Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v31i2.4676

INTERPRETING THE FUTURE OF COVID-19 WITH STATISTICAL FORECASTING MODELS

Muhammad Faisal Fahim[1*], Tayyab Raza Fraz[2], Ali Asghar Mirjat[1], Anzila Meer[3], Muhammad Gulzada[1], Hasnain Pasha[4], Amjad Ali[1], Albert John[1]

[1*]Bahria University Health Sciences Campus Karachi, Sailor St Cantonment, Karachi 74400, Sindh, Pakistan.
[2]Department of Statistics, University of Karachi, Circular Road, Karachi, Sindh, Pakistan.
[3]Department of Otorhinolaryngology (ORL), Universiti sains Malaysia USM kubang kerian kelantanMalaysia.
[4]Department of Optometry, POB Eye Hospital Karachi, C-15 Munawar Chowrangi Road, Block 12 Gulistan-e-Johar, Karachi, Karachi City, Sindh 74500.

***Corresponding Author:** Muhammad Faisal Fahim
*Bahria University Health Sciences Campus Karachi, Sailor St Cantonment,Karachi 74400, Sindh, Pakistan, faisalfahim88@hotmail.com

Abstract
Background: Sars-Cov2 is a deadly virus effected millions of peoples globally. Time series forecasting helps us to identify and plan things properly related to any particular disease or viruses. This is daily data of Covid-19 and researcher intended to find out best statistical model.

Objective: To evaluate best statistical model which forecast covid-19 data related to new cases in subcontinents of Pakistan.

Methods: This was an analytical observational design with daily data of new cases of COVID-19 among sub-continents of Pakistan. Data was imported from world health organization website. In this study statistical models applied were AR, MA, ARIMA, SETAR model by using threshold regression, ARCH effect, Simple GARCH model and Component GARCH model. Forecasting models used were AIC, MAPE,MAE and RMSE. Eviews version 12.0 used for data analysis.

Results: A total of 1146 observations for each country were taken for analysis. AR and MA model observed that Azerbaijan was significant at (1,0,1) model with AIC= 14.55, SBIC=14.57, HQC=14.56 and adjusted R^2=0.911. Bangladesh was significant at (1,0,2) model with AIC=15.55, SBIC=15.57, HQC=15.56 and adjusted R^2=0.958. Similarly, China was significant at (1,0,2) model, India was significant at (1,0,1) model, Iran found significant at (1,0,2) model. However, Pakistan, Sri Lanka and Kazakhstan were statistically significant at (1,0,1) model respectively.

Conclusion: These comprehensive long-term results showed best forecast models among different statistical models like AR, MA, ARIMA, SETAR, GARCH and component GARCH models with statistically significant findings. ARIMA and GARCH models showed best fit among all models to forecast pandemic new cases. Due to a globally controlled environment WHO announced that after 8th May 2023 global health emergency was ended and no further cases was reported therefore, we

have reported the databefore the ending emergency by WHO.

Keywords: Statistical Models, Forecasting, ARIMA model, SETAR model, ARCH model, GARCH model, Sars-Cov2.

Introduction

From the literature, it is revealed that the statistical modeling techniques had an effective and efficient role in the accurate predictions for any time series data. Researcher and forecasters can draw unbiased and consistent results based on different statistical models to identify the increasing or decreasing trend for any health science, medical, macro or microeconomic indicator [1]. Statistical model plays a crucial role in forecasting time series data for retrospective data. ARIMA model is one of the most comprehensive models used in several fields for prediction and forecasting of daily data sets. ARIMA model can be best fitted in short-term predictions. This model is purely based on AIC, BIC and HAC criteria and then it will forecastdata to run prediction estimates [2].

Auto-Regressive Integrated Moving Average is a statistical model that exhibits time series-based forecasts from previous data. The ARIMA time series model has the characteristics of strong short-term predictability and simplicity of previously coded data. ARIMA model is widely used in the forecast of various infectious diseases [3]. Time series forecasting plays a crucial role in forecasting health-related data sets and other fields. The time series forecasting of daily data can predict future values to understand the pattern and growth for any data set. ARIMA is a trending statistical model that draw conclusions taking linear assumptions in real world system [4]. Another literature reported the use of the ARIMA model in the pandemic to forecast future values. This model can deal with daily data sets in complex scenarios [5]. ARIMA is one of the common statistical forecasting models [6]. It followed previous data to forecast newly predicted values. This model was previously used to detect the best screening tool in terms of the prediction of various infectious disease including malaria, influenza and COVID-19 cases [7, 8].

Furthermore, after getting some predictions policies and precautions against covid-19 data. Research was conducted to predict pattern of covid-19 data by making novel space-time epidemic model framework by spatial statistical modeling [9]. They observed both long and short term forecast of infected people and death counts. Research showed that covid-19 is an infectious spreading disease and the biggest public health issue worldwide [10]. Statisticians thinks that there is a need to established some statistical models to describe disease pattern and find out statistical models which gives best forecast [10]. Study reported different statistical model such as Markov chain Monte Carlo (MCMC) stochastic process to find out disease transmissibility of covid-19 using logistic model [11]. A study conducted in China reported incidence of decreasing rate of covid-19using exponential adjustment model [12].

Regarding exponential growth model to get data driven technique in early phase of outbreak [13]. Research published Non-linear regression model applied to know short term forecast of cumulative number of covid-19 patients. For the prevention and evaluation measures and strategies against covid-19 statistical model forecast is necessary [14]. A study published on applied weibull distribution in respect to daily new and death cases of covid-19 to predict evolution pattern of SARS-Cov2. They stated that Weibull distribution used to assess lifetime models of time events [15]. Study reported new class of statistical model termed as new flexible extended–X (NFE-X) class of distributions. They also stated that this model without adding additional parameters results in avoiding rescaling problems. Secondly, next model was new flexible extended Weibull distribution. This new distribution investigates through graphical behavior by using density function [16]. Furthermore, it is reported distribution theory and introduced new statistical model to provide best fit of data related to COVID and other related events. They used modified Weibull distribution, which elaborated fitting power of exponential, Rayleigh linear failure rate. Globally different time

series model used to identify data gathering process [17].
In Pakistan, different articles have been published to know the forecast related to SARS Cov-2 corona virus cases, deaths and recoveries. It is presented that one month forecast for the number of cases, deaths and recoveries via time series method of Auto-Regressive Integrated Moving Average (ARIMA) Model [18]. It is also reported that a forecast of ten days for confirmed cases, deaths and recoveries of COVID patients by using vector autoregressive model with 5 lags. They identified unknown parameters of model using ordinary least square method [19]. Rationale of this study is to make a big data analytic forecast of covid- 19 of Pakistan using statistical models. Literature did not support complete statistical forecast results in Pakistan. Present study aims to evaluate best statistical model which forecast covid-19 data related to newcases in subcontinents of Pakistan.

Materials and Methods

Data related new cases of Covid-19 was downloaded from World Health Organization (WHO) website from March 2020 to 30th March 2023. WHO website regularly update data since COVID starts in Pakistan (March 2020). We have selected sub-continents of Pakistan including Azerbaijan, Bangladesh, China, India, Iran, Kazakhstan, Pakistan and Sri Lanka. WHO allowed all researchers to download data access freely so, no ethical approval was required from any source. In this study statistical models applied were AR, MA, ARIMA, Setar model by using threshold regression, ARCH effect, Simple GARCH model and Component GARCH model. Forecasting models used were AIC, MAPE, MAE and RMSE. **Autoregressive Model (AR)**

It is most common time series model in linear models. AR model can be observed easily from ordinaryleast square regression. An autoregressive model can be defined as,

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t \quad \ldots\ldots 1$$

Where "c" is an intercept parameter and ε_t is white noise random error with mean zero and variance 1. Y_t and Y_{t-1} are multiple regression lagged values where $\phi_1 Y_{t-1}$ is depending on last period values of Y_t.

Moving Average Model (MA)

In AR models we used past values to forecast in regression. Moving average model uses past past forecasterrors in regression model.
MA of order q is defined as,

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad \ldots\ldots 2$$

Where, "c" is an intercept parameter and ε_t is white noise random error with mean zero and variance 1. $\theta_1, \theta_2, \theta_3 \ldots$ are unknown parameters.

Autoregressive Moving Average (ARMA) Model

This ARMA model is the combination of AR and MA models. We can see the impact of previous lags along with residuals assumed for forecasting future values. There are two coefficients in ARMA model equation α (represent MA model) and β (represent AR model).

$$Y_t = \beta_1 * Y_{t-1} + \alpha_1 * \varepsilon_{t-1} + \beta_2 * Y_{t-2} + \alpha_2 * \varepsilon_{t-2} + \beta_3 * Y_{t-3} + \alpha_3 * \varepsilon_{t-3} + \ldots\ldots + \beta_k * Y_{t-k} + \alpha_k * \varepsilon_{t-k} \quad \ldots\ldots 3$$

Autoregressive Integrated Moving Average (ARIMA) Model & its Equation

ARIMA is a statistical model which analyze and forecast time series data. It also tells us to see and evaluate events that happened over a period of time. ARIMA model understand past data to predict

future data sense.It has three parameters “p” “d” & “q”.
Where, “p” refers to Auto Regressive (AR) terms “d” refers to non-seasonal differences which help us to make data stationary “q” refers to forecast terms i.e. Moving Averages (MA) The ARIMA (p,d,q) model equation is as follows, Yt= ϕ1 yt-1 + ϕ2 yt-2 + ………+ ϕp yt-p +ε1 − θ1εt-1 − θ2εt-2 — θqεt-q − εq

In above equation,
ϕ = *represents* AR terms
ε = *represents error terms*
θ = *represents Moving Average terms*
Yt = *represents d ordered difference in time series*

Model Accuracy Criteria
Akaike Information Criteria (AIC):
AIC is used to select a statistical model and relative quality of models with scoring method. It is an estimatorof out sample prediction error. Model selection equation is as follows,

$$AIC = -2ln\,(L) + 2k$$

Where,
L= Likelihood
k= number of parameters

Bayesian Information Criterion (BIC):
It is also used for scoring and selecting statistical model. It works under framework of maximum likelihoodestimator. The BIC statistics can be derived from this equation,

$$BIC = -2ln\,(L) + log\,log\,(N) * k$$

Where, L=Likelihood
N=number of examples in the training datasetk=number of parameters in the model

Forecast Evaluation Criteria:
Mean Absolute Percentage Error (MAPE):
The MAPE criteria is used to measure forecast accuracy. It is defined as sum of individual absolute errors (et) divided by the demand (d). we can say it is the mean of percentage errors. MAPE can be written in theform of equation as,

$$MAPE = \frac{1}{n}\sum \frac{|e|}{dt}$$

Mean Absolute Error (MAE):
MAE is used to forecast accuracy of statistical models. The value of absolute error is the absolute difference between forecasted and actual value. MAE gives an idea about expected error quantity from forecast on average.

$$MAE = \frac{1}{n}\sum |et|$$

Root Mean Squared Error (RMSE):
It is defined as the standard deviation of residuals i.e prediction errors. We can identify how far residualsare from the regression line. It is commonly used in forecasting of regression analysis.

$$RMSE = \sqrt{\frac{1}{n}\sum et^2}$$

Statistical Analysis:
Data was entered and analyzed in Eviews software version 12. Inferential statistics were reported. Initially Unit Root test was applied using Augmented Dickey Fuller (ADF) test to check stationary in the data. At first lag difference the data was found to have stationary. A normal line plot and a time series graph were also made. Statistical table was made to show all the Descriptive statistics. After that a corelogram was made to check significance of ACF and PACF lag differences. ARIMA models were run by using estimate equation least square method. P-value ≤ 0.05 considered to be statically significant. The given time series of different countries has to be examined first to determine whether or not they are stationary before applying the ARIMA model. The most effective way to use ARIMA models for time series forecasting is with stationary time series data. This data is found stationary. In this study, different Statistical Model Estimation and Forecast Methods were then applied to see the significance. P-value ≤ 0.05 considered to be statistically significant.

Results
In this data set researcher considered daily cases reported regarding COVID-19 among different sub continents of Pakistan from March 2020 to March 2023. Data was taken from World Health Organization (WHO) website which is updating on daily basis and can be freely accessible all over the world. This long duration data of three years was checked through different statistical model to predict best fit model. Statistical models used were Auto Regressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), SETAR model by using threshold regression, ARCH effect, Simple GARCH Model and Component GARCH Model. To see forecasting models initially used parameters were Akaike Information Criteria (AIC), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Descriptive statistics including mean, median, mode, skewness, kurtosis along with Jarque bera test among all countries. Median daily cases of Azerbaijan were 149 bangladesh was 528.5, china was 123, india was 12231, iran was 2561.5, kazakkhstan was 441.5, Pakistan was 592.5 and Sri Lanka was 57. Surprisingly, all countries showed significant probability (P-value< 0.0001) from Jarque bera test **Table 1**.

Unit root test of all countries in which Intercept, Trend and intercept both formats showed significant resultsP-value< 0.0001 **Table 2**.
Statistical model selection via AR and MA to assess best fit statistical model on daily cases of Covid-19. It was observed that Azerbaijan was significant at (1,0,1) model with AIC= 14.55, SBIC=14.57, HQC=14.56 and adjusted R^2=0.911. Bangladesh was significant at (1,0,2) model with AIC=15.55, SBIC=15.57, HQC=15.56 and adjusted R^2=0.958. Similarly, China was significant at (1,0,2) model, India was significant at (1,0,1) model, Iran found significant at (1,0,2) model. However, Pakistan, Sri Lanka and Kazakhstan were statistically significant at (1,0,1) model respectively **Table 3**.
SETAR model with threshold regression to assess best fit model. Azerbaijan fits its best model on (1,1,3) with AIC=14.57, SBIC=14.60, HQC=14.58 and adjusted R^2=0.908. Bangladesh fits best model on (1,1,2) with AIC=15.42, SBIC=15.45, HQC=15.43 and adjusted R^2=0.964. Similarly, China fitted its best model on (1,1,4), India best fitted model was (1,1,2), Pakistan best fitted model

observed was (1,1,3), Kazakhstan fit its best model on (1,1,2) and Sri Lanka fitted its best model on (1,1,4). **Forecasting** was also done after estimating SETAR model. In Azerbaijan, RMSE of (1,1,4) model was found 158.12, MAE was 137.61 and MAPE was 57.08 and concluded sufficient efficacy in terms of forecast. In Bangladesh, Competitive forecast accuracy observed on model (1,1,3) RMSE was 69.97, MAE was 51.12 and MAPE was 76.82. In China, best accurate model forecast observed was (1,1,3) with RMSE=107.16, MAE=751.56 and MAPE was 23.09. In India, best forecast model found was (1,1,2) with RMSE=242.74, MAE=184.55 and MAPE was 22.17. In Iran, best forecast model observed to be (1,1,4) with RMSE=374.15, MAE=298.94 and MAPE was 18.45. In Pakistan, best accurate model was (1,1,4) with RMSE=204.79, MAE=183.82 and MAPE was 10.24. In Kazakhstan, competitive model observed was (1,1,4) RMSE=117.78, MAE=89.27 and MAPE was 17.43. In Sri Lanka, best model accuracy observed in model (1,1,3) with RMSE=73.74, MAE=69.16 and MAPE=28.68 **Table 4**.

ARCH effect via heteroscedasticity test. Observed probabilities in all countries showed statistically significant results ($P< 0.0001$). **Table 5**.

Simple GARCH model up to 4 lags. Interestingly, all countries followed the same model (-1) which was best fitted among all 4 lags. The values of AIC, SBIC, HQC were observed closer to each other. Similarly, forecasting accuracy model was also checked through simple GARCH model. Azerbaijan showed best accuracy in model (-3), Bangladesh observed best accuracy forecast in (-2) model, China showed best forecast in model (-3), India observed best forecast model in (-2), Iran found best forecast in model (-4), Pakistan showed best forecast model (-1), Kazakhstan showed competitive forecast model as (-3) and Sri Lanka observed best forecast model in (-4) in terms of RMSE, MAE and RMSE values respectively **Table6**.

Component GARCH model. Similarly, after simple GARCH it clearly predicted that component GARCH model showed similar results in best fitted model selection. Model (-1) was statistically significant among all countries while checking up to 3 lags. However, only Sri Lanka showed significant at (-2) and fits its best model. On the contrary, forecasting accuracy model was also observed via Component GARCH model. Azerbaijan observed best forecast in model (-1), Bangladesh found best forecast in model (-2). China found best accuracy model in (-2), India showed best forecast model in (-1), Iran showed best forecast in model (-1), Kazakhstan found best forecast in model (-1), Pakistan showed best forecast model in model (-1) and Sri Lanka showed best forecast in model (-2) by predicting values related to RMSE, MAE and RMSE values respectively **Table 7**.

Table 1: Basic and Inferential Statistics observed among different countries

	AZERBAIJAN	BANGLADESH	CHINA	INDIA	IRAN	KAZAKHSTAN	PAKISTAN	SRI LANKA
Mean	687.7586	1679.234	81819.47	37023.80	6274.436	1238.012	1310.540	553.7784
Median	149.0000	528.5000	123.0000	12231.00	2561.500	441.5000	592.5000	57.00000
Maximum	7779.000	16230.00	6966046.	414188.0	50228.00	16442.00	8183.000	9962.000
Minimum	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Std. Dev.	1175.197	2811.892	543774.7	69995.55	8502.603	2137.613	1627.335	1020.062
Skewness	2.788357	2.856075	9.216465	3.288151	2.134191	3.394925	1.513452	3.601092
Kurtosis	12.20595	11.74784	94.23197	14.17803	7.668912	17.01870	4.717221	22.28183
Jarque-Bera	5860.043	5521.339	438205.7	8507.923	2024.235	12272.84	612.6143	21430.15
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	834939.0	2038590.	99328842	44946897	7617165.	1502947.	1590995.	672287.0
Sum Sq.Dev.	1.68E+09	9.59E+09	3.59E+14	5.94E+12	8.77E+10	5.54E+09	3.21E+09	1.26E+09
Observations	1146	1146	1146	1146	1146	1146	1146	1146

Table 2:- Result of unit root test of different countries

Countries	Augmented Dickey-Fuller test statistic(ADF)		Phillips-Perron Test Equation (PP)	
	(Intercept)	**Trend and Intercept**	**(Intercept)**	**Trend and Intercept**
Azerbaijan	-5.109 0.000	-5.046 0.000	-6.213 0.000	-6.222 0.000
Bangladesh	-4.118 0.000	-4.221 0.004	-4.035 0.001	-4.117 0.006
China	-5.221 0.000	-5.445 0.000	-5.308 0.000	-5.407 0.000
India	-3.956 0.001	-4.053 0.007	-3.470 0.009	-3.535 0.036
Iran	-3.378 0.011	-3.447 0.045	-4.489 0.000	-4.412 0.002
Kazakhstan	-5.888 0.000	-5.895 0.000	-3.878 0.002	-3.886 0.012
Pakistan	-3.644 0.005	-3.941 0.010	-3.373 0.012	-3.664 0.025
Sri Lanka	-3.608 0.005	-3.626 0.028	-10.859 0.000	-10.880 0.000

Table 3: Model selection by using AR and MA

Country	Model	Coefficient significant	AIC	SBIC	HQC	Adj. R2
Azerbaijan	(1,0,1)	YES	14.55	14.57	14.56	0.911
	(1,0,2)	YES	14.74	14.75	14.74	0.893
	(2,0,1)	YES	14.73	14.75	14.74	0.894
	(2,0,2)	YES	14.87	14.88	14.87	0.878
Bangladesh	(1,0,1)	YES	15.56	15.58	15.57	0.957
	(1,0,2)	YES	15.55	15.57	15.56	0.958
	(2,0,1)	YES	15.56	15.58	15.57	0.957
	(2,0,2)	YES	16.25	16.27	16.26	0.915
China	(1,0,1)	YES	24.96	24.97	24.96	0.986
	(1,0,2)	YES	25.16	25.17	25.16	0.983
	(2,0,1)	YES	25.5	25.51	25.5	0.976
	(2,0,2)	YES	26.26	26.28	26.27	0.949
India	(1,0,1)	YES	20.33	20.34	20.33	0.991
	(1,0,2)	YES	20.45	20.46	20.45	0.99
	(2,0,1)	YES	20.44	20.46	20.45	0.99
	(2,0,2)	YES	21.37	21.39	21.38	0.977
Iran	(1,0,1)	YES	17.87	17.88	17.87	0.953
	(1,0,2)	YES	17.84	17.86	17.85	0.954
	(2,0,1)	YES/NO 0.4902	17.84	17.86	17.85	0.954
	(2,0,2)	YES	18.42	18.43	18.42	0.919
Pakistan	(1,0,1)	YES	14.5	14.51	14.5	0.956
	(1,0,2)	YES	14.52	14.53	14.52	0.955
	(2,0,1)	YES	14.52	14.54	14.53	0.955
	(2,0,2)	YES	15	15.02	15.01	0.927
	(1,0,1)	YES	14.8	14.82	14.81	0.848

Sri Lanka	(1,0,2)	YES/NO 0.4086	15.24	15.25	15.24	0.767
	(2,0,1)	YES	15.16	15.18	15.17	0.782
	(2,0,2)	YES	14.92	14.94	14.93	0.83
Kazakhstan	(1,0,1)	YES	15.52	15.53	15.52	0.929
	(1,0,2)	YES	15.56	15.57	15.56	0.927
	(2,0,1)	YES	15.55	15.57	15.56	0.927
	(2,0,2)	YES	15.98	16	15.99	0.888

Table 4: SETAR model by using threshold regression model

						Forecast Criteria		
Country	**Model**	**AIC**	**SBIC**	**HQC**	**Adj. R^2**	**RMSE**	**MAE**	**MAPE**
Azerbaijan	(1,1,2)	14.633	14.668	14.646	0.908	457.86	363.3	73.13
	(1,1,3)	14.570	14.605	14.583	0.913	145.56	126.70	60.69
	(1,1,4)	14.600	14.635	14.613	0.911	158.12	137.61	57.08
Bangladesh	(1,1,2)	15.423	15.457	15.435	0.964	265.43	209.21	75.73
	(1,1,3)	15.527	15.562	15.540	0.960	69.97	51.12	76.82
	(1,1,4)	15.529	15.564	15.542	0.960	555.15	467.08	75.52
China	(1,1,2)	24.335	24.370	24.348	0.993	822.98	582.14	38.40
	(1,1,3)	24.201	24.236	24.214	0.993	107.16	751.56	23.09
	(1,1,4)	24.129	24.165	24.143	0.994	153.15	280.44	37.43
India	(1,1,2)	20.387	20.422	20.400	0.991	242.74	184.55	22.17
	(1,1,3)	20.449	20.484	20.462	0.991	418.14	340.79	38.51
	(1,1,4)	20.389	20.425	20.403	0.991	254.72	186.07	22.64
Iran	(1,1,2)	17.830	17.866	17.844	0.956	751.65	636.74	14.15
	(1,1,3)	17.774	17.810	17.788	0.958	760.29	630.68	13.79
	(1,1,4)	17.751	17.786	17.764	0.959	374.15	298.94	18.45`
Pakistan	(1,1,2)	14.533	14.568	14.546	0.955	258.76	234.39	13.23
	(1,1,3)	14.526	14.561	14.539	0.956	203.01	181.66	15.66
	(1,1,4)	14.453	14.578	14.556	0.955	204.79	183.82	10.24
Kazakhstan	(1,1,2)	15.259	15.295	15.273	0.948	194.57	148.22	15.85
	(1,1,3)	15.333	15.369	15.347	0.944	126.57	95.30	18.66
	(1,1,4)	15.361	15.396	15.374	0.942	117.78	89.27	17.43
Sri Lanka	(1,1,2)	14.757	14.792	14.771	0.862	153.47	107.34	32.53
	(1,1,3)	14.725	14.761	14.739	0.866	73.74	69.16	28.68
	(1,1,4)	14.703	14.738	14.716	0.869	128.39	115.90	44.64

Table 5: ARCH effect

Azerbaijan

Heteroskedasticity Test: ARCH

F-statistic	3149.840	Prob. F(1,1143)	0.0000
Obs*R-squared	840.1354	Prob. Chi-Square(1)	0.0000

Bangladesh

Heteroskedasticity Test: ARCH

F-statistic	7468.929	Prob. F(1,1143)	0.0000
Obs*R-squared	993.0323	Prob. Chi-Square(1)	0.0000

China

Heteroskedasticity Test: ARCH

F-statistic	17402.83	Prob. F(1,1143)	0.0000
Obs*R-squared	1074.432	Prob. Chi-Square(1)	0.0000

India

Heteroskedasticity Test: ARCH

F-statistic	45594.23	Prob. F(1,1143)	0.0000
Obs*R-squared	1116.998	Prob. Chi-Square(1)	0.0000

Iran

Heteroskedasticity Test: ARCH

F-statistic	6193.562	Prob. F(1,1143)	0.0000
Obs*R-squared	966.6147	Prob. Chi-Square(1)	0.0000

Kazakhstan

Heteroskedasticity Test: ARCH

F-statistic	7046.104	Prob. F(1,1143)	0.0000
Obs*R-squared	985.1858	Prob. Chi-Square(1)	0.0000

Pakistan

Heteroskedasticity Test: ARCH

F-statistic	6218.509	Prob. F(1,1143)	0.0000
Obs*R-squared	967.2192	Prob. Chi-Square(1)	0.0000

Sri Lanka

Heteroskedasticity Test: ARCH

F-statistic	225.6512	Prob. F(1,1143)	0.0000
Obs*R-squared	188.7776	Prob. Chi-Square(1)	0.0000

Table 6: Simple GARCH model:

Country	Model	AIC	SBIC	HQC	Adj. R^2	RMSE	MAE	MAPE
Azerbaijan	-1	11.857	11.879	11.866	0.889	29.03	24.84	36.41
	-2	12.060	12.085	12.071	0.862	29.13	23.20	37.85
	-3	12.155	12.177	12.163	0.858	28.22	21.88	29.05
	-4	12.253	12.275	12.261	0.812	35.87	30.64	19.53
Bangladesh	-1	12.217	12.239	12.226	0.956	3.83	2.59	32.62
	-2	12.677	12.699	12.686	0.911	3.03	2.36	40.25
	-3	12.815	12.837	12.824	0.875	3.28	2.32	33.37
	-4	12.970	12.992	12.978	0.845	3.25	2.39	37.24

China	-1	22.184	22.206	22.193	0.970	6420.19	6033.40	68.46
	-2	22.252	22.274	22.261	0.891	1869.89	1775.21	85.03
	-3	14.838	14.860	14.846	0.737	375.26	287.86	81.69
	-4	22.403	22.426	22.412	0.700	383.34	293.72	82.24
India	-1	17.411	17.433	17.420	0.990	5551.53	4777.58	55.40
	-2	18.121	18.143	18.130	0.975	5543.50	4772.62	55.36
	-3	18.357	18.379	18.365	0.960	6046.86	5303.88	63.30
	-4	18.506	18.528	18.514	0.940	6456.71	5756.59	69.99
Iran	-1	15.215	15.237	15.223	0.951	487.18	406.87	87.29
	-2	15.740	15.762	15.748	0.907	481.44	401.56	86.08
	-3	16.003	16.025	16.011	0.878	487.56	402.75	86.54
	-4	16.150	16.172	16.158	0.863	462.20	385.66	81.87
Pakistan	-1	13.224	13.246	13.232	0.953	342.55	40.66	42.33
	-2	13.566	13.588	13.575	0.921	358.61	55.25	47.56
	-3	13.736	13.758	13.744	0.895	396.45	60.24	48.95
	-4	13.823	13.846	13.832	0.872	388.21	70.52	50.36
Kazakhstan	-1	13.084	13.106	13.092	0.913	452.36	135.94	62.81
	-2	13.437	13.459	13.445	0.865	411.64	149.54	69.32
	-3	13.565	13.588	13.574	0.835	410.59	112.56	58.21
	-4	13.708	13.730	13.716	0.796	452.63	156.97	70.55
Sri Lanka	-1	11.402	11.424	11.410	0.756	2.47	1.96	66.52
	-2	11.624	11.655	11.636	0.748	2.96	2.55	33.13
	-3	11.916	11.947	11.928	0.756	2.30	1.76	64.62
	-4	11.885	11.916	11.896	0.696	2.05	1.46	50.81

Table 7: Component GARCH Model:

Country	Model	AIC	SBIC	HQC	Adj. R2	RMSE	MAE	MAPE
Azerbaijan	(-1)	11.939	11.970	11.951	0.877	38.35	31.90	41.34
	(-2)	12.072	12.103	12.084	0.862	39.11	31.66	42.55
	(-3)	12.188	12.219	12.199	0.856	39.35	32.90	40.15
Banglades h	(-1)	12.254	12.285	12.265	0.956	3.37	2.55	38.62
	(-2)	12.671	12.702	12.683	0.911	3.21	2.63	45.18
	(-3)	12.825	12.856	12.837	0.874	3.51	2.66	42.96
China	(-1)	15.629	15.660	15.641	0.338	334.77	248.32	70.89
	(-2)	16.235	16.266	16.246	0.243	319.62	218.65	54.15
	(-3)	17.315	17.346	17.327	-0.530	333.39	244.21	68.80
India	(-1)	17.456	17.487	17.467	0.990	1421.57	1102.84	14.61
	(-2)	18.171	18.202	18.183	0.975	2232.05	1776.29	23.10
	(-3)	20.206	20.237	20.218	0.960	2569.04	1973.62	25.24
Iran	(-1)	15.263	15.294	15.274	0.951	371.07	217.69	30.14
	(-2)	15.782	15.813	15.794	0.907	411.12	282.25	43.66
	(-3)	16.051	16.046	16.027	0.878	426.91	297.83	47.00
Kazakhsta n	(-1)	12.971	13.002	12.983	0.925	24.91	20.54	40.07
	(-2)	13.373	13.404	13.384	0.881	29.83	24.34	53.95
	(-3)	13.508	13.539	13.520	0.845	29.26	24.48	52.56
Pakistan	(-1)	13.088	13.119	13.100	0.953	15.26	9.96	33.71
	(-2)	13.574	13.605	13.586	0.920	22.12	16.97	61.31
	(-3)	13.737	13.768	13.748	0.895	25.88	20.25	70.41

Sri Lanka	(-1)	11.659	11.690	11.671	0.757	2.96	2.55	33.15
	(-2)	11.624	11.655	11.636	0.748	2.30	1.76	64.62
	(-3)	11.916	11.947	11.928	0.756	2.47	1.96	66.52

Discussion:

This research findings provide valuable insights into statistical modeling and prediction of COVID-19 cases in different subcontinents of Pakistan over a three-year period from March 2020 to March 2023. Various statistical models such as AR, MA, ARIMA, SETAR model, ARCH effect, simple GARCH model, and component GARCH model were used, combined with comprehensive evaluation indicators such as AIC, MAPE, MAE and RMSE to enhance the rigor of the analysis.

Comparing these results with similar studies reveals concordance and disagreement [20, 21]. Studies conducted in different countries have used similar statistical methods [22, 23], but may yield different parameter estimates due to changes in underlying data patterns and epidemiological dynamics. For example, while both studies may identify ARIMA models as suitable for forecasting, the specific order of the models may vary depending on the unique characteristics of the respective datasets [22, 23].

Similarly, findings about heteroskedasticity ARCH effect and GARCH modeling show commonalities across studies [24, 25]. The significant presence of the ARCH effect in all countries highlights the volatility and clustering of COVID-19 cases over time, highlighting the importance of employing robust statistical techniques to accurately capture such dynamics [25]. In addition, the consistency of the GARCH model in selecting the optimal lagging structure suggests that there is a common pattern of persistent fluctuations inglobal COVID-19 case data [26].

However, differences may occur when comparing the forecasting accuracy and model selection criteria of different studies. While some of the countries in this study have demonstrated superior forecast performance under specific models [5, 27], similar studies in other regions may find alternative models to be more effective [28, 29].

These differences may stem from differences in data quality, healthcare infrastructure, public health interventions, or socioeconomic factors that influence the dynamics of disease transmission. In addition, the application of threshold regression SETAR model has revealed insights into the potential nonlinearity and regime shift of COVID-19 across the subcontinent [30, 31]. While the SETAR model provides a flexible framework to capture this complexity, its effectiveness may vary depending on the underlying epidemiological pattern and the presence of different transmission mechanisms. Contrasting these findings with studies utilizing alternative regime transition models sheds light on the robustness andgenerality of the observed thresholds [32, 33].

Overall, while this study has made a significant contribution to understanding the spatiotemporal dynamics of COVID-19 transmission in Pakistan, combining its findings with similar studies from different geographical backgrounds has enriched our understanding of the multifaceted nature of the epidemic [23, 34, 35]. By synthesizing multiple research findings, policymakers and public health officials can gather actionable insights to inform evidence-based interventions and mitigate the impact of COVID-19 on globalhealth systems and society.

Similar literatures were published and reported the same findings but in very short time period. Either it was COVID era or in between the COVID era [36-38].

Conclusion:

This long-term analysis showed and proposed ARIMA as best statistical model among different statistical models related to new cases of COVID-19. ARIMA model can be used as effective theme to estimate and forecast predictions with respect to daily cases of COVID-19. Model accuracy can be checked through RMSE, MAE and MAPE because these tools also help to find accurate predictions. ARIMA and GARCH models showed best fit among all models to forecast pandemic new cases. Due to a globally controlled environment WHO announced that after 8th May 2023 global health emergency was ended and no further cases was reported therefore, we have reported the data before the ending emergency by WHO.

Acknowledgements

The authors would like to thank the WHO for allowing all researchers to download data access freely.

Conflict of interest

All authors declare no conflict of interest.

Funding

Authors has not obtained any financial support for this research.

References

1. WHO. *Coronavirus disease (COVID-19)*. 2024; Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1.
2. Katris, C., *A time series-based statistical approach for outbreak spread forecasting: Application ofCOVID-19 in Greece.* Expert Systems with Applications, 2021. **166**: p. 114077.
3. Zhihao, L., et al. *RESEARCH ON COVID-19 EPIDEMIC BASED ON ARIMA MODEL*. in *Journalof Physics: Conference Series*. 2021. IOP Publishing.
4. Al-Chalabi, H., Y.K. Al-Douri, and J. Lundberg. *Time Series Forecasting using ARIMA Model: A Case Study of Mining Face Drilling Rig*. in *12th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2018), Athens, Greece, November 18-22, 2018*. 2018. International Academy, Research and Industry Association (IARIA).
5. Ospina, R., et al., *An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil.* Mathematics, 2023. **11**(14): p. 3069.
6. Anwar, M.Y., et al., *Time series analysis of malaria in Afghanistan: using ARIMA models to predictfuture trends in incidence.* Malaria journal, 2016. **15**(1): p. 1-10.
7. He, Z. and H. Tao, *Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study.* International Journal of Infectious Diseases, 2018. **74**: p. 61-70.
8. Shoko, C. and P. Njuho, *ARIMA model in predicting of COVID-19 epidemic for the Southern Africaregion.* African Journal of Infectious Diseases, 2023. **17**(1): p. 1-9.
9. Kim, M., et al., *Methods, challenges, and practical issues of COVID-19 projection: A data scienceperspective.* Journal of Data Science, 2021. **19**(2).
10. Ghanim Al-Ani, B., *Statistical modeling of the novel COVID-19 epidemic in Iraq.* Epidemiologic Methods, 2021. **10**(s1): p. 20200025.
11. Shen, C.Y., *Logistic growth modelling of COVID-19 proliferation in China and its international implications.* International Journal of Infectious Diseases, 2020. **96**: p. 582-589.
12. Majumder, M.S. and K.D. Mandl, *Early transmissibility assessment of a novel coronavirus in Wuhan, China.* Social Science Research Network, 2020.
13. Zhao, S. and H. Chen, *Modeling the epidemic dynamics and control of COVID-19 outbreak in China.* Quantitative biology, 2020. **8**: p. 11-19.
14. Roosa, K., et al., *Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020.* Journal of clinical medicine, 2020. **9**(2): p. 596.
15. Moreau, V.H., *Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths.* Brazilian Journal of

Microbiology, 2020. **51**(3): p. 1109-1115.
16. Zuo, M., et al., *Comparison of COVID-19 pandemic dynamics in Asian countries with statistical modeling.* Computational and mathematical methods in medicine, 2020. **2020**.
17. Liu, X., et al., *Modeling the survival times of the COVID-19 patients with a new statistical model:A case study from China.* PloS one, 2021. **16**(7): p. e0254999.
18. Yousaf, M., et al., *Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan.* Chaos, Solitons & Fractals, 2020. **138**: p. 109926.
19. Khan, F., A. Saeed, and S. Ali, *Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan.* Chaos, solitons & fractals, 2020. **140**: p. 110189.
20. Mansour, M.M., et al., *Modeling the COVID-19 pandemic dynamics in Egypt and Saudi Arabia.*Mathematics, 2021. **9**(8): p. 827.
21. Konane, V.F. and A. Traore, *Statistical modeling and forecast of the corona-virus disease (COVID- 19) in Burkina Faso.* International Journal of Statistics and Probability, 2020. **9**(6): p. 76.
22. Sharma, V.K. and U. Nigam, *Modeling and forecasting of COVID-19 growth curve in India.*Transactions of the Indian National Academy of Engineering, 2020. **5**(4): p. 697-710.
23. Hwang, E. and S. Yu, *Modeling and forecasting the COVID-19 pandemic with heterogeneous autoregression approaches: South Korea.* Results in Physics, 2021. **29**: p. 104631.
24. Khan, M., et al., *COVID-19 Pandemic & Financial Market Volatility; Evidence from GARCH Models.* Journal of Risk and Financial Management, 2023. **16**(1): p. 50.
25. Elhini, M. and R. Hammam, *The impact of COVID-19 on the standard & poor 500 index sectors: A multivariate generalized autoregressive conditional heteroscedasticity model.* Journal of Chinese Economic and Foreign Trade Studies, 2021. **14**(1): p. 18-43.
26. Adenomon, M.O., B. Maijamaa, and D.O. John, *The effects of Covid-19 outbreak on the Nigerian Stock Exchange performance: Evidence from GARCH Models.* Journal of Statistical Modeling & Analytics(JOSMA), 2022. **4**(1).
27. Zeroual, A., et al., *Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study.* Chaos, solitons & fractals, 2020. **140**: p. 110121.
28. Dairi, A., et al., *Comparative study of machine learning methods for COVID-19 transmission forecasting.* Journal of Biomedical Informatics, 2021. **118**: p. 103791.
29. Martin-Moreno, J.M., et al., *Predictive models for forecasting public health scenarios: practical experiences applied during the first wave of the COVID-19 pandemic.* International Journal of Environmental Research and Public Health, 2022. **19**(9): p. 5546.
30. Morshed, N. and M.R. Hossain, *Causality analysis of the determinants of FDI in Bangladesh: fresh evidence from VAR, VECM and Granger causality approach.* SN business & economics, 2022. **2**(7): p. 64.
31. Jayadeva, S., *Systems in the subcontinent: Data, power, and the ethics of medical machine learningin India.* 2021.
32. Arenas, A., et al., *Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions.* Physical Review X, 2020. **10**(4): p. 041055.
33. Tkachenko, A.V., et al., *Persistent heterogeneity not short-term overdispersion determines herd immunity to COVID-19.* medRxiv, 2020: p. 2020.07. 26.20162420.
34. Wang, X., et al., *The spatiotemporal evolution of COVID-19 in China and its impact on urban economic resilience.* China Economic Review, 2022. **74**: p. 101806.
35. Alabdulrazzaq, H., et al., *On the accuracy of ARIMA based prediction of COVID-19 spread.* Resultsin Physics, 2021. **27**: p. 104509.
36. Khan, F.M. and R. Gupta, *ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India.* Journal of Safety Science and Resilience, 2020. **1**(1): p. 12-18.
37. Ding, G., et al., *Brief Analysis of the ARIMA model on the COVID-19 in Italy.* medRxiv, 2020: p. 2020.04. 08.20058636.

38. Ala'raj, M., M. Majdalawieh, and N. Nizamuddin, *Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections.* Infectious Disease Modelling, 2021. **6**:p. 98-111.