

Forecasting the population growth and wheat crop production in Pakistan with non-linear growth and ARIMA models

Muhammad Islam ¹, Farrukh Shehzad ², Samrat Ray ³, Mirza Waseem Abbas ¹

¹ Crop Reporting Service, Punjab, 58180, Pakistan

² The Islamia University of Bahawalpur, Bahawalpur, 63350, Pakistan

³ IIMS, Pune, 411001, India

Received 4 February 2023 ♦ Accepted 26 March 2023 ♦ Published 18 December 2023

Citation: Muhammad Islam, Farrukh Shehzad, Samrat Ray, Mirza Waseem Abbas (2023) Forecasting the population growth and wheat crop production in Pakistan with non-linear growth and ARIMA models. Population and Economics 7(3):172-187. <https://doi.org/10.3897/popecon.7.e101500>

Abstract

Food security as a major social concern and a global threat, requires better policy decisions based on empirical studies. This work presents a comparative statistical analysis of different methods to forecast wheat area, productivity, production, and population growth rate in Pakistan. Time series data from 1950 to 2020 were analyzed using various methods such as ARIMA, the compound growth exponential regression model (CGREM), Cuddy Della Valle instability index (CDVI), and decomposition analysis. The results show that CGREM performs better than other models. Periodic compound growth rates indicate that wheat area and yield decrease by about 67.0% and 40.0%, while the population decreases by 31.7%. For the period 2001-2020, the compound growth reaches the level of 0.60% for wheat area, 1.21% for yield, while it is high for the population and amounts to 2.22%. The overall compound growth rate for wheat area and yield (about 1.207%, 2.326%) is lower compared to the population (about 2.839%). The paper presents forecasts for wheat area, yield, and population in Pakistan will rise: 12.7%, 25.5%, 31.8% in 2030 and 43%, 97.8%, and 129% in 2050. The results of this study provide empirical evidence for the necessity of policy decisions addressing the problem of food security in Pakistan.

Keywords

CGREM, decomposition, growth rate, population and wheat crop, food security

JEL codes: C50; C51; C52; C53; J10; Q18; Q19

Introduction

1.1 Motivation and significance of the study

Food security has emerged as a major global concern. According to the Economic Survey of Pakistan 2019-2020, the agricultural growth rate was recorded at 4.0% in 2017-2018, 0.6% in 2018-2019, and approximately 2.7% in 2019-2020 (ESP 2020). The population growth rate in Pakistan was about 2.40% in 2018-2019 and 2017-2018, while it was 1.89% in 2016-17, indicating a rising short-term trend in population dynamics that could have adverse social effects on community welfare (ESP 2019). Pakistan's population growth rate is high compared to other South Asian countries, potentially leading to societal food conflicts (Islam 2022; Islam & Shehzad 2022). The rapid population growth necessitates a significant increase in production to meet the nutritional needs of the population.

The share of the agricultural sector is decreasing while the population is steadily increasing each year in Pakistan (Kumbhar et al. 2018). According to the United Nations Millennium Development Goals, food security has become a major social concern due to the current scenarios of population expansion. Given the direct relationship between food requirements and population growth, addressing this social dilemma is essential in both policy and research (Ahmad et al. 2017). The United Nations Economic and Social Commission for Asia and the Pacific (ESCAP UN 2009) reports that in 2005-2006, 16% of the world's population needed food, reaching up to 21%, especially in South Asian countries. With the expected population growth rate, Pakistan is forecast to be the 5th in the world in 2050 (compared to the current 6th place), potentially leading to disparities in food supply channels (Shah & Khalil 2017).

The work of the English economist R.T. Malthus (1766-1834), written in 1798, postulates, «...population growth will always tend to outrun the food supply and that betterment of humankind is impossible without strict limits on reproduction...» In the near future, the world cannot avoid issues of food scarcity (Abdulrahman 2013; Malthus 1986). Malthus was the first to address the issue of food scarcity, arguing that the rising global population would ultimately exceed the earth's capacity to feed it. Producing sufficient amounts of food for the growing population has become one of the prime social challenges globally. It is evident that the world is continuously becoming more populated, and various measures have been taken to address this issue.

World Health Organization (WHO), Food and Agricultural Organization (FAO), and many other international and national organizations address the issue of food insecurity (Abdulrahman 2013; Nelson et al. 2010). Mekuria (2018) reports that a growing population is a threat to food security, which can be addressed by increasing productivity. The global population is estimated to reach 9.0 billion by 2050, primarily contributed to by developing countries (Lutz et al. 1997; Nelson et al. 2010). Food production must increase by about 70% to eliminate the problem of food insecurity in developed countries and double that figure for the developing world (Kagan 2016). Pakistan, being a predominantly agricultural country, faces various food security concerns due to the highest population growth rate, especially among Asian countries. Water scarcity and adverse climate variations have also contributed to critical food concerns. Food availability is a basic pillar of food security, and for Pakistan, wheat crops stand first in terms of acreage and production. Pakistan ranks 99th out of 121 countries and has a serious level of hunger, according to the Global Hunger Index data for 2022. Tariq et al. (2014) assessed that due to the rapid population growth and climatic risks, the wheat crop per capita in Pakistan was 198 kg in 2014, and it is expected to be 105 kg

in 2031 and 84 kg in 2050. Pakistan's agriculture sector has growth potential, and a special comprehensive program for the development of different economic sectors should be implemented to ensure food security.

Various researchers, including the International Food Policy Research Institute (IFPRI), FAO, and many organizations, hypothesize that the high population growth trend, especially in South Asian countries, may lead to food conflicts in the region. Some African countries have been affected by severe food shortages. Food security means limited food availability and restricted access to food in society. The major concern of the current analysis is to statistically assess discrepancies between explosive population growth and productivity abilities in the context of the social dilemma of food security.

1.2 Problem statement and objective of the study

For a developing society, it is essential to assess future changes in population and agricultural production using statistical forecasting models with reliable and precise tools. A relevant and accurate statistical model enables us to make strong and accurate predictions for the future, supporting policy decisions in the sphere of the food security threat. The appropriate application of statistical crop yield prediction models is essential to assess how agricultural production responds to future challenges (Lobell & Burke 2010). This study aims to determine, evaluate, and compare the magnitude of changes in the population growth rate and food crop (wheat) availability and sustainability in Pakistan from 1950 to the present, addressing the prevailing social dilemma of food insecurity. The research employs statistical models to forecast wheat area, yield, and production for 2030 and 2050, along with a comparative analysis of population growth. Robustness measures, instability checks, and decomposition analysis with a comparison of compound growth rates are also applied using time series data covering the period 1950-2020.

Data and methods

2.1 Data collection and measuring scales

Secondary time-series data for the period 1950-2020 is collected for 71 years from the Pakistan Bureau of Statistics, Punjab Agriculture Marketing Information Service Department, Punjab Crop Reporting Service Agriculture Department, and various issues of the Economic Surveys of Pakistan, covering wheat crop area, productivity (yield), production, and population. These organizations are owned by the Government of Pakistan, providing reliable and authentic data sources for researchers worldwide. Four variables are employed: wheat crop area, productivity (yield), production, and population of Pakistan, measured as area in thousand acres ('000' acres), average yield of 40 kg. per acre (mds/acre), production in thousand tons ('000' tons), and population in millions, respectively.¹ R programming language and SPSS software are used to analyse the data.

¹ No modification has been carried out. Moreover the official links to the web-sites, from where data is collected as under: for wheat area: <http://www.amis.pk/Agristatistics/Data/HTML%20Final/Wheat/Area.html>, for wheat yield: <http://www.amis.pk/Agristatistics/Data/HTML%20Final/Wheat/Yeild.html>, for wheat production: <http://www.amis.pk/Agristatistics/Data/HTML%20Final/Wheat/Production.html>, for population in millions various issues of Economic Survey of Pakistan from its official web-site of Govt. of Pakistan Finance Division at https://www.finance.gov.pk/survey_2023.html

2.2 Time series Box-Jenkins methodology (ARIMA)

Box and Jenkins (1976) introduced the auto regressive integrated moving average (ARIMA) technique, based on a systematic iterative process used to select the best model for predicting unknown parameters. It involves a step-by-step estimation of parameters, identification, diagnostic checks, and finally, predicting unknown parameter values. Model fitting consists of determining three parameters: the order of the autoregressive component (p), the degree of differencing order (d), and the order of the moving average (q) (Makridakis & Hibon 1997; Sharma et al. 2009).

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (1)$$

where “ ϕ ” and “ θ ” are the autoregressive and moving average parameters respectively, “ X ” is the original series and the “ e ” is a series of the normally distributed residuals.

The time series Box-Jenkins Methodology (ARIMA) is used to predict wheat area, yield and production, and population of Pakistan based on data collected from 1950 to 2020¹.

2.3 Compound growth exponential regression model (CGREM)

The Compound Growth Exponential Regression Model (CGREM) is applied simultaneously with the ARIMA model to find the better model for wheat crop area, productivity, production, and population growth. The annual compound growth rates are measured by the following model (Dhakre & Sharma 2010; Kondal 2014):

$$y_t = y_0 [1 + r]^t \quad (2)$$

where “ y_t ” depicts the wheat area/yield/production/population, “ t ” denotes time period, “0” – initial year, “ r ” is the compound growth rate. The slope measures the relative change in the response variable, i.e. the absolute change accrued in the feature, and it measures the instantaneous rate of growth. Apply the log-transformation to the equation 2.

$$\ln(y_t) = \ln(y_0) + t\{\ln(1 + r)\}$$

$$\ln(y_t) = Y, \ln(y_0) = A, \ln(1 + r) = B$$

$$Y = A + Bt + \varepsilon, r = (\exp^B - 1) \cdot 100$$

To predict the parameter the following equation is applied.

$$y_p = [y_c(1 + B)^n] \quad (3)$$

where, “ y_p ” is the value of the response variable at the projected time, “ y_c ” denotes the actual/collected value of the response at time “ t ”, “ B ” is the regression slope of the line or regression coefficients, “ n ” is the total number of years (projection horizon), i.e. “ $t_p - t_c$ ”.

¹ The dataset is used as a training set.

2.4 Cuddy-Della-Valle instability index

Cuddy-Della-Valle instability index (CDVI) was developed by Cuddy and Valle in 1978 to measure the instability in time series data which is related to the trend specifics. CDVI attempts to adjust the coefficient of variation (CV) using R^2 . A low value of the CDVI indicates the low instability and vice-versa (Bezabeh 2016; Sihmar 2014).

$$CDVI = C.V \times (\sqrt{1 - R^2}) \quad (4)$$

$$C.V = \frac{s}{\bar{x}} \times 100 \quad (5)$$

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad (6)$$

where “C.V” stands for the coefficient of variation. The range of instability is as follows: low instability for $0 \leq CDVI \leq 15$, medium instability for $15 < CDVI \leq 30$ and high instability for $CDVI > 30$.

2.6 Decomposition analysis model

Production is a functional form of area and yield. The variation in production is due to changes in area and productivity. The relative contribution of area and productivity to the change in production is estimated by the decomposition analysis model (Dhakre & Sharma 2010; Murindahabi et al. 2018; Rehman et al. 2011). The change in production is represented as the sum of the following three effects: area effects, yield effects, and their interaction term.

Change in production = Yield effects + Area effects + Interaction effects

$$\Delta P = A_0 \Delta Y + Y_0 \Delta A + \Delta Y \Delta A \quad (7)$$

or

$$\text{Change in production} = \frac{(Y_c - Y_0) \times A_0}{P_c - P_0} \times 100 + \frac{(A_c - A_0) \times Y_0}{P_c - P_0} \times 100 + \frac{(Y_c - Y_0) \times (A_c - A_0)}{P_c - P_0} \times 100 \quad (8)$$

where Δ is for a change over time, “ P_0 ” – production at “ $x_{t=1950}$ ”, “ P_c ” – production at “ $x_{t=2020}$ ”, “ Y_0 ” – yield at “ $x_{t=1950}$ ”, “ Y_c ” – yield at “ $x_{t=2020}$ ”, “ A_0 ” – area at “ $x_{t=1950}$ ”, “ A_c ” – area at “ $x_{t=2020}$ ”.

2.7 Goodness of fit for better fitted model

Regression modeling is a crucial task in the applied statistical analysis. Gujarati and Porter (2009) highlight that crop yield depends on various predictors which makes the best regression model selection challenging. We address the following criteria of the fitting quality.

1. Lower value of mean square error (MSE) and higher value of coefficient of determination (R -square).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad (10)$$

$$R_{adj}^2 = 1 - \left(1 - R^2\right) \left(\frac{n-1}{n-k}\right) \quad (11)$$

2. Significance of the regression coefficient and the significance of the overall model which are determined by the t-statistic and F-statistic, respectively.

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \quad (12)$$

$$F = \frac{SSR_{eg} / k - 1}{SSR_{es} / n - k} \quad (13)$$

3. Lower value of the Akaike (1973) Information Criteria (Banks & Joyner 2017; Dziak et al. 2012; Gujarati 2003).

$$AIC = n \left\{ \ln \left(\frac{RSS}{n} \right) \right\} + 2k \quad (14)$$

where, “ k ” is a number of regressors including an intercept and “ n ” is a number of observations. The term “ $2k/n$ ” is defined as a penalty factor for AIC.

4. Lower value of the Schwarz (1978) information criterion (Gujarati 2003; Neath & Cavanaugh 1997).

$$SIC = n \left\{ \ln \left(\frac{RSS}{n} \right) \right\} + k \ln(n) \quad (15)$$

where, the term “ $((k/n) \ln(n))$ ” is defined as a penalty factor. “ k ” is a number of regressors including the intercept and “ n ” is a number of observations.

Results and discussions

3.1 Box-Jenkins ARIMA techniques

The R programming language is utilized to analyse seventy-one years of data from 1950 to 2020. The Hyndman-Khandakar algorithm (Hyndman & Khandakar 2008) is employed to select optimal predictive models for wheat crop area, productivity (yield), production, and population variables in Pakistan. The paper demonstrates plots of the auto-correlation function (ACF), partial auto-correlation function (PACF), and the plot of the differenced series. ACF indicates the value of parameter ‘ q ’ while the PACF indicates the value of parameter ‘ p ’.

Figures 1-4 display box plots, illustrating that the dataset is free from outliers for wheat area, yield, production, and population in Pakistan, respectively. Figures 5-8 present plots of the original data series, indicating an upward trend behavior for wheat area, yield, production, and population. It is evident that the series cannot be characterized as stationary, and the parameter should not be less than $d=1$. Figures 9-12 depict plots of the ACF, and Figures 13-16 show plots of the PACF with confidence bands for the same variables. Actual and predicted values of the time series are shown in Figures 17-20 for wheat area, yield, production, and population, respectively. Figures 21-24 illustrate out-sample forecast values for the fitted ARIMA with 80% and 95% confidence intervals (CI). Figures 25-28 display the

ACF for residual analysis, and Figures 29-32 show the PACF for residual analysis for wheat area, yield, production, and population.

There is no problem of severe autocorrelation in the residuals of the fitted ARIMA models. Table 1 outlines the best models for the variables as follows: ARIMA (0, 1, 0) for wheat area, ARIMA (1, 1, 2) for productivity, ARIMA (0, 1, 1) for production, and ARIMA (0, 2, 1) for population. The RMSE for the predictive models is found to be 553.3, 1.03, 996.7, and 0.46. Box-Ljung tests illustrate that there are no autocorrelation problems for the fitted ARIMA models.

Table 1. Detail of estimated parameters for ARIMA models

Determinants	ARIMA	AR(1)	MA(1)	MA(2)	AIC	SIC	RMSE	Box-Ljung p value
Wheat Area	(0, 1, 0)	----	----	----	1087.8	1092.3	553.3	0.43
Wheat productivity	(1, 1, 2)	0.943	-1.72	0.84	213.8	222.8	1.03	0.87
Wheat production	(0, 1, 1)	----	-0.57	----	1168.4	1175.1	996.7	0.55
Population	(0, 2, 1)	----	-0.60	----	95.5	99.7	0.46	0.98

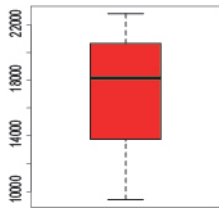


Figure 1. Box plot for wheat area

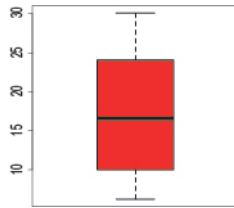


Figure 2. Box plot for wheat yield

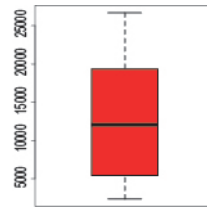


Figure 3. Box plot for wheat production

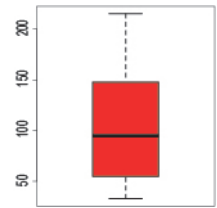


Figure 4. Box plot for population

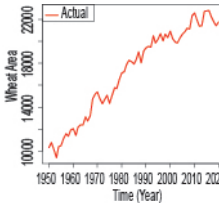


Figure 5. Time plot for wheat area

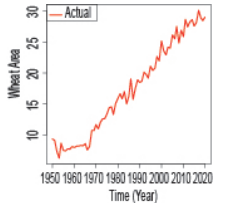


Figure 6. Time plot for wheat yield

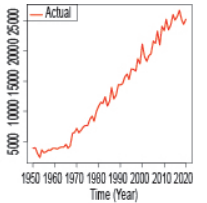


Figure 7. Time plot for wheat production

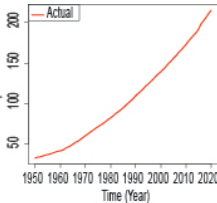


Figure 8. Time plot for population

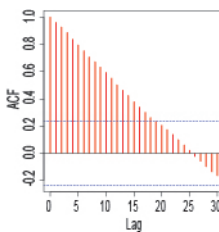


Figure 9. Plot of ACF for wheat area

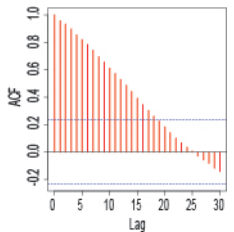


Figure 10. Plot of ACF for wheat yield

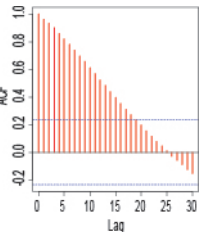


Figure 11. Plot of ACF for wheat production

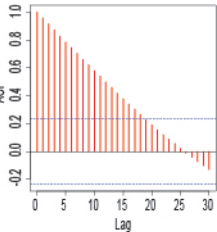


Figure 12. Plot of ACF for population

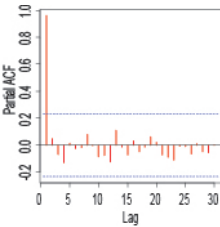


Figure 13. Plot of PACF wheat area

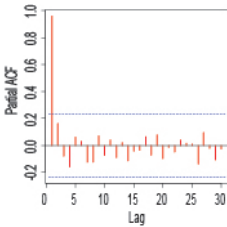


Figure 14. Plot of PACF wheat yield

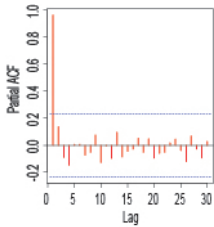


Figure 15. Plot of PACF wheat production

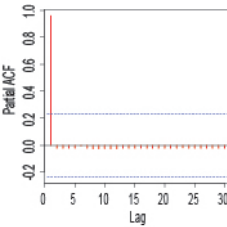


Figure 16. Plot of PACF population

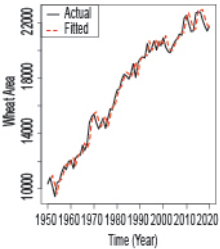


Figure 17. Plot of actual & predicted values ARIMA(0,1,0) for wheat area

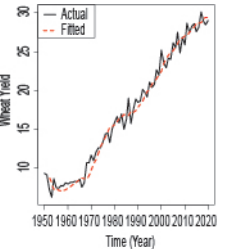


Figure 18. Plot of actual & predicted values ARIMA(1,1,2) for wheat yield

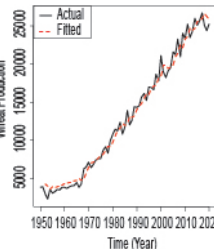


Figure 19. Plot of actual & predicted values ARIMA(0,1,1) wheat production

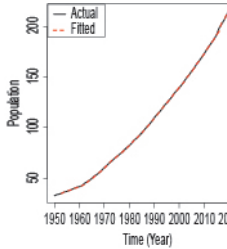


Figure 20. Plot of actual & predicted values ARIMA(0,2,1) for population

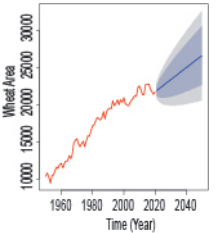


Figure 21. Forecasted area CI 80%, 95%

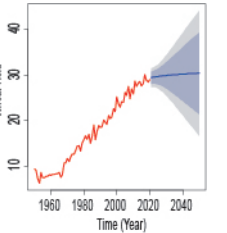


Figure 22. Forecasted yield CI 80%, 95%

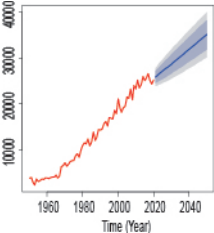


Figure 23. Forecasted production CI 80%, 95%

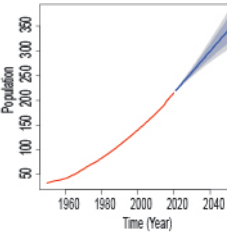


Figure 24. Forecasted population CI 80%, 95%

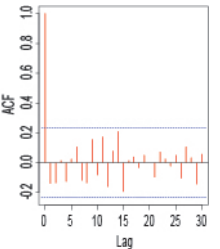


Figure 25. ACF (residual) wheat area

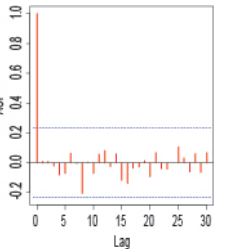


Figure 26. ACF (residual) wheat yield

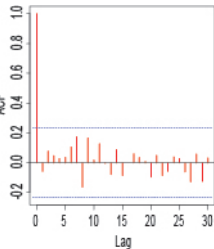


Figure 27. ACF (residual) wheat production

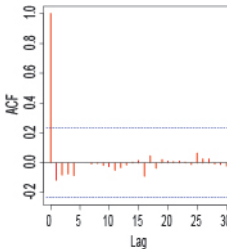


Figure 28. ACF (residual) population

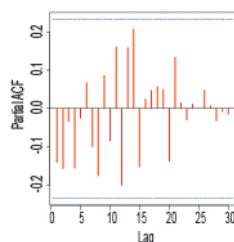


Figure 29. PACF (residual) wheat area

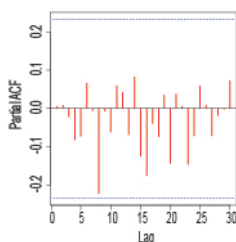


Figure 30. PACF (residual) wheat yield

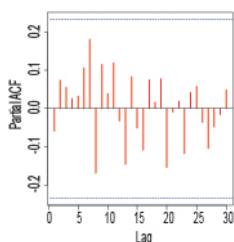


Figure 31. PACF (residual) wheat production

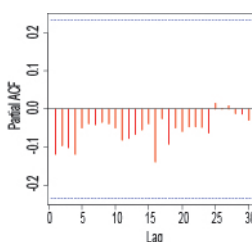


Figure 32. PACF (residual) population

3.2 Compound growth exponential regression model combined with ARIMA

Table 2 presents the performance measures for the Compound Growth Exponential Regression Model (CGREM). The values of adjusted R-square are high, and RMSE, AIC, and SIC tend to be low compared to baseline models. The regression coefficient and Compound Growth Rate (CGR) for area are found to be 0.012 and 1.207%, for productivity 0.023 and 2.326%, for production 0.034 and 3.458%, and for population 0.028 and 2.839%. The results are significant for all models and their regression coefficients.

The CGR models appear to be better-fitted models for time series analysis compared to the ARIMA models.

Production is a product function of area and yield, meaning that any change in the area and yield variables will automatically change production. They are driving factors of production that need to be predicted due to concerns about food availability in the future. Figures 33-35 indicate the patterns of the predicted CGREM and actual values for wheat area, yield, and population. We conclude that the area, yield, and population will increase by about 12.7%, 25.5%, and 31.8% up to 2030, and 43%, 97.8%, and 129% up to 2050. The figures show that the increase in population is estimated to be larger than the same increase in wheat area and yield.

Table 2. Compound growth exponential regression models and goodness of fit

Depen- dents	AdjR ²	RMSE	Sig	AIC	SIC	Regres- sion coeffi- cient	CGR	2020	2030 (Est)	2050 (Est)
Area	0.92	0.073	782.5**	-368.6	-364.4	0.012**	1.207%	21750.62	24506.2	31109.0
Yield	0.95	0.108	1302.5**	-313.0	-308.5	0.023**	2.326%	29.02	36.43	57.41
Pro- duction	0.95	0.159	1395.3**	-258.2	-253.2	0.034**	3.458%	25249.50	----	----
Popula- tion	0.99	0.051	8896.5**	-419.6	-415.9	0.028**	2.839%	215.25	283.70	492.87

** shows the area/productivity/production and population are significant at 5% level of significance

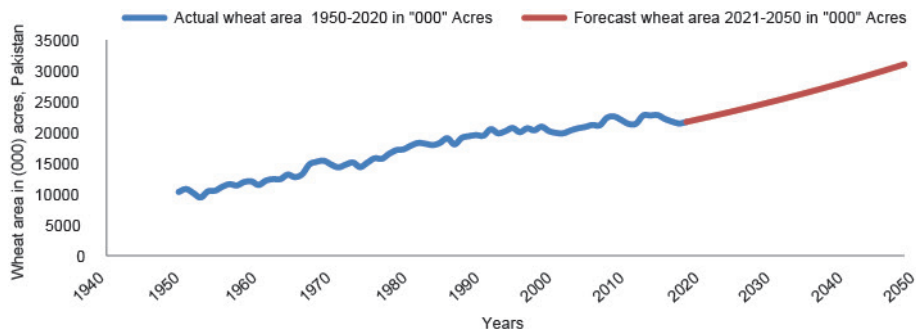


Figure 33. Actual and predicted wheat area in thousand acres

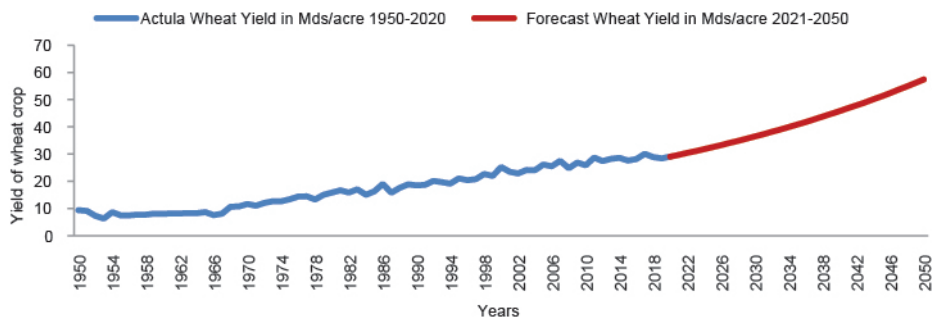


Figure 34. Actual and predicted wheat yield in Mds/acre

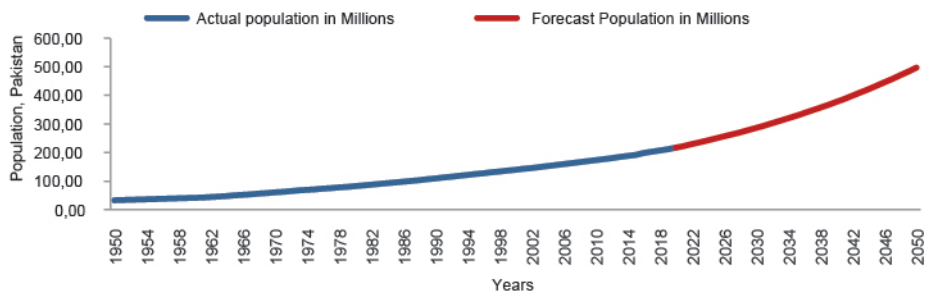


Figure 35. Actual and predicted Population in millions

3.3 Cuddy Della Valle instability index and compound growth rate

Table 3 indicates positive growth rates per annum for area (about 1.207%), for productivity (2.326%), for production (3.458%), and for population (2.839%). The relative Compound Growth Rate (CGR) gap reports that the population is growing at the rate of 135.2%, and 22.05% more than wheat area and yield up to 2050. The Cuddy Della Valle Index (CDVI) for investigating instability in area, productivity, and production is found to be 5.26%, 8.66%, 10.38%, and 9.08%, respectively. The coefficient of variation indicates high inconsistency for wheat production. The overall degree of instability lies in the low instability CDVI indexing region.

Table 3. Compound growth rate (CGR), Cuddy Della Valle instability index (CDVI) and coefficient of variation (CV) for wheat area, yield production and population

Items	Wheat Area	Wheat Productivity	Wheat Production	Population
CGR (%)	1.207	2.326	3.458	2.839
CDVI (%)	5.26	8.66	10.38	9.08
C.V (%)	23.52	43.3	59.97	52.46

3.4 Sporadic variations in compound growth rate

To study sporadic (periodic) variations in wheat area, yield, production, and population, the data is divided into three subsamples for the periods 1950-1975, 1976-2000, and 2001-2020. Table 4 and Figure 36 illustrate the periodic variation in growth rates of wheat area, yield, production, and population in Pakistan. They indicate that the Compound Growth Rate (CGR) of wheat area is 1.82%, 1.21%, and 0.60% during the specified periods, respectively. For other variables, these figures are 2.02%, 2.12%, 1.21% for wheat yield, and 3.87%, 3.36%, 1.71% for wheat production. The data reveals that the CGR for the population also exhibits a downward trend (3.25%, 2.74%, and 2.22%). It shows that the CGR for the population is still high at 2.22%, comparing with wheat production (1.71%), yield (1.21%), and area (0.60%) for the period 2001-2020.

The periodic changes of the Compound Growth Rate during different periods indicate that wheat area decreases by 67.0%, yield decreases by 40.0%, production decreases by 55.8%, and population decreases less by 31.7%. We interpret the results as follows: there is a mismatch between wheat production and the population of Pakistan, which seems to be a threat to food security in Pakistan.

Table 4. Periodic variation in the Compound Growth Rate (%) of determinates

Period	Wheat area	Wheat yield	Wheat production	Population
1950-1975	1.82	2.02	3.87	3.25
1976-2000	1.21	2.12	3.36	2.74
2001-2020	0.60	1.21	1.71	2.22
% age decrease	67	40	55.8	31.7

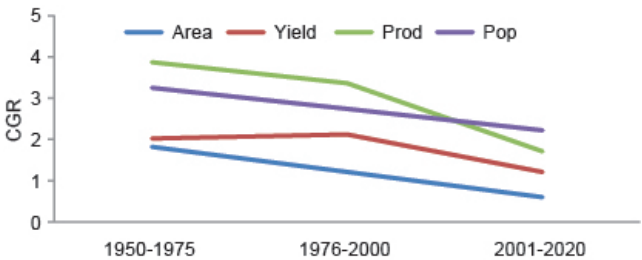


Figure 36. Periodic variations in CGR

3.5 Integrating the growth rate and decomposition analysis model

The decomposition analysis model is used by the researchers and policy makers to assess growth performance, particularly, to evaluate the contribution of area and productivity to the change in production. Projecting with the compound growth rate semi log model the expected increase in wheat area comparing to 2020 is 12.7% for 2030, 43% for 2050, the expected increase in wheat yield is 25.5% and 97.8% respectively for 2030 and 2050. The expected increase in population is 31.8% for 2030 and 129% for 2050, which shows population growth is expected to be higher compared to the same indicators for area and yield.

Table 5 shows that the productivity is the main contributor to the change of wheat crops production (about 38% vs. 20% for area). The interaction effect for production is 42% indicating the joint contribution of area and yield to production. However, low area and high productivity effects indicate the predominant role of the yield contribution towards production: yield compound growth rate (2.326%) is higher than area one (1.207%). The compound growth rate in population is higher than the growth rate in area ($CGR_{pop} > CGR_{area}$) and yield ($CGR_{pop} > CGR_{yield}$). Considering the current population growth rate () the population of Pakistan will reach 283.7 million in 2030 and 492.87 million in 2050 with the instability index (CDVI) at the level of 9.08%, while the area will reach 24506.2(000) acres in 2030 and 31109.0 (000) acres in 2050 with the growth rate 1.207% and instability index (CDVI) 5.26%, and yield will reach 36.43 mds/acre in 2030 and 57.41 mds/acre in 2050 with growth rate 2.326% and instability index (CDVI) 8.66%. Any increase in crop area will definitely affect the area of competitive crops. To overcome or combat the food security problem in Pakistan it is necessary to raise wheat crop productivity along with the controlling population in Pakistan.

Table 5. Decomposition analysis of area and productivity towards production

	Area effects (%)	Productivity effects (%)	Interaction effects (%)
Wheat	20	38	42

Conclusions

Pakistan is anticipated to confront several food security challenges due to its expected population growth until 2050. As of the Global Hunger Index data for 2022, Pakistan is ranked 99th among 121 countries. This study aims to assess the magnitude of population growth rates and the variables related to wheat crops, such as area, yield, and production, to gauge the level of food security concerns in Pakistan.

The study concludes that the Compound Growth Exponential Regression Model (CGREM) provides a better fit when compared to the ARIMA model. The CGREM exhibits lower values across various indicators (RMSE, AIC, and SIC) compared to the ARIMA model.

Periodic analysis of the Compound Growth Rate (CGR) indicates a significant decrease in wheat area, yield, and production by about 67.0%, 40.0%, and 55.8%, respectively, in con-

trast to the population of Pakistan, which only decreases by about 31.7%. Notably, in the period of 2001-2020, the CGR reached low levels for wheat area (0.60%), yield (1.21%), and production (1.71%), while remaining relatively high for population (approximately 2.22%), signaling a potential threat to food security in the region.

A noticeable gap exists between the growth rates of wheat area and population in Pakistan, with the CGR for wheat area at 1.207% and for population at 2.839%. The expected increases for wheat area are projected as 12.7% in 2030 and 43% in 2050, while for population, the projections indicate 31.8% in 2030 and 129% in 2050, highlighting a comparatively rapid population growth. Enhancing the productivity of wheat crops is identified as a key factor in addressing the imminent threat to food security. The CGR for yield is determined to be 2.326%, slightly lower than the CGR for population (2.839%). Projected increases for wheat yield are 25.5% in 2030 and 97.8% in 2050, compared to population projections of 31.8% in 2030 and 129% in 2050.

Wheat yield and productivity demonstrate lower instability compared to the population of Pakistan, with each determinant falling within the low-indexing region of CDVI instability. The decomposition analysis reveals that productivity contributes more significantly to production than the area (38% vs. 20%), aligning with the conclusion that yield plays a major role in wheat production.

According to the World Population Prospects report of the United Nations Department of Economic & Social Affairs, Pakistan's population is projected to reach 366 million¹ by 2050. However, this study estimates Pakistan's population to be around 492 million in 2050. A detailed comparison of various indicators, including growth rate, instability, and decomposition analysis results for wheat crops and population, leads to the conclusion that the government must consider issues of food sustainability when formulating agricultural and trade policies. The findings of this study provide robust empirical evidence supporting the imperative goal of increasing average yield to address the looming threat of food insecurity in Pakistan.

Novelty statement

This study contributes novel dimensions to the policymaking process in Pakistan. The forecasting of food crops and population involves the utilization of various methods, including the compound growth exponential regression model and ARIMA as a benchmark. The compound growth exponential regression model enhances prediction accuracy for long time series data of wheat area, yield, production, and population. The application of decomposition analysis, compound growth exponential regression, and instability analysis results enables us to evaluate and compare the extent of changes in yield and growth rates, an approach not previously utilized on Pakistan's data. The technique that combines the standard ARIMA approach with the compound growth rate model provides valuable insights for improving policies aimed at achieving the goals of food sustainability and food security.

¹ The value was corrected according to the source. (29 February 2024)

Author's contributions

- Muhammad Islam: Contributed to introduction, data analysis, descriptions, conclusions, methodologies and policy implications.
- Farrukh Shehzad: Involved in data analysis, supervision, descriptions, conclusions and policy implications.
- Samrat Ray: Contributed to revision, draft reading, analysis, descriptions, discussions and policy implications.
- Mirza Waseem Abbas: Involved in introduction, analysis, results, discussions and conclusions.

List of references

- Abdulrahman S (2013). Population growth and food security in Nigeria (2010-2012). *Arabian Journal of Business and Management Review (Nigerian Chapter)* 1(3): 1-13. <http://dx.doi.org/10.12816/0003624>
- Ahmad D, Chani MI, Humayon AA (2017). Major crops forecasting area, production and yield evidence from agriculture sector of Pakistan. *Sarhad Journal of Agriculture* 33(3): 385-96. <http://dx.doi.org/10.17582/journal.sja/2017/33.3.385.396>
- Akaike H (1973) Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60(2): 255-65. URL: <https://www.jstor.org/stable/2334537>
- Banks HT, Joyner ML (2017) AIC under the framework of least squares estimation. *Applied Mathematics Letters* 74: 33-45. <https://doi.org/10.1016/j.aml.2017.05.005>
- Bezabeh E (2016) Trends, Growth and Instability of Teff Production in Ethiopia. *International Journal of Recent Research in Life Sciences (IJRRLS)* 3(4):10-13. URL: <https://www.paperpublications.org/upload/book/Trends%20Growth%20and%20Instability-877.pdf>
- Box GE, Jenkins GM (1976) Time series analysis. Forecasting and control. *Holden-Day Series in Time Series Analysis*.
- Dhakre DS, Sharma A (2010) Growth analysis of area, production and productivity of maize in Nagaland. *Agricultural Science Digest* 30(2): 142-7. URL: <https://www.indianjournals.com/ijor.aspx?target=ijor:asd&volume=30&issue=2&article=017>
- Dziak JJ, Coffman DL, Lanza ST, Li R (2012) Sensitivity and specificity of information criteria. Technical Report 12-119. College of Health and Human Development, The Pennsylvania State University. URL: <https://bpb-us-e1.wpmucdn.com/sites.psu.edu/dist/f/84470/files/2019/03/12-119-2e90hc6.pdf>
- Gujarati DN (2003) Basic Econometrics. Fourth edition. McGraw-Hill/Irwin, NY. URL: <https://zalam-syah.staff.unja.ac.id/wp-content/uploads/sites/286/2019/11/7-Basic-Econometrics-4th-Ed.-Gujarati.pdf>
- Gujarati DN, Porter DC (2009) Basic econometrics. Fifth edition. McGraw-Hill/ Irwin, NY URL: https://www.cbpbu.ac.in/userfiles/file/2020/STUDY_MAT/ECO/1.pdf
- Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27(3): 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Islam M (2022) Integrating statistical and machine learning techniques to predict wheat production in Pakistan. A thesis submitted to the Islamia University of Bahawalpur for the award of the de-

- gree of doctor of philosophy in Statistics session 2017-2020. URL: <http://prh.hec.gov.pk/jspui/handle/123456789/21351>
- Islam M, Shehzad F (2022) A Prediction Model Optimization Critiques through Centroid Clustering by Reducing the Sample Size, Integrating Statistical and Machine Learning Techniques for Wheat Productivity. *Scientifica* 2022: 7271293. <https://doi.org/10.1155/2022/7271293>
- Kagan CR (2016) At the nexus of food security and safety: opportunities for nanoscience and nanotechnology. *American Chemical Society (ACS) Publications* 10 (3): 2985-6. <https://doi.org/10.1021/acsnano.6b01483>
- Kondal K (2014). Growth rate of area, production and productivity of onion crop in Andhra Pradesh. *Indian Journal of Applied Research*. 4(3): 4-6. URL: [https://www.worldwidejournals.com/indian-journal-of-applied-research-\(IJAR\)/fileview/March_2014_1492755436__02.pdf](https://www.worldwidejournals.com/indian-journal-of-applied-research-(IJAR)/fileview/March_2014_1492755436__02.pdf)
- Kumbhar AS, Magasi H, Kumbhar MI, Rind ZK (2018) Status of Population Growth and Food Sustainability in Pakistan. *Indian Journal of Science and Technology* 11(16): 1-10. <https://doi.org/10.17485/ijst/2018/v11i16/119048>
- Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology* 150(11):1443-52. <https://doi.org/10.1016/j.agrformet.2010.07.008>
- Lutz W, Sanderson W, Scherbov S (1997) Doubling of world population unlikely. *Nature* (387): 803-5. URL: <https://www.nature.com/articles/42935>
- Makridakis S, Hibon M (1997) ARMA models and the Box-Jenkins methodology. *Journal of Forecasting* 16(3): 147-63. [https://doi.org/10.1002/\(SICI\)1099-131X\(199705\)16:3%3C147::AID-FOR652%3E3.0.CO;2-X](https://doi.org/10.1002/(SICI)1099-131X(199705)16:3%3C147::AID-FOR652%3E3.0.CO;2-X)
- Malthus TR (1798) An essay on the principle of population (1798). The Works of Thomas Robert Malthus, Chapter 10. London, Pickering & Chatto Publishers, 1-139. URL: <https://la.utexas.edu/users/hcleaver/368/368MalthusPopCh10table.pdf>
- Mekuria W (2018) The link between agricultural production and population dynamics in Ethiopia: A review. *Advances in Plants & Agriculture Research* 8(4): 348-53. <https://doi.org/10.15406/apar.2018.08.00336>
- Murindahabi T, Li Q, Ekanayake E (2018) Economic Analysis of Growth Performance of Various Grains Crops During Agricultural Reform in Rwanda. *Journal of Economics and Sustainable Development* 9(2): 32-42. URL: <https://mpa.ub.uni-muenchen.de/90484/>
- Neath AA, Cavanaugh JE (1997) Regression and time series model selection using variants of the Schwarz information criterion. *Communications in Statistics – Theory and Methods* 26(3): 559-80. <https://doi.org/10.1080/03610929708831934>
- Nelson GC, Rosegrant MW, Palazzo A, Gray I, Ingersoll C et al. (2010) Food security, farming, and climate change to 2050: scenarios, results, policy options. *International Food Policy Research Institute*. <http://dx.doi.org/10.2499/9780896291867>
- Rehman FU, Saeed I, Salam A (2011) Estimating growth rates and decomposition analysis of agriculture production in Pakistan: pre and post sap analysis. *Sarhad Journal of Agriculture* 27(1): 125-31. URL: https://www.aup.edu.pk/sj_pdf/ESTIMATING%20GROWTH%20RATES%20AND%20DECOMPOSITION.pdf
- Schwarz G. Estimating the dimension of a model. *The annals of statistics* 6(2): 461-64. URL: <https://www.jstor.org/stable/2958889>
- Shah SAA, Khalil AK (2017) Growth and Variability of Major Food Crops Production in Khyber Pakhtunkhwa, Pakistan. *Journal of Applied and Advanced Research* 2(1): 6-12. <https://doi.org/10.21839/jaar.2017.v2i1.39>

- Sharma P, Chandra A, Kaushik SC (2009) Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City. *Environmental monitoring and assessment*. 157: 105–12. URL: <https://link.springer.com/article/10.1007/s10661-008-0520-2>
- Sihmar R (2014) Growth and instability in agricultural production in Haryana: A district level analysis. *International Journal of Scientific and Research Publications* 4(7): 1–12. URL: <https://www.ijsrp.org/research-paper-0714.php?rp=P312906>
- Tariq A, Tabasam N, Bakhsh K, Ashfaq M, Hassan S (2014) Food security in the context of climate change in Pakistan. *Pakistan Journal of Commerce and Social Sciences (PJCSS)* 8(2): 540–50. URL: <https://www.econstor.eu/handle/10419/188154>

Other sources of information

- Escap UN (2009) Sustainable agriculture and food security in Asia and the Pacific. UN ESCAP, Bangkok. URL: <https://repository.unescap.org/bitstream/handle/20.500.12870/2722/ESCAP-2009-FS-Sustainable%20Agriculture.pdf?sequence=4>
- ESP (2019) Pakistan Economic Survey 2018–2019. Finance Division, Government of Pakistan. URL: https://www.finance.gov.pk/survey_1819.html
- ESP (2020) Economic survey of Pakistan 2019–2020. Finance Division, Government of Pakistan. URL: https://www.finance.gov.pk/survey_1920.html

Information about the authors

- Muhammad Islam – PhD (Statistics), MA (Economics), Deputy Director (Stat), Crop Reporting Service, Agriculture Department, Punjab, Pakistan. E-mail: mislam6667@gmail.com
- Farrukh Shehzad – PhD (Statistics), Assistant Professor, Department of Statistics, the Islamia University of Bahawalpur. E-mail: fshehzad.stat@gmail.com
- Samrat Ray – PhD (Eco) from St Petersburg Polytechnic University, Russia , Affiliation Dean and Head of International Relations, IIMS, Pune, India. E-mail: s.ray@iimspune.edu.in
- Mirza Waseem Abbas – PhD (MS), Assistant Director (Stat), Crop Reporting Service, Agriculture Department, Punjab, Pakistan. E-mail wazeem.mirza@gmail.com