Revealing the Nature of Italian Life Expectancy

A Comparative Study of ARIMA Models Using the COVID-19 Shock

Girolamo Franchetti   Massimiliano Politano

University of Naples Federico II - Economics and Statistics Department

---

15-th Scientific Meeting Classification and Data Analysis Group

September 10th, 2025

Why This Study?

- Life Expectancy is a pivotal demographic indicator. In Italy, it has historically followed a nearly linear upward trend.
- Finding the appropriate stochastic model would give predictability of the process towards different influencing factors and events.
- The COVID-19 pandemic triggered an abrupt drop in 2020, disrupting this long-run pattern. Life expectancy fell by 1.08 years respect to 2019.

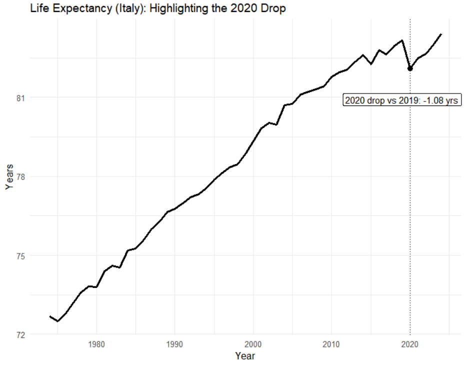


Figure: Life expectancy (Italy, 1974–2024): the 2020 COVID-19 drop

The Roadmap

- **It is possible to exploit a COVID-19-like artificial shock to check what model replicates better the process' behaviour (reactions).**
- **A tough task** to accomplish in safe way.
- **All models' reactions** start from 2020 with $e_{2020} = 82.08$ as observed.
- **Model-depended shocks:** ensures comparability according to models' features.

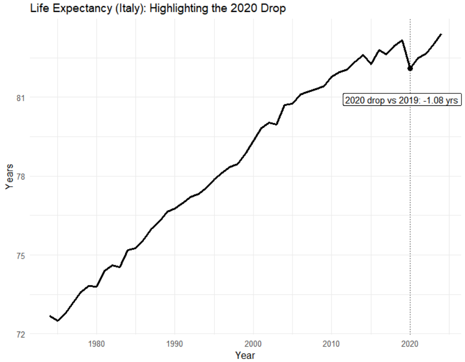


Figure: Life expectancy (Italy, 1974–2024): the 2020 COVID-19 drop

From Questions to Modeling Strategy

Main questions

- Is ARIMA an appropriate model in the presence of such severe shocks? If so, which specification is most robust to such disruptions?
- Is there a deterministic or stochastic trend?

From Questions to Modeling Strategy

Main questions

- Is ARIMA an appropriate model in the presence of such severe shocks? If so, which specification is most robust to such disruptions?
- Is there a deterministic or stochastic trend?

Five-step pipeline

- **Step 1:** Test stationarity using ADF and KPSS.
- **Step 2:** Use ACF/PACF to identify ARIMA components.
- **Step 3:** Identify and estimate possible ARIMA variants (with/without trend).
- **Step 4:** Evaluate model accuracy using error metrics and residuals.
- **Step 5:** Inject calibrated shock and analyse IRF propagation across models.

Key Insights

- Trend assumptions (deterministic vs. stochastic) crucially shape outcomes.
- Severe shocks could expose model limits and features.
- Results are acceptable for Italy only, each country has its own life expectancy path and appropriate statistic model.

Data: ISTAT Mortality Table Database from 1974 to 2024 (published in 2025) - entire population.

Step 1 — Stationarity Investigation

Objective: Test whether life expectancy is stationary or contains a unit root.

- **ADF test (with drift):** statistic -1.832 is *above* the 5% critical value -2.930 $\Rightarrow$ cannot reject unit root $\Rightarrow$ *non-stationary*.
- **KPSS test (trend-stationary):** statistic 0.208 is *above* the 5% critical value 0.146 $\Rightarrow$ reject trend-stationarity $\Rightarrow$ *non-stationary*.
- **Conclusion:** both tests agree that the series is **non-stationary**. Differencing once ($d = 1$) yields stationarity $\Rightarrow$ the series is an $I(1)$ process.

Test	Stat	5% Crit.
ADF (drift)	-1.832	-2.930
KPSS (trend)	0.208	0.146

Table: Unit root vs. trend-stationarity

- ADF: tests H_0: unit root (non-stationary).
- KPSS: tests H_0: trend-stationary.
- Consistent outcome: the series follows a stochastic trend.

Step 2 — ARIMA Structure Identification

- ACF slow decay; PACF spike at 1 $\Rightarrow$ nonstationary with AR(1) signal $\Rightarrow$.
- ARIMA models will have AR component of order 1.
- Necessary to check for MA component if order 1 or 0.

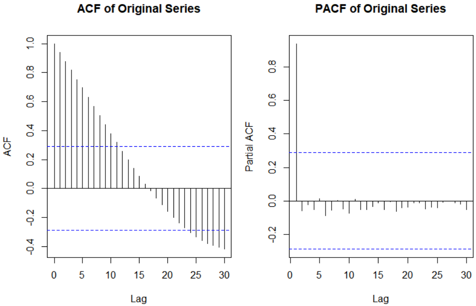


Figure: ACF and PACF of the *undifferenced* series

Step 3.1 - Model identification

1. Given Autoregressive and Integration component of order 1;
2. Then, ARIMA models will be of order (1,1,·);
3. So, test for MA component of order 0 and 1;
4. The same, for deterministic trend component (yes/no);
5. Models will be estimated on 1974–2019;
6. Then, forecast on 2020–2034.

Step 3.2 - ARIMA Coefficients

Compare models: ARIMA(1,1,0) vs. ARIMA(1,1,1), each with/without deterministic trend (NT / TR).

Model	AR(1)	MA(1)	σ^2	AIC
ARIMA(1,1,0) NT	0.367	—	0.089	22.13
ARIMA(1,1,1) NT	**0.9999**	-0.985	0.051	1.04
ARIMA(1,1,0) TR	-0.404	—	0.041	-13.21
ARIMA(1,1,1) TR	-0.230	-0.220	0.041	-11.83

Table: Estimated coefficients and fit statistics

Step 4.1 — Model Performance Comparison

- **In-sample (1974–2019):** trend-based models achieve the best statistical fit (AIC/BIC, LogLik).
- **Out-of-sample (2020–2024): ARIMA(1,1,1) NT** delivers the most accurate forecasts (lowest MAE/RMSE, highest pseudo-R^2).

Model	AIC	BIC	LogLik	RMSE	MAE	Pseudo-R^2
ARIMA(1,1,1) no trend	1.04	6.46	2.48	0.219	0.167	**0.768**
ARIMA(1,1,0) no trend	22.13	25.74	-9.06	0.292	0.227	0.679
ARIMA(1,1,1) with trend	-11.83	-6.41	**8.92**	**0.196**	**0.148**	-9.954
ARIMA(1,1,0) with trend	**-13.21**	**-9.59**	8.60	0.197	0.147	-10.854

Table: Model performance: AIC/BIC, LogLik = in-sample fit (1974–2019); RMSE/MAE = training errors; pseudo-R^2 = forecast accuracy (2020–2024).

Pseudo R^2

Step 4.2 — Residual Analysis

- **ACF/PACF diagnostics:** NT models display mild spikes at short lags; TR models appear closer to white noise.
- **Interpretation:** deviations are small and not systematic, thus residuals can still be regarded as approximately uncorrelated.
- **Takeaway:** despite minor signals, **ARIMA(1,1,1) NT remains adequate and robust for forecasting.**

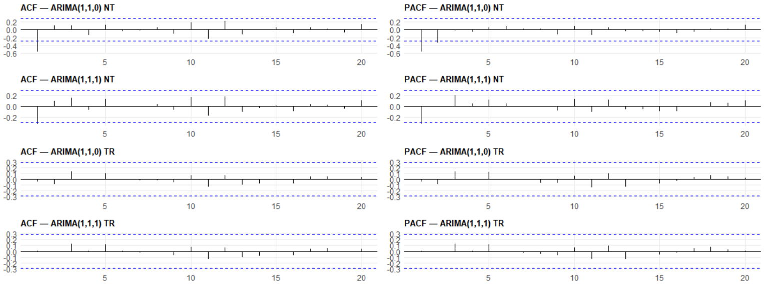


Figure: Residual ACF/PACF (all models)

Step 5.1 — Shock Calibration

- **TR baseline (OLS):** estimate linear trend on $t \leq 2019$ and extrapolate to 2020.

$$y_t = \alpha + \beta t + u_t, \quad \widehat{e}^{TR}_{2020} = \hat{\alpha} + \hat{\beta} \cdot 2020$$

- **NT baseline (ARIMA):** use the model's own forecast for 2020.

$$\widehat{e}^{NT}_{2020} = \text{ARIMA prediction at } t = 2020$$

- **Shock calibration:** deviation of observed 2020 value from the chosen baseline Values.

$$Y_{2020} = \begin{cases} e_{2020} - \widehat{e}^{TR}_{2020} & \text{for TR models} \\ e_{2020} - \widehat{e}^{NT}_{2020} & \text{for NT models} \end{cases}$$

Step 5.2 — Recursive Impulse Response (2020–2034)

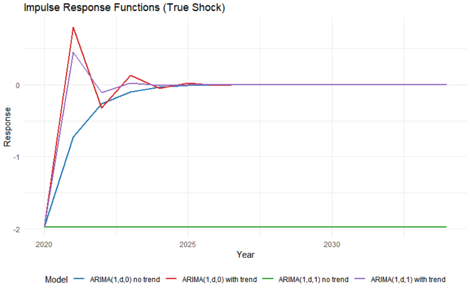
- **Shock 2020 (by baseline):**

$$Y_{2020} = \begin{cases} e_{2020} - \hat{e}^{TR}_{2020} & \text{(TR models)} \\ e_{2020} - \hat{e}^{NT}_{2020} & \text{(NT models)} \end{cases}$$

- **Propagation AR(1):**

$$Y_{2021} = \phi_{\text{model}} \cdot Y_{2020},$$
$$Y_t = \phi_{\text{model}} \cdot Y_{t-1}$$
$$\text{for } t = \{2022, \dots, 2034\}$$

- Y_t is the rensponse series to the shock (IRF) 2020–2034.



Impulse response: shock propagation to 2034

Step 5.3 — Model Response Comparison

- The calibrated shock ($\varepsilon^* \approx -1.97$ years) is broadly consistent with the observed decline in life expectancy reported by ISTAT (2021). Details
- Combine the recursive shock with the model's fitted values:

$$e^{for}_{model}(t) = \hat{e}_{model}(t) + Y_t$$

- Compare forecasts to observed data (2021–2024): below we report signed errors year-by-year and **absolute-error aggregates** (lower is better) .

Year	ARIMA(1,d,0) NT	ARIMA(1,d,1) NT	ARIMA(1,d,0) TR	ARIMA(1,d,1) TR
2021	0.350	-0.186	2.609	2.265
2022	0.485	-0.104	1.593	1.810
2023	0.196	-0.270	1.901	1.795
2024	-0.152	-0.429	1.582	1.629
Mean |error|	0.296	**0.247**	1.921	1.875
Sum |error|	1.183	**0.989**	7.685	7.499

Step 5.4 — Model Response Comparison

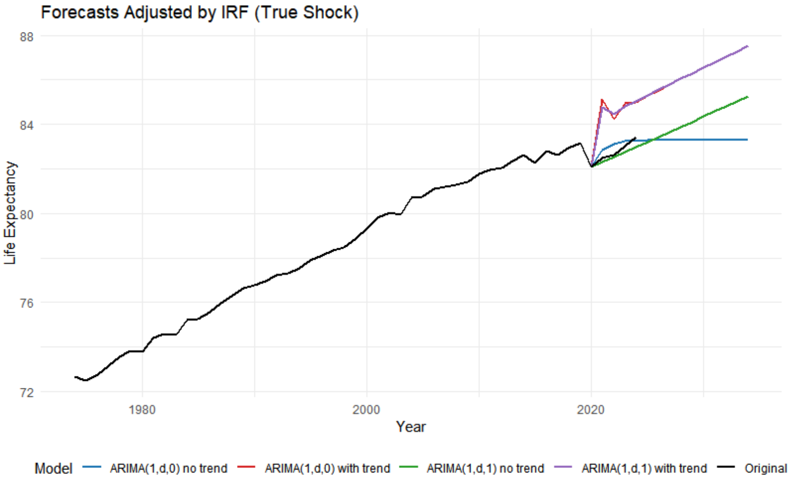


Figure: Forecast containing IRF propagation from 2020 to 2034

Summary of Findings

- **Choice problem:** it is necessary an equilibrium between fitting and forecasting power:
- **Deterministic trend component**
  - Excellent historical fit
  - Fail to capture severe breaks and relative response (e.g., COVID-19)
- **Lesson:** model choice not only AIC/BIC and MAE/RMSE, but also theoretical plausibility & robustness under shocks
- **COVID-19** = it seems temporary step back, not a slowdown in long-run growth

Final Insight: Models Must Handle Structural Breaks

- **ARIMA(1,1,1) without trend:**
  - Tracks historical dynamics
  - Incorporates structural shocks
  - Delivers realistic, policy-relevant projections
- ARIMA(1,1,1) without trend, **in Italian context**, is the most adequate to reply time series pattern.
- Here, this process seems to evolve, it does not revert mechanically.

Necessary to wait for new data

Final Insight: Models Must Handle Structural Breaks

- **ARIMA(1,1,1) without trend:**
  - Tracks historical dynamics
  - Incorporates structural shocks
  - Delivers realistic, policy-relevant projections
- ARIMA(1,1,1) without trend, **in Italian context**, is the most adequate to reply time series pattern.
- Here, this process seems to evolve, it does not revert mechanically.

Necessary to wait for new data

Thank You!

Actuarial Life Expectancy at Birth

Definition: The actuarial life expectancy at birth, denoted as e_0, represents the average number of years a newborn is expected to live, assuming current mortality rates remain constant throughout life.

Interpretation: It is calculated by dividing the total number of person-years lived by the cohort from birth onward by the initial number of births. This measure reflects longevity under stationary mortality conditions and is widely used in demographic and actuarial analyses.

Formula:

$$e_0 = \frac{T_0}{l_0}$$

Notation:

- e_0: Life expectancy at birth
- $T_0 = \sum_{x=0}^{\infty} L_x$: Total person-years lived from age 0 onward
- L_x: Person-years lived between ages x and $x+1$
- l_0: Number of births in the cohort (typically 100,000)

Return

Pseudo-R^2 (out-of-sample)

- Purpose: single, scale-free index of forecast accuracy.
- Definition (here):

$$\text{pseudo-}R^2 = 1 - \frac{\mathrm{MSE}(\hat{e}_t)}{\mathrm{Var}(e_t)}$$

  computed on the evaluation window (e.g., 2020–2024).
- Interpretation:
  - ≈ 1: excellent forecasts (small MSE vs variability of data)
  - ≈ 0: no gain vs variance baseline
  - < 0: worse than using the sample mean
- Note: not a regression R^2; it is an error–variance ratio used for model comparison.

Return

Forecast distance to observed values

Methodology:

- Compute model forecast $e^{for}_{model}(t)$
- Take observed data $e(t)$
- Compute the difference $e^{for}_{model}(t) - e(t) = \Delta e(t)$
- Obtain mean e sum of absolute errors $|\Delta \bar{e}(t)|$

Return

Shock Calibration (2020)

Model	Shock_2020
ARIMA(1,1,0) NT	-1.157
ARIMA(1,1,1) NT	**-1.305**
ARIMA(1,1,0) TR	-1.974
ARIMA(1,1,1) TR	-1.974

Notes. Shocks are defined as $e_{2020} - \widehat{e}_{2020}$, where the baseline is: (1) the ARIMA forecast at 2020 for NT models; (2) the OLS trend projection at 2020 for TR models. Differences reflect how each baseline interprets the break induced by COVID-19.

Return