TIME SERIES ANALYSIS USING ARIMA MODELS: AN APPROACH TO FORECASTING HEALTH EXPENDITURES IN USA

ABSTRACT

Many OECD countries are at the heart of the political agenda regarding rising healthcare spending and its long-term sustainability. The continuous rise in health expenditure exerts pressure on government budgets, health services and personal patient finance. This has led policy makers to implement reforms in order to mitigate pressures on these costs, as well as introduce programs and forecasting models to provide a support tool capable of adapting to issues that may arise. The purpose of this study is to investigate the best model to predict total health spending in the USA, a country with the highest global spending, using the Box-Jenkins methodology. Applying annual data for total US health expenditure from 1900 to 2017, resulted in the ARIMA (2,1,0) model with static forecasting being the most appropriate to predict these costs. Model estimation was achieved by the maximum likelihood-ML method and finally, the accuracy of the forecast was assessed based on certain criteria such as the root mean square error (RMSE), mean absolute percentage error (MAPE) and Theil’s inequality coefficient.

Keywords: ARIMA Model, Health Expenditure, Box-Jenkins, Forecasting
JEL Classification: C53, E27

RIASSUNTO

Analisi di serie temporale attraverso modelli ARIMA: un approccio per la previsione della spesa sanitaria negli USA

Molti paesi OCSE hanno al centro della loro agenda politica l’aumento della spesa sanitaria e la sua sostenibilità nel lungo periodo. Il continuo aumento della spesa per la sanità pubblica influenza i bilanci, i servizi sanitari così come la spesa sanitaria personale. Questo ha indotto la politica ad adottare riforme al fine di contenere questi costi e ad introdurre programmi e modelli...
di previsione per fornire strumenti in grado di gestire i problemi che ne possono derivare. Lo scopo di questa ricerca è trovare il modello migliore per prevedere la spesa sanitaria complessiva negli Stati Uniti, il paese con i costi maggiori, utilizzando la metodologia Box-Jenkins. Applicando dati annuali relativi alla spesa sanitaria totale degli USA dal 1900 al 2017, il modello a previsione statica più appropriato è risultato essere ARIMA (2,1,0). La stima del modello è stata effettuata col metodo di massima verosimiglianza e l’accuratezza della previsione è stata valutata tramite l’errore a radice quadratica (RMSE), l’errore a percentuale media assoluta (MAPE) e il coefficiente di disuguaglianza di Theil.

1. INTRODUCTION

A health system, as a macroeconomic unit is considered to be the healthcare system with total health spending. Health spending refers to any type of expense which is incurred for preventing and treating diseases and improving population and individual health. Health expenditure includes the three major categories of medical, hospital and pharmaceutical spending. In addition, expenditure for prevention, training, research, military health services, school medicine as well as operating costs are added to the three main categories with all of the above constituting total health expenditure. Needless to say, health expenses must adapt to developments and forecast future trends.

Health spending has grown over previous decades on a global scale. The factors which are fundamental to explaining the increase in health expenditure vary according to the time period investigated. In the short-term, health expenditure development is closely linked to a country’s state budget. In the medium-term, changes in technology play an important role in the development of health expenditure. Long-term risk factors, such as obesity and changes in the prevalence of chronic diseases, remain the primary causes of rising health expenses. Additional factors, which affect health expenditure include an ageing population, income growth, as well as changes in treatment practices (Astolfi et al., 2012).

Also, other factors which burden health costs and affect healthcare expenses are demographic factors, such as population size and structure. An ageing population and higher life expectancy seem to have an impact on health service consumption, since they are related to a reduction in the health status of the population. On the other hand, many scientists believe that an ageing
population is a secondary factor which has a fairly low impact on health costs due to medical
advances, enabling those individuals to remain in the labor market for longer periods of time.
Finally, changes in lifestyle consumption and the development of new therapeutic approaches
have led to greater health expenses. Changes in social standards and personal preferences via
health promotion and disease prevention programs have an impact on the demand for health
services and hence result in greater health expenditure.

According to the World Health Organization (WHO), the US is ranked first in the world for the
highest total health expenditure in GDP and per capita spending. Health expenses in the US are
funded by the public sector, private insurers and private payment providers. US employers cover
their workers insurance while private companies offer health insurance policies and usually
cover intrahospital care and general practitioner fees.

US social policy is realized (accomplished) through two federal programs, Medicare and
Medicaid, which are characterized as social health protection covering the elderly, deprived,
disabled and vulnerable population. The US health system also consists of health care programs
funded exclusively by personal income and employee and employer contributions. Almost all
physicians are freelancers, affiliated by various insurance companies or healthcare payers which
are autonomous organizations. Permanent civil servants simply do not exist, since there are no
state hospitals or health centers. State intervention is limited and takes place at two levels, that
of the central federal government and of the state. This intervention is basically limited to laying
down the frameworks and principles for health service operations, which are mostly private
carriers of a predominantly speculative nature.

Today in the US approximately 30% of the population are uninsured or under-insured, while in
other countries, in which there are National Health Systems, the percentage of uninsured people
ranges from 1% to 2% of the population. Today, more than 1000 insurance companies operate
and multiple health programs are offered to American citizens. The amount of cover, premium
and employee participation therein is negotiated between employers and employees.
The lack of a national social security institution and the rapid increase in health expenditure,
with continued widening of social inequalities, have led to the pursuit of policies to maintain low
costs of health services. These changes have been the pillars of reform at international level. The
most prevalent and well-known organizational cost containment policy is Health Care
Organizations (HMOs), which are groups of doctors who offer health services to a certain number of people registered for a specific period of time and are paid on a pre-defined basis.

The main purpose of “Obamacare” was to decrease public health spending by eliminating the number of individuals without healthcare insurance, which has reached twenty-eight million in the US. According to David Himmelstein, a professor at the City University of New York School of Public Health, research evidence argues that the greater the involvement of individuals in a health care system, the more expensive its administration costs. For the US health system, Mr. Himmelstein explains that two-thirds of his income comes from “federal taxes” and insurance contributions. The majority however, is channeled into the system, through private insurance companies and hospitals, in order to provide health services to insured people. On the other hand, he stresses that the personal contributions of insured people are constantly rising to cover the cost of expenses, therapies and medicine.

Increase in health expenditure must follow GDP growth. Otherwise, the continuous increase in health spending puts a strain on resources failing to satisfy other needs and demands of the population. Increase in health spending requires rational management in order to provide health services in the most efficient and effective way. The need for cost containment policies and control of health costs is therefore essential. The continuous rising cost of health services has led to the government adopting a series of measures in an attempt to control the cost and rationalization of using health system resources. Despite efforts having been made, the total cost of health spending has not reduced in any expenditure category.

The need for more accurate health expenditure forecasts in order to avoid the risk of uncertainty has led to the development and improvement of time series models over the previous decades. Extensive efforts have been made in the research community to develop and improve these models. One of the most important and widely used in time series models is the Box-Jenkins (1976) methodology.

The aim of this study is to construct the most appropriate model to investigate and forecast total US health expenditure. For this purpose, the ARIMA models and the Box-Jenkins methodology are used. Model forecasting capability is also examined using all of the forecasting criteria for both dynamic and static processes. The remaining sections of this study are organized as follows: Section 2 gives a brief literature review. Section 3 outlines a theoretical background about the
forecasting methodology and analysis. Section 4 analyzes data and descriptive statistics and the empirical results followed are described in Section 5. Section 6 proposes the forecasting methodology and finally, the last section concludes the study with some closing remarks.

2. LITERATURE REVIEW

The landscape for forecasting models in the health sector is dynamic and evolving. Several time series models have been proposed in the literature to project healthcare spending over time. These models are used to predict health expenses for individuals, groups or whole communities. These models also focus on specific categories of health expenditure, such as public spending, social or private insurance. Moreover, factors which have led to an increase in health expenditure based on the forecasting period are examined in the literature. In the short-term, the rise in health spending is largely linked to government budget decisions. In the medium term, technological changes play a more important role in explaining this increase (Di Matteo, 2005), while in the long term, risk factors such as changes in the prevalence of chronic diseases seem to be associated with this rise. In relation to USA the above studies forecast health expenses and some of these investigate the best model to predict health spending using methods and tools such as neural networks, VAR models as well as ARIMA models and the Box-Jenkins methodology.

To the best of our knowledge, no studies have been carried out using forecasting ARIMA models in the US. Other countries however, have used ARIMA models and other methods of forecasting which are mentioned below.

Lee and Miller (2002) used stochastic time series models to forecast the rise in health expenditure in the US, using the US health system for the ageing population (Medicare) with annual health spending as a GDP percentage from 2002 to 2075. It was predicted that by 2075 Medicare expenditures will have increased to 8% of GDP from the current 2.2%. This rise is attributed to an equivalent rise in per capita health spending and the ageing population.

To forecast health spending in Canada, Di Matteo (2010) used data from 1965 to 2008. With the least-squares method (OLS) and the generalized least-squares method (GLS), he predicted the future values of health expenditure and pointed out that per capita spending during this period rose.
Chaabouni and Abednnadher (2013) used Tunisian annual health spending data from 1961 to 2008 to conduct forecasts using neural networks and an autoregressive distributed lag model and proved that such networks are more accurate in forecasting. They also concluded that the Tunisian health expenditure is rising in a non-analogous manner in relation to GDP.

Zhao (2015) employed annual health spending data for 34 member states of OECD and appraised the performance of forecasting methods based on three criteria: accuracy, precision and certainty. Using these criteria he assessed the performance of univariate, multivariate, static and dynamic panel data and ARIMA models, as well as VAR models and showed in contrast to Getzen and Poullier (1992) that simple statistical models (e.g. smoothing) and time series models provide better forecasts against complex micro panel data models.

Yue et al. (2015) used the autoregressive moving average (ARIMA) models to analyze and forecast health expenditure, and in particular, hospital costs for respiratory illnesses in Shanghai, China. Using monthly data from January 2012 to December of the same year showed that the ARIMA (0,1,1)(0,1,1)$_{12}$ and ARIMA (0,1,1) models proved to be the best for this projection.

Peijun (2016) analyzed and predicted China’s total health expenditure. His results have shown that the ARIMA (5,1) model is the most appropriate to forecast total expenditure as a percentage of GDP.

3. METHODOLOGY

3.1 Theoretical Background

Box and Jenkins (1976) proposed a method called the ARIMA or Box-Jenkins method for constructing time series models. The Box-Jenkins methodology involves controlling the stationarity of a time series and converting a non stationary series into a stationary one, by using the following differences:

$$\Delta y_t = y_t - y_{t-1}$$  

which with the lag order operator L can be written as:

$$\Delta y_t = (1 - L)y_t$$

$$\Delta y_t = y_t - y_{t-1}$$

(1)

(2)
An integrated model using $d$ number of differences to convert into a stationary model is given as:

$$\Delta^d y_t = (1-L)y_t$$  \hspace{1cm} (3)

The new stationary series is determined by the autoregressive integrated moving average ARIMA ($p,d,q$) model. The general form of this model is:

$$A(L)(1-L)^d y_t = \delta + \Theta(L)\epsilon_t$$  \hspace{1cm} (4)

where

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$$ \hspace{1cm} (5)

$$\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q$$ \hspace{1cm} (6)

d: is the order of differencing

$p$: is the order or parameters of the autoregressive process

$q$: is the order or parameters of the moving average process

The Box-Jenkins method is used for short time forecasting and includes four stages: 1) Model identification, 2) Estimation, 3) Diagnostic checking and 4) Forecasting.

- **Identification**: Once the appropriate value $d$ is set to convert a non stationary series into a stationary one, we then have to identify the order (i.e., $p$ and $q$) of the autoregressive and moving average terms in order to capture the salient dynamic characteristics of the data. This leads to the use of graphical procedures (plotting the ACF and PACF, etc.).

- **Estimation**: The estimation of the parameters $\alpha_1, \ldots, \alpha_p$ of the autoregressive process (AR) and the parameters $\theta_1, \ldots, \theta_q$ of the moving average process (MA) is based on the partial autocorrelation functions and autocorrelation functions, respectively. If the series is an autoregressive (AR) procedure, the coefficients can be estimated via the least-squares method. If the series includes moving average terms (MA) or is a mixed (ARMA) process, therefore non linear estimation methods, such as maximum likelihood, using the arithmetic optimization of algorithms can be used to estimate the parameters.

In a sample of $n$ independent and identically distributed observations $y_1, y_2, \ldots, y_n$, derived from a distribution with unknown probability density function $f_0(\cdot)$, the function $f_0$ belongs to a specific distribution family, which is given by:

$$\{f(\cdot | \theta), \theta \in \Theta\}$$ \hspace{1cm} (7)

so that
\[ f_0 = f(\theta_0) \tag{8} \]

where

\( \theta \) : the vector of parameters for this distribution family.

\( \theta_0 \): the unknown actual value of the parameter vector.

The aim is to find an estimator \( \hat{\theta} \) as close as possible to the actual value \( \theta_0 \). For an independent and equally distributed sample, the joint density function for all observations is:

\[ f(y_1, y_2, ..., y_n | \theta) = f(y_1 | \theta) \times f(y_2 | \theta) \times ... \times f(y_n | \theta) \tag{9} \]

If we consider that the observations \( y_1, y_2, ..., y_n \) are the constant parameters and that \( \theta \) is the function variable we therefore obtain the likelihood function (Mai et al., 2014):

\[ L(\theta) = L(\theta | y_1, y_2, ..., y_n) = f(y_1, y_2, ..., y_n | \theta) = \prod_{i=1}^{n} f(y_i | \theta) \tag{10} \]

and the loglikelihood function:

\[ LL(\theta) = \ln L(\theta | y_1, y_2, ..., y_n) = \ln L(y_1, y_2, ..., y_n | \theta) = \sum_{i=1}^{n} \ln f(y_i | \theta) \tag{11} \]

The average loglikelihood is given by:

\[ \hat{c} = \frac{1}{n} \sum_{i=1}^{n} \ln f(y_i | \theta) \tag{12} \]

where

\( \hat{c} \) : is the expected loglikelihood in a single observation.

The maximum likelihood estimates \( \theta_0 \) by finding a value of \( \theta \) that maximizes the function \( L(\theta | y_1, y_2, ..., y_n) \) in Eq. (10) and defines the maximum likelihood estimator (\( \hat{\theta}_n \)) of \( \theta_0 \) as:

\[ \hat{\theta}_n = \arg \max_{\theta \in \Theta} L(\theta | y_1, y_2, ..., y_n) \tag{13} \]

or

\[ \hat{\theta}_n = \arg \max_{\theta \in \Theta} LL(\theta) \tag{14} \]

since the logarithm operator is an ascending monotonous function.
3.2 Covariance Matrix Estimation and Hessian Estimation

The distribution of the maximum likelihood estimator $\hat{\theta}_n$ from a parameter vector $\theta_0$ can be approximated from a multivariate normal distribution with mean $\theta_0$ and covariance matrix as:

$$V_n = \frac{1}{n} \left( Var[\nabla_\theta \ln(f(y | \theta_0))] \right)^{-1}$$

where $\ln(f(y | \theta_0))$ is the loglikelihood of an observation from the sample, evaluated at the true parameter $\theta_0$ and the gradient $\nabla_\theta \ln(f(y | \theta_0))$ is the vector of first derivatives of the loglikelihood (Taboga, 2012; Newey and McFadden, 1994).

To estimate asymptotic covariance matrix of Eq. (15) the Hessian estimation is used and calculated as:

$$\tilde{V}_n = \left( -\frac{1}{n} \sum_{i=1}^n \nabla_{\theta \theta} \ln(f(y_i | \hat{\theta}_n)) \right)^{-1}$$

where the Hessian matrix $\nabla_{\theta \theta} \ln(f(y_i | \theta_0))$ is the matrix of second-order partial derivatives of the loglikelihood function. Under some regularity conditions, the Hessian estimator $\tilde{V}_n$ is a consistent estimator of $V_n$.

3.3 Maximum Loglikelihood Optimization and BFGS Method

The estimation of maximum loglikelihood can be expressed as a nonlinear optimization problem:

$$\min_{x \in \mathbb{R}^p} f(x)$$

where

$$f(x) = -\mathcal{L}(x) \quad \text{and} \quad f(x) = -\mathcal{L}(x)$$

and we use $x$ instead of $\theta$ in order to follow conventional notation to optimization. The optimization algorithm tries to reliably converge into a local minimizer from an arbitrary starting point.

Every iteration $k$ of the secant method uses current iteration information to define the matrix $H_k$ as well as $H_{k+1}$. The matrix $H_{k+1}$ is calculated to satisfy the equation:
where

\[ H_{k+1}d_k = y_k \quad (18) \]

\( H_k \) is an approximation of the Hessian matrix that is renewed in every iteration \( k \),

\[ d_k = x_{k+1} - x_k \]

and

\[ y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \]

In this way, \( H_{k+1}d_k \) in Eq. (18) is a finite differencial approach to the derivative of \( \nabla f(x) \) in the direction of \( x_{k+1} - x_k \). To define uniquely \( H_{k+1} \), an additional condition is imposed so that among all the symmetrical matrixes which satisfy the secant equation, the one closest to the current matrix \( H_k \) is selected:

\[
\min_{H=H^T, H d_k = y_k} \| H - H_k \|_w
\]

(19)

where

\( T \) is the transpose operator

\[ \| \|_w : \] is the weighted Frobenius norm:

\[ \| A \|_w = \| W^{1/2} A W^{1/2} \|_F \]

in which \( \| \|_F \) is defined by

\[ \| C \|_F = \sqrt{\sum_{i \leq j \leq n} c^2} \]

for a constant \( c \).

The weight \( W \) can be selected as a matrix satisfying the condition \( W y_k = d_k \). This condition allows the above problem to be solved. The unique solution is:

\[
H_{k+1} = H_k - \frac{H_k d_k d_k^T H_k}{d_k^T H_k d_k} + \frac{y_k y_k^T}{y_k^T d_k}
\]

(20)

Eq. (20) is the BFGS method and refers to the initials which stand for Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970) and is one of the most well-known Hessian matrix approximation methods to solve non linear least-squares problems.
• **Diagnostic checking:** when the model fits precisely with the above estimation, the next step is to undertake diagnostic checking of the model. Diagnostic checking refers to statistical tests for the significance of the coefficients, the behavior of the residuals as well as the order of the model. If the estimated model satisfactorily represents the process from which the data originates, the residuals should then behave as white noise:

\[ \varepsilon_t \sim N(0, \sigma^2) \]

i.e. the residuals should not autocorrelate. The autocorrelation of the residuals is tested by the statistics \( Q \) of Ljung–Box (1978, 1979) and is defined as:

\[
Q_{LB} = n(n + 2) \sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{n-s} \tag{21}
\]

where

- \( n \): is the number of observations
- \( m \): is the number of autocorrelation coefficients and usually \( m = \sqrt{n} \)
- \( \hat{\rho}_s \): is the sampled residual autocorrelation

To check the normality of residuals, we use the Shapiro-Wilk test (1965) as well as a graph of the standardized residuals.

In order to compare the interpretative capacity of alternative models which differ in both parameter number and sample size, three criteria are used: Akaike, Schwartz and Hannan-Quinn. The model which returns the lowest value for these three criteria is selected. If all autocorrelations and partial autocorrelations are low in value, the model is considered adequate and forecasts are generated.

• **Forecasting:** The forecasting process for the future time series values follows completion of the above steps, based on the most appropriate model which has emerged from the previous stages. The forecasting of ARIMA models is evaluated both in-sample and out-of-sample analysis. The optimum predicted value is assessed by the Mean Squared Error (MSE). Other indicators usually used for measuring forecasting performance is the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Theil (U-Theil) (1961) Inequality Coefficient. These indexes are given as follows:

\[
MSE = \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \tag{22}
\]

\[
MAE = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_t - y_t| \tag{23}
\]
Theil Inequality Coefficient is as follows:

\[ U = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t)^2 + \frac{1}{T} \sum_{t=1}^{T} (y_t)^2}} \quad 0 \leq U \leq 1 \tag{25} \]

where

- \( y_t \): is the actual value of the endogenous variable \( y \) in time \( t \).
- \( \hat{y}_t \): is the revised value of the endogenous variable \( y \) in time \( t \).
- \( T \): is the number of observations in the simulation (of a sample).

If the Theil Inequality Coefficient \( U=0 \), the actual values of the series are equal to the estimated values \( y_t = \hat{y}_t \) for all \( t \). This case presents a perfect fit between the actual and predicted values.

Alternatively, if \( U=1 \), there is no such correct forecasting for the sample being investigated.

In addition, using Chow test (1960), we investigate the model forecast after 2008 (the year of the international financial crisis and a lack of liquidity in the stock markets of all developed countries).

4. DATA AND EMPIRICAL RESULTS

The data analyzed in this study derives from the Historical Statistics database of the United States covering the period from 1900 to 2017. The data is transformed into logarithms in order to avoid heteroskedasticity as well as asymmetric distributions problems. Figure 1 shows the rate of total US health expenditure (in their logarithmic form) for the period in levels and Table 1 displays the estimation of the variable in relation to time to determine whether or not a trend exists.

In Figure 1, we observe that total US health spending presents an upward trend throughout the period of study. Table 1 shows that there is a trend in the model we estimate. Hence, we can say that the series under investigation is non-stationary. Figure 2 represents the autocorrelation and partial autocorrelation plots confirming whether stationarity exists.
**FIGURE 1 – The Rate of Total US Health Expenditure**

![Figure 1](image)

**TABLE 1 - Estimation of Total Health Expenditure from 1900 to 2017**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-3.544352</td>
<td>0.071741</td>
<td>-49.40487</td>
<td>0.0000</td>
</tr>
<tr>
<td>TREND</td>
<td>0.095667</td>
<td>0.001060</td>
<td>90.27096</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared | 0.985965    | Mean dependent var | 2.052190 |

Adjusted R-squared | 0.985844 | S.D. dependent var | 3.295771 |

S.E. of regression | 0.392132 | Akaike info criterion | 0.982369 |

Sum squared resid  | 17.83706 | Schwarz criterion | 1.029330 |

Log likelihood    | -55.95977 | Hannan-Ouinn criter. | 1.001436 |

F-statistic       | 8148.845  | Durbin-Watson stat   | 0.030285 |

Prob(F-statistic) | 0.000000  |

*Notes: dependent variable is lhealth. Method used is Least Squares. Sample is from 1900 to 2017. Included observations are 118.*
From the autocorrelation plot in Table 2, we notice that the autocorrelation coefficients decline slowly, indicating that the series is non-stationary.

Therefore, the next step is to reapply the above tests to detect the existence of stationarity of the time series in the first difference. Figure 2 and Table 3 show total US health expenditure in the first difference and their estimation, respectively.

**Figure 2 - The Autocorrelation and Partial Autocorrelation Plots of Total Health Expenditure from 1900 to 2017**

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.979</td>
<td>0.979</td>
<td>119.02</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.935</td>
<td>-0.026</td>
<td>335.65</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.913</td>
<td>-0.009</td>
<td>439.24</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.891</td>
<td>-0.024</td>
<td>538.62</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.867</td>
<td>-0.024</td>
<td>633.75</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.844</td>
<td>-0.016</td>
<td>724.63</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.820</td>
<td>-0.023</td>
<td>811.26</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.794</td>
<td>-0.017</td>
<td>894.83</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.771</td>
<td>-0.023</td>
<td>971.91</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.727</td>
<td>-0.023</td>
<td>1115.9</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.699</td>
<td>-0.015</td>
<td>1181.0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.674</td>
<td>-0.020</td>
<td>1243.6</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.649</td>
<td>-0.012</td>
<td>1301.6</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.619</td>
<td>-0.012</td>
<td>1367.0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.590</td>
<td>-0.014</td>
<td>1486.3</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.576</td>
<td>-0.010</td>
<td>1542.2</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.552</td>
<td>-0.014</td>
<td>1496.0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.528</td>
<td>-0.016</td>
<td>1537.0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.516</td>
<td>-0.013</td>
<td>1592.7</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.497</td>
<td>-0.013</td>
<td>1626.7</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.437</td>
<td>-0.019</td>
<td>1666.8</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.405</td>
<td>-0.022</td>
<td>1692.0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.382</td>
<td>-0.022</td>
<td>1714.6</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.359</td>
<td>-0.003</td>
<td>1734.6</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.328</td>
<td>-0.014</td>
<td>1759.2</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.312</td>
<td>-0.014</td>
<td>1787.9</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.285</td>
<td>-0.021</td>
<td>1791.1</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.241</td>
<td>-0.020</td>
<td>1820.5</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.217</td>
<td>-0.021</td>
<td>1808.9</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.193</td>
<td>-0.026</td>
<td>1816.1</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.144</td>
<td>-0.020</td>
<td>1624.5</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3 - The Rate of Total US Health Expenditure in First Difference**
The results from Figure 3 show that there is stationarity in first difference. In addition to this, the absence of a trend shown in Table 3 confirms the stationarity of the series in first difference. Following this, series stationarity is confirmed from the autocorrelation and partial autocorrelation plots in the Figure 4 below.

In Figure 4, the autocorrelation plot in first difference shows that the autocorrelation coefficients decline at a rapid rate, indicating that the series is stationary. To further confirm the stationarity of total US health expenditure in first difference we use Dickey-Fuller (1979, 1981) and Phillips–Perron (1988) unit root tests.

The results from Table 3 confirm that the time series is stationary in first difference. Therefore, for the ARIMA (p,d,q) model under investigation the value d = 1.

### TABLE 2 - Estimation of Total Health Expenditure in First Difference from 1900 to 2017

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.075663</td>
<td>0.012589</td>
<td>6.010245</td>
<td>0.0000</td>
</tr>
<tr>
<td>TREND</td>
<td>0.000185</td>
<td>0.000185</td>
<td>1.022389</td>
<td>0.3087</td>
</tr>
</tbody>
</table>

R-squared 0.009008, Mean dependent var 0.086833
Adjusted R-squared 0.000390, S.D. dependent var 0.067662
S.E. of regression 0.067649, Akaike info criterion -2.532014
Sum squared resid 0.526289, Schwarz criterion -2.484797
Log likelihood 150.1228, Hannan-Ouinn criter. -2.512845
F-statistic 1.045279, Durbin-Watson stat 0.980048
Prob(F-statistic) 0.308743

Notes: dependent variable is dlhealth. Method used is Least Squares. Sample is from 1901 to 2017. Included observations are 117.
FIGURE 4 - The Autocorrelation and Partial Autocorrelation Plots of Total Health Expenditure in First Difference from 1900 to 2017

TABLE 3 - Summary of Augmented Dickey–Fuller and Phillips–Perron Unit Root Tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF</th>
<th>P-P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>C, T</td>
</tr>
<tr>
<td>lhealth</td>
<td>0.081851(1)</td>
<td>-1.850906(1)</td>
</tr>
<tr>
<td>dlhealth</td>
<td>-6.1552(0)***</td>
<td>-6.1288(0)***</td>
</tr>
</tbody>
</table>

Notes: *, ** and *** show significance at 1%, 5% and 10% levels respectively. The numbers within parentheses followed by ADF statistics represent the lag length of the dependent variable used to obtain white noise residuals. The lag lengths for ADF equation were selected using Schwarz Information Criterion (SIC). Mackinnon (1996) critical value for rejection of hypothesis of unit root applied. The numbers within brackets followed by PP statistics represent the bandwidth selected based on Newey West (1994) method using Bartlett Kernel. C=Constant, T=Trend. 7. d=First Differences, l=log transformation.

Table 4 presents the descriptive statistics for total US health spending both in levels and first difference.
TABLE 4 - Descriptive Statistics of US Total Health Expenditure in Levels and First Difference

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
<th>Std. Dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Jarque-Bera</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>lhealth</td>
<td>2.052</td>
<td>1.571</td>
<td>7.346</td>
<td>-2.813</td>
<td>3.295</td>
<td>0.156</td>
<td>1.608</td>
<td>9.999</td>
<td>0.006</td>
</tr>
<tr>
<td>dlhealth</td>
<td>0.086</td>
<td>0.081</td>
<td>0.402</td>
<td>-0.036</td>
<td>0.067</td>
<td>1.283</td>
<td>6.733</td>
<td>100.09</td>
<td>0.000</td>
</tr>
</tbody>
</table>

From the table above, we observe that US total health spending in levels and first difference does not follow normal distribution.

The form of the ARIMA (p,d,q) model is then determined using the results of the autocorrelation and partial autocorrelation plot in Table 2. The results from Figure 4 and Table 3 showed that the series is stationary in first difference and therefore the value of parameter d equals 1. Parameters p and q of the ARIMA model can be determined by the coefficients of partial autocorrelation and autocorrelation, respectively, by comparing them to the critical value 
\[ \pm \frac{2}{\sqrt{n}} \approx \pm 0.184 . \]

From the values of the autocorrelation and partial autocorrelation coefficients in the plot in Figure 4 we can observe that the value of p will be between 0<p<2 and respectively, the value of q will be between 0<q<3. Using the above values, we select the best ARIMA (p,1,q) model from the lowest values of the AIC, SC, and HQ criteria. Table 5 gives the values of p and q.

The results from Table 5 indicate that according to the Akaike (AIC) and Hannan-Quinn (HQ) criteria the ARIMA (2,1,0) model is the most appropriate, whereas according to the Schwartz criterion (SC) the ARIMA (1,1,0) model is the most suitable. Tables 6 and 7 present the estimations for the above models.
### TABLE 5 - Comparison of the ARIMA $(p,1,q)$ Models via the AIC, SIC and HQ Test Criteria

<table>
<thead>
<tr>
<th>ARIMA model</th>
<th>AIC</th>
<th>SC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>dlhealth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2,1,0)</td>
<td>-2.825750</td>
<td>-2.731317</td>
<td>-2.787412</td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>-2.814992</td>
<td>-2.720559</td>
<td>-2.776654</td>
</tr>
<tr>
<td>(2,1,1)</td>
<td>-2.812896</td>
<td>-2.694854</td>
<td>-2.764972</td>
</tr>
<tr>
<td>(1,1,2)</td>
<td>-2.812469</td>
<td>-2.694428</td>
<td>-2.764546</td>
</tr>
<tr>
<td>(1,1,0)</td>
<td>-2.805105</td>
<td>-2.734280</td>
<td>-2.776350</td>
</tr>
<tr>
<td>(0,1,2)</td>
<td>-2.797276</td>
<td>-2.702843</td>
<td>-2.758937</td>
</tr>
<tr>
<td>(2,1,2)</td>
<td>-2.796532</td>
<td>-2.654882</td>
<td>-2.739024</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>-2.795468</td>
<td>-2.653818</td>
<td>-2.737960</td>
</tr>
<tr>
<td>(2,1,3)</td>
<td>-2.793584</td>
<td>-2.628326</td>
<td>-2.726491</td>
</tr>
<tr>
<td>(0,1,3)</td>
<td>-2.788587</td>
<td>-2.670545</td>
<td>-2.740663</td>
</tr>
<tr>
<td>(0,1,1)</td>
<td>-2.693803</td>
<td>-2.622978</td>
<td>-2.665049</td>
</tr>
<tr>
<td>(0,1,0)</td>
<td>-2.522966</td>
<td>-2.475749</td>
<td>-2.503796</td>
</tr>
</tbody>
</table>

Tables 6 and 7 show no existing problems in the two models that were estimated. We can therefore continue with diagnostic checking of these models. Figures 5 and 6 indicate the tests for the existence of conditional heteroskedasticity (ARCH process test $(q)$) for these two models, from the squares of the residuals.
### Table 6 - Estimation of the ARIMA (2,1,0) Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.084827</td>
<td>0.013142</td>
<td>6.454736</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.413354</td>
<td>0.090304</td>
<td>4.577371</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.192877</td>
<td>0.090846</td>
<td>2.123117</td>
<td>0.0359</td>
</tr>
<tr>
<td>SIGMASQ</td>
<td>0.003230</td>
<td>0.000422</td>
<td>7.648147</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared     0.288387  Mean dependent var 0.086833
Adjusted R-squared 0.269494  S.D. dependent var 0.067662
S.E. of regression 0.057831  Akaike info criterion -2.825750
Sum squared resid 0.377918  Schwarz criterion -2.731317
Log likelihood 169.3064  Hannan-Ouinn criter. -2.787412
F-statistic 15.26470  Durbin-Watson stat 1.964699
Prob(F-statistic) 0.000000

Inverted AR Roots .69  -.28

**Notes:** dependent variable is dlhealth. Method used is ARMA Maximum Likelihood (BFGS). Sample is from 1901 to 2017. Included observations are 117. Convergence achieved after 4 iterations. Coefficient covariance computed using observed Hessian.
### TABLE 7 - Estimation of the ARIMA (1,1,0) Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.085772</td>
<td>0.010847</td>
<td>7.907799</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.510300</td>
<td>0.079585</td>
<td>6.412034</td>
<td>0.0000</td>
</tr>
<tr>
<td>SIGMASQ</td>
<td>0.003357</td>
<td>0.000439</td>
<td>7.648332</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.260522
Adjusted R-squared 0.247549
S.E. of regression 0.058693
Sum squared resid 0.392716
Log likelihood 167.0986
F-statistic 20.08142
Prob(F-statistic) 0.000000
Inverted AR Roots .51

Notes: dependent variable is dlhealth. Method used is ARMA Maximum Likelihood (BFGS). Sample is from 1901 to 2017. Included observations are 117. Convergence achieved after 4 iterations. Coefficient covariance computed using observed Hessian.

The results from Figures 5 and 6 show that all autocorrelation and partial autocorrelation coefficients are not statistically significant. Therefore, the ARCH or GARCH process is rejected in both models studied.
The value of the maximum likelihood (LL) in these two models is then checked to find the most appropriate model.
Table 8 - ARIMA (2,1,0) and ARIMA (1,1,0) Model Estimation

<table>
<thead>
<tr>
<th>ARIMA model</th>
<th>(2,1,0)</th>
<th>(1,1,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>169.3064</td>
<td>167.0986</td>
</tr>
</tbody>
</table>

From the above Table 8, we observe that the maximum likelihood value is greater for the ARIMA (2,1,0) model and therefore we select this model as the best for forecasting. We then check the normality of the residuals with the Shapiro-Wilk test (1965) and the plot of standardized residuals in Table 9 and Figure 7, respectively.

Table 9 - Shapiro-Wilk Test of the Standardized Residuals

<table>
<thead>
<tr>
<th>W</th>
<th>0.938</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value (Two-tailed)</td>
<td>&lt; 0.035</td>
</tr>
<tr>
<td>alpha</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Figure 7 - The Plot of Standardized Residuals between the ± 3s Control Limits

Notes: the outliers, points 47 and 67, correspond to the years 1947 and 1967, respectively.
Table 9 and Figure 7 both show that the residuals do not follow normal distribution (some are outside the ± 3s limits).

5. **FORECASTING METHODOLOGY**

In this section of the study we present the results for the ARIMA (2,1,0) model. To predict health expenditure we use one-step ahead static and dynamic forecasting. Static forecasting extends the recursion forwards from the end of the sample estimation, allowing one-step ahead projection both in structural samples and innovations.

Having selected the form of the ARIMA (2,1,0) model, the graphs for the actual and predicted model values as well as the innovations for dynamic and static forecasting, respectively, are presented in Figures 8 and 9. In the same tables we notice some statistical indicators such as the Root Mean Squared Error, the Mean Absolute Error and Theil’s Inequality Coefficient to see the predictability of the model. In addition to this, in the same tables, the mean value prediction is presented in a wide ± 2SE confidence interval.

**Figure 8 - Dynamic Forecast – Comparative Statistics**
**Figure 9 - Static Forecast – Comparative Statistics**

**Table 10 - Chow Forecast Test**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>df</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-statistic</td>
<td>0.207334</td>
<td>(10, 103)</td>
<td>0.9952</td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td>2.711770</td>
<td>10</td>
<td>0.9874</td>
</tr>
</tbody>
</table>

**F-test summary**

<table>
<thead>
<tr>
<th></th>
<th>Sum of Sq.</th>
<th>df</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test SSR</td>
<td>0.007457</td>
<td>10</td>
<td>0.000746</td>
</tr>
<tr>
<td>Restricted SSR</td>
<td>0.377918</td>
<td>113</td>
<td>0.003344</td>
</tr>
<tr>
<td>Unrestricted SSR</td>
<td>0.370461</td>
<td>103</td>
<td>0.003597</td>
</tr>
</tbody>
</table>

**LR test summary**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted LogL</td>
<td>169.3064</td>
<td>113</td>
</tr>
<tr>
<td>Unrestricted LogL</td>
<td>170.6623</td>
<td>103</td>
</tr>
</tbody>
</table>

*Notes:* the specification is dlhealth c ar(1) ar(2). Test predictions for observations are from 2008 to 2017. Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample.
From the above table we observe that the coefficients in both models are stable. Therefore, the increase in health expenditure continued to rise in the US following the 2008 financial crisis. In continuation, the graphs for the actual and predicted model values for dynamic and static forecasting following the crisis of 2008, are presented in Figures 10 and 11, respectively.

**Figure 10 - Dynamic Forecast after 2008 – Comparative Statistics**

![Dynamic Forecast after 2008 – Comparative Statistics](image1)

**Figure 11 - Static Forecast after 2008 – Comparative Statistics**

![Static Forecast after 2008 – Comparative Statistics](image2)
The results from Figures 10 and 11 show that the Mean Absolute Error, the Root Mean Squared Error (RMSE) and Theil’s Inequality Coefficient criteria are again lower in static forecasting. In Figure 12, which follows, the actual and fitted values of static forecasting are presented. The following fan chart in Figure 13 joins a line for the actual past results after 2008 and 4 periods of future prediction by showing ranges for possible values of future data.

**Figure 12 - Actual and Fitted Values of Static Forecasting**

**Figure 13 – Fan Chart for the Health Expenditure Forecasts**

*Notes:* the different shades of color represent the different confidence intervals. The widest band reflects the 90% confidence interval. The edges of the shaded area were derived by adding to (and subtracting from) the central estimated projection the average absolute value of past forecast errors.
Figure 12 shows that the fitted values closely follow the actual observations with a quite good fit and a sense of the uncertainty around our central economic forecast until 2021 is given in Figure 13.

6. CONCLUSION

The continuous rise in health spending forces governments to rationally manage the structural functionality of the health system in order to provide services in the most efficient, effective and equitable way. What is more, in the framework of the social state and the non-negotiable human right to health, these services must be socially just. Considering that meeting the health needs of the population requires the presence of a rational health system, it becomes obvious that the inputs of the system and the procedures followed when consuming resources are those that will produce good or poor quality results for both the contributor’s system, and the wider community.

The use of ARIMA models is an excellent tool to predict health costs and capture this trend. The main objective of this study is to find the most appropriate model for projecting health spending in the US. The results showed that US health spending could be better constructed and predicted using the ARIMA (2,1,0) model. With the Box-Jenkins methodology, the form of the ARIMA (2,1,0) model was estimated by the non-linear maximum likelihood optimization method, using the arithmetic optimization of the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS), which is one of the best known approaches for solving nonlinear least squares problems. To estimate the asymptotic matrix of the coefficient’s covariance, the Hessian matrix method was used. To predict the power of the model, both the dynamic and static procedure were used with statistical forecasting criteria. The forecasting results showed that the estimated value of health expenditure is close to the actual value. This result has shown that the suitability of the ARIMA (2,1,0) model can be used to predict health spending in the US in the coming years with static forecasting, over the short-term period. Obviously, this result may be affected by changes in both the time period and the sample data size.
REFERENCES


*Biometrika*, 65(2), 297-303.

Ljung, G.M. and G.E.P. Box (1979), “The Likelihood Function of Stationary Autoregressive-
Moving Average Models”, *Biometrika*, 66(2), 265-270.


Likelihood Estimation”, CIRRELT Report No. 64.

North Holland, Elsevier.


Model”, *Medicine and Society*, (03).

*Biometrika*, 75(2), 335-346.


Samples)”, *Biometrika*, 52(3-4), 591-611.

Amsterdam.

CreateSpace Independent Publishing Platform.

Hospital Admissions and Hospitalization Expenses for Respiratory Diseases”, *Chinese 
Journal of Health Statistics*, (02), 197-200.

Zhao J. (2015), “Forecasting Health Expenditure: Methods and Applications to International 