Forecasting the PhDs-Output of the Higher Education System of Pakistan

Ghani ur Rehman, Dr. Muhammad Khalil Shahid, Dr. Bakhtiar Khan Khattak and Syed Fiaz Ahmed

Center for Emerging Sciences, Engineering and Technology, Islamabad,
Department of the Management Sciences, Gomal University D.I. Khan

Abstract: Today, forecasting of research and development (R&D) professionals is an important function of management in the field of higher education (HE). In this research study the PhDs output of the higher education system (HES) of Pakistan from 1960 to 2014 is autoregressive integrated and moving average-ARIMA(2,1,2) modeled for next two decades (2035):-

$$\text{ARIMA}(2,1,2): Y_t = 0.05771 + 0.6139Y_{t-1} - 0.7839Y_{t-2} + \varepsilon_t + 1.1336\varepsilon_{t-1} - 0.9580\varepsilon_{t-2}$$

The model showed that if the present outputs pace persists, then it perhaps PhDs output in Pakistan will be reached up to 12176 annually in 2035.

Appropriateness of the ARIMA(2,1,2) forecast model was clarified by the criteria's and tests; AIC , SC(BIC), DW, R Square, Log Likelihood, LM Test, ARCH test, JB Statistic, JB Statistic P-Value, RMSE, MAE, PAPE and TIC. Beside these diagnostic checks some graphical techniques such as time plots of PhDs outputs series at different levels, correlograms at different levels for ACF & PACF of residuals, and histograms etc were also used to confirm further suitability of the developed model.

Keywords: HES, PhDs Output, ARIMA Model and Forecasting.

1. INTRODUCTION

Forecasting are playing very pivotal role in managing organizations. Across the world many developed and developing countries are measuring their Higher Education Systems (HESs) for its performance from different aspects [1].

Forecasting of well educated workforce is an important management function in the area of higher education. The Higher Education System (HES) PhDs-Outputs output (the R&D professionals characterized for R&D activities across the world) forecast provide basis for long term planning both to manage it at organizational (university) level and as well as at national (overall HES level [2].

In developed for past more than fifty years it has been considered an integral part of the plans of economic advancement, whereas in developing countries forecasting of highly educated workforce has gained special importance in the last ongoing two decades [2, 3, 4, 5, 6].

In subcontinent the practice of professional workforce planning was first time initiated in the ending era of British rulers in 1946, when there in the subcontinent the amount of health (medical) professionals was forecasted for the next 25 years through techniques at that time [2]. In subsequent years, after dissolving the subcontinent in the two countries India and Pakistan this practice was continued.

Like in Pakistan after independence in 1947, in the newly born country just after three months the first educational conference was held in Karachi from 27th to 1st December 1947. In this conference the Muslims leader of subcontinent Quaid-e-Azam, Muhammad Ali Jinaah highlighted that R&D professionals and workforce in the field of science and technology are very important for Pakistan. So, he states that "---- Education does not merely mean academic education. There is immediate and urgent need for giving scientific and technical education to our people in order to build up our future economic life and to see that our people

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

take to science, commerce, trade and particularly well planned industry. -----"[7].

Currently, in Pakistan there are many organizations and several other departments that are monitoring education systems at various levels in Pakistan. In these all the Higher Education Commission (HEC) of Pakistan is the one that has put great intention to promote higher education in Pakistan since the date of its establishment in year-2002. Several programs were launched "to reverse the brain drain and attract the best talent back to Pakistan" [8].

In 2002 the HEC had forecasted that in next five years PhDs outputs in Pakistan will be increased from 250 to 1500 per annum [8]. However, yet that date PhDs outputs in Pakistan is 1249 annually as shown by Fig-1. It declares that HEC has not still achieved the target (1500) of 2007 that was forecasted in year 2002 [9].

Across world several developing countries have realized that increase in the PhDs output is a key for to turn their economies in smart economy. For the reason, the countries like India and Malaysia have planned to foster their higher education systems (HES) for PhDs output. So, India has forecasted their PhDs output up to 20,000 annually for the year-2020 and Malaysia has planned by itself to achieve the target of 60000 PhDs in year-2023 [10].

There is no doubt that in the last decade the Higher Education System (HES) of Pakistan has greatly increased their PhDs outputs. But it needs great emphasization like china. Such as in the last decade (2002-2010) from PhDs outputs aspects China has greatly emphasized on their PhDs outputs and grow their PhDs outputs from 14368 PhDs to 27677 (doubled) in two years only (2005), and then in 2010 increased their PhDs up to 48987 as shown by Fig-2. This China has surpassed the US by PhDs outputs in the last era [11].

Today, across the world from R&D aspects the HES PhDs outputs are characterizes as the top level qualification. The PhDs professionals are called the high level R&D man power in the field of S&T. because they are the individuals which are distinguished for their knowledge, expertise and skillness [12].

The PhDs individuals past decade data from year 1998 to 2007 of the HESs of India and Pakistan it shows that Pakistani universities PhDs production is not only less than India total PhDs but its total is less PhDs than India female PhDs in the entire 10 years period as compiled in Table-1 & 2 [13, 16]. Even though the Pakistan last year-2014 total PhDs outputs is less than the India female except the two year1989 & 2001 [9].

A research study, it shows that out of 156 public and private institutions (universities and Degree awarding institutions i.e. DAIs etc) in Pakistan only 40% institutions are producing PhDs. Beside this 60 % of the institutions are fail to produce even a single PhDs graduate [14].

The PhDs professional are considered an important asset in current era, because they are playing very significant role in the smart economies. The researchers focus on finding the major trends of HE PhDs outputs. For the reason that in current the PhDs-outputs of HESs are working like foundation for development of the nations. In the last two decades that is why, that several publication are found on R&D professionals trend [2, 10, 12, 15, 16]. There are many forecast techniques are in use for planning of the human workforce such as linear & nonlinear models, first, second and higher degree regressive models, logistic regressive model, exponential mode and so on [17, 18, 19]. The box Jenkins methodology that is called ARIMA is also applied to model the time series data of human population and many other specific areas human workforce [20].

The main objective of this research paper is to forecast the PhDs output of the HES of Pakistan by using ARIMA forecast models. The Eview-5 software and Minitab 14 statistical software were applied for ARIMA forecast model fitting to determine the HES of Pakistan PhDs production trend for the next two decades (year-2035).

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

2. METERIALS AND MEHTODS

The PhDs outputs data used in this research are consist of the last 55 years (1960 -2014) of the HES of Pakistan. Almost data (from year 1960 to 2009) was taken from the HEC Annual Report 2012-2013, and the last five years data (from 2010-2014) was retrieved from the HEC online "PhD country directory" [9].

There are numerous research studies on educational planning, and in this connection many techniques and tools are in use. Thus the PhDs outputs data can be modeled and forecasted by many ways. Forecasting's have many types, such as (i) Auto-Regressive (AR) Model, (ii) Moving Average (MA) Model, (iii) Auto-Regressive Moving Average (ARMA) Model and (iv) Auto Regressive Integrated Moving Average (ARIMA). In these four ARIMA is the one that very effective, because it is comprises of all the features of the remaining three. For the reason the ARIMA approach was applied for forecasting the PhDs outputs. The ARIMA model is also called the Box-Jenkins (BJ) Methodology, labeled with the developer's (Box and Jenkins) names those who first time introduced the model in 1976 [21].

The ARIMA model fitting is a one effective technique and tool of forecasting, however it was not found anywhere by any researcher that one done forecast of the HES PhDs outputs using this approach. According the Auto-Regressive Moving Average model with "P" autoregressive and "q" moving average terms is denoted by ARMA (p, q) model and is described by the relation as under [20, 22]:

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \P$$

It is important, to note, when

(i). Moving Average term q=0, then, the ARMA (p, q) modal =AR (p) model

(ii). Autoregressive term p=0, then, the, ARMA (p, q) modal =MA (p) model

(iii). In the ARMA (p, q) modal time series data integrated by transformations, such as differentiation and logging of time series data for to stationeries the data. The stationary time series data fitted forecast model is termed as ARIMA (p, d, q) forecast model. The term d means order of differentiation.

In this forecasting technique, the following main steps are involved:

2. (a) Stationarity process:

In forecasting the PhDs outputs time series data, stationarity in the series is essential. To check stationarity of the PhDs outputs (PHD_PROD) data, many approaches are in use. In this study two techniques were used, mathematical assessment and visual inspection. In visual inspection, time plots and correlograms (of the autocorrelation functions-ACF and partial autocorrelation functions-PACF at different lags) were used. For mathematical assessment parametric test (Augmented Dickey-Fuller-ADF Unit Root Test) was applied here.

2. (b) Identification process:

In identification process of the ARIMA forecasting model, the three terms; p (AR term), d (Difference), and q (MA term) are of great importance. In a suitable ARIMA model the correlograms are playing vital role, to select a right p, d, and q term, because the ACF and PACF values and spikes presented by the correlograms shows that what will be the possibly right term. Such as in the correlogram (Fig-12), almost spikes of both ACF and PACF at different lags are in limits (95% confidence limits). The values of AC (ACF), PAC (PACF), Q-Stat and Prob (P) are reasonable. Values of the terms p & q, normally means the small integers 0, 1, 2, 3, 4 or 5, and for term d, typically, the integers are 0, 1 or 2.

2. (c) Estimation process:

In this process, on the basis of identified terms p, d, and q of identified ARIMA forecast models, the parameters are estimated. Such as Akaike Information Criterion (AIC), Schwarz or Basic Information Criterion (SIC or BIC), Durbin-Watson (DW), Residual (R) Square, Adjusted R-Square, Lagrange Multiplier (LM) Test, Autoregressive Conditional Heteroskedasticity (ARCH) test, Jarque-Bera (JB) test Statistic, ARMA Structure, Mean Square Error (MAE) Root Mean Square Error (RMSE), Mean Absolute Percent Error(MAPE) and Theil Inequality Coefficient (TIC) etc, for to measure the goodness of fit of selected model [23, 24].

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

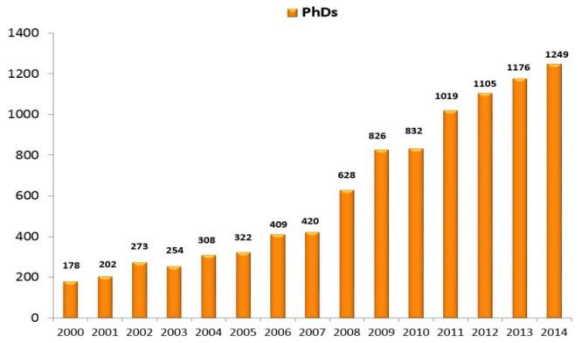


Figure-1, Source: http://www.hec.gov.pk/InsideHEC/Divisions/QALI/Others/Statistics/Pages/YearWiseSummary2.aspx

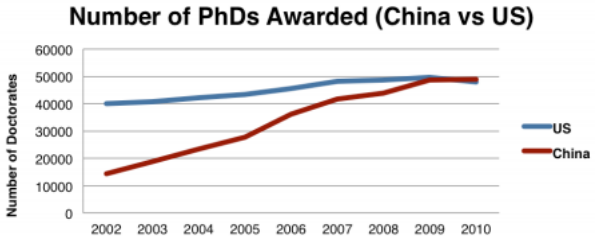


Figure-2, Source: http://blog.inomics.com/chinese-phds-vs-us-phds

Year	Male	Female	Total
1998	1258	644	1902
	66.14%	33.86%	100.00%
1999	2799	1426	4225
	66.25%	33.75%	100.00%
2000	3672	1854	5526
	66.45%	33.55%	100.00%
2001	1845	880	2725
	67.71%	32.29%	100.00%
2002	3352	1711	5063
	66.21%	33.79%	100.00%
2003	3958	1769	5727
	69.11%	30.89%	100.00%
2004	3729	1949	5678
	65.67%	34.33%	100.00%
2005	2810	1325	4135
	67.96%	32.04%	100.00%
2006	3522	1946	5468
	64.41%	35.59%	100.00%
2007	3319	1793	5112
	64.93%	35.07%	100.00%
Total PhDs	15297	30264	45561

Table-1, Source: Anitha Kurup, 2010, pp. Table-1, Page-18

Year	Male	Female	Total
1998	176	36	212
	83.02%	16.98%	100.00%
1999	143	58	201
	71.14%	28.86%	100.00%
2000	138	38	176
	78.41%	21.59%	100.00%
2001	142	62	204
	69.61%	30.39%	100.00%
2002	202	72	274
	73.72%	26.28%	100.00%
2003	202	70	272
	74.26%	25.74%	100.00%
2004	212	97	309
	68.61%	31.39%	100.00%
2005	232	94	326
	71.17%	28.83%	100.00%
2006	318	90	408
	77.94%	22.06%	100.00%
2007	316	116	432
	73.15%	26.85%	100.00%
Total PhDs	2081	733	2814

Table-2, Source: HEC Annual Report: 2008-2009, p. 108

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

2. (d) Validation through Diagnostic Checks:

In the validation process, the estimation process estimated parameters are used for a hypothesized (parsimonious) ARIMA model. Quite often the final results of the estimation processes are tabulated (as showed in Table-10). The following diagnostic checks are used:-

A parsimonious ARIMA Model		
Parameters		Diagnostic Checks
C (Intercept) Coef C Prob.(P-Valve)		<0.05
AR	Coef. $\emptyset_1$ P Valve	<0.05
	Coef. $\emptyset_2$ P Valve	<0.05
	Coef. $\emptyset_3$ P Valve	<0.05
MA	Coef. θ_1 P Valve	<0.05
	Coef. θ_2 P Valve	<0.05
	Coef. θ_3 P Valve	<0.05
AIC		Smallest value
SC (BIC)		Smallest value
DW		0-4, the best is Close to 2
R Square		0-1, the best max R Square
Log Likelihood		Max Likelihood Value
LM Test		P Value should be > 0.05
ARCH test		P Value should be > 0.05
JB Statistic		JB Value should be <5.99
JB Statistic P-Value		P Value should be > 0.05
RMSE		Smallest value
MAE		Smallest value
PAPE		Smallest value
TIC		Smallest value

Table-3

Summary for goodness of an economical ARIMA model are as under:

(i) The model coefficients are statistically significant.

(ii) The model coefficients are Stationary and Invertible.

(iii) The residuals are independently and normally distributed.

(iv) The model Durban Watson value is less than 4 and is close to 2.

(v) The model having no serial correlation and ARCH effect.

(vi) The model has low AIC, SC, RMSE and TIC values.

(vii) Ljung-Box statistic (Q-stat) represents that there is no pattern in the AC & PAC functions.

2. (e) Forecasting process of the Suitable ARIMA model.

In this step from the general ARIMA model equation given below different are estimated by putting the values of the estimated parameters in the equation:

ARIMA General Model:

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Then different graphs and tables are generated, to show adequacy of the suitable ARIMA model. In last forecast values are generated for the required period.

3. Results and Discussions

After putting the HES of Pakistan PhDs output data (PhDs time series) in forecasting analysis, the following results were obtained for ARIMA forecast Models:

3. (a). Stationarity Process:

(i) Results:-

Augmented Dickey-Fuller (ADF) Unit Root Test of PhDs series at Level

Null Hypothesis: PHD_PROD has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		5.636126	1.0000
Test critical values:	1% level	-3.557472	
	5% level	-2.916566	
	10% level	-2.596116	

*MacKinnon (1996) one-sided p-values.

Table-4

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

Time Plot of PhDs series at Level

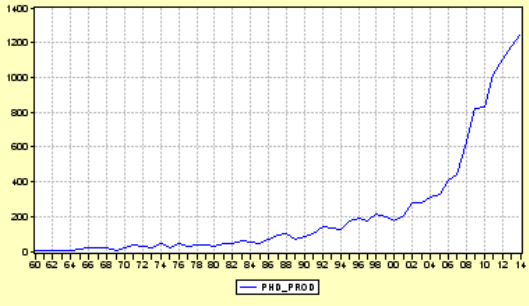


Figure-3

Correlogram of PhDs series at Level

Correlogram of PHD_PROD

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.705	0.705	28.839	0.000
		2	0.463	-0.068	41.506	0.000
		3	0.297	-0.006	46.839	0.000
		4	0.216	0.052	49.706	0.000
		5	0.182	0.043	51.774	0.000
		6	0.162	0.024	53.457	0.000
		7	0.152	0.030	54.959	0.000
		8	0.147	0.033	56.404	0.000
		9	0.142	0.022	57.770	0.000
		10	0.135	0.021	59.041	0.000
		11	0.130	0.024	60.252	0.000
		12	0.126	0.019	61.406	0.000
		13	0.121	0.017	62.500	0.000
		14	0.115	0.013	63.505	0.000

Figure-4

Interpretation:

In mathematical assessment the ADF Unit Root Test at Level of the PhDs series (Table-4) shows that p value (probability) is 1 (p>0.05), so, the series is not stationary.

Visual inspection of Time Plot (Fig-3) for PHD_PROD series from 1960 to 2014, at Level (original series) exhibits upwards trend, it means the series is not stationary. The correlogram (Fig-4) results for ACF and PACF at level also exhibits continuous decay that means the series is not stationary.

(ii) Results:-

Augmented Dickey-Fuller (ADF) Unit Root Test of PhDs series at 1st Difference

Null Hypothesis: D(PHD_PROD) has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.330416	0.6086
Test critical values:	1% level	-3.565430	
	5% level	-2.919952	
	10% level	-2.597905	

*MacKinnon (1996) one-sided p-values.

Table-5

Time Plot of PhDs series at 1st Difference

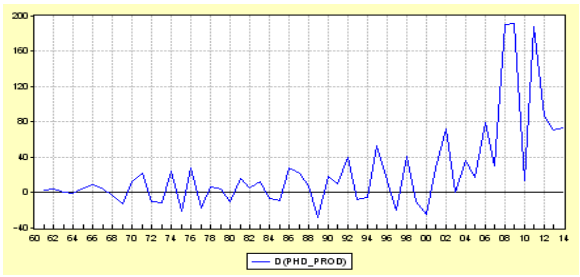


Figure-5

Correlogram of PhDs series at 1st Difference

Correlogram of D(PHD_PROD)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.376	0.376	8.0668	0.005
		2	0.351	0.244	15.222	0.000
		3	0.387	0.243	24.115	0.000
		4	0.129	-0.145	25.129	0.000
		5	0.070	-0.112	25.432	0.000
		6	0.146	0.091	26.775	0.000
		7	0.072	0.067	27.110	0.000
		8	0.082	0.048	27.550	0.001
		9	0.086	-0.032	28.051	0.001
		10	0.092	0.031	28.632	0.001
		11	0.083	0.030	29.116	0.002
		12	0.088	0.031	29.676	0.003
		13	0.056	-0.031	29.904	0.005
		14	0.074	0.019	30.317	0.007

Figure-6

Interpretation:

In mathematical in assessment the ADF Unit Root Test at 1st difference of the series shows that P value is 0.6068 (p>0.05) as given in the Table-5, it means the series is not stationary.

Visual inspection of PHD_PROD series at 1st difference the Time Plot (Fig-5) exhibits that there is apparently shift in the mean and dispersion over the period of time, it appears that the series is not stationary. The correlogram (Fig-6) results for ACF and PACF at 1st difference show improper pattern and decay, thus the series is not stationary.

(iii) Results:-

Augmented Dickey-Fuller (ADF) Unit Root Test of PhDs series at 2nd Difference

Null Hypothesis: D(PHD_PROD,2) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.50638	0.0000
Test critical values:	1% level	-3.565430	
	5% level	-2.919952	
	10% level	-2.597905	

*MacKinnon (1996) one-sided p-values.

Table-6

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

Time Plot of PhDs series at 2nd Difference

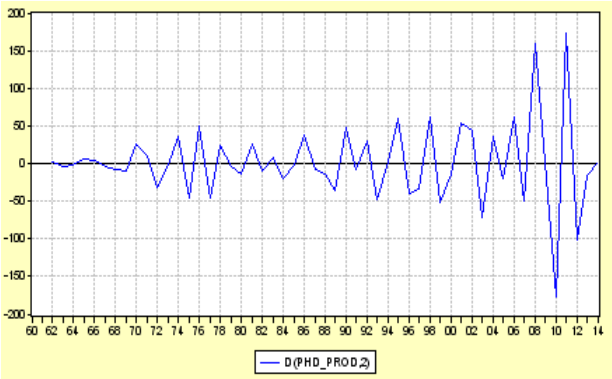


Figure-7

Correlogram of PhDs series at 2nd Difference

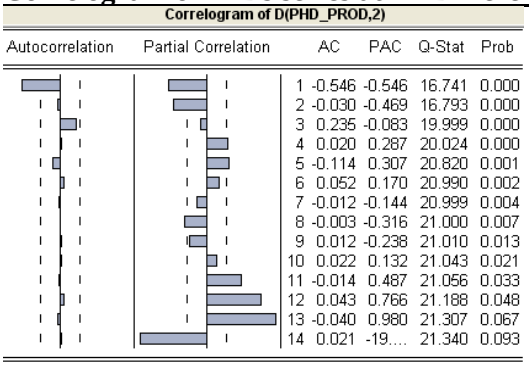


Figure-8

Interpretation:

In mathematical assessment, the ADF Unit Root Test results (Table-6) at 2nd difference seems realistic (as p=0.0000), the result shows that the series may be stationary.

Visual inspection of Time Plot (Fig-7) for PHD_PROD series at 2nd difference exhibits that variability in both the mean and standard deviation (dispersion) over the period of time is increasing, it shows that the series is not stationary. The correlogram results (Fig-8) at 2nd difference for ACF and PACF, show that lags pattern & values of both, ACF and PACF are unrealistic, so, it seems that the series is not stationary.

(iv) Results:-

Augmented Dickey-Fuller (ADF) Unit Root Test Log (PHD_PROD) series at level

Null Hypothesis: LOG(PHD_PROD) has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.508430	0.8807
Test critical values:	1% level	-3.568308	
	5% level	-2.921175	
	10% level	-2.598551	

*MacKinnon (1996) one-sided p-values.

Table-7

Time Plot of log (PHD_PROD) series at level

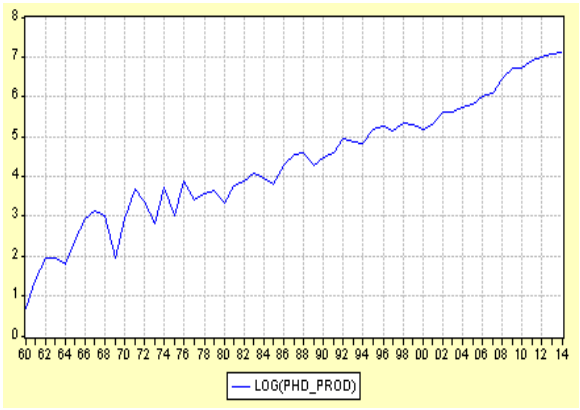


Figure-9

Correlogram log (PHD_PROD) series at level

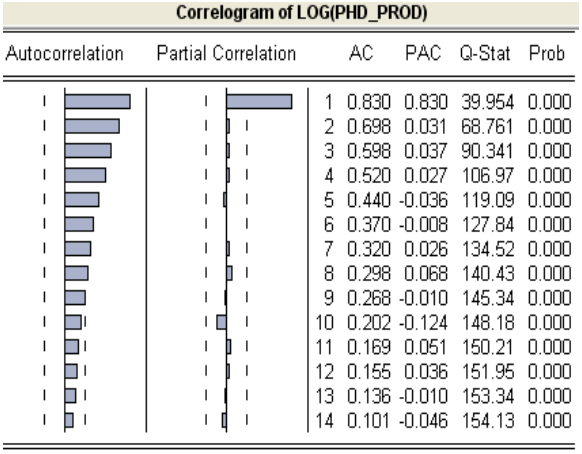


Figure-10

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

Interpretation:

Mathematical assessment through ADF Unit Root Test results (Table-7) at level of the series (by taking log) shows that p=0.8807 (p>0.05), it means that the series is not stationary.

Visual inspection of Time Plot (Fig-9) for log (PHD_PROD) series at level shows increasing trend over the period of time. It means the series is not stationary. The correlogram (Fig-10) at level for ACF and PACF of the series shows inappropriate lags pattern and values of the ACF and PACF, so, it concludes the series is not stationary.

(v) Results:-

Augmented Dickey-Fuller (ADF) Unit Root Test Log (PHD_PROD) series at 1st Difference

Null Hypothesis: D(LOG(PHD_PROD)) has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.985469	0.0000
Test critical values:	1% level	-3.565430	
	5% level	-2.919952	
	10% level	-2.597905	

*MacKinnon (1996) one-sided p-values.

Table-8

Time Plot of log (PHD_PROD) series at 1st Difference

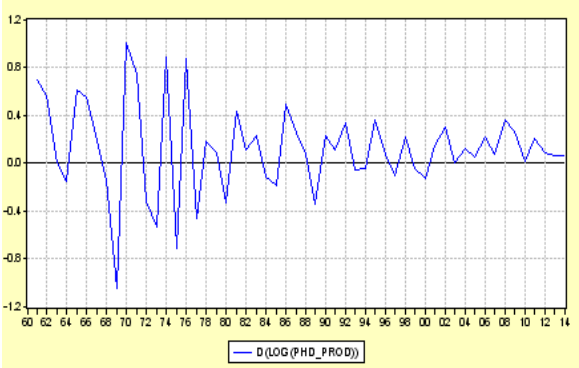
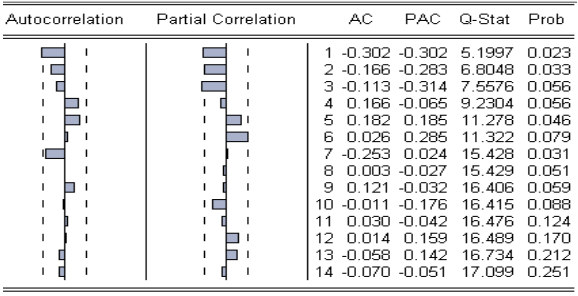


Figure-11

Correlogram log (PHD_PROD) series at 1st Difference

Correlogram of D(LOG(PHD_PROD))



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.302	-0.302	5.1997	0.023
		2	-0.166	-0.283	6.8048	0.033
		3	-0.113	-0.314	7.5576	0.056
		4	0.166	-0.065	9.2304	0.056
		5	0.182	0.185	11.278	0.046
		6	0.026	0.285	11.322	0.079
		7	-0.253	0.024	15.428	0.031
		8	0.003	-0.027	15.429	0.051
		9	0.121	-0.032	16.406	0.059
		10	-0.011	-0.176	16.415	0.088
		11	0.030	-0.042	16.476	0.124
		12	0.014	0.159	16.489	0.170
		13	-0.058	0.142	16.734	0.212
		14	-0.070	-0.051	17.099	0.251

Figure-12

Interpretation:

Mathematical assessment through ADF Unit Root Test results (Table-8) at 1st difference of the series (Log (PHD_PROD) shows that p=0.000, it seems the series is stationary.

Visual inspection of Time Plot (Fig-11) for log (PHD_PROD) series at 1st difference shows that shift in central tendency and dispersion is decreasing over the period of time. It indicates that the series is stationary. The correlogram (Fig-12) at 1st difference for ACF and PACF of the series Log (PHD_PROD) shows that lags pattern and values of the ACF and PACF are suitable, so, it appears that the series is stationary.

The above discussion and results it exposed that the Log (PHD_PROD) series at 1st difference is suitable ARIMA model fitting. So, the PhDs output series of Log (PHD_PROD) at 1st difference was used in the ARIMA forecast modeling.

3. (b). Identification process:

Identification process of a parsimonious ARIMA model is a fun. In which, the correlogram (Fig-13) of the Log (PHD_PROD) series at 1st difference (stationary time series) was used, because all other Correlograms, shown by Fig-5, Fig-7, Fig-9 and Fig-11, indicate that there values are not within limits. This correlogram, almost spikes of both ACF and PACF at different lags were in limits (95% confidence limits). The values of AC (ACF), PAC (PACF), Q-Stat and Prob. (P) were reasonable. So, from the series log (PhD_Prod) at 1st difference following suitable terms were identified as shown by Table-9.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

Appropriate values of p and q for the ARIMA models of PhDs outputs of the HES of Pakistan logarithmic series data:-

p and q values for the ARIMA models of PhDs_outputs											
Sr.#	ARIMA			Sr.#	ARIMA			Sr.#	ARIMA		
	p	d	q		p	d	q		p	d	q
1	3	1	3	3	1	1	2	5	1	1	0
2	2	1	2	4	1	1	1	6	0	1	1

Table-9

3. (c) Estimation process:

After identification of the suitable p, d, and q terms, the estimation of different parameters were carried out. The estimated parameter with goodness of fit of the identified ARIMA (p, d, q) forecast models are tabulated the in Table-11.

3. (d) Validation of the Parsimonious ARIMA Model:-

Now, it is quite easy to validate and choose a parsimonious ARIMA model from the estimated parameter given in Table-11 by applying the diagnostic checks of Table-3. In light of the diagnostic checks, the values of the parameters of all the models given the Table-10 it shows that the ARIMA (2,1,2) model is most parsimonious model for forecast of the PhDs outputs of the HES of Pakistan. Because in this model:-

(i) All the coefficients are statistically significant.
(ii) All coefficients are stationary and invertible.
(iii) Residuals are independent and normally distributed.
(iv) Durban Watson is less than 4 and close to 2.
(v) There is no serial correlation and ARCH effect.
(vi) The AIC, SC, RMSE and TIC values are appropriate.
(vii) There is no pattern in the Ljung-Box statistics of ACF & PACF.

Beside this there is no model in the Table-11 that has fulfilled the maximum diagnostic checks of the Table-3.

4. (e) Forecasting of the suitable model.

By putting the values of estimated parameters in general form of equation, our suitable ARIMA model for forecasting is as under:

$$\text{ARIMA}(2,1,2): Y_t = 0.05771 + 0.6139Y_{t-1} - 0.7839Y_{t-2} + \varepsilon_t + 1.1336\varepsilon_{t-1} - 0.9580\varepsilon_{t-2}$$

The ARIMA(2,1,2) model ACF of residual, PACF of Residual and four residuals in aggregate at 1st difference for the logarithmic PHD_PROD series i.e. D(LOG(PHD_PROD)) are shown by Fig-13, 14 and 15 respectively. The ARMA structure at 1st difference for inverse roots of AR/MA polynomial shown by Figure-16 indicates that there is no root outside the circle. This clarified that the model is stationary and suitable for forecast.

Consequently, the forecast values from the suitable ARIMA model are calculated for next two decades (years: 2015-2035) and are given in Table-11. The forecasting results it indicates that the PhDs output of HES of Pakistan will be 12176 in 2035. The HES outputs on 95% confidence interval, it displays that the PhDs-output would be 1672 (lower limit) and 88689 (upper limit) in the year-2035.

Graphical representation of the projected & predicted trend at 1st difference for the LOG (PHD_PROD) is shown by Figure-17.

ACF of residual

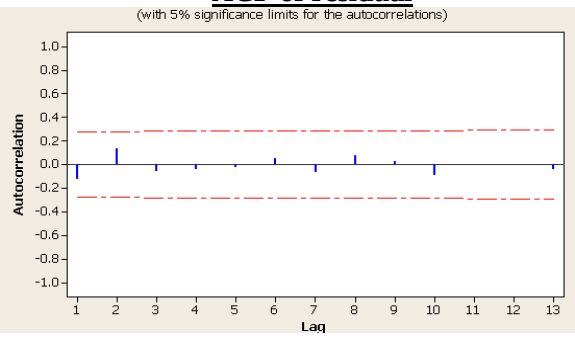


Figure-13

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

PACF of Residual

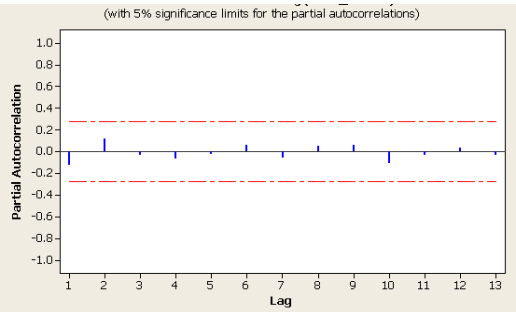


Figure-14

ARMA structure

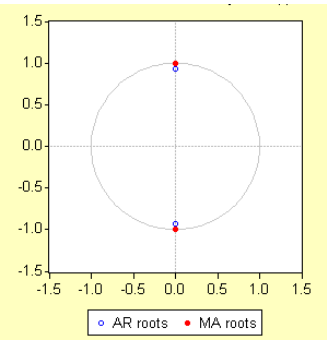


Figure-16

Four Residuals in Aggregate

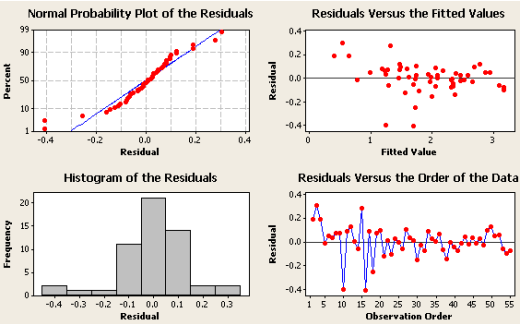


Figure-15

Projected & Predicted Trend

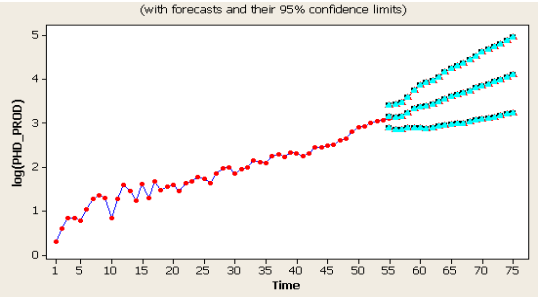


Figure-17

Year	Projection of PhDs Output in Pakistan with logarithmic value	Projection of PhDs output in Pakistan without logarithmic value	The projected PhDs output lower limit	The projected PhDs output upper limit
2015	3.14159	1385	758	2534
2016	3.13187	1355	693	2647
2017	3.14832	1407	699	2833
2018	3.22376	1674	750	3738
2019	3.31489	2065	766	5566
2020	3.36942	2341	755	7259
2021	3.38918	2450	743	8075
2022	3.41628	2608	762	8924
2023	3.47514	2986	822	10847
2024	3.54775	3530	885	14073
2025	3.60390	4017	923	17484
2026	3.63917	4357	945	20077
2027	3.67452	4726	984	22707
2028	3.72629	5325	1056	26859
2029	3.78808	6139	1142	32997
2030	3.84314	6969	1216	39950
2031	3.88623	7695	1275	46448
2032	3.92723	8457	1343	53245
2033	3.97635	9470	1440	62281
2034	4.03207	10766	1557	74462
2035	4.08550	12176	1672	88689

Table-10

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

ARIMA General Model: $Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$							
Parameters		ARIMA(3,1,3)	ARIMA(2,1,2)	ARIMA(1,1,2)	ARIMA(1,1,1)	ARIMA(1,1,0)	ARIMA(0,1,1)
C (Intercept)	Coef	0.09324	0.05771	0.055835	0.039425	0.06676	0.0473
C Prob.(P-Valve)		0.000	0.000	0.000	0.000	0.003	0.000
AR	Coef. $\emptyset_1$	0.0842	0.6139	-0.2017	0.1509	-0.3151	-
	P Valve	0.817	0.000	0.822	0.456	0.020	-
	Coef. $\emptyset_2$	-0.8519	-0.7839	-	-	-	-
	P Valve	0.000	0.000	-	-	-	-
	Coef. $\emptyset_3$	-0.1993	-	-	-	-	-
	P Valve	0.520	-	-	-	-	-
MA	Coef. θ_1	0.6764	1.1336	0.3631			-
	P Valve	0.060	0.000	0.676	0.000	-	0.6608
	Coef. θ_2	-0.7251	-0.9580	0.2736	-	-	-
	P Valve	0.021	0.000	0.627	-	-	-
	Coef. θ_3	0.3118	-	-	-	-	-
	P Valve	0.236	-	-	-	-	-
AIC		0.678203	0.709010	0.512706	0.508379	0.853185	0.575541
SC(BIC)		0.791839	0.821582	0.624232	0.619905	0.927536	0.649208
DW		2.635539	2.797646	1.524287	1.974019	2.228861	1.466982
R Square		0.278042	0.240199	0.379966	0.382644	0.094951	0.330560
Log Likelihood		-14.29417	-15.43425	-10.58670	-10.47204	-20.60941	0.316848
LM Test		0.020124	0.002371	0.092780	0.866130	0.004012	0.096450
ARCH test		0.071460	0.002576	0.420633	0.804482	0.210512	0.573925
JB Statistics value		7.087902	4.599458	1.516173	1.178208	9.979573	0.675787
JB Statistic P-Value		0.028899	0.100286	0.468562	0.554824	0.006807	0.713271
RMSE		0.372120	0.401998	0.465898	0.373045	0.380019	0.380917
MAE		0.270695	0.273881	0.313663	0.266976	0.273407	0.275045
MAPE		108.3113	103.9639	105.4738	95.91464	104.1298	96.61484
TIC		0.687561	0.690288	0.773816	0.765220	0.756944	0.775384

Table-11

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)

4. Conclusion and Recommendations

The analysis results show that the ARIMA(2,1,2) model is a suitable forecast model. This model projects that the PhDs output of HES of Pakistan would be reached to 12175 PhDs per annum in the subsequent two decades (in year_2035).

By comparison, the forecasted PhDs outputs of the ARIMA(2,1,2) for year 2017 is 1407, and is closed to 1500 of the HEC of Pakistan planned PhDs for the year 2007. This prediction shows Pakistan is delay at that time about for three years from the target of 2007 that was planned in 2002 to achieve it in next fivers [8].

This situation of the PhDs outputs in Pakistan demands reengineering of the HES to move it beyond the minimal outputs (the system that currently working) to a maximal outputs (to a system that promote the production of PhDs outputs positively) and keep the R&D professional vital [25].

Beside that this situation compels the HES management and policy makers to forge ahead on the current universities infrastructure and systems for PhD studies of PhDs outputs in Pakistan.

The reengineering process of the current universities education Pakistan is not a matter to be left alone to managements. It needs such R&D professionals that not only change the HES policies but they turn the HES of Pakistan in a practical and result oriented R&D system according the contemporary era. So, that to evolve the PhDs outputs in Pakistan in such a way to meet the needs of Pakistan as well as they serve in other countries to commute in bilateral revenues.

Thus, the forecast of the PhDs production of HES is very significant for policy makers and policy making in Pakistan. The forecast of the PhDs is equally important for the researchers and organizations that are involved anywhere in the future planning.

References

[1] Chase, R. b., & et.al. (2006). Operations Management (11th Edition ed.). McGraw-Hill.

[2] Agrawala, R., Nanda, S., Rao, D. R., & Rao, B. (2013). Integrated Approach to Human Resource Forecasting. Agricultural Economics Research Review , 26 (2), 173-184.

[3] Ozay, M. (1977). The role of manpower planner in developing countries. Industrial Relations Journa , 7 (3), 66-74.

[4] Psacharopoulos, G. (1984). Assessing training priorities in developing countries. International Labour Review , 5, 569-583.

[5] Psacharopoulos, G. (1991). From manpower planning to labour market analysis. International Labour Review , 130 (4), 459-474.

[6] Willems. (1996). Manpower Forecasting and Modelling Replacement Demand. Maastricht: An Overview.Research Centre for Education and the Labour Market, Maastricht University, Maastricht.

[7] MIED. (1947). Doc ID 166, Proceedings of the First Educational Conference. Karachi: Ministry of the Interior Education Division, Government of Pakistan.

[8] HEC. (2003). HEC Annual Report 2002-2003. Islamabad: HEC Printshop.

[9] HEC. (2015). PhD Country Directory. Retrieved from PhD Country Directory, HEC, Islamabad, Pakistan: http://www.hec.gov.pk/InsideHEC/Divisions/QALI/Others/Statistics/Pages/YearWiseSummary2.aspx

[10] The Group of Eight. (2013). The Changing PhD. Northbourne, Australia: The Group of Eight.

[11] Sharova, A. (2013). Chinese PhDs vs US PhDs, INOMICS Blog. Retrieved from blog.inomics.com: http://blog.inomics.com/chinesephdsvsus-phds/

[12] Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A., & Yahia, M. (2011). The PhD Factory. Nature, International Weekly Journal of Science , 276-279.

[13] HEC. (2009). Annual Report 2008-2009. Islamabad: Higher Education Commission - Pakistan.

[14] Sheikh, M. A. (2014). 60 % Pakistani Universities Fail to Produce PhD Scholars. Retrieved from worldtimes.com.

http://jworldtimes.com/Page/60_Percent_Pakistani_Universities_Fail_to_Produce_PhD_Scholars

[15] ASC. (2010). The Role of PhDs in the Smart Economy. Advisory Science Council for Science, Technology & Innovation.

[16] Kurup, A., & Arora, J. (2010). Trends in Higher Education. Bangalore: National Institute of Advanced Studies.

[17] Shryock, H., & Seigel, J. (1973). The Methods and Materials of Demography. Washington: U.S. Bureau of the Census, Governoment Printing Office.

[18] Agarwal, U. (2000). *Population Projections and their Accuracy.* Dehli: B.R. Publication Corporation.

[19] Jan, B., Ali, I., & Shah, S. (2007). Selecting of Mathematical Model for Projections of NWFP Population. The Journal of Humanities and Social Sciences , XV (2), 69-78.

[20] Zakria, M., & Muhammad, F. (2009). Forecasting the Population of Pakistan using ARIMA Models. Pak J. Agri. Sci. , 46 (3), 214-223.

[21] Box, E. P., & Jenkins, M. G. (1976). Time Series Analysis: Forecasting and Control (Revised ed.). Oakland: CA: Holden-Day.

[22] Verbeek, M. (2005). A guide to modern econometrics. England: John Wiley & Sons.

[23] QMS, LLC. (2005). EViews 5 User's Guide. EViews 5 User's Guide . Irvine:CA, US: Quantitative Micro Software, LLC, United States of America.

[24] Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). New York: McGraw-Hill/Irwin.

[25] Salvendy, G. (2001). Handbook of Industrial Engineering: Technology and Operations Management. New York: John Wiley & Sons, Inc.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-9/MRR.19)