrates the best predictive accuracy across all criteria, with the lowest MAE and RMSE and a pseudo-R^2 of 0.841. This model exhibits robustness to structural breaks and aligns more closely with the actual path of life expectancy in the wake of COVID-19.

In contrast, deterministic trend models—despite their superior in-sample statistical fit—underperform in the post-2020 window. Their forecasts systematically overestimate recovery, resulting in higher prediction errors and even negative pseudo-R^2 scores. This discrepancy underscores the limitation of assuming full mean reversion in the presence of a persistent exogenous shock.

In summary, while deterministic trend models may appear statistically appealing under stable conditions, they prove less reliable in disrupted environments. More flexible specifications—especially ARIMA models with both autoregressive and moving average terms but without imposed trend—offer improved forecasting performance and greater adaptability to real-world structural changes.

Table 3. Comparison of Forecasted and Observed Life Expectancy Values (Decimal Format)

Year	Observed	AR(1,d,0) NT	AR(1,d,1) NT	AR(1,d,0) TR	AR(1,d,1) TR
2020	82.08	82.34	82.19	82.22	82.17
2021	82.50	82.57	82.43	82.66	82.49
2022	82.64	82.76	82.66	83.09	82.81
2023	83.03	82.92	82.89	83.53	83.12
2024	83.42	83.04	83.10	83.96	83.43

6 Comparative Model Evaluation

In this section, we systematically compare the four ARIMA model specifications along multiple performance dimensions, incorporating both in-sample fit and out-of-sample forecast accuracy [14]. The objective is to understand not only which model best approximates the historical data, but also which one generalizes effectively under structural instability.

6.1 In-Sample Model Fit

We begin with an evaluation of statistical fit over the estimation sample using conventional information criteria: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the log-likelihood [10,15,19]. These metrics are presented in Table 4.

Table 4. Model Diagnostics and Performance Metrics

Model	AIC	BIC	LogLik	MAE	RMSE	Pseudo-R^2
ARIMA(1,d,1) no trend	1.03	6.45	2.49	**0.158**	**0.182**	**0.841**
ARIMA(1,d,0) no trend	22.13	25.74	−9.06	0.285	0.320	0.510
ARIMA(1,d,1) with trend	−11.83	−6.41	8.92	1.658	1.699	−12.83
ARIMA(1,d,0) with trend	**−13.21**	**−9.59**	**8.60**	1.677	1.725	−13.27

From the table, it is clear that models incorporating deterministic trends (particularly ARIMA(1,d,0) with trend) yield the best in-sample fit, as reflected in the lowest AIC and BIC values. However, this statistical superiority comes at a cost when forecasting under conditions of structural disruption.

6.2 Out-of-Sample Accuracy and Structural Robustness

When comparing forecast performance over the COVID-19 period (20202024), the models with deterministic trends exhibit marked degradation in accuracy.

Their high MAE and RMSE values, along with large negative pseudo-R^2 scores, suggest a systematic overestimation of recovery—an artifact of assuming full reversion to a pre-shock deterministic path.

Conversely, the **ARIMA(1,d,1) without trend** model achieves the lowest forecast errors and highest pseudo-R^2 among all specifications. This model accommodates the persistence of the 2020 shock while still capturing short-term dynamics, providing a more realistic depiction of the partial rebound in life expectancy observed in recent years.

6.3 Residual Diagnostics

Further support for model adequacy is provided by residual autocorrelation analysis. Figure 5 displays the ACF and PACF of residuals for all four models.

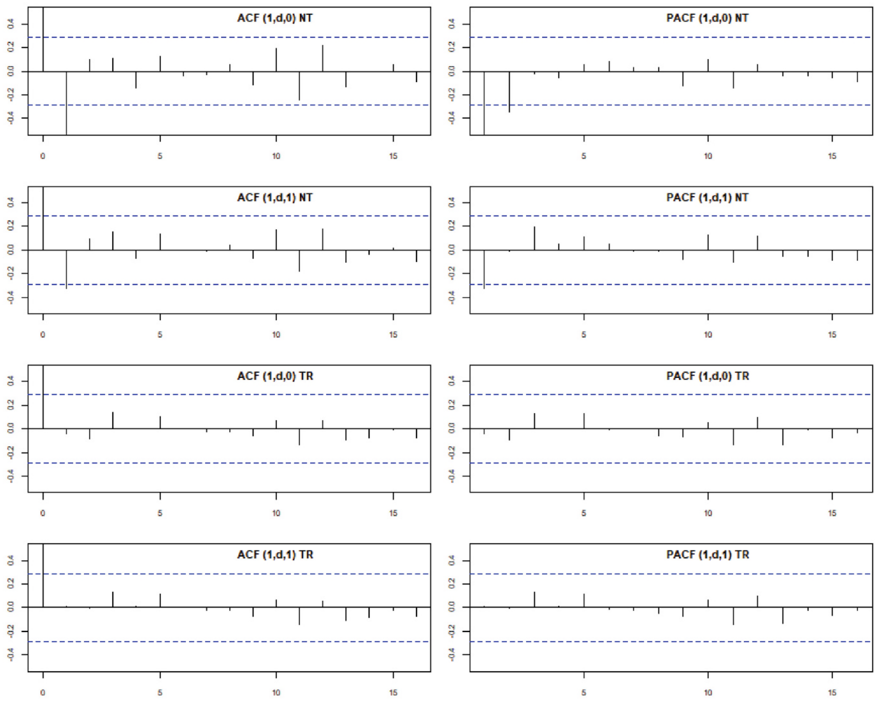

Fig. 5. ACF and PACF of Residuals from All Models

The residuals from the ARIMA(1,d,1) without trend model show no significant autocorrelation, indicating that the model successfully captures the underlying structure of the data. In contrast, the residuals from trend-based models display low-level serial dependence, suggesting that important dynamics may have been omitted or misspecified due to the imposition of a deterministic structure.

6.4 Summary of Findings

The comparative analysis reveals a fundamental trade-off between in-sample fit and structural adaptability. Models optimized for historical data—particularly those with deterministic trends—can fail to anticipate or incorporate structural changes such as pandemics. More flexible models, especially ARIMA(1,d,1) without trend, offer a better balance by accommodating both persistent level shifts and short-term corrections.

This reinforces the principle that model selection should not rely solely on information criteria or in-sample fit but must also account for theoretical plausibility and empirical robustness in the presence of shocks.

7 Conclusion

This study provides a comprehensive comparative analysis of ARIMA model specifications applied to Italian life expectancy data over the period 19742024, with a particular focus on the response to an exogenous structural shock—the COVID-19 pandemic. By evaluating model behavior across five key dimensions—trend specification, differencing strategy, dynamic adjustment, forecast accuracy, and residual diagnostics—we offer both empirical insights and methodological guidance for modeling demographic time series under uncertainty.

Our results demonstrate that model assumptions regarding the nature of the trend—deterministic versus stochastic—have profound implications for performance. While deterministic trend models (particularly ARIMA(1,d,0) with trend) achieve excellent in-sample fit, they exhibit poor forecast accuracy in the wake of structural disruption, often overestimating post-shock recovery. In contrast, the **ARIMA(1,d,1) without trend** model consistently delivers superior out-of-sample performance, lower forecast errors, and more plausible impulse responses, aligning closely with the observed partial rebound in life expectancy following the 2020 decline.

These findings emphasize the value of flexibility and parsimony in time series modeling. Models that allow for both short-run dynamics and persistent shocks—without imposing rigid deterministic structure—are better suited to capturing real-world demographic processes, especially under conditions of instability.

Beyond empirical forecasting, this work contributes to a broader theoretical understanding of demographic systems as stochastic and evolving processes. In a world increasingly shaped by non-linear shocks—pandemics, climate events, geopolitical crises—modeling frameworks must be equipped not only to fit the past but to anticipate structurally plausible futures.

In this sense, the ARIMA(1,1,1) without trend emerges not just as a statistically adequate model, but as a conceptually coherent and policy-relevant tool for demographic forecasting. It enables decision-makers to internalize uncertainty, evaluate alternative scenarios, and respond more effectively to systemic risks.

We believe these results hold relevance beyond the Italian case, offering a transferable modeling approach for national statistical agencies, health planners, and policy institutions grappling with the demographic consequences of a volatile global environment. Future work may extend this framework to multivariate models, explore regime-switching structures, or integrate high-frequency health and mortality indicators to enhance predictive power.

Ultimately, this study affirms the power of structurally grounded, empirically validated models in navigating the complexity of demographic change—and underscores the need for resilient forecasting tools in an era of uncertainty.

References

1. Alshehri, A., El-Halawany, F., Abu-Zinadah, H.: Prediction of life expectancy in saudi arabia by 2030 using arima models. J. Phys: Conf. Ser. **1978**(1), 012060 (2021)
2. ArunKumar, K.E., Kalaga, D.V., Kumar, C.M.S., Chilkoor, G., Kawaji, M., Brenza, T.M.: Forecasting the dynamics of cumulative covid-19 cases for top-16 countries using statistical machine learning models: Arima and sarima. Appl. Soft Comput. **103**, 107161 (2021)
3. Adeleye, D.A., Muhajarine, N.: Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, holt-winters exponential smoothing and arima models. BMC Med. Res. Methodol. **20**(1), 1–13 (2020)
4. Cerqueira, V., Torgo, L., Mozetič, I.: Evaluating time series forecasting models: an empirical study on performance estimation methods. Mach. Learn. **109**(11), 1997–2028 (2020). https://doi.org/10.1007/s10994-020-05910-7
5. Torri, T., Vaupel, J.W.: Forecasting life expectancy in an international context. Int. J. Forecasting **28**(2), 519–531 (2012)
6. Box, G.E.P., Pierce, D.A.: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Am. Stat. Assoc. **65**(332), 1509–1526 (1970)
7. Christiano, L.J., Eichenbaum, M., Evans, C.L.: Monetary policy shocks: What have we learned and to what end? Handb. Macroecon. **1**, 65–148 (1999)
8. Clements, M.P., Hendry, D.F.: Economic forecasting: some lessons from recent research. Econ. Modelling **18**(2), 323–339 (2001)
9. Diebold, F.X.: Forecasting in Economics, Business, Finance and Beyond. University of Pennsylvania, 2015. Online textbook. https://www.sas.upenn.edu/~fdiebold/ForecastingBook.html
10. Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. J. Bus. Econ. Stat. **13**(3), 253–263 (1995)
11. James, D.: Hamilton. Princeton University Press, Time Series Analysis (1994)
12. Husin, W.Z.W., Aziz, N.A.b.: Box-jenkins model for forecasting malaysia life expectancy. Int. J. Acad. Res. Bus. Soc. Sci. **11**(12), 11980 (2021)
13. Liu, Z., Zhu, Z., Gao, J., Xu, C.: Forecast methods for time series data: a survey. IEEE Access **9**, 91896–91912 (2021)

14. Lütkepohl, H.: New introduction to multiple time series analysis. Springer, Heidelberg (2005)
15. Marcellino, M., Stock, J.H., Watson, M.W.: A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. J. Econometrics **135**(1–2), 499–526 (2006)
16. Pandey, A., Chhikara, R.: Analysis of life expectancy using various regression techniques. In: 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), p. 9362914 (2020)
17. Perone, G.: An arima model to forecast the spread and the final size of covid-2019 epidemic in Italy. medRxiv (2020)
18. Perone, G.: Comparison of arima, ets, nnar and hybrid models to forecast the second wave of covid-19 hospitalizations in Italy. medRxiv (2020)
19. Pesaran, M.H., Schuermann, T., Vanessa Smith, L.: Forecast scenarios for global gdp using model averaging. Int. J. Forecasting **22**(4), 527–542 (2006)