

# Modeling of the Pandemic risk in Lebanon

Zaher Khraibani

**Abstract**— The main objective of this article is to study the impact of the influenza pandemic on the Lebanese population, which provides a new methodology and a technique of constructing a model for influenza pandemic. Another objective is to study also the impact of influenza pandemic on the portfolio of the company insurance by determining the mortality experience from such event in order to evaluate the pandemic losses of the portfolio. The results presented in this report are based on both quantitative and qualitative analysis to the extent permitted by the available data. Moreover, the results performed in the analysis were affected by major levels of uncertainty. This is obvious due to its relation to the future mortality and various incidence experiences taking place, as well as the future composition of the portfolio by using a different statistical method.

Flu pandemics have occurred throughout history. There have been four different types since 1918, each with different characteristics [1]. In 1918-1919 illness of the Spanish Flu came on quickly.

**Index Terms**—Extreme Risk, Pandemic, Lebanon, Times Series, Mortality Excess, Life Insurance.

## I. INTRODUCTION

In February 1957, an identification of a new flu virus took place in the Far East. In early 1968, a detection of a new flu virus occurred in Hong Kong. This pandemic witnessed high mortality rates for ages over 65 making it the mildest flu pandemic in the 20th century. The same virus re-occurred in 1970 and 1972; it was similar to 1957 flu. In spring 2009, a new flu virus H1N1 (swine flu) spread quickly across the United States and the world. The first three cases of swine flu discovered took place in Lebanon on July 1, 2009 according to Lebanon Health Ministry. Moreover, young people in 74 countries were highly affected by this pandemic. The epidemiology is by definition, the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems. Second, endemic is a disease that occurs at an expected constant level in a certain population. Regarding epidemics, it is an endemic that exceeds the normal expectations. Also, the pandemic is an epidemic occurring worldwide, crossing international boundaries and usually affecting far a higher number of people. For these reasons our objective is to study the pandemic risk especially the influenza on the Lebanese population and their impact on the company insurance. So, the fatal infectious diseases and the various devastating implications they may have, such diseases could lead to a major epidemic and eventually trigger the level for the notes to be highly exceeded. Besides, taking into consideration the

Influenza epidemic of 1918 that resulted in a dramatic increase in deaths in a given year due to that specific influenza epidemic [3].

Modeling a severe epidemic must be done by following a certain approach, and based on that chosen approach we can assume through that the number of deaths resulting from an infectious disease outbreak. Thus, due to the hazardous threats presented by the influenza virus throughout history, that guided the importance of developing methods and assumptions for the country pandemic model.

The modeling process starts with a random variable that determines if an epidemic has occurred in any given year. If an epidemic has occurred, another random variable is sampled so as to determine the degree of severity of the epidemic in terms of percentages of excess mortality. The assumptions and methodology used in the Pandemic Model are discussed in further detail below.

## II. MODELING FREQUENCY

The Pandemic Model assumes a probability that any given year will contain an epidemic is 7.62%. This is based on an assumed frequency of 32 influenza epidemics in a 420 years' period.

When taking the actual pandemic frequency, there are no assurances that this frequency won't be higher or lower than this figure. Besides, when considering the dependence between the frequency assumption and our model, here the frequency assumption should be carefully considered when evaluating how the model contributes to the various modeled scenarios under which both a trigger event and a loss to the notes might take place.

The Country Pandemic Model will model frequency by taking 10,000 random number between 0 and 1, then says if this random is less than 7.62%, we should consider a pandemic year (1) else normal year (0) [2].

if (rand() < 7.62%, 1, 0)

Various studies of influenza experience have shown that influenza epidemics have all been worldwide events in the past century. United States (US) has the most comprehensive set of data for analyzing influenza experiences contrary to Lebanon where the data was sparse and did not contain sufficient granularity for modeling purposes. Thus, the Country Pandemic Model was constructed using the US data, and by assuming that both countries have the same percentage excess mortality, this model was applied on Lebanon to estimate the number of deaths resulting from a pandemic.

## III. DATA POINTS

We collect the detailed record of severity of mortality from the available influenza epidemics 1918-1920 (H1N1 Spanish Flu), 1957-1958 (H2N2 and H3N2 Asian Flu), 1968-1969

(H3N2 Hong Kong Flu), 1977-1978 (H1N1 Russian Flu), and 2009 (H1N1 Flu). In addition to these points, the 2003 SARS epidemic in Hong Kong was included as another data point. SARS considered as an epidemic first originated in China in 2002. SARS (severe acute respiratory syndrome) is a viral respiratory illness [4].

It's important to note that our calculations carry a greater number of significant digits than those shown in some of the discussion below, so some rates shown will differ slightly from what the reader may calculate using rounded values.

For all the considered Influenza epidemics, an excess mortality rate per 1000 lives was calculated and represented as 2015 term. This was done by multiplying the excess mortality rate (EMR) of each pandemic year by the 2015 total mortality rate (2015TMR) over pandemic year total mortality rate (TMR) ratio.

**Table 1:** 1957 Adjusted Influenza US Excess Mortality Rate per 1000 Lives.

|  |         |       |
|--|---------|-------|
| 1957 Actual US TMR (per mil)           | A=      | 9.59  |
| 1957 Actual Influenza US EMR (per mil) | B=      | 0.384 |
| 2015 Actual US TMR (per mil)           | C=      | 7.319 |
| 2015 Actual TMR/1957 Actual TMR Ratio  | D= C/A= | 76.4% |
| 1957 Adjusted Influenza EMR (per mil)  | F= B*D= | 0.293 |

Table 1 presents this calculation for the 1957 pandemic knowing that the same calculation is applied to the rest data points.

**Table 2.** 1918 Adjusted Influenza US Excess Mortality Rate per 1000 Lives.

|  |                    |   |
|--|--------------------|---|
| 1918 Actual US TMR (per mil)           | A=                 | 18.10   |
| 1918 Actual Influenza US EMR (per mil) | B=                 | 5.290   |
| 2015 Actual US TMR (per mil)           | C=                 | 7.319   |
| 2015 Actual TMR/1918 Actual TMR Ratio  | D= C/A =           | 40.4%   |
| 1918 Adjusted Influenza EMR (per mil)  | E= B*D =           | 2.139<br><i>placed at the 3.13 percentile level</i> |
| Modern Health and Medicine Adjustments | F=                 | 60%   |
| 1918* Adjusted Influenza EMR (per mil) | G= F*E + (1-F)*B = | 3.40<br><i>Placed at the 0.10 percentile level</i>  |

Table 2 considers the 1918-20 Influenza epidemics; an excess mortality rate of 5.29 per 1000 lives indicates an increase of 29.23% over 18.10 per 1000, which is the total mortality rate of 1918 as recorded by the Centers for Disease Control and Prevention (CDC). Like the other data points, the 1918 pandemic will be subjected to the above adjustment. In

addition, another adjustment on 1918 was to assume that there is 60% broad improvement in mortality between 1918 and 2015 that results in 3.40 extra deaths per 1000 lives [12].

IV. MODELING SEVERITY

In developing a model for the severity of an infectious disease epidemic, a pandemic severity curve was plotted by fitting excess mortality rate points against the probability of achieving the excess mortality rates given a disease event has occurred.

As a first step in fitting the severity curve, the 1918\* Adjusted influenza epidemic was assumed to be the worst possible epidemic that may occur. So, we decide to place this data point at the 0.10 percentile level.

To Allow the model simulate events that are worse than the mortality implied by the adjusted 1918\*, a separate curve was attached at the 0.5 percentile level. This curve has been chosen so that there is no maximum limit of excess mortality that can be generated by the Pandemic Model. The resulting pandemic severity curve is therefore a construction of two curves corresponding to two components:

- Main Component represents more probable outcomes of excess mortality associated with more probable epidemic events. It models severity for percentiles higher than 0.5 percentile level.

- Extreme Component represents extremely severe levels of excess mortality significantly beyond historically experienced worst cases such as the 1918 epidemic event. It models severity for percentiles lower than 0.5 percentile level.

A. Fitting the Main Component

It is assumed that the 1918 Influenza epidemic is the worst out of the 32 Influenza epidemics that are assumed to have occurred over the last 420 years. Hence, this event is plotted at the 3.13 percentile (= 1/32) level.

The 2009 H1N1 has the least increase in excess mortality for which reliable data is available. Thus, this event is plotted at the 100 percentile level indicating that if an epidemic took place its severity will be at least like that of 2009.

1957, 1977, 1968 and 2003 SARS epidemics were equally spaced between the 1918 epidemic and 2009 epidemic. The table below shows the severity distribution.

**Table 3:** Severity Distribution.

| Pandemic | Percentile | 2015 Adjusted EMR (Per Mil) |
|----------|------------|-----------------------------|
| 1918*    | 0.10%      | 3.40                        |
| 1918     | 3.13%      | 2.139                       |
| 1957     | 22.50%     | 0.293                       |
| 1968     | 41.88%     | 0.106                       |
| 1977     | 61.25%     | 0.09                        |
| 2003     | 80.63%     | 0.073                       |
| 2009     | 100%       | 0.011                       |

Table 3 presents the adjusted 2015 EMR severity distribution. To fit the severity distribution using “Microsoft

Excel”, we will generate 10,000 scenarios having the severity distribution in table 3. Using “Easy Fit 5.4 Professional” software we will fit these 10,000 scenarios and look for the distribution that best fits. The Kolmogorov Smirnov goodness of fit test indicates that the “Exponential 2P” distribution is the best fit for the actual distribution.

Using the software Easyfit we will generate 10,000 “Exponential 2P” random numbers. Then we sort them from largest to smallest to take the bottom 9,950 scenario. These 9,950 scenarios will be divided by the 2015TMR to reflect the excess mortality percentage (EMP) and represent the percentiles higher than 0.5 percentile level.

The Exponential 2P probability density function provided by EasyFit is [5]:

$$f(x) = \lambda e^{-\lambda(x-\gamma)} \quad (1)$$

Where;

$\lambda$ = Continuous inverse scale parameter ( $\lambda > 0$ ).

$\gamma$ = Continuous location parameter

Domain:  $\gamma \leq x < +\infty$ .

### B. Fitting the Extreme Component

The extreme component (at percentile levels below 0.5%), utilizes a tangent function that approaches the severity axis asymptotically. The formula for the tangent curve is as follows [14].

$$ExcessMortalityPercentage = TAN(90 - Percentile * d) * \pi / 180 \quad (2)$$

Where;

percentile = The percentile level

$d = 9976.8$

The  $d$  factor ensures that the tangent function and the exponential function produce the same percentage increase in mortality at the 0.5 percentile level.

So, to calculate the  $d$  factor we will put the two functions equal at the 0.5 percentile level and we goal seek finding  $d$ .

Now, we take 10,000 random numbers between 0 and 1 to generate 10,000 scenarios having the Tangent distribution. We sort them from largest to smallest and we take the top 50 to represent the percentile levels less than 0.5%.

### C. Severity Distribution

And known as the country excess mortality percentage distribution. It is determined by merging the top 50 Tangent scenarios with the bottom 9950 Exponential scenarios to produce a combined Tangent-Exponential distribution that represents the excess mortality percentage distribution of the country (EMP). Thus, for a pandemic year we will be sampling a certain EMP from the combined distribution. Doing this, will give the rise of a convoluted Frequency-Severity distribution where we will apply the percentile function and determine the number of deaths from pandemic in Lebanon for different percentile levels.

## V. FREQUENCY-SEVERITY CONVOLUTION

As we already mentioned, modeling frequency was by generating 10,000 random number between 0 and 1 then relying on the assumption that there is a probability 7.62% of

a pandemic in a certain year we modeled the frequency by giving 1 for the pandemic year (if the random number is less than 0.0762) and 0 for a normal year (if the random number is greater than 0.0762). For Severity, we generate another 10,000 random numbers and using the index function we sample a certain ED% from the combined Exponential-Tangent distribution when the frequency models gives 1 (pandemic year).

$$IF(Freq=1, NDEX(SeverityDist., ROUND((SeverityRand.) * 10000, 0), 1), 0) \quad (3)$$

Convolution is obtained when giving each year a severity if the frequency model indicates a pandemic year.

We can notice, till now the Country Pandemic Model is dealing with a population that contains all ages [13]. Later on, our study will go deep and study the impact of pandemics in 10-years age groups.

The main target of this section was to estimate the number of deaths that may result from a pandemic on the Lebanese population. Knowing that the Lebanese population is 4,330,735 for 2015 we get the following result shown in the table below.

$$PandemicDeaths = ExcessMortalityRate * LebanesePopulation(4)$$

### Influenza Pandemic Deaths in Lebanon 2015

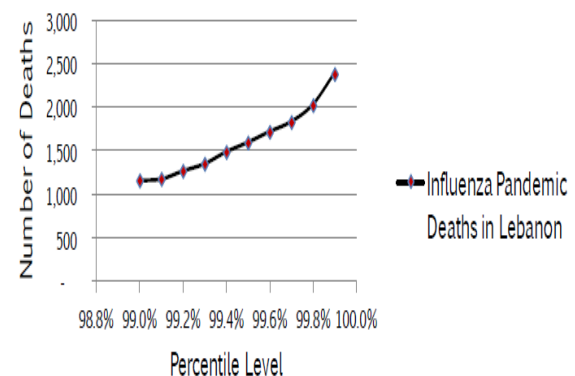


Figure 1: Number of deaths from Influenza Pandemic in Lebanon 2015.

Figure 1 shows the number of pandemic deaths in Lebanon as a function of the percentile levels.

## VI. TIME SERIES

It was actually worth it to perform such a projection for the pandemic deaths in Lebanon for 5 years ahead. Thus, relying on the available data for Lebanon population and total mortality rates from 2000 to 2015, we will be forecasting the population and total mortality rates for years 2016 to 2020.

### A. Autoregressive (AR) Models

A real valued stochastic process  $(Y_t)$  is said to be an autoregressive process of order  $p$ , denoted by  $AR(p)$  if there exist  $a_1, \dots, a_p \in \mathbb{R}$  with  $a_p \neq 0$ , and a white noise  $(\epsilon_t)$  such that

$$Y_t = a_1 Y_{t-1} + \dots + a_p Y_{t-p} + \epsilon_t \quad (4)$$

Where,

- $t \in \mathbb{Z}$
- $\epsilon_t \sim N(0, \sigma^2)$

□  $p$  is a non-negative integer

The value of a  $AR(p)$  process at time  $t$  is, therefore, regressed on its own past  $p$  values plus a random shock [6].

The identification of  $AR(p)$  models from the autocorrelation function (ACF) and the partial autocorrelation function (PACF) is as follows:

- ACF: declines in geometric progression from its highest value at lag  $p$
- PACF: cuts off abruptly after lag  $p$  [7].

**B. Moving Average (MA) Models.**

The moving average process of order  $q$  is denoted  $MA(q)$  and defined by [8]

$$X_t = \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (5)$$

□  $\theta_1, \dots, \theta_q$  are fixed constants

□  $\theta_0 = 1$

□  $\epsilon_t$  sequence of independent random variables of mean 0 and variance  $\sigma^2$ .

The identification of  $MA(q)$  models from the autocorrelation function (ACF) and the partial autocorrelation function (PACF) is as follows:

- ACF: cuts off abruptly after lag  $q$
- PACF: declines in geometric progression from its highest value at lag  $q$  [7].

**C. Autoregressive Moving Average (ARMA) Models.**

Moving averages  $M(q)$  and autoregressive  $AR(p)$  processes are special cases of so called autoregressive moving averages. Let  $(\epsilon_t), t \in \mathbb{Z}$  be a white noise,  $p, q \geq 0$  integers and  $a_0, \dots, a_p, b_0, \dots, b_q \in \mathbb{R}$ . A real valued stochastic process  $(Y_t), t \in \mathbb{Z}$  is said to be an autoregressive moving average process of order  $p, q$ , denoted  $ARMA(p, q)$ , if it satisfies the equation [6].

$$Y_t = a_1 Y_{t-1} + \dots + a_p Y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (6)$$

**D. Population Forecasting Model.**

The first time series “population”, ACF and PACF indicates that the time series is an ARMA (1, 2) model since its ACF vanishes after Lag 2 and its PACF vanishes for orders greater than 1.

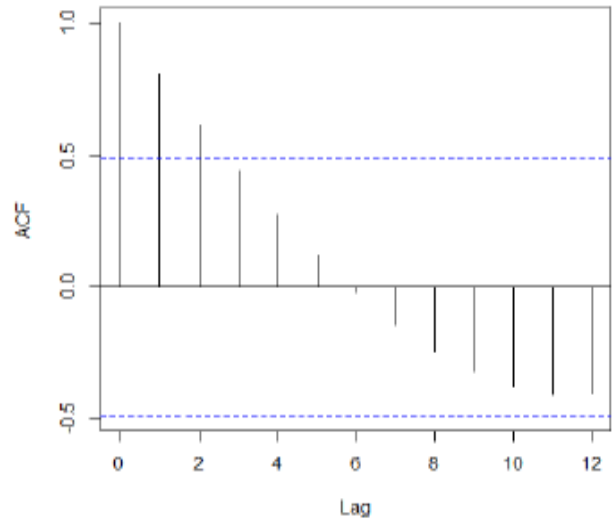


Figure 2: Population Autocorrelation Function (ACF).

ACF vanishes for orders greater than Lag 2, thus indicating it is a moving average of order 2.

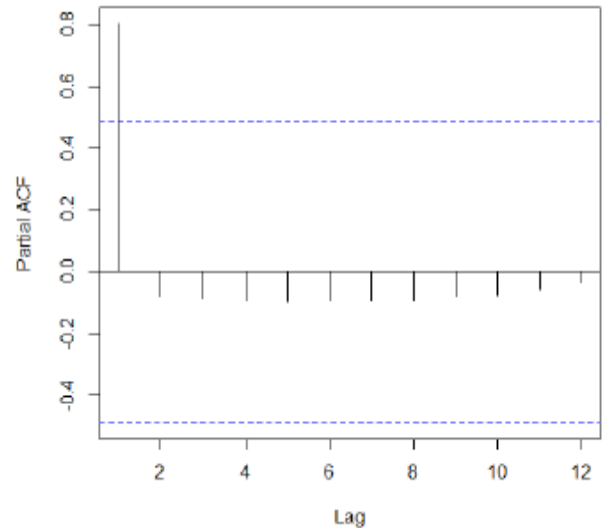


Figure 3: Population Partial Autocorrelation Function (PACF).

PACF vanishes for orders greater than Lag 1, thus indicating that it is an autoregressive model of order 1.

There is a possibility that the population is an ARMA (1, 1). To make sure that our choice is an ARMA (1, 2) we check the Akaike Information Criterion (AIC) and we choose the model with smallest AIC. The ARMA (1, 2) AIC is 386.47 less than 390.75 the ARMA (1, 1) AIC.

Hence, population is an ARMA (1, 2).

**D.1. Forecast**

The main point is to predict the population in the years 2016 till 2020. So, for ARMA (1, 2) model we will get the following results.

**Table 4:** Lebanese Population Projection for 5 years ahead.

| Year | Population |
|------|------------|
| 2016 | 4,392,209  |
| 2017 | 4,410,640  |
| 2018 | 4,410,083  |
| 2019 | 4,409,527  |
| 2020 | 4,408,971  |

Table 4 presents the 5 years' projection of the Lebanese population for years 2016, 2017, 2018, 2019, and 2020.

*E. Total Mortality Rate Forecasting Model*

The second time series "Total Mortality Rate", ACF and PACF indicates that there is a possibility of an ARMA (1, 1) or AR (1).

ACF vanishes for orders greater than Lag 1, thus indicating it is a moving average of order 1.

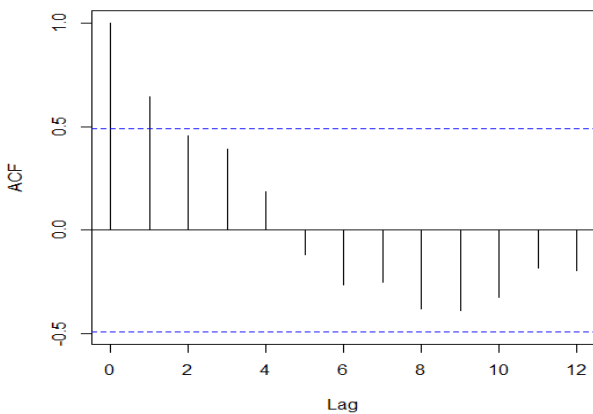


Figure 4: Total Mortality Rates Autocorrelation Function (ACF.)

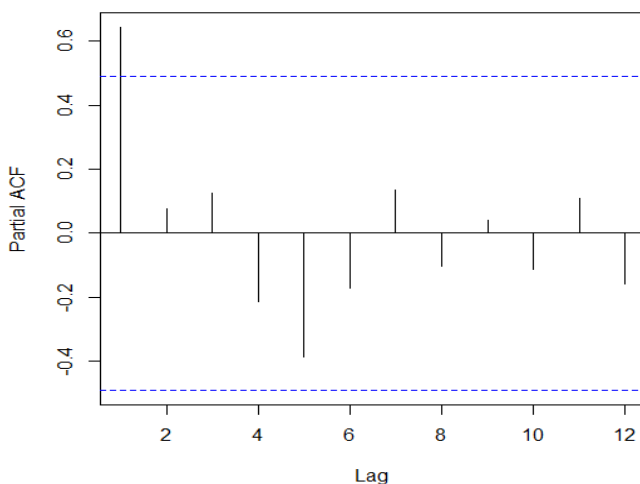


Figure 4: Total Mortality Rates Partial Autocorrelation Function (PACF)

PACF vanishes for orders greater than Lag 1, thus indicating that it's an autoregressive model of order 1.

To decide whether the Total Mortality Model is ARMA (1, 1) or AR (1) we will rely on the aic of each model. The aic of the AR (1) model is 19.54 less than 21.32 the aic of the ARMA (1, 1).

Thus, Total Mortality Rate is an autoregressive model of order 1 AR (1).

*E.1.Forecast.*

The main point is to predict the Total Mortality Rates in the years 2016 till 2020. So, for AR (1) model we will get the following results.

**Table 5:** Lebanese Total Mortality Rates Projection for 5 years ahead.

| Year | Total Mortality Rates |
|------|-----------------------|
| 2016 | 5.77788               |
| 2017 | 5.71029               |
| 2018 | 5.66695               |
| 2019 | 5.63916               |
| 2020 | 5.62134               |

Table 5 presents the 5 years' projection of the Lebanese total mortality rate for years 2016, 2017, 2018, 2019, and 2020.

VII. COMPANY PANDEMIC MODEL

In this section, we'll be considering the Lebanese company Life Portfolio used in mortality analysis Term Life and Saving Plan. Loss Model will estimate the Life Portfolio loss in case of influenza pandemic for year 2015.

We'll be using the following variables data to construct our study.

□ Policy Number(It's used to identify the policies we are working on), Status (This variable gives the situation of the policy in the study year. For each policy there is a status that identifies the policy situation, it can be: in force (A), canceled (C), surrendered (S), lapsed (A), born dead (BD), or expired (X), Date of Birth(One of the important variables in this article). Used to calculate the age of the insured person as of 2015), Sum at Risk (is the main variable in this study. For term life policies, the sum at risk is the sum insured of these policies. Whereas Saving Plan policies make things both complex and interesting at the same time, as the sum at risk for each policy differs from one year to another. Sum at Risk calculation depends on the cash value of each policy and reflects the amount of money that the company loses in case of the insured death. The portfolio consists of 31,420 term and saving policies of different statuses. In our study we will consider the policies with non-zero sum at risk. To put that into context, we will be focused only on the in force policies. Thus, the Loss Model will be estimating the expected loss from 13,187 policies in case of an influenza pandemic. For the 13,187 policy we find the age of the policyholders per policy as of 2015. Then we count the existing number of insured population in each age group of the 10-year age groups mentioned previously. The result is shown in the table below.

Table 6. Age group Portfolio Population

| AGE GROUP | POPULATION |
|-----------|------------|
| 1-        | 0          |
| [1-4]     | 0          |
| [5-14]    | 0          |
| [15-24]   | 495        |
| [25-34]   | 4,143      |
| [35-44]   | 5,902      |
| [45-54]   | 2,217      |
| [55-64]   | 414        |
| [65-74]   | 16         |
| [75-84]   | 0          |
| 85+       | 0          |

A. Sum at Risk Distribution

It is important to note that the currency is in US dollar. The sum at risk amounts distribution give a high weight on values less than 20,000 USD compared to values that are greater than 120,000 USD.

The table below shows the weights for different sum at risk ranges.

Table 7. Descriptive Analysis.

| Sum at Risk             |               |
|-------------------------|---------------|
| Mean                    | 31,993        |
| Standard Error          | 502           |
| Median                  | 19,103        |
| Mode                    | 14,161        |
| Standard Deviation      | 57,619        |
| Sample Variance         | 3,319,994,396 |
| Kurtosis                | 1,014         |
| Skewness                | 23            |
| Range                   | 3,186,181     |
| Minimum                 | 44            |
| Maximum                 | 3,186,225     |
| Sum                     | 421,893,117   |
| Count                   | 13,187        |
| Largest(1)              | 3,186,225     |
| Smallest(1)             | 44            |
| Confidence Level(95.0%) | 984           |

To construct the Loss Model of the company, we consider the Excess Mortality Rate distribution of the Lebanese country. It is important to recall that the country Excess Mortality Rate distribution was a combination between the Exponential 2P distribution and the Tangent function.

We generate 10,000 random numbers between 0 and 1, and for each random scenario we set 1, if the random is less than

0.0762 indicating a pandemic year, and 0 for a normal year. If the scenario gives a normal year than there is no pandemic loss and the model will generate 100 scenarios of 0 losses. Whereas, for a scenario indicating a pandemic year the model will generate 100 scenarios of different losses. As a result, our model will be generating 1,000,000 scenarios to reflect the pandemic losses.

The titles bellow gives the methodology of constructing the Loss Model by setting the required assumptions and the work path.

B. Loss Distribution.

Loss Model assumes that the portfolio loss distribution follows a lognormal distribution of mean  $\mu$  and standard deviation  $\sigma$ . Let  $L$  be the random variable representing the portfolio loss. Then,  $L \sim \ln(\mu, \sigma)$ .

Hence, the expected value of  $L$  is

$$E[L] = e^{\mu + \frac{\sigma^2}{2}}$$

And, the variance of  $L$  is

$$\sigma^2 = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Our assumption will let us generate for each pandemic year of the 1000 scenarios, 100 scenarios of different losses. Thus for each pandemic year we need to know its  $\mu$  and  $\sigma$  in order to generate 100 pandemic loss following  $\ln(\mu, \sigma)$ . Moment Matching

To determine the parameters of the lognormal distribution  $\mu$  and  $\sigma$ , we will set the expected loss of the portfolio equal to the  $E[L]$  of the lognormal distribution, and the variance of the portfolio equal to the  $\sigma^2$  of the lognormal distribution to get the following system of 2 equations and 2 unknowns.

$$\sum_{k=1}^N EMR_k * SAR_k = e^{\mu + \frac{\sigma^2}{2}}$$

$$\sum_{k=1}^N EMR_k * \left( SAR_k - \sum_{i=1}^N EMR_i * SAR_i \right)^2 = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Where,

$N$ : The number of in force policies

$EMR_k$ : Company Excess mortality rate of the  $k$ th policy

$SAR_k$ : Sum at risk of the  $k$ th policy

VIII. EXTREME VALUE THEOREM

EVT is a powerful and yet fairly robust framework for studying the tail behavior of a distribution. There are two families of extreme value distributions corresponding to two approaches to sampling extreme events [9].

A. Block Maxima Approach

The Block Maxima (BM) approach uses the generalized extreme value (GEV) distribution. The BM approach assumes breaking up a sequence into blocks of size  $n$  (with  $n$  large enough), and extracting only the maximum observations  $M_i, i \in \{1, \dots, n\}$  from each block regardless of whether the second largest event in a block exceeds the largest events of other blocks.

The generalized extreme value (GEV) distribution is given by [10]:

Parameters

$K$ – Shape parameter

$\sigma$ – Scale parameter ( $\sigma > 0$ )

$\mu$ – Location parameter

And we define the domain by  $1 + k \frac{x-\mu}{\sigma} > 0$  for  $k \neq 0$  and  $-\infty < x < +\infty$  for  $k=0$

With a probability density function,

for  $k \neq 0$

$$f(x) = \frac{1}{\sigma} \exp\left(-\left(1 + kz\right)^{\frac{1}{k}}\right) \left(1 + kz\right)^{-\frac{1}{k}}$$

For  $k=0$

$$f(x) = \frac{1}{\sigma} \exp(-z - \exp(-z))$$

Where,  $Z = \frac{x-\mu}{\sigma}$

### B. Peak over Threshold Approach

The Peak over Threshold approach uses the generalized Pareto distribution (GPD). In contrast to BM model, POT model uses more natural way of determining whether an observation is extreme. All values greater than the given threshold are considered as extremes.

The Generalized Pareto distribution (GPD) is given by [11]:

Parameters

$K$ – Shape parameter

$\sigma$ – Scale parameter ( $\sigma > 0$ )

$\mu$ – Location parameter

The domain is defined by:  $\mu \leq x < +\infty$  for  $k \geq 0$  and  $\mu \leq x \leq \mu - \frac{\sigma}{k}$  for  $k < 0$ . The probability density function:

$$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{k(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}} \\ \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \end{cases}$$

In this article, we use the peak over threshold approach since previous studies showed that the extreme estimator from the POT approach has a better performance from the BM approach for the same set of data [21-23].

### C. Choice of Threshold

The sample size of the Peak over Threshold approach is determined by the selected threshold. A high threshold will likely lead to a large variance as there few observations over the threshold, while taking a lower threshold likely incorporates observations with ordinary values into extremes and then the asymptotic assumption becomes less valid.

The “Kurtosis method” (proposed by Pierre Patie) [12] was adopted to choose the threshold,

*Kurtosis Method*

Step I: Calculate the sample Kurtosis  $K_n$ , sample skewness  $\mu_n$  and sample variance  $S_n^2$ .

Step II: If  $K_n \geq 3$ , then remove the Country’s EMR maximizing  $(EMR_i - \mu_n)^2$  from the sample;

Step III: Repeat the first and second step until the Kurtosis less than 3; Step IV: Choose the largest Country’s EMR from the rest of the sample point as the threshold.

For the 760 sampled country EMR, and after the 6th iteration we get the following result:

**Table 8.** Kurtosis Method Results

| Kurtosis Method (The Last 2 steps) |                        |          |          |                      |                  |
|------------------------------------|------------------------|----------|----------|----------------------|------------------|
| EMR Country All Ages               | (EDRi-Sn) <sup>2</sup> | Step     | 7        | EMR Country All Ages | Threshold 0.6934 |
| 6.7774                             | 45.5818                | Kurtosis | 477.2812 | 0.6934               | Kurtosis 1.9231  |
| 0.6934                             | 0.4454                 | Skewness | 0.0260   | 0.6678               | Skewness 0.0019  |
| 0.6678                             | 0.4119                 |          |          | 0.5736               |                  |
| 0.5736                             | 0.2998                 |          |          | 0.5498               |                  |
| 0.5498                             | 0.2744                 |          |          | 0.5481               |                  |
| 0.5481                             | 0.2726                 |          |          | 0.5311               |                  |
| 0.5311                             | 0.2551                 |          |          | 0.5291               |                  |
| 0.5291                             | 0.2531                 |          |          | 0.5253               |                  |
| 0.5253                             | 0.2493                 |          |          | 0.5218               |                  |
| 0.5218                             | 0.2458                 |          |          | 0.5218               |                  |
| 0.5218                             | 0.2458                 |          |          | 0.5215               |                  |
| 0.5215                             | 0.2455                 |          |          | 0.5055               |                  |
| 0.5055                             | 0.2299                 |          |          | 0.5038               |                  |
| 0.5038                             | 0.2282                 |          |          | 0.5035               |                  |
| 0.5035                             | 0.2280                 |          |          | 0.4971               |                  |

This table shows that the chosen threshold is the country’s EMR 0.6934 per 1000 lives.

## IX. APPLICATION

For each pandemic scenario of the 10,000 scenarios, we will consider a severe pandemic with mortality shape “W” if the company EMR is greater than or equal to 1918 EMR. Else, the pandemic scenario will be considered as moderate as 1957 with mortality shape “U”.

We construct a VBA Macro that reads the 10,000 simulated scenarios and for each pandemic year the macro will read whether the profile of the pandemic is “U” or “W”.

If the pandemic year mortality shape is “U”, the macro considers it as moderate as the 1957 pandemic. Then, the macro moves on the 13,187 policies and match the age of each policy with the age group that it belongs to. So, if the  $k$ th policy is of age 38, the macro will read the 1957 EMR of the age group [35-44] and apply to it the calculations mentioned above. Thus, for a moderate pandemic “U”, the company EMR of age 38 for the  $k$ th policy is equal to

$$\text{Company EMR} * 1957 \text{ EMR} [35-44] \div 1957 \text{ EMR All Ages}$$

If the pandemic scenario mortality shape is “W”, the macro considers it as severe as the 1918 pandemic. Then, the macro moves on the 13,187 policies and match the age of each policy with the age group that it belongs to. So, if the EMR policy is of age 38, the macro will read the 1918 EMR of the age group [35-44] and apply to it the calculations mentioned above. Thus, for a severe pandemic “W”, the company EMR of age 38 for the EMR policy is equal to

$$\text{Company EMR} * 1918 \text{ EMR} [35-44] \div 1918 \text{ EMR All Ages}$$

For each pandemic year of the 10,000 simulated scenarios, the macro will move on the portfolio policy by policy and for each policy will read its age and calculate its EMR as we just

explained. In addition to that, and for each policy of the 13,187 policies, the macro will read its sum at risk and multiply it by the calculated EMR of this policy. Repeating that for each policy and summing up the results will give us the expected loss of the portfolio from the 1st detected pandemic year. And in the same way, we calculate the standard deviation of the portfolio from the 1st detected pandemic year. Now, after determining the portfolio expected loss and standard deviation for the 1st simulated pandemic, the macro will make a moment matching between the portfolios expected loss and standard deviation with the parameters of the lognormal distributions and solves the system mentioned in the previous section. After calculating the lognormal parameters of the 1st detected pandemic year, the macro will simulate the 1st 100 scenarios having lognormal distribution of mean  $\mu_1$  and standard deviation  $\sigma_1$  to reflect the loss of the 1st detected pandemic. The same calculations are applied on the new detected pandemics.

Let us consider the 1st detected pandemic and show how the macro will deal with it.

Given:

- $Threshold=0.6934$
- $1957\ EMR_{AllAges}=0.2930$
- $1918\ EMR_{AllAges}=2.1392$
- $CountryEMR=0.0371$

Solution:

As the Country’s EMR is below the threshold, we will consider the pandemic year as moderate as the 1957 pandemic and we give it a mortality shape “U” and the company’s EMR will be 55% from the country’s EMR.

$$CompanyEMR=55%*CountryEMR$$

$$=55%*0.0371=0.0204$$

When the macro detects this scenario, which is in fact the 1st scenario indicating a pandemic of the 10,000 simulated scenario, the macro will enter a loop that moves on the 13,187 policies, one after the other, and read for each policy its age in order to calculate its EMR. We will run the macro for a sample of 4 policies of different age groups to see how the calculations are done.

**Table 9.** Sample from the Life Portfolio

| Policy Number | Status   | Age | Age Groups | Country EMR 1957 | Sum at Risk (USD) |
|---------------|----------|-----|------------|------------------|-------------------|
| 1             | In Force | 32  | [25-34]    | 0.0486           | 174,040           |
| 2             | In Force | 38  | [35-44]    | 0.0810           | 31,706            |
| 3             | In Force | 46  | [45-54]    | 0.1701           | 9,341             |
| 4             | In Force | 56  | [55-64]    | 0.3565           | 4,500             |

The Macro calculates the Company’s EMR by age in this way:

- $CompanyEMR(age_1,32)=0.0204*\frac{0.0486}{0.2930} = 0.0033$
- $CompanyEMR(age_2,38)=0.0204*\frac{0.0810}{0.2930} = 0.0056$
- $CompanyEMR(age_3,46)=0.0204*\frac{0.1701}{0.2930}=0.0118$
- $CompanyEMR(age_4,56)=0.0204*\frac{0.3565}{0.2930}=0.0248$

Then, the Macro calculates the Loss of each policy in this way:

- $L1=CompanyEM(age_1 * SAR_1) = (0.0033/1000)*174,040\$ = 0.5743 \$$
- $L2=CompanyEM(age_2 * SAR_2) = (0.0056/1000)*31,706 \$ = 0.1775 \$$
- $L3=CompanyEM(age_3 * SAR_3) = (0.0118/1000)*9,341 \$ = 0.1102 \$$
- $L4=CompanyEM(age_4 * SAR_4)=(0.0248/1000)*4,500 \$ =0.1116 \$$

Summing up all the losses will give the portfolio expected loss.

$$\sum_{i=1}^4 0.5743 + 0.1775 + 0.1102 + 0.1116 = 0.9736 \$$$

For the portfolio standard deviation, it is calculated in this way:

$$\sum_{i=1}^4 companyEMR(age_i) * (SAR_i - 0.9736)^2 = 107116369$$

Hence, the portfolio standard deviation is equal to 10350 \$

Now the Macro solves the system in the previous section to determine the lognormal parameters.

$$e^{\mu+\frac{\sigma^2}{2}} = 0.9736$$

$$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = 107116369$$

$$\mu = \ln(0.9736) - \frac{\sigma^2}{2} = \ln(0.9736) - \frac{18,5429}{2} = -9.2928$$

$$\sigma^2 = \ln\left(1 + \frac{107116369}{0.9736^2}\right) = 18.5429$$

Therefore, the macro will generate 100 scenarios that have a lognormal distribution of parameters  $\mu = -9.2928$  and  $\sigma = 4.3061$  thus reflecting the loss from the 4 policies.

In this chapter, we have considered the Lebanese company Life Portfolio used in mortality analysis Term Life and Saving Plan. Loss Model estimate the Life Portfolio loss in case of influenza pandemic for year 2015.As a result, we estimated the losses from Influenza pandemic and we placed the required capital that this company must put aside to handle this risk.

## X. CONCLUSION

To put all what has been written together, throughout this article we have conducted our work by considering the Lebanese country population and identifying the influenza pandemic impact on this population. This was done through calculating the number of deaths for such event for 2015. Besides, our model was designed to give a 5 years ahead provision for the number of deaths that may result from such events. We did also study the impact of influenza pandemic



on Lebanese insurance companies, by taking a life portfolio from a Lebanese company. We studied the mortality analysis of this portfolio to discover how the life portfolio mortality would be in normal cases. The portfolio's required capital from normal cases is 1,640,897 USD. Then, we studied the mortality experience of the same life portfolio in case of influenza pandemic and we calculated the required capital from influenza pandemic of the considered portfolio. The portfolio's required capital from influenza pandemic is 38,132 USD. Eventually, we calculated the required capital of this portfolio, through both identifying the losses of the portfolio from normal cases and extreme influenza pandemic cases. Thus, the company should put 1,679,029 USD as a capital to cover itself from normal and extreme cases.

As we mentioned in our study, our pandemic model was constructed using data from US. And by considering that Lebanon and United States have the same excess mortality percentage; our model was applied in Lebanon. So, as this study provides the methodology and the technique of constructing a model for influenza pandemic, it will be worth constructing our model on another country.

#### REFERENCES

- [1] Federal Government, U.S Department of Health and Human Services, FLU.GOV
- [2] Society of Actuaries, Article from Reinsurance News July 2013 - Issue 76.
- [3] Stefano Lazzari and Klaus Stohr, "Avian Influenza and Influenza pandemics [editorial]", Bulletin of the World Health Organization, April 2004, 82 (4), p. 242.
- [4] WHO, Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003.
- [5] Mathwave data analysis and simulation-article exponential 2P distribution.
- [6] A First Course on Time Series Analysis-by Chair of statistics, University of Wurzburg. Version 2012.August.01-Copyright © 2012 Michael Falk.
- [7] Chatfield, C., 2004. The analysis of time series, an introduction, sixth edition: New York, Chapman & Hall/CRC.
- [8] P.J. Brockwell and R.A. Davis Book, Time Series: Theory and Methods, Springer Series in Statistics (1986).
- [9] Using Extreme Value Theory Approaches to Forecast the Probability of Outbreak of Highly Pathogenic Influenza in Zhejiang, China. Jiangpeng Chen, Xun Lei, Li Zhang, Bin Peng\*.
- [10] Mathwave data analysis and simulation-article generalized extreme value distribution.
- [11] Mathwave data analysis and simulation-article generalized Pareto distribution.
- [12] Patie P. Estimation of value at risk using extreme value theory. 2000 Mar 23 [cited 10 June 2014]. In: Talks in financial and insurance mathematics [Internet]. Lausanne: Eidgenossische Technische Hochschule Zürich 1855.
- [13] Xu L, Zhang Q. Modeling agricultural catastrophic risk. Agriculture and Agricultural Science Procedia.2010; 1: 251-257.