Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v16n2p473

Investigation on life expectation at birth process nature.

By Giovanna Di Lorenzo, Girolamo Franchetti, Massimiliano Politano

14 October 2023

Electronic Journal of Applied Statistical Analysis
Vol. 16, Issue 2, 2023, 473-486
DOI: 10.1285/i20705948v16n2p473

Investigation on life expectation at birth process nature.

An analysis of shock response to the COVID19 outbreak

Giovanna Di Lorenzo * , Girolamo Franchetti†‡ , and Massimiliano Politano§

University of Naples "Federico II" - Economics and Statistics Department, , 249-251 Vicinale Cupa Cintia Street, 80126 Naples (NA), Italy

14 October 2023

Abstract: This framework consists in an application of an ARIMAX model in order to analyze the impact of COVID19 pandemic on the life expectancy Malia (2021) at birth process with a Impulse Response Function. More precisely we model the process, then we emulate the shock observed on this process to check if it is stable and trend reverting. In this way we decompose the time series of the life expectancy at birth in a deterministic trend component and a residual part where we will model with an ARIMAX model with order 1 in auto-regressive component and with order 1 for the integration component. We will show that residual could be considered as white noise and, then, that the process is estimated to be stable and trend reverting in few years. Finally we compare forecast of the model without shock with the one with shock. Then, we complete the comparison looking at the observed life expectancy at birth in order to see if the realty is following the theoretical framework.

keywords: Time Series Analysis, Stochastic Processes, ARIMAX Models, Life Expectation.

*Email: giovanna.dilorenzo@unina.it
†Corresponding author
‡Email: girolamo.franchetti@unina.it
§Email: massimiliano.politano@unina.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

Once the pandemic is over it is necessary to take advantage of this calm period to think about the effects of COVID19 pandemic. This pandemic has been a shock under different aspects for people (Alradaideh et al., 2022; Keser and Deveci Kocakoc, 2021). When we analyze a shock on a process we have to see its duration. If the shock is transitory this event lasts of a finite number of time periods, if it is permanent it lasts for infinite time periods. Once a shock is happened we need to see if this shock has temporary effects or permanent. If a shock has temporary effects it means that, given a shock, the process gets far from the original trend but it will revert to it. If a shock has permanent effects it means that, although the shock ended, it causes a structural change in the process' path. This shock, that is clearly temporary in its duration because this shock lasted for 1 year, in 2019. Now we need to understand if it has permanent or temporary effects on the life expectancy at birth of people. Nowaday, it is not clear, as we will see in the following sections, if this temporary shock has also temporary effects too. It could have permanent effects on the process' structure as we can see in Noymer et al. (2022) where the authors suggest that the response is different according to the economical condition of the country such that short-term effects are subject to wealthy countries. It could, otherwise, have long-term effects as explained in Carannante et al. (2023) where the authors explains that there are different levels of frailty and for the more frailty people there is a structural change in the process; for the less vulnerable ones, there are short-term effects. There are some researchers, as you can read in Islam et al. (2021), who sustain just that COVID19 has temporary effects on life expectations regardless other features of countries. We do the analysis on the Italian population looking at data from actuarial tables from ISTAT data platform[1] such that we obtain the time series of the biometric function e_0, the life expectation at birth according to which we do the investigation. Starting database available on OSF.io storage [2]. In our work we have three specific research questions to answer. We start the investigation to understand how is severe the effect of COVID19 pandemic outbreak. Then, we move on to find the best model that theoretically represents the process according to data, then, we will search for the best fitting model to data to describe the process. Finally, once we have the best theoretical process we will observe how it reacts in case of a shock similar to the COVID19 outbreak one, so, we observe how much it is shock-resistant. Our framework is defined as follows: in section 2 we define the data that we use for the investigation, then in section 3 we define the model that we want to use. So, there is section 4 where there is the residuals analysis. Once we have faced those arguments we go on the analysis of the response of the process to an impulse that emulates the COVID19 shock happened in 2019 in section 5. Finally we do a presentation of the results in 7 and then the discussion about them. This framework consists in studying the time series of e_0 in order to have a theoretical process that shows its pattern and its shape Šimpach et al. (2013), Singh and Hasija (2021). Then, how it reacts to a shock as the one of COVID19 in 2019 and

[1] `http://dati.istat.it/`
[2] `https://osf.io/5gsz6/`

we compare it with the observed time series.

2 Data description

Let's start with the description of data used in our analysis. Table 1 shows the e_0 time series features: We have taken into account the actuarial mortality tables from 1974 to

Observation period	1974 to 2022
Geographic zone	Italy
Gender	Total
Biometric function	Life Expectation e_0
Age chosen	0 Years Old

Table 1: Life expectancy at birth time series features

2022 from ISTAT database platform filtering for the biometric function e_x and selecting only 0 years old as age. Then, we obtain the time series of life expectancy at birth that is exhibited in the graph 1 We can look at 2019 observation where clearly there is the

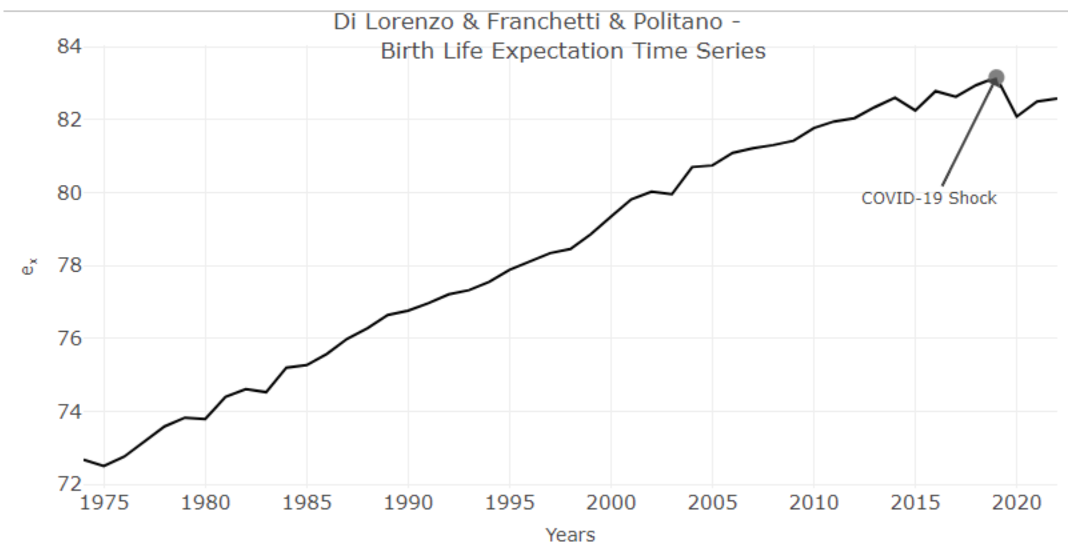

Figure 1: Life expectancy at birth time series

shock due to COVID19 pandemic. Here, life expectation from birth dropped by about 83 years to around 82. Then, it is exhibited a pattern suggests, but not in a clear way, a resumption to the original trend. That is one of the reasons that leads us to ask if this shock caused a temporary or permanent effect of the process' structure.

3 Model presentation

So given data presented in section 2 we can present the model used in this framework. We wanted to model the time series using an ARIMAX model to describe the theoretical model in order to have a reliable forecast in the next years of the time series. The time horizon is by 1974 to 2019. Then, Once we have developed the estimation, we compare the forecast by 2020 to 2030 and the impulse response of the process to the observed data from the actuarial mortality table. The model used is an ARIMAX(1,1,1) with a trend as external regressor.

$$e_{0,t} = \alpha + \rho \cdot e_{0,t-1} + \gamma \cdot t + \delta \cdot (\epsilon_t + \theta\epsilon_{t-1}) \quad (1)$$

The reason of controlling the linear trend is obvious from the graph 1, there is clearly a linear positive trend during the years that we need to control in order to make the non-stationary process into stationary one and to have a reliable forecast of the data that takes it into account. We can justify this choice looking at table 1 that exhibits a regression of the time series on a linear trend:

$$e_0 = \alpha + \beta \cdot t + \hat{e}_0 \quad (2)$$

Coefficient	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	72.391307	0.090567	799.31	$< 2e-16***$
t	0.248238	0.003355	73.98	$< 2e-16***$

Table 2: Linear trend regression

With Residual standard error: 0.3021 on 44 degrees of freedom; Multiple R-squared: 0.992; Adjusted R-squared: 0.9918; F-statistic: 5473 on 1 and 44 DF; p-value: $< 2.2e-16$. Here $\hat{e}_0$ stands for the residuals of the process once the trend is removed. The relative graph is visible in figure 2.

So, as we can see in table 2 trend is significative at 5%. So, we go on with the de-trending procedure and start to work on residuals of this model.

We have now to understand if the information set available in a observation data affects the observation of the next year one. First of all, we need to check if the residuals are stationary and then we execute the augmented Dickey-Fuller test on the residuals, in table 3 the results:

Augmented Dickey-Fuller Test
Dickey-Fuller = -0.20717, Lag order = 3, p-value = 0.99
alternative hypothesis: stationary

Table 3: ADF on $\hat{e}_0$

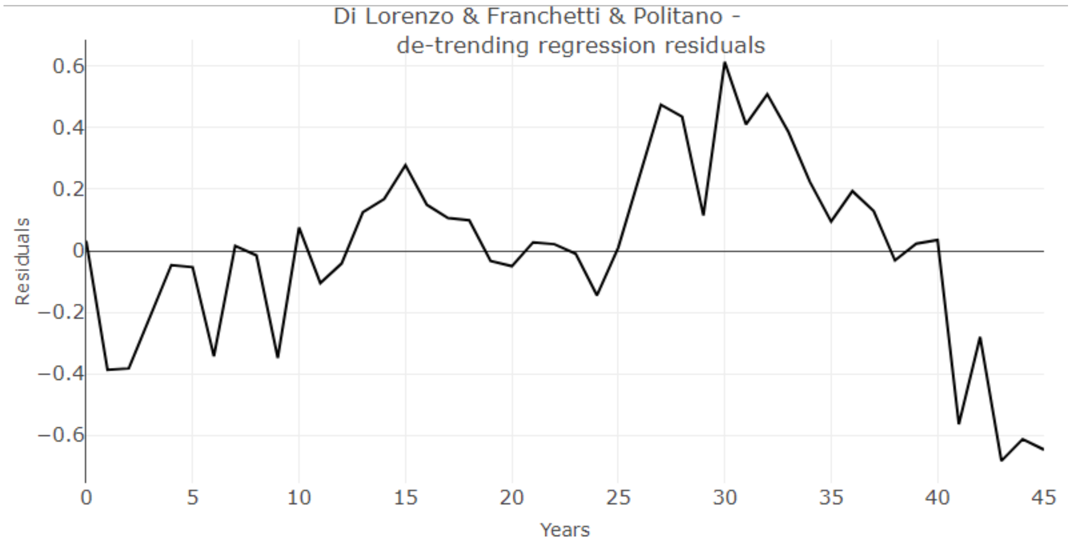

Figure 2: $\hat{e}_0$ time series

So, we need to go on with the order 1 integration, so the first difference, in order to see if the time series becomes stationary. Then if we do the first difference of the residuals $\hat{e}_0$, so $\Delta\hat{e}_0$, we obtain the results in table 4:

Augmented Dickey-Fuller Test
Dickey-Fuller = -3.5224, Lag order = 3, p-value = 0.05025 alternative hypothesis: stationary

Table 4: ADF on $\Delta\hat{e}_0$

Here, we need to be careful to go for a conclusion about the stationarity of the time series. We have a p-value very close of the critic value at 5% probability and then we decide to conclude that the time series is stationary because it can be considered in this manner at 1% of confidence but also because of it is around the critical value, the 5-th percentile. This justify the order 1 of the integration.

Once we go on with the model we can look how lagged values of the process affect the present one. We start looking at the Autocorrelation Function ACF that we can see in figure 3.

We can conclude that a lagged value of one year affects the current one but we wish to go in a deeper investigation about the auto-correlation of the process. So we do another regression, visible in table 5, that consists in do a similar analysis regressing the time series on three lagged observations.

With Residual standard error: 0.1911 on 37 degrees of freedom; Multiple R-squared: 0.2942; Adjusted R-squared: 0.218; F-statistic: 3.857 on 4 and 37 DF; p-value: 0.01021.

For this reason we conclude that only the observation lagged by one year affects

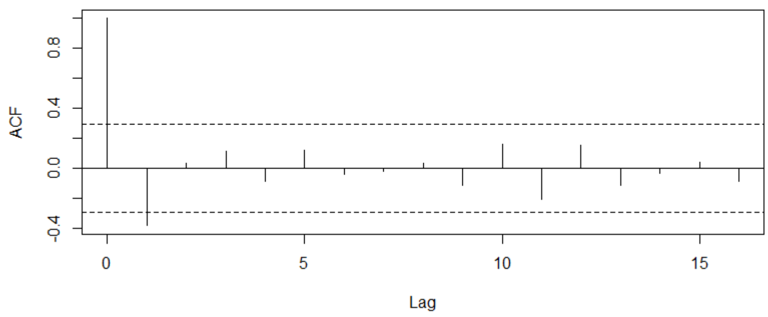

Figure 3: ACF on $\hat{e}_0$

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	0.142925	0.074584	1.916	0.06308 .
lag(1)	-0.587175	0.164422	-3.571	0.00101 **
lag(2)	-0.243488	0.182636	-1.333	0.19062
lag(3)	0.015714	0.155971	0.101	0.92030
t	-0.006197	0.002730	-2.270	0.02913 *

Table 5: AutoRegression of $\hat{e}_0$

the current one, then we have the justification for the order 1 of the auto-regressive component of the ARIMAX model. Finally, we can look at the Partial Auto Correlation of the time series in order to check if any previous shock affects the current one, so we obtain the graph 4.

This graph leads us to conclude that any possible previous shock, so only the immediately preceding shock happened in the past, not explained with the lagged value of the process, affect the current observation. So this justify the order 1 of the average mean component of the ARIMAX model. Then, once all this test are executed, the results suggest us that a good model could be the ARIMAX(1,1,1) model with trend regressor to model the life expectancy at birth process. I other words, each realization in a given year is affected by the entire information present in the previous year, so, each year contains the information set of the entire period prior to the observed one plus the information realized in that moment. For this reason we run the model and obtain the results that need to be White Noise. We test for this in the next section.

We conclude this section pointing out that it is used the statistical software *R* with IDE *RStudio* to compute the calculus. The main package used in the scripts are *tseries* for time series analysis, *lmtest* for linear regression models tests, *forecast* to compute the forecasts and *portes* for portmanteau tests.

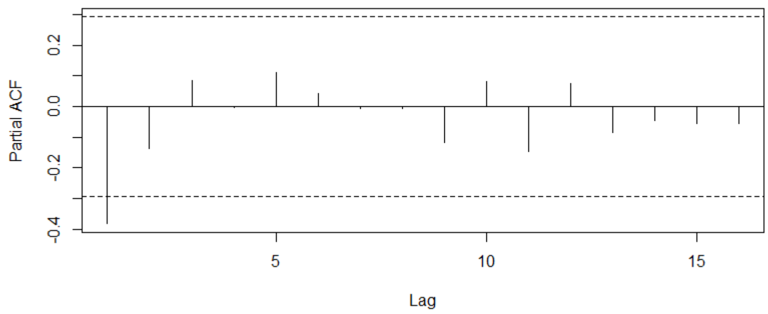

Figure 4: PACF on $\hat{e}_0$

4 Residuals study

In this section we need to test if the residuals of the ARIMAX model are white noise. Let's call them η_t. In the figure 5 it is exhibited the residuals η_t of the model.

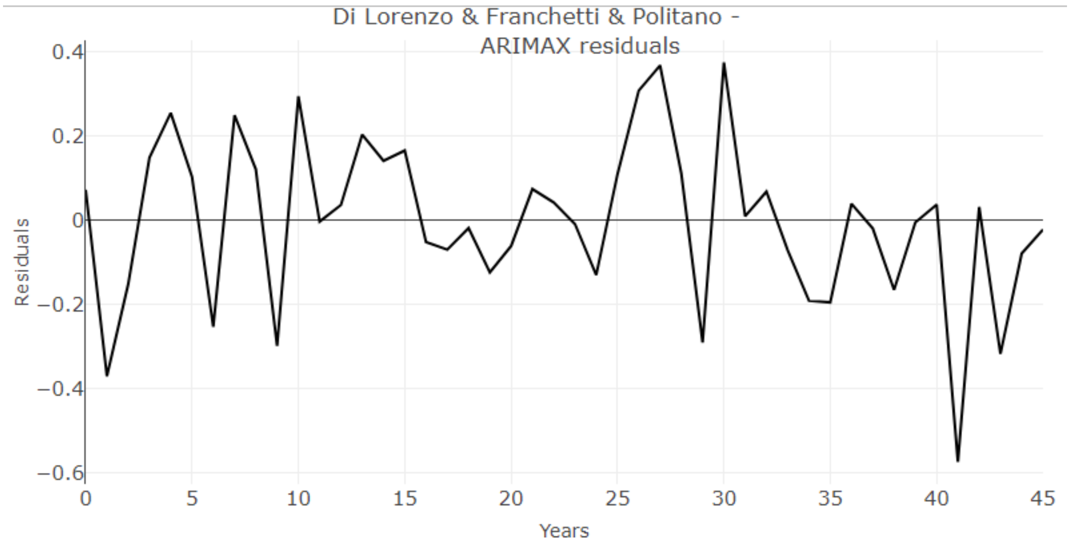

Figure 5: η_t time series

Then we start to check for stationarity of eta through the Augmented Dickey-Fuller Test and the KPSS one. Results in tables 6 and 7.

By the test we can conclude that the residuals are stationary, a very important property of white noise. So we now test for the dependence of the residuals to the previous lags following the Box-Pierce test and the Ljung-Box one.¡ in the family of Portmanteau Test. In table 8 the results for both tests and for 5, 10, 15, 20, 25 snd 30 lag.

By this table it is possible to conclude that observations are serially independent, so we can conclude that for each year the distributions of values are independent, another

Augmented Dickey-Fuller Test
Dickey-Fuller = -5.2192, Lag order = 1, p-value = 0.01 alternative hypothesis: stationary

Table 6: ADF on η_t

KPSS Test for Trend Stationarity
KPSS Trend = 0.12787, Truncation lag parameter = 9, p-value = 0.08357

Table 7: KPSS on eta_t

white noise feature confirmed.

Then we can go on test for homoschedasticity of eta_t. We use the White version of the Breush-Pagan test, the Studentized one, to check for this. In table 9 we can see the results.

By the table, we can conclude that η_t is homoschedastic. So another feature of the white noise is verified. We pass now to the testing is the mean is effectively equal to zero. So we proceed with the traditional Student's t test in order to test for the hypothesis. In table 10 we can see the result of the test.

Effectively the test give us the opportunity to accept the null hypothesis that the mean of the time series is zero, so the last feature of white noise is satisfied.

In conclusion we check for the normality of the model's residuals and then we run the traditional Shapiro-Wilk test for normality which result is visible in table 11 and figure 6.

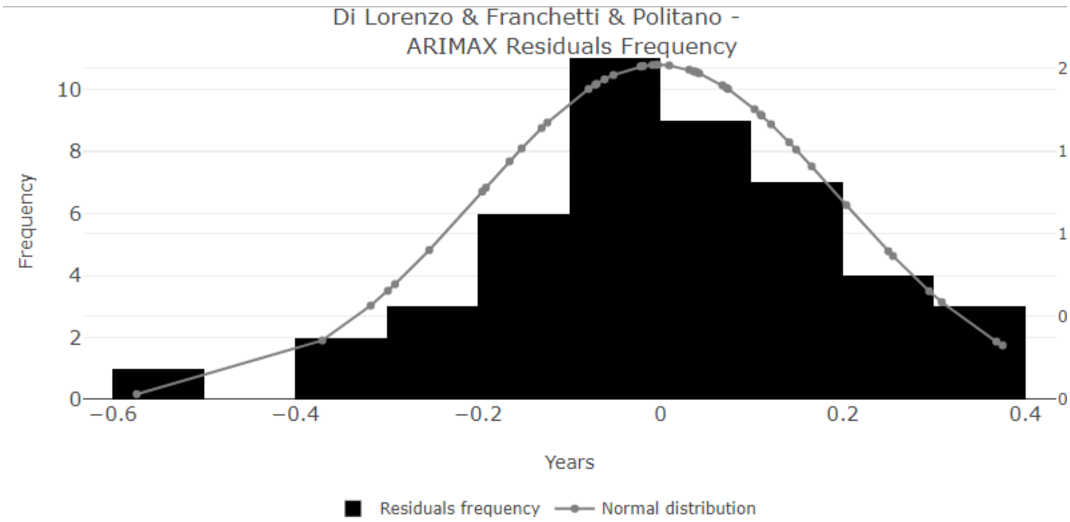

Figure 6: Probability distribution of η_t

	BP		LB	
lags	statistic	p-value	statistic	p-value
5	0.8352104	0.9490509	0.8352104	0.9490509
10	1.6768737	0.9820180	1.6768737	0.9820180
15	3.0006777	0.9900100	3.0006777	0.9900100
20	3.9560861	0.9930070	3.9560861	0.9930070
25	5.0686785	0.9970030	5.0686785	0.9970030
30	6.3861707	0.9960040	6.3861707	0.9960040

Table 8: Portmanteau test on η_t

studentized Breusch-Pagan test
BP = 0.021488, df = 1, p-value = 0.8835

Table 9: White Test on η_t

This result let us conclude that normality is a feature satisfied for the time series of the model's residuals. So we can conclude that the residuals of the ARIMAX model are white noise and then the model fits good the data. So we can go with the core of our work: to see what would happen if a shock as COVID19 one on the life expectation by the birth.

5 Impulse Response

We can now focus on our question: *how the life expectation at birth would react if a shock as COVID19 one would happen?*. Let's see first the coefficients of the ARIMAX model in table 12.

We can see that the absolute value of the auto-regressive component is less than 1. For this reason we can say that the model is stable and, then, we already know that, following the model, the process tends to come back to the mean. Now, we know that the mean of the original process is a increasing function of the time. For this reason, thanks to this model, it is possible to estimate the future lifespan expectations considering the process' mean increasing during the time. For this reason we can analyse the response of the process to the shock. We start with the forecast in 10 years of the life expectation following the coefficients of the model. So, we emulate a shock such that it happens the same thing happened in reality: the life expectancy drops by 83.163 in 2019 to 82.084 in 2020. So we now apply the negative gap between the expectation in 2020, 83.40959, and the observed value in 2020. Then, we can look how the process reacts and how much time it needs to come back to the mean. The graph in figure 7 exhibits the dynamics.

Clearly, it takes six years for the process to return to the mean. Then, the model exhibits a stability such that in case of a shock (first lag) of this severity it is necessary

One Sample t-test alternative hypothesis: true mean is not equal to 0
t = -0.073267, df = 45, p-value = 0.9419
95 percent confidence interval: -0.06076869 0.05650269
sample estimates: mean of x -0.002133001

Table 10: Test t on η_t mean

Shapiro-Wilk normality test
W = 0.9802, p-value = 0.6141

Table 11: Shapiro-Wilk test on η_t

to wait six years to return to the original pattern. In the graph it is not considered the mean as function of the trend, but just a zero mean. If we integrate in the data the linear forecast given before we can obtain a time series that exhibits the complete dynamics of the variable. The graph is shown in figure 8 in section 6.

6 Results discussion

Let's see the graph of e_0 dynamics in figure 8.

Clearly, we can see that the variable takes six years to come back to the original pattern that exhibits a positive trend. The black line shows the dynamics as the combination of the the linear positive trend and the impulse response to the shock. The grey line is the observed dynamics until 2022 and, then the dotted one is the forecast by 2020. First of all, we can see that forecast is aligned with the post-shock pattern at fourth lag and we can see that in the 2029, so after 10 years by the outbreak, estimation gives us the possibility to expect that the expectancy of life span is around 85.5 years. Furthermore, despite the shock and event without it, there is the same estimation. this is a very huge clue about the rapid absorbing capability of the studied process. Then we can conclude by the model results that if a shock as COVID19 one happens, life expectation bears a drop but, then, it will come back in around six years to the original pattern such that the effects of the shock will be fully absorbed is short time. Once this results is confirmed we can compare the black line with the grey one by 2020 and then we can see that the direction suggest a comeback to the pattern but at the same time it seems that if it is true the comeback will slower respect to what the model suggests.

7 Conclusion

Once we have developed the whole framework we can now get the conclusions of our analysis. COVID19 pandemic is a relevant shock that dropped the expectation of a human life duration at birth of one and a half year. There are different thoughts about

ar1	ma1	t
-0.2272254	-0.2269637	0.2365536

Table 12: ARIMAX model coefficients

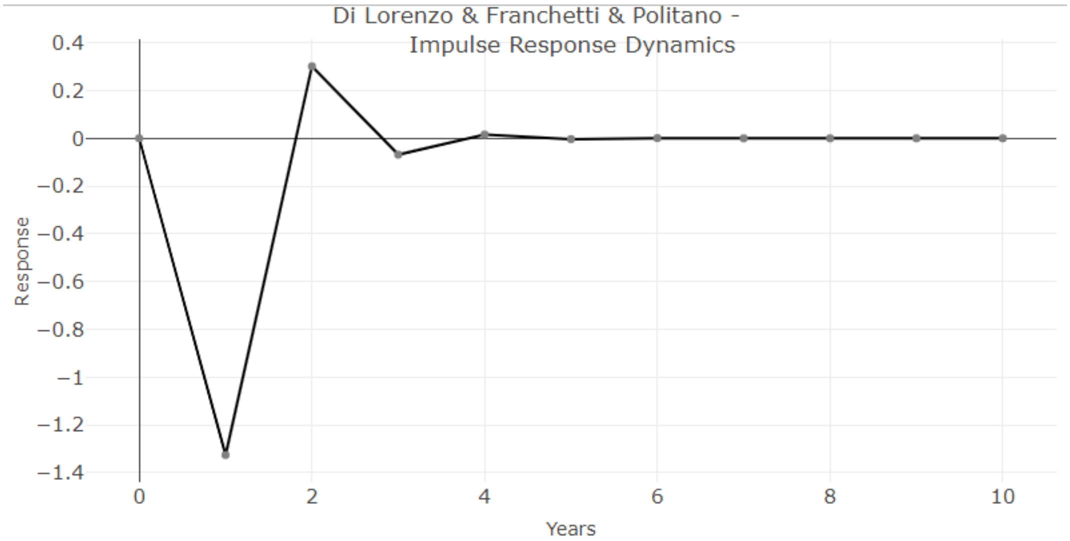

Figure 7: Impulse response of e_0 with zero mean

the effect of COVID19 on life expectation. What it is important is first of all to understand that all the analysis is based on mortality estimations that lead to the definition of mortality tables. For this reason it is necessary to be aware about the the set of all other events that cause an increase of mortality in a community. So we can just say that it has occurred a shock in 2019 leading to a drop in life expectancy in 2020 where the main cause of this shock is COVID19 outbreak. Going on of course we need to say that this is a model that in our opinion, following our tests, leads to reliable results and then the forecast that we can see in the end of the framework. So our conclusions and interpretations are based on this particular specification. Then, we can look at the behaviour of the real data until 2022 and the theoretical values. Here we can conclude only that, following the features of life expectancy's process, it is expected that in six years the effects of pandemic will be absorbed in the mean. looking at the real data the variable is growing up suggesting this absorbing path. There could be some doubts about that because in 2022 there is a very slightly increase of expectancy but not much instead of 2021 where it is stronger. So we just want to say that our analysis give the opportunity to expect the return to the original pattern in six years of something more looking how slow is the return in 2022. Furthermore we can also say that the estimation shows that there is an important gain, in ten lags, of 2.5 years more of life duration and this regardless the 2019 outbreak. Anyway we need to underline that, supposing the COVID19 shock acts like a benchmark as the intensity cap of the shock, this behaviour happens only if a shock happens with a distance of at least six years. Finally we can

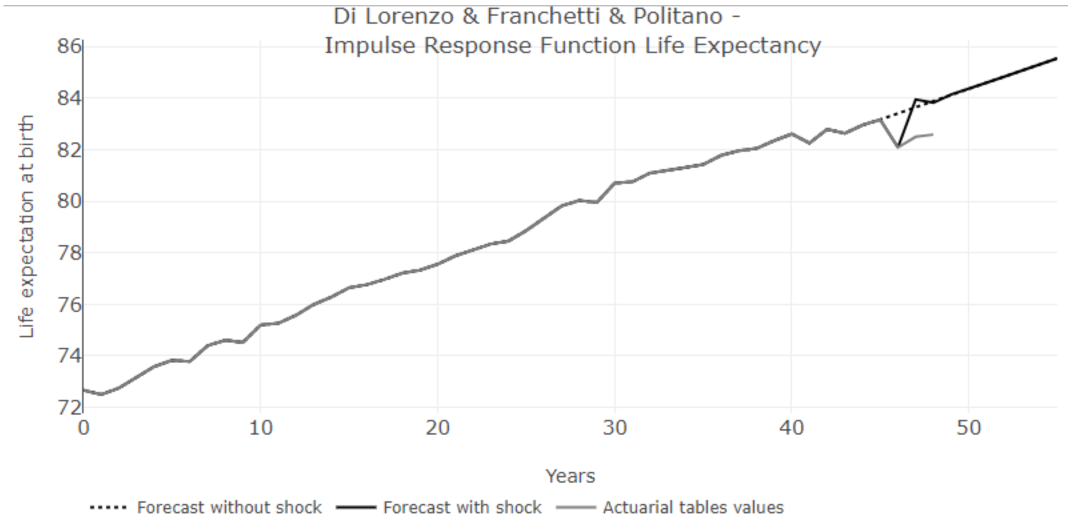

Figure 8: e_0 dynamics

also underline that the nature of this process is peculiarly shock-resistant because of the strong tendency to increase during the time thanks to the faster and faster development in technology and medics Shang (2016) but also thanks to the timely government intervention to contain the pandemics and also to recover the situation prior to the outbreak. Indeed, this process shows that if in a given future day another shock would happen its effect remain for short time giving support to the so notorious phenomena of "rectangularization" of the survival function. Anyway we have to wait for future data on mortality tables to clarify if the real behaviour is more o less consistent with the expectations, for now we can just have some suggests based on the information available until 2022 of a sufficiently quick absorbing of the shock.

8 Declarations

Conflict of Interest Statement: The authors declare no conflicts of interest related to this research.

Research Grant Disclosure: This research received no external research grant funding.

References

Alharbi, R. and Alnagar, D. K. (2022). Forecasting life expectancy at birth using time series and holt's method in saudi arabia according to saudi vision 2030. *Advances and Applications in Statistics*, 79:121–130.

Alradaideh, A. A. M. et al. (2022). Happiness among uae undergraduates one year since covid 19. *Electronic Journal of Applied Statistical Analysis*, 15(2):277–289.

Alshehri, A., El-Halawany, F., and Abu-Zinadah, H. (2021). Prediction of life expectancy in saudi arabia by 2030 using arima models. *Journal of Physics: Conference Series*, 1978(1):012060.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control (5th ed.).* John Wiley & Sons.

Cao, X., Hou, Y., Zhang, X., Xu, C., Jia, P., Sun, X., and Wang, Y. (2020). A comparative, correlate analysis and projection of global and regional life expectancy, healthy life expectancy, and their gap: 1995-2025. *Journal of global health*, 10(2).

Carannante, M., D'Amato, V., and Haberman, S. (2023). Effect of covid-19 frailty heterogeneity on the future evolution of mortality by stratified weighting. *Journal of Demographic Economics*, 89(3):513–532.

Carter, L. R. (1996). Forecasting us mortality: A comparison of box-jenkins arima and structural time series models. *The Sociological Quarterly*, 37(1):127–144.

Di Fonzo, T. and Lisi, F. (2005). *Serie storiche economiche.* Roma: Carocci.

Husin, W. Z. W. and Aziz, N. A. (2021). Box-jenkins model for forecasting malaysia life expectancy. *International Journal of Academic Research in Business and Social Sciences.*

Hussein, E. M. and Mohamed, E. E. (2015). Trend of life expectancy in sudan. *Int. J. Sci. Res.*, 4(2):2375–9.

Islam, N., Jdanov, D. A., Shkolnikov, V. M., Khunti, K., Kawachi, I., White, M., and Leon, D. A. (2021). Effects of covid-19 pandemic on life expectancy and premature mortality in 2020: time series analysis in 37 countries. *BMJ*, 375.

Keser, I. K. and Deveci Kocakoc, İ. (2021). Functional cluster and canonical correlation analysis of eu countries by number of daily deaths and stringency index during covid-19 pandemic. *Electronic Journal of Applied Statistical Analysis*, 14(1):197–216.

Levine, R. S., Foster, J. E., Fullilove, R. E., Fullilove, M. T., Briggs, N. C., Hull, P. C., and Hennekens, C. H. (2016). Black-white inequalities in mortality and life expectancy, 1933–1999: implications for healthy people 2010. *Public health reports.*

Lucchetti, R. (2015). Appunti di analisi delle serie storiche.

Malia, R. (2021). Forecasting indonesia's life expectancy during the covid-19 period 2021-2026. *Int J Econ Manag Stud*, 8(2):105–109.

Noymer, A., Garenne, M., and White, K. M. (2022). Life expectancy losses and bounce-backs during the covid-19 pandemic. *Nature Human Behaviour*, 6:1613–1614.

Ntamjokouen, A., Haberman, S., and Consigli, G. (2014). Modeling multi-population life expectancy: a cointegration approach. *Insurance markets and companies: analyses and actuarial computations*, 5(2):12–23.

Pitacco, E. (1989). *Elementi di matematica attuariale delle assicurazioni sulla vita.* Trieste: LINT.

Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics.*

Singh, A. and Hasija, S. (2021). Modeling and forecasting life expectancy in india: A systematic approach. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 7(2):200–214.

Torri, T. and Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28(2):519–531.

Šimpach, O., Dotlačilová, P., and Langhamrová, J. (2013). Logistic and arima models in the estimation of life expectancy in the czech republic. In *Proceedings of the 31st International Conference of Mathematical Methods in Economics*, pages 915–920.