MODELING AND FORECASTING THE TOTAL NUMBER OF CASES AND DEATHS DUE TO PANDEMIC

Nasrullah Khan1 | Asma Arshad2 | Muhammad Azam3 | Ali Hussein Al-marshadi4 | Muhammad Aslam4

1Department of Statistics, College of Veterinary and Animal Sciences, Jhang, University of Veterinary and Animal Sciences Lahore, Lahore, Pakistan
2Department of Statistics, National College of Business Administration and Economics, Lahore, Pakistan
3Department of Statistics and Computer Science, University of Veterinary and Animal Sciences Lahore, Lahore, Pakistan
4Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence
Muhammad Aslam, Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia.
Email: aslam_ravian@hotmail.com

Abstract
The COVID-19 pandemic has appeared as the predominant disease of the 21st century at the end of 2019 and was a drastic start with thousands of casualties and the COVID-19 victims in 2020. Due to the drastic effect, COVID-19 scientists are trying to work on pandemic diseases and Governments are interested in the development of methodologies that will minimize the losses and speed up the process of cure by providing vaccines and treatment for such pandemics. The development of a new vaccine for any pandemic requires long in vitro and in vivo trials to use. Thus the strategies require understanding how the pandemic is spreading in terms of affected cases and casualties occurring from this disease, here we developed a forecasting model that can predict the no of cases and deaths due to pandemic and that can help the researcher, government, and other stakeholders to devise their strategies so that the damages can be minimized. This model can also be used for the judicial distribution of resources as it provides the estimates of the number of casualties and number of deaths with high accuracy, Government and policymakers on the basis of forecasted value can plan in a better way. The model efficiency is discussed on the basis of the available dataset of John Hopkins University repository in the period when the disease was first reported in the six countries till the mid of May 2020, the model was developed on the basis of this data, and then it is tested by forecasting the no of deaths and cases for next 7 days, where the proposed strategy provided excellent forecasting. The forecast models are developed for six countries including Pakistan, India, Afghanistan, Iran, Italy, and China using polynomial regression of degrees 3–5. But the models are analyzed up to the 6th-degree and the suitable models are selected based on higher adjusted R-square ($R^2$) and lower root-mean-square error and the mean absolute percentage error (MAPE). The values of $R^2$ are greater than 99% for all countries other than China whereas for China this $R^2$ was 97%. The high values of $R^2$ and Low value of MAPE statistics increase the validity of proposed models to forecast the total no cases and total no of deaths in all countries. Iran, Italy, and Afghanistan also show a mild decreasing trend but the number of cases is far higher than the decrease percentage. Although India is expected to have a consistent result, more or less it depicts some other biasing factors which should be figured out in separate research.

KEYWORDS
COVID-19 pandemic, mean absolute percentage error, polynomial regression, root mean square error, $R^2$
1 | INTRODUCTION

In this advanced area of science, man is enjoying the luxurious life that seems to be a dream in the early Twenty century. Science, technology, and artificial intelligence have dramatically improved every field of life. Similarly, the health sector has witnessed an intense improvement in methodologies, medicine, and Treatment. With all this advancement, humankind is facing the challenge to find out a treatment for new viruses, Parasites, and Microbes. In December 2019, an unprecedented disease similar to pneumonia-like clinical symptoms emerged in Wuhan, the biggest city of central China. It is said to be connected with wet-market dealing of victims with fish, bats and poultry. Further research revealed that it has very clear clinical characteristics of fever, cough, fatigue, loss of smell or taste, sore throat, labored breathing, and pneumonia. Recent studies further demonstrated that age is a predisposing factor in victimizing through this virus so older people with underlying disease conditions such as diabetes and hypertension are at higher risk of advanced symptoms which can lead to a casualty. The disease was later named COVID-19 by World Health Organization and the virus responsible for the disease was named SARS-CoV2.

As of May 15th, 2020, the disease has been reported in 216 countries causing infections in >4.4 Million people and >0.35 million deaths. In the absence of effective vaccination and therapeutics, social distancing and movement restriction through partial and complete lockdowns remained the key approach to tackle this pandemic. On one hand, movement restriction has negative impacts on the economy while easing restrictions may increase the number of cases. Therefore, understanding the future trends of this disease through a statistical model could be highly useful in policymaking. In this regard, Pandey developed a mathematical SIER model and a regression model for forecasting the total confirmed cases in India. Tan et al. used a method of differential equations system to forecast the cumulative cases counts in China. Recently, Li et al. has conducted a prospective study to forecast the number of future COVID-19 confirmed cases by observing the correlation between internet researches estimated results and daily actual reported cases. Remuzzi and Yang et al. accessed the situation of Italy and suggested the possible trend of the disease to overcome the pandemic. Yang et al. uses a modification of the SEIR model to forecast the disease patterns along with the AI method for forecasting the trend of the epidemic of COVID-19 in China. Panwar et al. developed a mathematical model using CF and ABC nonsingular derivative to model the COVID-19. Hu et al. using modeling techniques access the risk of COVID transmission in train passengers. Anzum and Islam have given a mathematical modeling approach for the production rate of COVID with the effect of policies and behaviors.

In the current study, using the probabilistic approach polynomial regression model is used to model the future scenario of any pandemic in terms of the number of deaths and the number of cases using R programming software. We applied a polynomial regression approach to model COVID-19’s prevalence in Pakistan and 5 other countries (India, Afghanistan, Iran, Italy, and China) and forecasted by using a polynomial regression model of degrees 3-5. Recently, Li et al. proposed the chaotic cloud quantum bats algorithm algorithm for optimization problems. Zhang and Hong forecasting modeling approach for electric loads using a support regression vector is discussed. The SAR modeling approach and using the CF and ABC nonsingular approach are mathematical models using the deterministic approach to model the number of cases, in the current study we assumed that the numbers of cases and number of deaths follow random behaviors, we applied the probabilistic approach to the model number of cases and number of deaths. Furthermore, these models are analyzed up to the 6th-degree and the suitable models are selected based on higher adjusted R-square (R²) and lower root-mean-square error (RMSE) & the mean absolute percentage error (MAPE) which satisfied these statistical checks at their maximum. The total no of cases and total no of deaths in Pakistan, India, Afghanistan, Iran, Italy, and China are from an open data source of Our World in Data and John Hopkins University, United States. In the study we have discussed the possible modeling strategy in case of a pandemic, for this, the data of the COVID-19 dataset of John Hopkin University repository is taken and we have selected 6 countries that cover approximately 42% of the world population and are neighboring countries to China from where the Pandemic as firstly reported whereas to test the efficacy of our proposed model for predicting the total no of effective cases and the number of deaths R², and MAPE are used. The short time forecast comprised of 7 days of the week is presented to assist the Government and Medical practitioners to formulate their policies and tackle with COVID-19 challenge of the century. Verity et al. provided the model-based analysis for COVID-19. Some more details can be seen in Ruan, Verity, and Tang et al. The rest of the formation of the paper is as under in Section 2 the applied methodology is discussed in detail, the result and discussion are carried out in Section 3 for each of the six countries. Section 4 provides the interpretation of the proposed methodology.

2 | PROPOSED METHODOLOGY AND RESULTS

Time series data provided by Our World in Data and John Hopkins University, United States has been used for analysis and forecasting of the total no of cases and total no of deaths in Pakistan, India, Afghanistan, Iran, Italy, and China. In the current study, we used the data of total no cases and total no deaths from the date when the first time a case is reported in these countries from the time of emergence to mid-May 2020. The understudy countries cover approximately 42% of the world population and they are neighboring or adjoining to neighboring countries to China from where this pandemic started, as it is supposed. These factors are the main motivation for our research. The key limitation of this study is that this modeling strategy can be used only where data is exponentially increasing where the infection rate is very high as it is in the case of pandemics used to see.

In the following section, the models used for analysis and forecast purposes are illustrated.
2.1 | Polynomial regression model

The multiple regression model with \( k \) independent variables regressing on the dependent variable \( Y \) is defined as

\[
Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k
\]

Where \( i = 1, 2, \ldots, n \) represents the number of observations in the sample. The \( \beta_0 \) is the intercept of multiple regression lines while \( \beta_1, \beta_2, \ldots, \beta_k \) are regression coefficients related to \( k \) predictors of the regression line.\(^{27}\) The polynomial regression model is the extension of the multiple regression model where the \( k \) predictor \( X_1, X_2, \ldots, X_k \) is replaced with \( X_1^2, X_2^2, \ldots, X_k^2 \), respectively, to study pandemics areas of research like COVID-19 which have an exponential spread by each passing unit of time. Hence, the polynomial regression model becomes

\[
Y = \beta_0 + \beta_1X_1 + \beta_2X_2^2 + \ldots, \beta_kX_k^2
\]

Where \( i = 1, 2, \ldots, n \) represents the number of observations in the sample. The \( \beta_0 \) is the intercept of the polynomial regression model and \( \beta_1, \beta_2, \ldots, \beta_k \) are polynomial regression coefficients related to \( k^{th} \) power of predictor \( X \).\(^{27}\) The interpretation of the regression coefficients of the polynomial regression model is the same as of multiple regression. The mean square error (MSE) is an unbiased estimator of \( \sigma^2 \) and is defined as

\[
MSE = \frac{\text{SSE}}{df} = \frac{\sum_{i=1}^{n}(Obs_i - Pred)^2}{n - (k - 1)}
\]

The quantity MSE measures the average square of the difference between observed and predicted values and is used as how well a model is fitted to data. The quantity \( \text{RMSE} = \sqrt{MSE} \) as although a biased estimator of \( \sigma \) but is widely used to measure the size of the error.

The MAPE is a widely used statistic to compare the accuracy of different models to compare the relative performance of models.\(^{28}\) This measure MAPE is defined as:

\[
\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{\text{Obs}_i - \text{Pred}_i}{\text{Obs}_i} \right|
\]

The Coefficient of Determination \( R^2 \) and adjusted \( R^2 \) are used to measure how much variation a model explains of the total variation. It ranges from 0 to 1, i.e., \( 0 \leq R^2 \leq 1 \) and is defined as:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(\text{Obs}_i - \text{Pred})^2}{\sum_{i=1}^{n}(\text{Obs}_i - \text{Mean})^2}
\]

Whereas the adjusted \( R \)-square (adj. \( R^2 \)) is defined as

\[
\text{adj. } R^2 = R^2 - \frac{(1 - R^2)k}{n - (k + 1)}
\]

2.2 | Algorithm for model selection

We applied the cyclical model approach for the selection of an appropriate model for the forecasting of the number of cases and the number of deaths. In the first step, we proposed a model, and after checking the data validity and assumption of the model we fitted the model, the efficacy statistic MAPE, \( R^2 \), and adj. \( R^2 \) for each of the proposed models is computed, if the proposed selected model outperforms the existing model then that is selected and the previous is dropped. This process is replicated till the selection of the most appropriate model. The flow chart of the process is illustrated in Figure 1.

2.3 | Results and discussion

The spread of the COVID-19 pandemic is undeniable as is evident from the drastic increase in the number of confirmed patients and death count in the whole world as shown in Figure 2 and 3. By the time of writing this manuscript, >4 million are the victims of this disease with approx. The mortality rate varies from country to country and climate to climate. None of the countries is saved from its effects but more or less in some countries due to various factors.
In this study, the discussion is comprised of the dataset for the number of confirmed COVID-19 patients and the number of deaths reported till mid of May 2020 from the emergence of the COVID-19 pandemic through authentic sources for the respective six countries specifically: Afghanistan, Pakistan, India, China, Iran, and Italy. As the target is to analyze the COVID-19 situation in Pakistan so its neighboring countries are selected for the comparison relative to the countries with severe spread of the COVID-19 pandemic in China and Italy. This study is an effort to give a forecast about the number of deaths and the expected number of new COVID-19 confirmed cases in the next weeks with statistical modeling. It is important to break the strength of the rumors spreading around through the encouraging effective facts and figures based on analytical output in favor of a decreased number of deaths and the expected number of COVID-19 cases as well. Analysis of the data is performed through Polynomial Regression up to 6th degree to fit the best model on the basis of lower value of RMSE and MAPE along with the higher value of $R^2$ as a measure of the reliability and accurate forecast obtained through fitted polynomial regression model with the suitable degree. The analysis is performed and presented in such a way that one can understand by depicting results separately in the percentage for individual countries along with most concerning China for elaborating details about the sensitivity of the conditions at the risk of 1% level of significance.

Available data about these countries is analyzed through the Polynomial Regression Model technique to peep into the depth of data. Pakistan is the target forecast country while the rest of the countries are included to observe the difference on relative grounds. In Table 1 the polynomial regression analysis for the confirmed COVID-19 cases counts is presented along with the fitted model coefficients. The possible significance of each coefficient with respect to individual countries is observed at 1% significance level. Similarly, in Table 2 COVID-19 deaths count is analyzed by fitting a polynomial regression model and related information about coefficients and possible significance is presented. By putting the values of coefficients in above-mentioned equation along with the time period as an independent variable provides the predicted number of cases and deaths. Similarly, forecasts about the future cases and deaths are provided in Tables 1 and 2, respectively. In the following paragraphs, individual countries along with the fitted models have been discussed (Table 2a).

### 2.3.1 Pakistan

The dataset of Pakistan has been taken since the emergence of the pandemic COVID-19 till mid-May 2020. The data are fitted with the 4th-degree of the polynomial regression model and found significant except the intercept. As the best-fitted model is chosen with the lowest value of RMSE & MAPE and the higher value of $R^2$ among all other fitted polynomial regression models with varying degrees for the forecast of the number of cases regarding outbreak and the spread of COVID-19 pandemic disease among humans in the upcoming week (see Table 1). The mathematical form of polynomial regression equation in the case of Pakistan to get a forecast about the total number of confirmed cases is as follows:

$$Y = \beta_0 + \beta_1X + \beta_2X^2 + \beta_3X^3 + \beta_4X^4 + e$$

So the fitted polynomial regression model is stated as: $\hat{Y} = 273.91 - 107.33X + 8.9784X^2 - 0.1852X^3 + 0.002382X^4$. Here an estimate of $\beta_0$ which is 273.91 explains that the total number of cases of COVID-19 patient independent from time period are initially around 274. Similarly, estimated values of the rest of the regression coefficients, i.e., $\beta_1$ through $\beta_4$ are reported ($-107.33, 8.9784, -0.1852$, and $0.002382$) respectively explain the impact of each day as linear, quadratic, cubic and quadrant in the additive model. In addition, some of the factors have an additive or negative impact on the composite output (prediction or forecast). Now one can use this fitted model to get the desired forecast about the total number of COVID-19 cases for the upcoming days or period. But it is recommended that the forecast period should not be far from the period of the fitted model. Pakistan’s forecasted data is depicting the significant increase in upcoming days which is apparently because of ease in lockdown and opening of markets 4 days a week. Similarly, data on the total number of deaths that occurred in patients of the COVID-19 pandemic is analyzed and results are presented in Table 2. Here, a most suitable polynomial regression model with degree 3 is chosen on the basis of the lowest value of RMSE & MAPE and the higher value of $R^2$ among all other fitted polynomial regression models for the forecast of death counts regarding outbreak and the spread of COVID-19 pandemic disease among humans in the upcoming days or period (see Table 2). The mathematical form of polynomial regression equation in the case of Pakistan to get a forecast about the total number of deaths is as follows:

$$Y = \beta_0 + \beta_1X + \beta_2X^2 + \beta_3X^3 + e$$

Therefore, the equation of the fitted polynomial regression model is: $\hat{Y} = -3.199 + 1.291X - 0.119X^2 + 0.003X^3$. Here an estimate of $\beta_0$ which is $-3.199$. This negative sign shows that there was initially no impact on the total number of deaths due to COVID-19 independent from the time period. But as time passed, an impact is observed in the form of estimated values of the rest of the regression coefficients, i.e., $\beta_1$ through $\beta_3$ are reported (1.291, 0.119, and 0.003) respectively, explain the impact of each day as linear, quadratic, and cubic in the additive model. In addition additive as well as
negative impact on various factors is observed on the composite output (predicted or forecasted deaths). Now a fitted model is ready to be utilized to get the desired forecast about the total number of COVID-19 deaths for the upcoming days or period.

Both of these fitted models explain the current scenario of Pakistan making it obvious that the current situation is under control but it could be dangerous or alarming if the government fails to maintain the current situation of lockdown in the country. So, necessary actions taken well in time are the need of the hour to mitigate the spread of this pandemic among the masses.

### 2.3.2 | India

the neighboring country of Pakistan is selected for the comparison due to the climatic, cultural, and infrastructural similarities. India’s data has been

| Country     | Estimate | SE     | t Value | p Value |
|-------------|----------|--------|---------|---------|
| Pakistan    | Intercept = $\beta_0$ | 273.907 | 144.757 | 1.892   | 0.063   |
|             | Poly (1) = $\beta_1$ | -107.329 | 26.362  | -4.071  | 0.01    |
|             | Poly (2) = $\beta_2$ | 8.978    | 1.416   | 6.341   | 0.01    |
|             | Poly (3) = $\beta_3$ | -0.185   | 0.028   | -6.550  | 0.01    |
|             | Poly (4) = $\beta_4$ | 0.002    | 0.000   | 12.735  | 0.01    |
| India       | Intercept = $\beta_0$ | 267.300  | 230.472 | 1.160   | 0.249   |
|             | Poly (1) = $\beta_1$ | -91.723  | 29.548  | -3.104  | 0.003   |
|             | Poly (2) = $\beta_2$ | 7.732    | 1.116   | 6.932   | 0.01    |
|             | Poly (3) = $\beta_3$ | -0.230   | 0.016   | -14.684 | 0.01    |
|             | Poly (4) = $\beta_4$ | 0.002    | 0.000   | 30.398  | 0.01    |
| Afghanistan | Intercept = $\beta_0$ | 50.838   | 32.721  | 1.554   | 0.125   |
|             | Poly (1) = $\beta_1$ | -18.875  | 6.203   | -3.043  | 0.003   |
|             | Poly (2) = $\beta_2$ | 1.612    | 0.347   | 4.647   | 0.01    |
|             | Poly (3) = $\beta_3$ | -0.041   | 0.007   | -5.675  | 0.01    |
|             | Poly (4) = $\beta_4$ | 0.001    | 0.000   | 10.582  | 0.01    |
|             | Poly (5) = $\beta_5$ | -2651.252| 1005.239| -2.637  | 0.010   |
| Iran        | Intercept = $\beta_0$ | 1069.996 | 228.728 | 4.678   | 0.01    |
|             | Poly (1) = $\beta_1$ | -109.041 | 16.059  | -6.790  | 0.01    |
|             | Poly (2) = $\beta_2$ | 5.231    | 0.465   | 11.252  | 0.01    |
|             | Poly (3) = $\beta_3$ | -0.078   | 0.006   | -13.258 | 0.01    |
|             | Poly (4) = $\beta_4$ | 0.000    | 0.000   | 13.980  | 0.01    |
|             | Poly (5) = $\beta_5$ | -11749.338| 2136.873| -5.498  | 0.01    |
| Italy       | Intercept = $\beta_0$ | -11749.338| 2136.873| -5.498  | 0.01    |
|             | Poly (1) = $\beta_1$ | 4165.197 | 396.887 | 10.495  | 0.01    |
|             | Poly (2) = $\beta_2$ | -341.070 | 22.718  | -15.014 | 0.01    |
|             | Poly (3) = $\beta_3$ | 9.675    | 0.536   | 18.068  | 0.01    |
|             | Poly (4) = $\beta_4$ | -0.098   | 0.006   | -17.736 | 0.01    |
|             | Poly (5) = $\beta_5$ | 0.000    | 0.000   | 16.174  | 0.01    |
| China       | Intercept = $\beta_0$ | 19561.665| 3096.381| 6.318   | 0.01    |
|             | Poly (1) = $\beta_1$ | -5063.443| 447.607 | -11.312 | 0.01    |
|             | Poly (2) = $\beta_2$ | 272.330  | 19.918  | 13.673  | 0.01    |
|             | Poly (3) = $\beta_3$ | -4.385   | 0.364   | -12.031 | 0.01    |
|             | Poly (4) = $\beta_4$ | 0.030    | 0.003   | 10.142  | 0.01    |
|             | Poly (5) = $\beta_5$ | 0.000    | 0.000   | -8.565  | 0.01    |

**TABLE 1** Polynomial regression analysis for COVID-19 cases counts dataset (model coefficients)
**TABLE 2** Polynomial regression analysis for COVID-19 deaths counts dataset (model coefficients)

| Country   | Intercept = $\beta_0$ | Estimate | SE  | t Value | p Value |
|-----------|-----------------------|----------|-----|---------|---------|
| Pakistan  | -3.199                | 3.294    | -0.971 | 0.335   |
|           | 1.291                 | 0.378    | 3.416 | 0.001   |
|           | -0.119                | 0.012    | -10.178 | 0.01    |
|           | 0.003                 | 0.000    | 33.250 | 0.01    |
| India     | 3.279                 | 11.665   | 0.281 | 0.779   |
|           | -2.013                | 1.496    | -1.346 | 0.181   |
|           | 0.217                 | 0.057    | 3.840 | 0.01    |
|           | -0.007                | 0.000    | -9.054 | 0.01    |
|           | 0.000                 | 0.000    | 19.453 | 0.01    |
| Afghanistan | 0.755                | 1.264    | 0.597 | 0.552   |
|           | -0.115                | 0.151    | -0.763 | 0.448   |
|           | -0.004                | 0.005    | -0.847 | 0.400   |
|           | 0.001                 | 0.000    | 10.691 | 0.01    |
| Iran      | 383.983               | 50.806   | 7.558 | 0.01    |
|           | -114.987              | 7.992    | -14.387 | 0.01    |
|           | 7.117                 | 0.371    | 19.208 | 0.01    |
|           | -0.084                | 0.006    | -13.083 | 0.01    |
|           | 0.000                 | 0.000    | 8.413  | 0.01    |
| Italy     | -1969.319             | 358.019  | -5.501 | 0.01    |
|           | 644.370               | 66.496   | 9.690  | 0.01    |
|           | -48.476               | 3.806    | -12.736 | 0.01    |
|           | 1.265                 | 0.090    | 14.101 | 0.01    |
|           | -0.012                | 0.001    | -12.905 | 0.01    |
|           | 0.000                 | 0.000    | 11.014 | 0.01    |
| China     | 790.834               | 161.565  | 4.895  | 0.01    |
|           | -186.224              | 23.356   | -7.974 | 0.01    |
|           | 9.295                 | 1.039    | 8.944  | 0.01    |
|           | -0.147                | 0.019    | -7.705 | 0.01    |
|           | 0.001                 | 0.000    | 6.613  | 0.01    |
|           | 0.000                 | 0.000    | -5.760 | 0.01    |

**TABLE 2a** Comparison of proposed with exponential and linear regression model

| Country   | Exponential regression | Linear regression | Proposed polynomial regression |
|-----------|------------------------|-------------------|-------------------------------|
|           | $R^2$ | Adj. $R^2$ | Mape | $R^2$ | Adj. $R^2$ | Mape | $R^2$ | Adj. $R^2$ | Mape |
| Pakistan  | 0.8621 | 0.8602 | 853.8 | 0.7908 | 0.7879 | 0.5871 | 0.9995 | 0.9995 | 0.2422 |
| India     | 0.9615 | 0.9611 | 1132.2 | 0.6007 | 0.5969 | 0.787 | 0.9995 | 0.9995 | 0.9099 |
| Afghanistan | 0.9191 | 0.9179 | 144 | 0.751 | 0.7474 | 0.6213 | 0.9989 | 0.9988 | 0.6465 |
| Iran      | 0.6939 | 0.6903 | 4426.1 | 0.973 | 0.9727 | 0.3035 | 0.9959 | 0.9957 | 0.5506 |
| Italy     | 0.811 | 0.8092 | 7239.3 | 0.9174 | 0.9166 | 0.4782 | 0.9985 | 0.9985 | 0.4446 |
| China     | 0.5945 | 0.5915 | 5415.5 | 0.7422 | 0.7403 | 0.4205 | 0.9747 | 0.9737 | 0.2781 |
taken since the start of the pandemic COVID-19 till mid-May 2020. The data are fitted with the 4th-degree polynomial regression model and found significant except the intercept as in the case of Pakistan. Similar conditions for the selection of the final model are imposed to get the desired model, i.e., \( \hat{Y} = 267.30 - 91.723X + 7.732X^2 - 0.230X^3 + 0.002X^4 \). Here an estimate of \( \beta_0 \) which is 267.30 explains that a total number of cases of COVID-19 patients without an impact of the time periods are initially around 267. Similarly, estimated values of the rest of the regression coefficients, i.e., \( \beta_1 \) through \( \beta_4 \) are reported (≈91.723, 7.732, 0.230, and 0.002) respectively, explain the impact of each day as linear, quadratic, cubic, and quadrant in the additive model. In addition, some of the factors have additive or negative impacts on the composite output (prediction or forecast). Now one can use this fitted model to get the desired forecast about the total number of COVID-19 cases for the upcoming days or period. India's forecasted data is depicting the continuous and significant rise in upcoming days similarly, data on the total number of deaths that occurred in patients of the COVID-19 pandemic is analyzed and results are presented in Table 2. An appropriate model fitted to get a forecast about the total number of deaths is as follows: \( \hat{Y} = 3.279 - 2.013X + 0.217X^2 - 0.007X^3 \). Here is an estimate of \( \beta_0 \) which is 3.279. This shows that initially almost 3 deaths were reported due to COVID-19 irrespective of the time period. Similarly, the estimated values of the rest of the regression coefficients, i.e., \( \beta_1 \) through \( \beta_3 \) are reported (≈2.013, 0.217, and 0.007) respectively, explain the impact because each successive day is linear, quadratic, and cubic in the composite model. Furthermore, additive/negative impacts on various factors are observed on the response.

2.3.3 | Afghanistan

Afghanistan is another neighboring country of Pakistan on the west and has a serious impact on the spread of the pandemic in Pakistan, if not get controlled. That's why to have an eye on its ground realities, it is selected for the comparison. An appropriate fitted model from the data about the total number of cases is \( \hat{Y} = 50.838 - 18.875X + 1.612X^2 - 0.041X^3 + 0.001X^4 - 2.651.252X^5 \). Intercept is found insignificant while the remaining three coefficients are found significant with \( p < 0.01 \). The second model is fitted on the total number of deaths that occurred due to COVID-19. This fitted model is described in mathematical form as: \( \hat{Y} = 0.755 - 0.115X - 0.004X^2 + 0.001X^3 \). The first three coefficients are found insignificant while the fourth coefficient is observed significantly. From the fitted model, it is clear that there is a low impact of this pandemic on Afghanistan. The reason may be the unavailability of testing kits and other tools to judge the accurate number of affected ones.

2.3.4 | Iran

The facts and figures of Iran have been analyzed for the period of the emergence of COVID-19 till 15th May 2020 and significantly revealed drastic results through a suitable model whose results are tabulated in Table 1 for the cases and Table 2 for the deaths, to efficiently forecast the upcoming number of cases and the deaths due to the COVID-19. The estimated model \( \hat{Y} = 1069.996 - 109.041X + 5.231X^2 - 0.078X^3 + 0.000X^4 \). All coefficients are found significant. Similarly, a fitted model for the forecast of the number of deaths is obtained as: \( \hat{Y} = 383.983 - 114.987X + 7.117X^2 - 0.084X^3 + 0.000X^4 \). These models present higher values of the total number of cases and deaths reported at the initial stage as compared to Pakistan, India, and Afghanistan.

2.3.5 | Italy

Italy is included in the analysis because of the higher number of cases and deaths in Europe. The data is analyzed for the same span of time from the emergence of COVID-19 to Mid-May, 2020 and results are analyzed with a fitted polynomial regression model of degree 5 as: \( \hat{Y} = -11749.338 + 4165.197X - 341.070X^2 + 9.675X^3 - 0.098X^4 + 0.000X^5 \). Similarly, another model for the forecast of the total number of deaths in Italy is \( \hat{Y} = -1969.319 + 644.37X - 48.476X^2 + 1.265X^3 - 0.012X^4 + 0.000X^5 \). The above mentioned models present a vulnerable situation of Italy at the initial stage in the form of a larger number of cases as well as a total number of deaths in the country. This indicates that no preventive measures were taken after the outbreak of this pandemic in China. But the forecasts show that the increase in the total number of cases and deaths has been slowed down due to very strict measures taken by the government of Italy.

2.3.6 | China

It is included in the analysis because cases of COVID-19 appeared first time in China. The data is analyzed for the same period of time from the start of COVID-19 to Mid-May, 2020 and results are analyzed with a fitted polynomial regression model of degree 5 as: \( \hat{Y} = 19561.665 - 5063.443X + 272.33X^2 - 4.385X^3 + 0.03X^4 + 0.000X^5 \). Similarly, another model for the forecast of the total number of deaths in China is \( \hat{Y} = 790.834 - 186.224X + 9.295X^2 - 0.147X^3 + 0.001X^4 + 0.000X^5 \). The above mentioned models present a vulnerable situation of China at the initial stage in the form of a larger number of cases (around 19562) as well as a total number of deaths (around 791) independent of the time period in the country. This shows how rapidly Wuhan city of China was affected by this pandemic. The forecasts show that the increase in the total number of cases and deaths has been almost stopped due to the very strict lockdown by the government of China.

In Tables 1 and 2, a declining trend in coefficients estimated value shows that there is a significant effect of all preventive and the treatment measures taken during lockdown conditions to overcome the outbreak of pandemic virus COVID-19, proved a success story in this regard.
3 | INTERPRETATION OF FITTED MODEL TRENDS

Figures 4 and 5 show the impact of prevalence with the number of cases and the deaths per day for all respective periods mentioned along the x-axis. The y-axis represents the frequency of the counts with the blue color for the observed dataset and the red color for the estimated best-fitted trend of respective polynomial regression models of the appropriate degrees. Each adjusted $R^2$ value is greater than 0.99, revealing the accuracy and goodness-of-fit of the model for each country respectively with the appropriate degree to forecast. Interpretation of the derived results, about the prevalence of cases per day, is discussed above in detail subject to the parametric behaviors of the polynomial regression models; which is also shown in Tables 1–2 and Figures 6–7 the coefficients graphical understanding. The data is analyzed using the R language.

3.1 | Forecast of total number of cases and deaths

The forecasting of the number of cases and Deaths can be performed by using polynomial regression modeling.

Tables 3 and 4 present the projected total number of cases and deaths for the upcoming week from May 16–22, 2020. These forecasts are based on fitted polynomial regression models. Along with forecasts, the data within parenthesis gives 95% confidence interval limits. These limits give an overview that our forecasted values may lie somewhere between these limits with high confidence. Strict measures can slow down the increase in cases while the ease in
lockdown or other such measures may cause an increased opening of
the markets etc. The data is hereby arranged in ascending order of
prevalence forecast report in the respective countries. The least case
counts are a forecast in Afghanistan, followed by Pakistan, India,
China, Iran; and the highest cases are in Italy for the coming week
shown through Figures 8 and 9. Comparative Forecast Trend Analysis
Relative to China is hereby presented to forecast the proportionate
increase of case counts for all countries under study based on facts
and figures caused by the significant outbreak and the prevalence of
pandemic virus COVID-19, and are mentioned in Table 5 and
Figure 10. There is a drastically increasing trend of COVID-19 con-
firmed cases in Italy depicting an average proportion of 295.4%,
166.6% (Iran), and 127.7% (India) relative to China. But in Pakistan,
the number of cases emergence forecast on the average around
59.3% with a slow increase in comparison with Italy. However, a
slightly increasing trend of prevalent COVID-19 confirms cases in
Afghanistan are 9.0% as an average relative to China. In contrast, a
significant declining trend itself is occurring in China of 96.4% at a
base period of May 16, 2020. However, this declining impact may be
attributed to the significant effect of the preventive and treatment
measures taken during lockdown conditions to overcome the out-
break of pandemic virus COVID-19.

Similarly, a comparative Forecast Trend Analysis based on the
proportion of fatalities computed for each country relative to China is
presented in Table 6 and Figure 11. However, the analysis performed
in Table 2 and shown in Figure 5 discussed in detail in the previous
section, proved that there is a drastically significant outbreak with the
fatalities of pandemic COVID-19 forecast and can be viewed in Figures 5
and 11. The highest no. of fatalities/deaths (30,274–32,154) per day is expected to occur in Italy; which shows

**FIGURE 5** Fitted curve against the observed data with adjusted $R^2$ for total no of death
FIGURE 6  Behavior of regression coefficients of no of cases of COVID-19

FIGURE 7  Behavior of regression coefficients of no of deaths due to COVID-19
an increasing trend in fatalities/deaths in Italy with an average proportion of 682.9%, in Iran 148.6% and India 71.8%, in Pakistan 21.4% relative to China. However, a slightly increasing trend of prevalent deaths forecasts an average proportion of 3.6% in Afghanistan relative to China. In contrast, a slightly declining trend itself is occurring in China 98.1% on average with a base period of May 16, 2020. However, this declining impact may be attributed in favor of the significant effect of all preventive and treatment measures taken during lockdown conditions to overcome the outbreak of pandemic virus COVID-19.

The comparison of the proposed methodology with the existing Simple linear and Exponential regression model reveals that the polynomial regression model does better in terms of $R^2$, Adjusted $R^2$, and MAPE. The model significance was determined by F-statistic at $p < 0.05$ is considered as a significant model and only such models are considered for forecasting of no of cases and deaths.

### Table 3: COVID-19 average forecast cases of next week (95% confidence interval limits)

| May 2020 | Afghanistan | Pakistan | India | China | Iran | Italy |
|----------|-------------|----------|-------|-------|------|-------|
| 16       | 5906        | 39973    | 86610 | 82546 | 119277 | 226532 |
|          | (5819–5993) | (39590–40357) | (86005–87215) | (74452–90640) | (116625–121930) | (220921–232143) |
| 17       | 6277        | 42156    | 91172 | 81792 | 122789 | 228716 |
|          | (6175–6380) | (41706–42605) | (90495–91849) | (72609–90974) | (119552–126025) | (222116–235316) |
| 18       | 6667        | 44437    | 95909 | 80903 | 126727 | 231165 |
|          | (6547–6788) | (43912–44961) | (95154–96664) | (70511–91296) | (122807–130647) | (222437–238894) |
| 19       | 7077        | 46820    | 100825 | 79870 | 131135 | 233917 |
|          | (6937–7217) | (46212–47428) | (99985–101666) | (68143–91597) | (126426–135843) | (224915–242919) |
| 20       | 7507        | 49308    | 105926 | 78679 | 136056 | 237012 |
|          | (7345–7670) | (48607–50009) | (104993–106858) | (65489–11870) | (130447–141665) | (226584–247440) |
| 21       | 7959        | 51904    | 111214 | 77319 | 141538 | 240492 |
|          | (7771–8146) | (51100–52709) | (110182–112247) | (62531–92107) | (134910–148166) | (228478–252505) |
| 22       | 8432        | 54613    | 116696 | 75776 | 147630 | 244399 |
|          | (8217–8646) | (53696–55530) | (115556–117835) | (59252–92301) | (139855–155405) | (220631–258167) |

### Table 4: COVID-19 average forecast deaths next week (95% confidence interval limits)

| May 2020 | Afghanistan | Pakistan | India | China | Iran | Italy |
|----------|-------------|----------|-------|-------|------|-------|
| 16       | 148         | 860      | 2850  | 4741  | 6815 | 31214 |
|          | (145–151)   | (852–869) | (2819–2880) | (4319–5163) | (6681–6949) | (30274–32154) |
| 17       | 155         | 902      | 3001  | 4722  | 6848 | 31370 |
|          | (151–158)   | (892–911) | (2966–3035) | (4243–5201) | (6695–7002) | (30264–32476) |
| 18       | 162         | 945      | 3157  | 4697  | 6880 | 31535 |
|          | (158–166)   | (934–955) | (3119–3196) | (4155–5239) | (6705–7056) | (30240–32830) |
| 19       | 169         | 989      | 3320  | 4665  | 6911 | 31712 |
|          | (164–173)   | (977–1000) | (3277–3363) | (4053–5276) | (6711–7111) | (30203–33220) |
| 20       | 176         | 1034     | 3489  | 4625  | 6941 | 31904 |
|          | (171–181)   | (1021–1047) | (3441–3536) | (3937–5313) | (6714–7168) | (30157–33651) |
| 21       | 184         | 1081     | 3664  | 4578  | 6970 | 32115 |
|          | (178–189)   | (1067–1095) | (3611–3716) | (3806–5349) | (6714–7226) | (30102–34128) |
| 22       | 192         | 1129     | 3845  | 4522  | 6998 | 32350 |
|          | (186–198)   | (1114–1145) | (3787–3902) | (3660–5384) | (6710–7285) | (30043–34656) |
3.2 Forecast of death counts per COVID-19 cases for the next week (%age)

Conclusive data analysis of forecasts for each country in the upcoming week is performed with the death rates per COVID-19 cases. The average death rate among COVID-19 cases in Pakistan is 2.1% which is the least one amongst the others such as Afghanistan has 2.4%, India has 3.3%, Iran has 5.3%, China 5.8%, and Italy attains the highest expected Forecast for the deaths is 13.5%. The results are mentioned in the percentage in Table 7 for ease of understanding. On the other hand, it can also be seen that whether the expected number of cases is greater in Pakistan but still Table 7 depicts the decreasing percentage from 2.2% to 2.1% which can be taken as an overwhelming for the effectiveness of the preventive measures to control the epidemic. Iran, Italy, and Afghanistan also show a mild decreasing trend but the number of cases is far higher than the decrease percentage. However, India is expected to have a consistent result, more or less it depicts some other biasing factors which should be figured out in separate research.

| May 2020 | Afghanistan | Pakistan | India | China | Iran | Italy |
|----------|-------------|----------|-------|-------|------|-------|
| 16       | 7.2%        | 48.4%    | 104.9%| 100.0%| 144.5%| 274.4%|
| 17       | 7.7%        | 51.5%    | 111.5%| 99.1%  | 150.1%| 279.6%|
| 18       | 8.2%        | 54.9%    | 118.5%| 98.0%  | 156.6%| 285.7%|
| 19       | 8.9%        | 58.6%    | 126.2%| 96.8%  | 164.2%| 292.9%|
| 20       | 9.5%        | 62.7%    | 134.6%| 95.3%  | 172.9%| 301.2%|
| 21       | 10.3%       | 67.1%    | 143.8%| 93.7%  | 183.1%| 311.0%|
| 22       | 11.1%       | 72.1%    | 154.0%| 91.8%  | 194.8%| 322.5%|
| Average  | 9.0%        | 59.3%    | 127.7%| 96.4%  | 166.6%| 295.4%|
4 | CONCLUSIONS

Polynomial regression of appropriate degree found a best-fitted forecast model approach for prediction of no effective cases and no of deaths in case of any pandemic, as here for COVID-19 prevalent cases as well as fatalities/deaths for each country, respectively. The value of adjusted $R^2 > 0.99$, can be described as a benchmark for the accuracy and goodness-of-fit of the fitted or proposed model for each country pandemic forecast. The average death rate per COVID-19 prevalent cases in Pakistan is at the least 2.1%, following Afghanistan (2.4%), India (3.3%), Iran (5.3%), China (5.8%), and the highest death rate is 13.5% in Italy. A slightly declining trend in deaths among COVID-19 cases is occurring in China 98.1% itself with a base period of May 16, 2020, showing a decline in forecasted fatalities of 4741 with 95% confidence limits as ($L = 4319$, $U = 5163$) but still with a 4% increase. The number of fatalities/deaths in Italy is forecasted on May 16, 2020, is 31 214 with 95% confidence limits as ($L = 30274$, $U = 32154$); which shows an average proportion of 658.4% times relative to that in China. The respective datasets also show that the severity of the pandemic in Pakistan occurred after the months in comparison with other countries. As it can be easily seen in March for other countries, whereas in Pakistan it started at that time. This declining impact of analysis in China and Italy is attributed to the significant effect of the preventive and treatment measures taken during lockdown conditions to overcome the outbreak of pandemic virus COVID-19. So, it can be concluded to opt for such methodologies for prediction of no of cases and casualties, so that the sensible and early steps can be taken to avoid economic losses and human life can be protected by making wise decisions using modeling based upon data-driven methods. The limitation of the study is the data for the number of cases reported and the number of deaths is assumed random, the case reported count or death count may be influenced by the Government prevention steps, Society behaviors, Environment, Disease pattern and, another unexplored effect. More detailed investigation/studies could be carried out to see the influence of these factors on the counts of reported and death cases. The current study using neutrosophic statistics can be extended in future research.

ACKNOWLEDGMENTS
We are thankful to the editor and reviewers for their valuable suggestions to improve the quality of the paper. This work was supported by King Abdulaziz University.

CONFLICT OF INTERESTS
There are no conflict of interests.
AUTHOR CONTRIBUTIONS
Nasrullah Khan, Asma Arshad, Muhammad Azam, Ali Hussein AL-Marshadi, and Muhammad Aslam wrote the paper.

DATA AVAILABILITY STATEMENT
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID
Muhammad Aslam http://orcid.org/0000-0003-0644-1950

REFERENCES
1. Li Q. An outbreak of NCIP (2019-nCoV) infection in China—wuhan, Hubei province, 2019–2020. China CDC Weekly. 2020;2(5):79-80.
2. Rodríguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al. Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. Travel Med Infect Dis. 2020;34:101623.
3. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA. 2020;323(11):1061-1069.
4. COVID-19/Team R. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(12):343-346.
5. Organization WH. WHO Director-General’s remarks at the media briefing on 2019-nCoV on 11 February 2020. [Internet] World Health Organization; 2020.
6. CSG of the International, C.S.G. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5:1.
7. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet. 2020;395(10228):931-934.
8. Kraemer M, Yang CH, Gutiérrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020;368(6490):493-497.
9. Arenas A, Cota W, Gómez-Gardeñes J, et al. Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. medRxiv. 2020. doi:10.1101/2020.04.06.20054320
10. Maier BF, Brockmann D. Effective containment explains sub-exponential growth in recent confirmed COVID-19 cases in China. Science. 2020;368:742-746.
11. Pandey G. SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv preprint arXiv:2004.00958. 2020. doi:10.1101/2020.04.04.20049825
12. Tan G, Li X, Niu R, et al. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. arXiv preprint arXiv:2002.12298. 2020;1751:147200.
13. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill. 2020;25(10):2000199.
14. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lancet. 2020;395:1225-1228.
15. Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165-174.
16. Panwar VS, Uduman PS, Gómez-Aguilar J. Mathematical modeling of coronavirus disease COVID-19 dynamics using CF and ABC non-singular fractional derivatives. Chaos, Solitons Fractals. 2021;145:110757.
17. Hu M, Lin H, Wang J, et al. Risk of coronavirus disease 2019 transmission in train passengers: an epidemiological and modeling study. Clin Infect Dis. 2021;72(4):604-610.
18. Anzum R, Islam MZ. Mathematical modeling of coronavirus reproduction rate with policy and behavioral effects. medRxiv. 2021;16:20133330. doi:10.1101/2020.06.16.20133330
19. Li M-W, Wang YT, Geng J, Hong WC. Chaos cloud quantum bat hybrid optimization algorithm. Nonlinear Dynamics. 2021;103(1):1167-1193.
20. Zhang Z, Hong W-C. Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads. Knowledge-Based Systems. 2021;228:107297.
21. Ritchie H. Coronavirus Source Data. last accessed on, 2020.
22. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20:533-534.
23. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020;20:669-677.
24. Ruan S. Likelihood of survival of coronavirus disease 2019. Lancet Infect Dis. 2020;20:630-631.
25. Verity R, et al & Dighe, A.(2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis, 30243-30247.
26. Tang B, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. J Clin Med. 2020;9(2):462.
27. Aczel AD, Sounderpandian J. Complete business statistics. 1999: Irwin/McGraw Hill Boston, MA.
28. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. John wiley & sons; 2008.

How to cite this article: Khan N, Arshad A, Azam M, AL-Marshadi AH, Aslam M. Modelling and Forecasting the Total Number of Cases and Deaths Due to Pandemic. J Med Virol. 2022;94:1592-1605. doi:10.1002/jmv.27506