Indian Journal of Economics and Business
Vol. 20 No. 4 (December, 2021)

http://www.ashwinanokha.com/IJEB.php

Time series Model selection via stepwise regression to predict GDP Growth of Pakistan

Nasim Akhter[1]Arshad Mahmood Malik[2], Fouzia Jamshaid*[3], Maryam Yasir[4], Aneeda Ayub[4]& Nigah Hussain[2]

1.Assistant Professor, Department of Economics, Mirpur University of Science and Technology, Mirpur AJK, Pakistan.Email: nasim.eco@must.edu.pk

2. Department of Economics PMAS-Arid Agriculture University Rawalpindi, Pakistan

3. Assistant Professor Department of Economics, Allama Iqbal Open University Islamabad. Email: fouzia.jamshaid@aiou.edu.pk*

4. Lecturer, Department of Economics, Mirpur University of Science and Technology, Mirpur AJK, Pakistan.

Received July 15, 2021
Revised August 30, 2021
Accepted September 15, 2021
Published December 15, 2021

Abstract: Agriculture, Industry and Services are the major sectors of Pakistan economy having fourteen sub-sectors. The analysis of sectoral development has an important underpinning in the policy making. GDP growth is used to measure the growth of an economy and higher GDP growth indicates the vigorous growth of economy. GDP growth is influenced by the growth of sub-sectors which are often correlated with each other and may create a problem in building appropriate forecasting model. Therefore, selection of the appropriate predictor variables is important for accurate prediction of future GDP growth. The objective of this study was to attain an appropriate time series model to predict the GDP growth of Pakistan economy. To accomplish the study time series data ranging 1950 to 2020 were used. Stationarity analysis was carried out by applying Augmented Dickey- Fuller and Philips - Perron unit root tests. Correlation analysis was made to find positive and negative correlations among study variables. Stepwise regression, forward selection and backward elimination was applied on fourteen sub-sectors as independent variables and GDP as dependent variable. The model was selected with nine significant variables and the selected model was statistically evaluated through some diagnostic tests to check the presence of multicollinearity, autocorrelation, heteroscedasticity, and stability of the model.

Key words: Economy, Agriculture, Industry, Services, Gross Domestic Product, time series, stationarity analysis, Augmented Dickey- Fuller, Philips - Perron. Correlation, Stepwise regression, forward selection, backward elimination, multicollinearity, autocorrelation, heteroscedasticity, stability, sectoral development, Policy making.

1. Introduction

Economically Pakistan faced an uneven journey and economy has experienced many ups and downs since independence. The economic history of Pakistan indicates that during different regimes no serious efforts have been made to enhance the economic growth. In the beginning, the economy faced internal political instability, assorted foreign investment and confrontation with India. During the first four decades after independence, the economy's growth rate was 6 percent per annum. Although during this period economy faced a speedy population enlargement but its per capita income became twofold, the rise in prices and poverty turned down to 18 percent from 46 percent by late 1980. This vigorous economic appearance was upheld all the way through consecutive civilian and military commands during 1950s, 1960s, 1970s and 1980s in anticipation of the decade of 1990s (Pakistan Economic Survey, various issues).

The economy of Pakistan has three key sectors namely primary sector /agriculture sector, secondary sector/industrial sector and tertiary sector/services sector have been major contributors towards GDP growth in Pakistan economy. The primary sector/ agriculture having four sub-sectors, is the backbone of the economy and it contributes 19.41 percent to Gross Domestic Product (GDP) by employing 44 percent of labor force. The secondary sector/industry having four sub-sectors contributes 19.19.4 percent to Gross Domestic Product (GDP) and 20 percent of total labor force has been employed by industrial sector**.** The tertiary sector/services having six sub-sectors has emerged as a new growth powerhouse in Pakistan economy by contributing 61.39 in Gross Domestic Product and has a tremendous role for the economic development (GoP, 2020-21)). In this study the key variables are Gross Domestic Product, fourteen sub-sectors of agriculture, industry and services sectors in the perspective of economic growth.

The dynamics of economic growth have been emphasized after World War-II, when independent states emerged which were colonies earlier (Easterly, 2001).The aspiration of developing nations was to formulate such policies through which they can be in line with developed nations by accelerating the pace of economic growth. The change in Gross Domestic Product (GDP) in percentage form is a measure of economic growth. So, in the framework of observed fact an important question arises that what should be the intensity and frame of economic growth in the country and to improve the existing growth dynamics which factors should be influenced from policy perspective? The GDP growth of Pakistan economy is influenced by the growth of fourteen sub-sectors. By keeping in view this aspect present study is conducted to attain a suitable time series model through which the GDP growth of Pakistan economy can be predicted.GDP growth forecasting is a vital element in planning and sustainable development of a country and main objective of stepwise regression is to identify the correct predictor variables, which have an important influence on the response variable.

Numerous researches theoretically and empirically have found several reasons for GDP to the economy to grow (Wu et al., 2010; Hassan, et al., 2011; Lee, 2011). Economic growth can be attributed to efficiency of investment, the well-functioning financial system. An efficient financial system can be considered necessary but not sufficient condition for steady economic growth (Ang, 2008; Hassan, 2011). A large number of factors influencing GDP growth have been discussed in literature available but current study analyses via stepwise regression, the effect of sub-sectors to predict the GDP growth of Pakistan. By using same technique Chin and Fitrianto (2013) found the impact of private consumption, exports of goods and services and interest rate to the Malaysia's GDP. Mehmood (2012) acknowledged significant factors for Pakistan and Bangladesh's GDP and, Bobinaite et al. (2011) for Lithuaniaâs GDP. Akter et al. (2012) by considering seventeen factors investigated via factor analysis the momentous factors that contribute to Bangladesh's GDP.

2. Data

Annual time series secondary data ranging 1950 to 2020were used.

Table 1. Data Sources.

Particulars	Sources
1.A Hand Book of Statistics on Pakistan economy 2. Pakistan Economic Survey (Various Issues) 3.Agricultural statistics (various editions) 4.World Development Indicators	▪ State Bank of Pakistan ▪ Finance Division, GoP ▪ Government of Pakistan ▪ World Bank

3. Methods

The data obtained were in nominal terms which was converted into real terms by using GDP deflator, as the real GDP is a good indication of the economic development. Growth rates from real term data of GDP and fourteen sub sectors were calculated.

Table 2. Variable Description

Variable	Description	Abbr.
Gross Domestic Product (GDP)	The market value of all final goods and services which are produced by normal residents and non-residents of a country within the geographical territory of a country, during a specified period of time usually a year.	Y
Crops	Crops include rice, cotton, wheat, sugarcane and maize onion, potatoes, tomatoes, pulses, chilies fruit, vegetables, nursery plants, nuts, cereal growing, tobacco, barley, sesame, jawar, rapeseed, mustard, and gram.	X_1
Livestock	Rearing poultry, cattle, mules, poultry products, asses, buffaloes, goat, sheep, camels and dairy farming.	X_2
Fisheries	Production of fishing on commercial basis and the hunting of aquatic animals (sea-squirts, turtle, sea scamp and aquatic vertebrate).	X_3
Forestry	The provision of timber and fuel wood for domestic and commercial use, the conservation of wildlife habitats, conservation of water resources, management of bio-diversity, watershed management. the soil erosion control is part of forestry.	X_4
Mining and quarrying	It includes multiple activities as the extraction of all solids (coal and ores), liquids (crude petroleum) and gases (natural gases).	X_5
Manufacturing	Transformation of physical and chemical material into new product, it comprises of large scale, small-scale and slaughtering industry.	X_6
Construction	The process in which building, or infrastructure assembling is undertaken having three types' i-e building construction, heavy/civil construction and industrial construction.	X_7
Electricity and gas distribution	The production / generation of electricity, its collection than transmission and finally the distribution to final consumer, whether the use is for domestic or commercial.	X_8
Transport, storage and communication	Transportation means are roads, railway, air and water to carry out goods and passengers.	X_9
Wholesale and retail trade	It is consisting of two components; wholesaling the merchandized, the individuals / retailers merchandised, hotels and restaurants.	X_{10}
Finance and insurance	It includes the banking system, insurance company's development, financial institutions, cooperative societies leasing and mudarabas.	X_{11}
Ownership of dwellings	It includes the owner-occupiers and landlords of dwellings.	X_{12}
Public administration and defense	It includes administration of financial and fiscal affairs, health care, primary, secondary, post-secondary and special re-training programmers, recreational and cultural services, potable water supply programmes, agricultural affairs, environment protection programmes, forestry affairs, fishing & hunting affairs, electricity affairs, mining and mineral resources affairs, hotel, restaurant and tourism affairs, tourism promotion, regulation, , trade inspection, general labor affairs transportation & communication affairs, research and development affairs, unemployment reduction policies, military and defense affairs and civil defense forces.	X_{13}
Community services	It includes the activities performed by a person for the benefit of himself or the community, education, health care and real estate.	X_{14}

Source. Pakistan economic survey (various issues)

3.1 Unit root tests and order of integration

Prior to estimation it is obligatory to ensure stationarity of time series data. Numerous researchers as Nelson and Plosser(1982) and Hall (1978) concluded that time series variables have random pace. Granger and Newbold (1974); Granger (1986); Phillips (1986) and Ohanian (1988) analyzed that regression results will be spurious in case of non-stationary data. The stationarity of variables was checked by applying Augmented Dickey- Fuller (1979) and Philip Perron (1990) test. The test was conducted for each variable separately and lag length was selected by using Akaike Information Criterion (AIC). Following three regression equations have been used by Dicky and Fuller to investigate unit root issue.

$y_t = \gamma\, y_{t-1} + \mu_t,$ No intercept and trend (none) (1)

$y_t = \alpha + \gamma\, y_{t-1} + \mu_t$ With intercept (2)

$y_t = \alpha + \beta t + \gamma\, y_{t-1} + \mu_t$ With intercept and trend (3)

Yt is variable of interest , α is the intercept and βt is the trend and μ_t is error term. In above regression equations the Yt is time series. The parameter of interest is γ (where, γ = ρ-1, if γ = 0, then ρ = 1). If γ=0 it indicates that it contains a unit root-the time series is non-stationary and contrary if the γ < 0 the series is stationary.

H0. γ = 0 the series has unit root, series is not stationary or integrated of order one and it indicates a long run relationship.

H1. γ ≠1 or γ < 0. The series has no unit root; series is stationary and it indicates no long run relationship.

If the data are not stationary at levels than stationarity is checked by applying following equations.

$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{q} \theta_i \Delta y_{t-1}$ (Without intercept and trend) (4)

$\Delta y_t = a + \gamma y_{t-1} + \sum_{i=1}^{q} \theta_i \Delta y_{t-1}$ (With intercept) (5)

$\Delta y_t = a + \beta t + \gamma y_{t-1} + \sum_{i=1}^{q} \theta_i \Delta y_{t-1}$ (With intercept and trend) (6)

If the variables are differenced stationary than series are integrated of order one or I (1). A series (time series) is integrated of order one if after differencing once it became stationary. So if H0 is rejected than first differenced stationary is confirmed I (1)

A non-parametric method used by Phillips and Perron (1990) to correct the serial correlation of the disturbances. This test was a generalization of the Dickey-Fuller procedure. The PP test based upon the long run variance of residuals. The equations are same as were used by the DF. While for hypothesis testing the critical values of MacKinnon (1991) were used.

3.2. Model specification - stepwise regression (forward selection and backward elimination)

FollowingDrapper and Smith (1998) stepwise multiple regression analysis was conducted on GDP growth as the dependent variable with the crops, livestock, fisheries, forestry, mining and quarrying, manufacturing, construction, electricity, wholesale and retail trade, transport, storage and communication, ownership of dwellings, finance and insurance and housing services as the independent variables to analyze each predictor variable whereas controlling the effects of the other predictor variables (Nawaz et al., 2011; Malik and Naeem, 2011; Field, 2013).Stepwise regression is an effective and efficient technique for large data (Bager et al., 2017).The aim of employing the stepwise regression was to find the correlations between variables to determine the significance of the relationships between variables and to found whether independent variables have the ability to predict GDP growth. Through stepwise regression the insignificant predictors/ independent variable were removed which have not the ability to predict the dependent variable (Field, 2013). The inclusion of any variable in the model was made through forward selection. The variable which was significant at earlier stage was eliminated later depending upon the nature of relationship with new variable inducted through backward elimination. A correlation matrix was constructed and X_{10} (wholesale and retail trade) was found most correlated so model $Y= \beta_0+\beta_1 X_{10} +\mu_t$ was specified.

The partial F-values of the entire predictor variables were checked by putting turn by turn in the regression equation $Y= \beta_0+\beta_1 X_{10} +\mu_t$ The X_9 having highest partial F-values was included in the model

in next step as $Y= \beta_0+\beta_1 X_{10}+\beta_2 X_9 +\mu_t$. The overall significance of the model was observed through the improvement of R^2. The partial F-values for both the variables in the equation were checked. If the calculated partial F remained greater than tabulated F-value it means both the variables are significant and would retain in model. The F-Statistic has the,

H_0= All the coefficients of the regression model are zero.

H_1= At least one of the coefficients is non-zero. If the p-value ≤ 0.05 it means the null hypothesis should be rejected that the coefficients are zero. If the p-value ≥ it means the null hypothesis should be accepted that the coefficients are zero the inclusion of any variable in the model is known as the ***Forward Selection***

A predictor which was significant at any earlier stage may be not significant at later stage due to the nature of relationship with the new variable included in the model. So that variable was eliminated from the model on the basis of partial F value. This process of the elimination of variable is known as the ***Backward Elimination.***

Finally, the process stops when in the last equation no variables can be removed and no new, next best candidate variable can hold its place in the equation. Now the explanatory variables entered the regression have their effect on the R^2 usually recorded. Stepwise regression analysis is a combination of certain aspects of forward selection and backward elimination methods. Stepwise selection method investigates all the variables by including in models and excludes any variable on the basis of F-value.

The final model was selected after **twenty-eight steps**. The model is as given below.

$$Y_t=\beta_0+\beta_1 \text{ wholesale \& retail trade}gr_t+\beta_2 \text{ transport storages\& communication } gr_t+\beta_3 \text{ crops}gr_t+\beta_4 \text{ manufacturing}gr_t+\beta_5 \text{ livestock}grt+ \beta_6 \text{ community services } gr_t+ \beta_7 \text{ public administration \& defence}gr_t+\beta_8 \text{ electicity, gas generation and distribution}gr_t +\beta_9\text{finance and insurance}grt+\mu t$$

After renaming the variables model was in the following form

$$Y= \beta_0+\beta_1 X_{10}+\beta_2 X_9+\beta_3 X_1+\beta_4 X_6+ \beta_5 X_2+\beta_6 X_{14}+ \beta_7 X_{13}+ \beta_8 X_8+ \beta_9 X_{11}+\mu t$$

$$Ygr_t=\beta_0+\beta_1 w\&rtgr_t+\beta_2 ts\&cgr_t+\beta_3 crgr_t+\beta_4 mfggr_t+\beta_5 lsgr_t+\beta_6 csgr_t+ \beta_7 \text{ p.ad \&def } gr_t+ \beta_8 \text{ eg \&gdgr}_t+\beta_9 f\&igr_t+\mu t \quad (7)$$

Where

X_1 = cgr_t = crops' growth (percent)

X_2 = $lsgr_t$= livestock 's growth (percent)

X_6 = $mfggr_t$ = manufacturing 's growth (percent)

X_8 = $eg \&gdgr_t$ =electricity generation and gas distribution 's growth (percent)

X_9 = $ts\&cgr_t$=transport storage and communication 's growth (percent)

X_{10} = $w\&rtgr_t$=wholesale and retail trade 's growth (percent)

X_{11} = $f\&igr_t$ = finance and insurance 's growth (percent)

X_{13} = $p.ad \&def\ gr_t$= public administration and defense growth (percent)

X_{14} = $csgr_t$= community service's growth (percent)

Y =Ygr_t =gross domestic product' growth (percent)

β_0 = Intercept if ($\beta_1=\beta_3=\beta_4=\beta_5 =\beta_6 =\beta_7 =\beta_8 =\beta =0$)

β_s = parameters which have to be estimated

μ_t=error term(assumed with zero mean and constant variance).

grt = growth and t indicates that data is time series

Model was specified by adopting ***forward selection and backward elimination*** approach and was estimated by using Minitab.

3.3 .Statistical diagnostic tests

During statistical and econometric analysis in many cases the researchers remained not sure about the models that whether models were correctly specified and estimated or not. The results of specified models based upon some statistical assumptions and results were come true only if

assumptions were hold. During analysis some diagnostic tests have been applied to check the accuracy of models specified.

4.Results and discussion.

Table 3. Data description of sub-sectors included in stepwise regression model.

Variables	Mean	Median	Maximum	Minimum	std. dev.	skewness	Kurtosis	jarqueber a	probability	Sum	ssq. dev.
X_1	2.936	2.700	24.100	-15.100	7.071	0.081	3.338	0.410	0.815	205.500	3449.561
X_2	4.329	3.500	27.700	-5.200	5.385	1.494	7.554	86.520	0.000	303.000	2000.823
X_3	6.634	1.850	139.000	-34.000	21.443	3.715	22.660	1288.418	0.000	464.400	31727.52
X_4	11.899	2.550	591.600	-43.200	72.380	7.487	60.517	10302.890	0.000	832.900	361477.7
X_5	**13.486**	5.850	440.700	-57.600	53.808	7.271	58.485	9595.928	0.000	944.000	199772.6
X_6	6.284	6.600	20.500	-13.500	5.687	-0.378	4.723	10.329	0.006	439.900	2231.413
X_7	6.367	5.550	36.700	-15.800	10.682	0.716	4.097	9.495	0.009	445.700	7873.094
X_8	9.621	6.450	105.800	-24.700	18.876	2.101	11.597	267.062	0.000	673.500	24585.54
X_9	6.636	4.650	39.900	-10.500	8.643	2.269	9.464	181.903	0.000	464.500	5154.301
X_{10}	5.749	5.350	46.600	-6.300	6.402	3.637	24.956	1560.292	0.000	402.400	2828.255
X_{11}	9.251	8.700	45.300	-18.100	10.836	0.707	5.085	18.512	0.000	647.600	8101.555
X_{12}	4.075	3.450	42.000	-13.400	5.678	3.964	30.870	2448.782	0.000	285.250	2224.604
X_{13}	5.384	4.450	57.400	-16.200	8.885	2.797	18.212	766.237	0.000	376.900	5447.013
X_{14}	5.986	5.900	29.700	-8.900	3.973	2.391	21.839	1101.834	0.000	419.000	1089.166
Y	5.070	4.800	26.000	-1.500	3.676	2.492	16.221	582.315	0.000	354.900	932.607

Table 3 indicates the data description of growth rates of fourteen sub sectors used for stepwise regression. The mining and quarrying (X_5)remained high on average during the study period. The series $X_1, X_2, X_3, X_4. X_5, X_7, X_8, X_9, X_{10}, X11, X_{12}, X_{13}, X14$ and Y are positively skewed which indicate that most values lie on the right side of the mean. While the series $, X_6,$is negatively skewed which indicate that most values lie on the left side of the mean.

4.1. Unit root test and order of integration

Table 4. Stationarity analysis.

Variables		AT Levels		
		without intercept and trend	with intercept	with intercept.& trend
X_1	t-statistics	-9.0184(0.0000)*	-8.4942(0.0000)*	-8.4221(0.0000)*
	1%	-2.6021	-3.5401	-4.1130
	5%	-1.9460	-2.9092	-3.4839
	10%	-1.6134	-2.5922	-3.1700
X_2	t-statistics	-1.5538(0.1121)	-7.5628(0.0000)*	-8.4015(0.0000)*
	1%	-2.6034	-3.5383	-4.1104
	5%	-1.9462	-2.9084	-3.4827
	10%	-1.6133	-2.5917	-3.1693
X_3	t-statistics	-7.6904(0.0000)*	-8.3539(0.0000)*	-8.3594(0.0000)*
	1%	-2.6021	-3.5383	-4.1104
	5%	-1.9460	-2.9084	-3.4827
	10%	-1.6134	-2.5917	-3.1693
X_4	t-statistics	-7.9674(0.0000)*	-8.2160(0.0000)*	-8.2556(0.0000)*
	1%	-2.6021	-3.5383	-4.1104
	5%	-1.9460	-2.9084	-3.4827
	10%	-1.6134	-2.5917	-3.1693
X_5	t-statistics	-7.4858(0.0000)*	-7.9169(0.0000)*	-7.9326(0.0000)*
	1%	-2.6021	-3.5383	-4.1104
	5%	-1.9460	-2.9084	-3.4827
	10%	-1.6134	-2.5917	-3.1693
X_6	t-statistics	-1.6342(0.0960)	-7.2431(0.0000)*	-7.6177(0.0000)*
	1%	-2.6040	-3.5383	-4.1104
	5%	-1.9463	-2.9084	-3.4827
	10%	-1.6132	-2.5917	-3.1693
X_7	t-statistics	-5.8142(0.0000)*	-7.3793(0.0000)*	-8.3470(0.0000)*
	1%	-2.6021	-3.5383	-4.1104
	5%	-1.9461	-2.9084	-3.4828
	10%	-1.6135	-2.5918	-3.1694
X_8	t-statistics	-6.5507(0.0000)*	-8.1775(0.0000)*	-8.4722(0.0000)*
	1%	-2.6022	-3.5384	-4.1104
	5%	-1.9461	-2.9084	-3.4828
	10%	-1.6135	-2.5918	-3.1694
X_9	t-statistics	-5.5066(0.0000)*	-7.7473(0.0000)*	-7.720010(0.0000)*
	1%	-2.6022	-3.5384	-4.1104
	5%	-1.9461	-2.9084	-3.4827
	10%	-1.6135	-2.5918	-3.1694
X_{10}	t-statistics	-0.9171(0.3150)	-7.2555(0.0000)*	-7.2189(0.0000)*
	1%	-2.6047	-3.5402	-4.1130
	5%	-1.9467	-2.9092	-3.4839

	10%	-1.6133	-2.5922	-3.1700
X_{11}	t-statistics	-3.1390(0.0000)*	-7.6891(0.0000)*	-8.0883(0.0000)*
	1%	-2.6027	-3.5383	-4.1104
	5%	-1.9461	-2.9084	-3.4827
	10%	-1.6133	-2.5917	-3.1693
X_{12}	t-statistics	-6.8983(0.0000)*	-7.7385(0.0000)*	-7.8940(0.0000)*
	1%	-2.602185	-3.538362	-4.110440
X_{13}	5%	-1.946072	-2.908420	-3.482763
	10%	-1.613448	-2.591799	-3.169372
	t-statistics	-6.1567(0.0000)*	-8.0517(0.0000)*	-7.9872(0.0000)*
	1%	-2.6021	-3.5383	-4.1104
	5%	-1.9460	-2.9084	-3.4827
	10%	-1.6134	-2.5917	-3.1693
X_{14}	t-statistics	-1.241312(0.1948)	-8.578051(0.0000)*	-8.550506(0.0000)
	1%	-2.6040	-3.5383	-4.1104
	5%	-1.9463	-2.9084	-3.4827
	10%	-1.6132	-2.5917	-3.1693
Y	t-statistics	-0.9462(0.3030)	-7.8382(0.0000)*	-7.7924(0.0000)*
	1%	-2.6040	-3.5383	-4.1104
	5%	-1.9463	-2.9084	-3.4827
	10%	-1.6132	-2.5917	-3.1693

Note. *symbolizes the level of significance (one percent) at which the null hypothesis was rejected. The values in parenthesis signifies the Mackinnon critical at which the unit root hypothesis can be accepted or rejected.

Researcher's calculations using EViews.

The results of Augmented Dickey Fuller (ADF) test presented in table 4 signposts that the variables included in the model were found stationary at levels so stepwise regression was applied.

4.2. Steps to build model

- **Correlation analysis**

Correlation matrix was constructed between the dependent variable and independent variables to explore the linear relationship. The stepwise regression starts with the exploration of correlations by viewing the r-square and p-value(Williams, Credle, & McLain., 2018).The results of correlation analysis lts have been presented in table 5. All the independent variables except X_{12} were positively correlated with dependent variable. So all were the candidates to predict the GDP of Pakistan.

Table 5. Correlations (Pearson) matrix.

	X_1X_2 X_3X_4	X_5	X_6 X_7X_8 X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}

X_2	-0.117												
X_3	0.057	0.058											
X_4	0.128	0.563	-0.043										
X_5	0.138	0.504	-0.209	0.929									
X_6	0.035	0.201	0.009	0.303	0.354								
X_7	-0.104	-0.154	-0.183	-0.159	-0.036	0.294							
X_8	0.328	-0.229	0.143	-0.028	0.058	-0.058	-0.033						
X_9	0.207	0.203	0.29	0.499	0.404	0.19	-0.035	0.04					
X_{10}	0.274	0.372	-0.105	0.739	0.8	0.583	0.077	0.074	0.38				
X_{11}	0.061	0.215	0.027	0.383	0.422	0.333	0.155	-0.034	0.185	0.507			
X_{12}	-0.023	-0.044	0.621	-0.3	-0.463	-0.104	-0.1	-0.009	0.257	-0.318	-0.257		
X_{13}	-0.204	-0.133	0.037	-0.004	-0.005	0.1	0.159	-0.065	0.101	0.096	0.11	0.167	
X_{14}	0.136	0.304	-0.364	0.647	0.769	0.246	-0.022	0.079	0.117	0.698	0.402	-0.553	0.092
Y	**0.476**	**0.413**	**0.066**	**0.682**	**0.721**	**0.602**	**0.080**	**0.142**	**0.551**	**0.873**	**0.477**	**-0.127**	**0.157**
	0.601												

Researcher's calculations using MINITAB

Table 6 indicates that most correlated variable with dependent variable Y (GDP) was X_{10}(wholesale and retail trade) i.e., r = 0.842. After comparing F_{10} with F-tab it was found significant because $F_{10} = 150.84 > F_{tab}(1, 62, 0.95) = 3.99589$. *So* $Y=f(X_{10})$, $Y= \beta_0+\beta_1 X_{10} +\mu_t$. *was constructed.* Following Forward selection all variables were introduced one by one in specified model to check their partial F-ratios and on the basis of these partial F -ratios the decision for the inclusion of next variable in model was made.

Table 6. An overview of step-wise regression models(specified and estimated)

	Specified and estimated models	*S*	*R-Sq (%)*	*R^2 (adj) (%)*	*F*	*P*
Model#01	*Y= β0+β1 X10 +μt*					
	$\hat{Y}$= 2.187+ 0.501X_{10}	1.80	76.1	75.7	216.9	0.000
Model#02	*$Y = \beta_0+\beta_1 X_{10}+\beta_2 X_9 +\mu_t$*					
	$\hat{Y}$= 1.784 + 0.445 X_{10}+ 0.109 X_9	1.59	81.7	81.2	150.20	0.000
Model#03	*Y= $\beta_0+\beta_1 X_{10}+\beta_2 X_9+\beta_3 X_1 +\mu_t$*					
	$\hat{Y}$= = 1.376+ 0.415 X_{10}+ 0.097 X_9+ 0.119 X_1	1.37	86.6	86.0	142.23	0.000
Model#04	*Y= $\beta_0+\beta_1 X_{10}+\beta_2 X_9+\beta 3 X_1+\beta_4 X_6+\mu_t$*					
	$\hat{Y}$= 1.231 + 0.342X_{10}+ 0.099 X_9 + 0.133 X_1 + 0.130 X_6	1.24	89.2	88.5	134.41	0.000
Model#05	*Y= $\beta_0+\beta_1 X_{10}+\beta_2 X_9+\beta_3 X_1+\beta_4 X_6+ \beta_5 X_2+\mu_t$*					
	$\hat{Y}$= 0.918 + 0.300 X_{10} + 0.092X_9+ 0.155 X_1+ 0.137 X_6+ 0.114 X_2	1.17	91.4	90.7	136.8	0.000
Model#06	*Y= β0+β1 X10+β2 X9+β3 X1+β4 X6+ β5 X2+β6 X14+μt*					
	$\hat{Y}$=0.296 + 0.218 X_{10}+ 0.103 X_9+ 0.161X_1+ 0.163 X_6+ 0.109 X_2 + 0.142X_{14}	*1.05*	*92.4*	*91.7*	*129.23*	0.000
Model#07	*Y=β0+β1 X10+β2 X9+β3 X1+β4 X6+ β5 X2+β6 X14+ β7 X13μt*					
	$\hat{Y}$= 0.039+ 0.203 X_{10} + 0.092 X_9+ 0.189 X_1 + 0.160 X_6+ 0.142 X_2+0.128 X_{14}+ 0.068 X_{13}	*0.88*	*94.8*	*92.2*	*162.04*	0.000
Model#08	*Y= β0+β1 X10+β2 X9+β3 X1+β4 X6+ β5X2+β6X14+ β7 X13+ β8 X8+μt*					
	$\hat{Y}$= -0.152 + 0.200X_{10}+ 0.092X_9+ 0.181 X_1 + 0.163X_6+ 0.153 X_2 +0.123 X_{14} + 0.069 X_{13} + 0.011 X_8	*0.86*	*95.1*	*94.4*	*148.33*	0.000
Model#09	*Y= β0+β1 X10+β2 X9+β3 X1+β4 X6+ β5 X2+β6 X14+ β7 X13+ β8 X8+ β9 X11+μt*					
	$\hat{Y}$ = -0.197 + 0.189 X_{10} + 0.092X_9+ 0.182 X_1+ 0.161 X_6+ 0.152 X_2+ 0.118 X_{14}+ 0.068X_{13} + 0.012 X_8 + 0.016 X_{11}	***0.85***	***95.3***	***94.6***	***134.5***	**0.000**

Researcher 's calculations using MINITAB

Table 6 indicates that at each step not only the R^2 increased with the inclusion of additional variable (the R^2 was 76.1 percent at the first step but at 28th step it raised upto 95.3 percent), R^2 (Adj.) also increased which urged that inclusion of variables in the model was significant. Result indicated that error decreased with the inclusion of additional variables. The probability value of F-statistic indicated that variables included in the model are jointly, positively and significantly affecting dependent variable GDP growth.

After proceeding 28 steps of forward selection and backward elimination the final selected "appropriate Model" is given below

*Y=f(**X1,X2, X6,X8, X9 , X10, , X11, X13,X14**)* (8)

The estimated model

$\hat{Y}$ == -0.197 +0.182 X_1 +0.152 X_2+ 0.161 X_6 0.012 X_8 + 0.092X_9 + 0.189 X_{10} +0.016 X_{11} + 0.068X_{13} + 0.118 X_{14}(9)

Table 7: Estimations of regression model.

Analysis of Variance (ANOVA)					
Source	Df	SS	MS	F	Significance
Regression	9	889.464	98.8293	134.56	0.000
Error	60	44.067	0.7345		
Total	69	933.531			

Variables		Coefficients	St. error	T-statistics	P-value
Intercept		-0.197	0.262	-0.75	0.454
Crops	X_1	0.183	0.0174	10.51	0.000
Livestock	X_2	0.153	0.023	6.65	0.000
Manufacturing	X_6	0.162	0.0239	6.78	0.000
Electricity and gas dist.	X_8	1.012	0.00598	2.01	0.049
Trans.Stor&communi.	X_9	0.092	0.0136	6.78	0.000
Wholesale & retail	X_{10}	0.190	0.0333	5.71	0.000
Finance and insurance	X_{11}	0.016	0.0112	1.47	0.045
Public administration	X_{13}	0.069	0.0126	5.47	0.000
Community services	X_{14}	0.119	0.0395	3.01	0.004
R-Sq= 95.3		R-Sq(adj)= 94.6		S= 0.857	

Researcher 's calculations using MINITAB

Table7 indicates that predictor variables crops (X_1), livestock (X_2), manufacturing (X_6), electricity and gas distribution (X_8), transport storage and communication (X_9), wholesale and retail trade (X_{10}), finance and insurance (X_{11}), public administration and defense (X_{13}), community services (X_{14}) are best predictor variables for the Pakistan economy. All variables in the model were found significant and depicted positive correlation and significant contribution towards the GDP growth. The coefficients of the variables described that one percentage point change in crops (X_1) will cause 0.183 percentage point change in overall GDP growth , one percentage point change in livestock (X_2) will cause 0.153 percentage point change in overall GDP growth, one percentage point change in manufacturing(X_6) will cause 0.162 percentage point change in overall GDP growth, one percentage point change in electricity and gas distribution (X_8) will cause 0.012 percentage point change in overall GDP growth, one percentage point change in transport storage and communication (X_9) will cause 0.092 percentage point change in overall GDP growth, one percentage point change in wholesale and retail trade (X_{10}) will cause 0.190 percentage point change in overall GDP,one percentage point change in finance and insurance (X_{11}) will cause 0.016 percentage point change in overall GDP, one percentage point change in public administration and defense (X_{13}) will cause 0.069 percentage point change in overall GDP growth, community services (X_{14}) will cause 0.119 percentage point change in overall GDP growth.

Fisheries, forestry, mining & quarrying, construction and ownership of dwelling were found weak sub-sectors of the economy. The constructed model is an appropriate model for the Pakistan economy. All the sub-sectors positively and significantly correlated with GDP growth.

5. Statistical diagnostic tests for the validity of step wise regression model

The selected model (9) was statistically evaluated through some diagnostic tests to check the presence of autocorrelation, multicollinearity, heteroscedasticity and stability of the model. They are given below,

- **Autocorrelation test**

  H_0. No autocorrelation exists among residuals

H_1. Autocorrelation exists among residuals

Table 8.Results of Breusch-Godfrey -LM test.

F-statistic	0.206682	Prob.	0.8139
Obs*R-Squared	0.495356	Prob.Chi-square	0.7806

Researcher 's calculations using EViews.

Table 8 depicted the p value (0.7806) of obs*R-sq. is insignificant so accept H_0.

- **Multicollinearity/ Variance inflation factor criterion**

Table 9. Results of Variance inflation factor criterion

Variables	X1	X2	X6	X8	X9	X10	X11	X13	X14
Centered VIF	1.4171	1.4357	1.7308	1.1972	1.2987	4.2678	1.3802	1.1787	2.3196

Table 9 indicates that centered VIF value of all variables is less than 5 so no multicollinearity exists.

- **Heteroscedasticity test**

H_0. No heteroscedasticity exists among residuals.

H_1. Heteroscedasticity exists among residuls

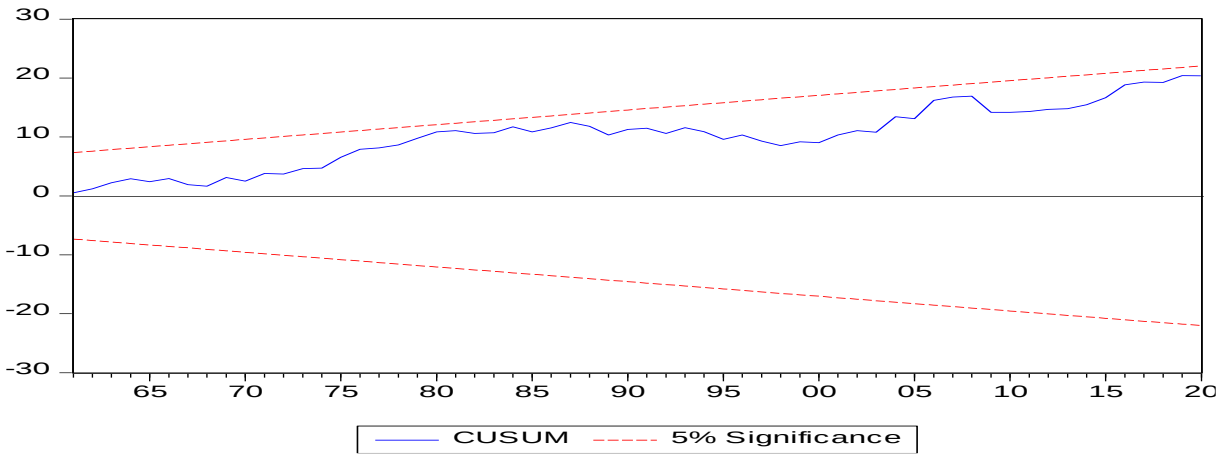
Table10. Results of Breusch-Pagan-Godfrey test.

F-statistic	1.069484	Probability of F(9,60)	0.3982
Obs*R-Squared	9.677147	Probability of Chi-Square(9)	0.3772
Scaled explained SS	12.11971	Probability of Chi-Square(9)	0.2066

Researcher's calculations using EViews.

Table 10 depicted that P-value of Obs*R-squared is insignificant (0.4961), so accept H0.

- **Model Stability Test**



Constructed by the researcher using EViews.

Figure 1. Model stability test Cumulative Sum of Recursive Residuals

This test has been applied on estimated model the basis for the CUSUM (Brown et al., 1975) was the cumulative sum of recursive residuals. In this test the cumulative sum was plotted along with 5 percent critical lines/bands. The decision was made about the stability of model with this plotting if the plotted cumulative sum gone outside the two critical lines/bands the model will be instable. In figure 1 it was demonstrated the two dotted lines or bands with percent significance level and the curve within the line's indications the stability of the model.

Stepwise regression demonstrated that analysis have no issue as autocorrelation, multicollinearity, and heteroscedasticity. Conclusions was drawn from it that statistically R^2= 95.3 percent portrayed that a very good fit. From economic point of view about 95.3 percent of total variations in GDP were explained by the independent variables included in the model and remaining contributed by other factors. Overall, the regression model was statistically significant F-statistic 134.56 (0.0000) indicates that sub-sectors included in model jointly effect GDP growth.

Conclusion

The objective of current study was to fit a suitable time series model to predict the GDP growth of Pakistan. Through stepwise regression analysis a suitable and appropriate model was constructed. The results revealed that crops, livestock, manufacturing, electricity and gas distribution, transport storage and communication, wholesale and retail trade, finance and insurance, public administration and defense, community services are best predictors for the Pakistan economy and wholesale and retail trade was observed as most significant sub sector towards GDP growth. All variables in the model were found significant and depicted positive correlation and significant contribution towards the GDP growth. *Fisheries, forestry, mining & quarrying, construction, and ownership of dwelling* were found weak sub-sectors of the economy.

References

Akhter, Y., Mahsin, M. and Mohaimin, M. Z. (2012). An Application of Factor Analysis on Gross Domestic Product Data of Bangladesh. *Bangladesh e-Journal of Sociology*, 9(1):6-18.

Ang, J. B. (2008). What are the mechanisms linking financial development and economic growth in Malaysia? Economic Modelling, 25(1):38-53.

Bager, A., Roman, M., Algedih, M. & Mohammed, B. (2017). Addressing multicollinearity in regression models: a ridge regression application. Online at https://mpra.ub.uni muenchen.de/81390/ MPRA Paper No. 81390.

Bobinaite, V., Juozapaviciene, A. and Konstantinaviciute, I. (2011). Regression Analysis of Gross Domestic Product and its Factors in Lithuania. Economics and Management, 16:116-126.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society: Series B (Methodological), 37(2), 149-163.

Chin, H. W. and Fitrianto, A. (2013). Statistical Approach for Modeling Malaysia's Gross Domestic Product. Middle-East Journal of Scientific Research, 15(4):606-612.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.

Drapper, N.R., Smith, H. (1998). Applied regression analysis, 3rd edn. 327-355

Easterly, W. (2001). The lost decades: developing countries' stagnation in spite of policy reform 1980–1998. Journal of Economic Growth, 6(2), 135-157.

Field, A. (2013). Discovering statistics using IBM SPSS statistics.

GoP. (1960-2021). Pakistan Economic survey((Various issues).Islamabad: Government of Pakistan Retrieved from www.gop.com.k.

Granger, C. J. (1986). Developments in the study of co-integrated economic variables. Oxford Bulletin of economics and statistics, 48(3), 213-228.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. Journal of political economy, 86(6), 971-987.

Hassan, M. K., Sanchez, B. and Suk Yu, J. (2011). Financial development and economic growth: New evidence from panel data. The Quarterly Review of Economics and Finance, 51(1):88-104.

Hassan, M. K., Sanchez, B. and Suk Yu, J. (2011). Financial development and economic growth: New evidence from panel data. The Quarterly Review of Economics and Finance, 51(1):88-104.

Lee, J. (2011). Export specialization and economic growth around the world. Economic Systems, 35(1):45-63.

MacKinnon, J. G. (1991). Critical values for cointegration tests. In Eds.), Long-Run Economic Relationship: Readings in Cointegration.

Malik, M. E., & Naeem, B. (2011). Role of spirituality in job satisfaction and organizational commitment among faculty of institutes of higher learning in Pakistan. African journal of business management, 5(4), 1236-1244.

Mehmood, S. (2012). Effect of Different Factors on Gross Domestic Product: A Comparative Study of Pakistan and Bangladesh. Academy of Contemporary Research Journal, V I(I):11-22.

Nawaz, A., Khan, S., & Khan, H. (2011). Stepwise Regression of Demographics to Predict e-Learning Problems & User-Satisfaction in HEIs of Khyber Pakhtunkhwa (KPK) Pakistan. Global Journal of Computer Science and Technology, 11(2), 0975-4172.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. Journal of monetary economics, 10(2), 139-162.

Ohanian, L. E. (1988). The spurious effects of unit roots on vector autoregressions: A Monte Carlo study. Journal of Econometrics, 39(3), 251-266.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. Journal of Business & Economic Statistics, 8(2), 153-162.

Phillips, P. C. (1986). Understanding spurious regressions in econometrics. Journal of econometrics, 33(3), 311-340.

Williams, Z. (2018). The negative, positive, and increasing correlation of historic interest rates with current market performance. Journal of Management and Innovation, 4(2), 11-30.

Wu, Sh. Y., Tang, J. H. and Lin, E. S. (2010). The impact of government expenditure on economic growth: How sensitive to the level of development? Journal of Policy Modeling, 32(6): 804-817.