https://doi.org/10.52131/joe.2024.0601.0197

*i*RASD Journal of Economics

Volume 6, Number 1, 2024, Pages, Pages 109 – 127

Journal Home Page:
https://journals.internationalrasd.org/index.php/joe

Hybrid ARIMA-IIS Approach for Wheat Yield Forecasting: An Integrated Approach

Hanzala Zulfiqar[1], Rizwan Ahmad[2], Umar Shahzad[3]

[1] Agriculture Econometrician, Ministry of National Food Security and Research, Islamabad, Pakistan.
Email: hanzalazulfiqar41@gmail.com
[2] PhD (Econometrics) Scholar, Pakistan Institute of Development Economics, Islamabad, Pakistan.
Email: rizvi2041@gmail.com
[3] Data Officer, Ministry of National Food Security and Research, Islamabad, Pakistan.
Email: umarshahzad.eco@gmail.com

ARTICLE INFO

Article History:
Received: December 11, 2023
Revised: March 08, 2024
Accepted: March 09, 2024
Available Online: March 10, 2024

Keywords:
ARIMA
IIS
ARIMA(x)
ADF
ML
MSE

JEL Classification Codes:
B41, C4, C32, D4

Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

This study explores the application of a hybrid approach, combining the Impulse Indicator Saturation (IIS) method with an ARIMA(x) model, to forecast wheat yield. The IIS method is employed to find potential impulse responses, which are then integrated into the ARIMA(x) framework. The IIS method captures the potential joint effects of the weather, climate and other inputs on the data generating process of the wheat yield time series. The performance of the hybrid ARIMA(x) model is compared with that of the standalone ARIMA model using various error metrics, including Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD). Additionally, model selection criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Schwarz Bayesian Information Criterion (SBIC), and Hannan-Quinn Criterion (HQ) are used to identify the optimal model for forecasting. The training data of wheat yield from 1948-2018 was used to fit both the ARIMA and ARIMA(x) models, while the remaining observations until 2023 are used for model validation. The results of the study reveal that the hybrid ARIMA(x) model exhibits superior forecasting ability, demonstrating lower error metrics compared to the standalone ARIMA model. Notably, the ex-ante forecasts for the 2023-24 period predict a wheat production of 29.916 million tons using the ARIMA(x) model and 29.656 million tons using the ARIMA (2,1,2) model. These findings underscore the efficacy of the hybrid approach in enhancing production forecasting accuracy, thereby serving as a valuable basis for early warning systems to address potential demand and supply gaps in wheat production.

Corresponding Author's Email: rizvi2041@gmail.com

Citation: Zulfiqar, H., Ahmad, R., & Shahzad, U. (2023). Hybrid ARIMA-IIS Approach for Wheat Yield Forecasting: An Integrated Approach. *IRASD Journal of Economics*, *6*(1), 109–127. https://doi.org/10.52131/joe.2024.0601.0197

1. Introduction

Predicting wheat yield is crucial for ensuring food security and managing agricultural resources effectively. Wheat is a staple crop globally, serving as a primary food source for millions of people. The ability to forecast wheat yield accurately enables farmers, policymakers, and researchers to make informed decisions about planting strategies, resource allocation, and market trends. By analyzing historical data and using advanced modeling techniques, stakeholders can optimize crop yields and contribute to stable food prices, sustainable agricultural practices, and overall economic stability.

Various statistical techniques has been developed and used extensively in the literature for the forecasting purpose, of these, the crop weather model Fisher (1925), and crop weather model and their use in the yield prediction Baier (1975) are the earliest attempts. Later remote sensing-based approaches Saeed, Saeed, Zakria, and Bajwa (2000) and simultaneous equation-based approaches also got the momentum (Basso & Liu, 2019; Peng et al., 2020). There are also models for wheat yield prediction based on the physiological characteristics of the crops (Bian et al., 2022; Li, Zhang, & Shen, 2017). However, these approaches may not be suitable at the macrolevel due to the economic and data availability constraints. Whereas, thanks to the easy use cases, the statistical models can be affectively deployed for the forecasting tasks.

The ARIMA based modelling got its familiarity with the novel paper by the Box & Jenkins (1970). The auto regressive integrated moving average models (ARIMA) stands as a frequent modelling approach used to forecast the wheat yield (Bian et al., 2022; Hamid, Pinckney, Gnaegy, & Valdes, 2015; Iqbal, Bakhsh, Maqbool, & Ahmad, 2005; Karim, Awal, & Akter, 2010; Masood, Raza, & Abid, 2018; Muhammad, 1989; Sapkota, Singh, Neely, Rajan, & Bagavathiannan, 2020; Shtewy, Hamzah, & Alwan Alsharifi, 2020; Singh, Darji, & Parmar, 2015; Yücel & Erkan, 2020). The ARIMA is a univariate model, meaning it considers only a single time series (e.g., historical wheat yield data) to make forecasts. The ARIMAX (ARIMA with exogenous regressors) model, an enhanced version of the ARIMA model, has been leveraged well to incorporate the quantitative responses of the crops. Thus, within the framework of the univariate family it allows to incorporate the other factors that may have influence on the crop yield and thus improved forecasts.

Currently, common approaches to yield forecasting can be categorized into three main groups: field surveys, dynamic process-based crop simulation models, and statistical regression-based models. Field surveys, still widely used in yield forecasting systems, involve on-the-ground assessments of crop growth by experienced farmers or farm managers (Erciulescu, Cruze, & Nandram, 2019). These assessments reflect farmers' insights into how environmental and human factors are influencing final yields. Through interviews, such as phone interviews, numerous farmers' experiences can be gathered to provide an overall assessment of yield prospects in a specific region. However, the field survey method is often time- and labour-intensive and offers short lead times for decision-making. As a result, significant effort has been dedicated to obtaining timely and reliable yield forecasts from the other two methods. Crop simulation models can depict crucial physical and physiological processes, capturing the complex interplay between crops, soil, weather, and management practices. Consequently, they can typically deliver satisfactory end-of-season yield forecasts with the necessary input data and parameters.

However, when employing crop models for in-season yield forecasting, a major constraint arises from the uncertainty surrounding weather conditions between the forecasting date and maturity (Basso & Liu, 2019). The previous ARIMA based studies has used a single series to predict the wheat yield, although it is the classical practice but, given the weather and environment conditions together with the intricacies in the data. The abrupt shocks and changes in the wheat yield due to these environment and climate changes issues may have propounding impacts on the data generating process of the modelled series and this implicitly affecting the wheat yield. Translated into the statistical notation ignoring or failure to model the relevant variables may cause biased results. Therefore, it is imperative that the factors affecting wheat yield may properly be modelled. But in the univariate family, this can only be addressed through the ARIMAX approach.

The available literature on the wheat yield modelling in the world, and specifically in case of Pakistan using the ARIMA approach has not utilized the strength of the ARIMAX model despite data intricacies and breaks most common in time series data. It is important to stress here that using the multiple regression and/ or the multivariate approaches the effects of numerous factors such as weather, environment, climate, water, fertilizer, various growing practices and phenological stages can be studied explicitly. But due to the data limitations, quality of data and the cost associated with the data these factors are not possible to measure and model. Therefore, the effects of these factors can be jointly captured by the abrupt changes in the yield and/ or production. These abrupt changes are again statistically measured using the data driven approach, known as the Impulse Indicator Saturation (Hendry & Kinnison, 1999).

The present paper thus adds to the existing literature on the wheat yield forecasting in several ways. In the first place it employs a hybrid modelling approach to improve the forecast accuracy in comparison to the simple ARIMA. Hybrid models are utilized in forecasting and various other fields due to their ability to enhance prediction accuracy, robustness, flexibility, and bias reduction (P. Feng et al., 2020; Khaki, Pham, & Wang, 2021; Pandit, Sawant, Mohite, Rajpoot, & Pappula, 2023; Paswan, Paul, Paul, & Noel, 2022). By combining different models, hybrid approaches can leverage the strengths of each component model, resulting in more accurate predictions than any single model could achieve alone. Furthermore, the combination of models can make the overall forecasting approach more resilient to changes in data patterns or model assumptions. Hybrid models can also be tailored to specific forecasting tasks, allowing for greater flexibility in addressing diverse types of data and problems. Additionally, these models can help mitigate bias that may be present in individual models, leading to more reliable forecasts. Overall, the use of hybrid models offers a powerful approach for handling complex data and improving the accuracy and reliability of forecasts. Second, it employs IIS Hendry and Kinnison (1999) approach to identify the potential outlier and structural breaks that may affect the data generating process of the underlying wheat yield, and finally this will be the updated study on the wheat yield forecasting in case of Pakistan that will help the Ministry of National Food Security & Research and other policy institutions related to the policy decisions of the import & export of the wheat.

In the heartland of South Asia, where the Indus River weaves through vast plains and fertile soil, Pakistan stands as a pivotal player in the global agricultural landscape. One of the fundamental cornerstones of its agricultural prowess is the cultivation of wheat, a vital commodity that sustains the nation's economy, feeds its people, and contributes significantly to the daily diet intake. Pakistan is the 8^{th} largest producer of the wheat in the world and 3^{rd} in the south Asia. In the food year 2022-23 Pakistan has produced 28.18 MMT wheat which is historically highest. Wheat is among the major crop of the Pakistan and cultivated by the approximately 80% of the farmers in Rabi season. It was grown on 22.34 million acres during 2022-23 with the record historic production of 28.18 MMT of which 21.225 was contributed by the Punjab the leading producer by area and Production and feeding 75% wheat needs of the population the second largest producer is Sindh with the production of 3.940 MMT in 2022-23 and with the total share in production around 14%. KP and Balochistan contributed around 11% in the national production of wheat. Frequent extreme climate events, such as drought, heat, and frost have caused severe wheat yield losses during the last decades. As such, previous studies in Pakistan have given more attention to the univariate analysis when developing yield forecasting models and hence ignoring the important covariates in the univariate framework. Therefore, this study has opted the hybrid approach with the IIS to account for the potential years of breaks in the yield of wheat and fairly has the following objectives. 1) to develop a hybrid model of wheat yield forecast using the ARIMAX approach, 2) using the Impulse indicator Saturation identifying the potential years of abrupt changes in the wheat yield, 3) Comparison of the simple ARIMA and hyb-ARIMA approach using the errors metrics.

Rest of the study is organized as, next section is about the detailed literature review of the wheat yield forecasting then, data and modelling procedure is outlined finally results and discussions.

2. Literature Review

The literature on wheat yield forecasting in Pakistan draws its traces back to 1970's and 1980's, most of the studies conducted at the time were of the multiple linear regression type where different weather, meteorological, fertilizer and other inputs related variables were used (Azhar, Chaudhry, & Shafique, 1973; Qureshi, 1963; Salam, 1981; Salem, 1989). The focus of the studies at that time were on the role of inputs in the production of the wheat. With the introduction of the second order techniques of the time series analysis Box, Jenkins, Reinsel, and Ljung (2015) the focus was diverted towards the univariate analysis due to the increased accuracy of the forecast. A notable study on the ARIMA model is by the Iqbal et al. (2005). According to this study the production of the wheat was 29.77 MMT in 2022. The identified ARMA process was the ARIMA (2,1,2). The study was conducted at the national level data. The study also concluded with the limited availability of the weather and meteorological data.

Saeed et al. (2000) also use the ARIMA approach to forecast the wheat production in Pakistan. The time series used contained data of 1986-2013. The identified ARMA process was ARIMA (2,2,1). This study suggested the long term forecast of 15 years with the 95% CI. Using the same ARMA approach Amin, Amanullah, and Akbar (2014) identified the ARIM(1,2,2) model to forecast the wheat production in Pakistan, the findings of this study suggested that wheat production in Pakistan will be double in 2060 as compared to the 2010. Very few studies are available in the ARIMA modeling of the wheat yield in Pakistan. The Box-Jenkins method has been extensively used in the wheat yield/ production forecasting in the international scenarios. Masood et al., (2019) used the ARIMA for wheat forecasting in India. The study revealed that the ARIMA (2,1,2) model is the best model for forecasting wheat. Feng, An, Chen, and Wang (2020) study the potential of world wheat under the global warming conditions using the ARIMA model.

Singh et al. (2015) forecasted the wheat production and yield of the Ahmedabad district of the region of Gujrat state in India. The ARMA model suitable for this forecasting was ARIMA(0,1,1). Mishra, Sahu, Dhekale, and Vishwajith (2015) used the ARIMA model to predict the yield of the wheat and rice in the South Asian Association for the Regional Cooperation (SAARC), Karim et al. (2010) utilized regression modeling to forecast wheat production in Bangladesh districts, employing seven model selection criteria. They identified the ARIMA(2,1,2) for the forecasting of wheat in different states of Bangladesh. Boruah, Roy, Mahanta, and Barhoi (2020) used ARIMA (2,1,4) to forecast the wheat production in India. There are few studies which incorporate the exogenous variables into the univariate framework. Given current advancement in the technology, the hybrid ARIM models that is ARIMA-WNN and ARIMA-ANN were used in the forecasting of the wheat production (Melekşen & EYDURAN, 2017; Mishra et al., 2015; Ray, Rai, Ramasubramanian, & Singh, 2016; Saeed et al., 2000). They integrated the ARIMA models with the advanced methods of the predicting and found that these modifications are important and have better predictions when using with the ARIMA process.

In a recent application of the hybrid-ARIMA to predict the wheat yield Saha, Chakraborti, Barua, Gyatsho, and Ghosh (2023) utilized various hybrid models based on ARIMAX and compared their performance with both their standalone versions and the ARIMA model for forecasting yields of major Rabi crops in India. The results indicate that the combination of ARIMAX and LSTM modeling has yielded superior forecasts compared to other time series models. Sudheer et al. (2022) also use the ARIMA-ANN to forecast the sugarcane production in case of India. The ARIMA methos is not only limited to the wheat yield. Production forecasting. It is also equally applicable in other agriculture food crops forecasting.

In this context, it is noteworthy to mention the few applications of the ARIMA model for forecasting agricultural products. Hossain, Abdus Samad, and Ali (2006) utilized the ARIMA model to forecast the prices of three different varieties of pulses, namely motor, mash, and mung, in Bangladesh using monthly data from January 1998 to December 2000. Rachana, Suvarna, and Sonal (2010) forecasted the production of pigeon peas in India using annual data from 1950-1951 to 2007-2008. Mandal (2005) forecasted sugarcane production in India, while Khin, Chong, Shamsudin, and Mohamed (2008) forecasted the price of natural rubber in the world market. Rachana et al. (2010) also used ARIMA models to forecast pigeon pea production in India. Badmus and Ariyo (2011) forecasted the area of cultivation and production of maize in Nigeria using the ARIMA model. They estimated ARIMA (1, 1, 1) and ARIMA (2, 1, 2) for cultivation area and production, respectively. In summary given the limited literature on the application of ARIMA modelling in prediction for wheat yield in Pakistan. The available literature focusses on the modelling of univariate series ignoring other related factors that may have profound impact on the wheat yield while in available literature in the world, few articles are available which incorporate hybrid approach to forecast the agriculture crops which are based on the machine learning hybridization. Therefore, the literature lacks that incorporated the strengths of the ARIMA(x) into the forecasting of wheat yield. Therefore, current hybrid approach using the Impulse Indicator Saturation is the first attempt to utilize the hybrid ARMA(x) to forecast the wheat yield in Pakistan.

3. Data and Methodology

The data for the wheat yield (ton/hectare) is obtained from the Agriculture Statistics of Pakistan from the Ministry of National Food Security and Research (MNFS&R) spanning 1947 to 2023. The wheat production main relies on the area cultivated, which is decided by the Federal Committee on Agriculture (FCA). Therefore, given the area cultivated of the wheat an indirect approach to forecast production of wheat is utilized. The analysis is done on the national level as well as for the four states of Pakistan.

3.1 Auto Regressive Integrated Moving Average

In statistics and econometrics, and in a particular time series analysis, an auto regressive integrated moving average (ARIMA) is generalization of an auto regressive moving average (ARMA) models (Box et al., 2015). The ARIMA models are applied where, the data series show some evidence of the non-stationarity. Therefore, the integrated part of the process shows the level of differencing, used to apply to get the series stationary. Given the data series x_t, where ‘t’ is an integer index and the x_t are real numbers, keeping this setting an ARMA (p.q) model is given by.

$$X_t - \alpha_1 . X_{t-1} - \cdots - \alpha_p . X_{t-p'} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (1)$$

This can be equivalently written as,

$$\left(1 - \sum_1^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t \quad (2)$$

Where, L is the lag operator, the α_i are the parameters of the autoregressive part of the model, and the θ_i are the parameters of the moving average part and ϵ_t are the random error terms. $\epsilon_t \sim iid(0, \sigma^2)$, that is independently, and identically distributed variables sampled from a normal distribution with the mean zero and unit variance. Now to accommodate the integrated part of the ARIMA models. Let us assume that the series is differenced *"d"* times to make it stationary. It means that the polynomial on the left-hand side of the equation (2) has a unit root that is a factor $(1-L)$ of the multiplicity nature. Therefore, the polynomial can be written as.

$$\left(1-\sum_1^{p'}\alpha_i\, L^i\right)(1-L)^d X_t = \left(1-\sum_1^{p'-d}\phi_i L^i\right)(1-L)^d \quad (3)$$

The $ARIMA\,(p,d,q)$ process expresses this polynomial factorization property with the $p = p'$ and this is given by the

$$\left(1-\sum_1^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\epsilon_t \quad (4)$$

This is known as the case of the $ARIMA(p,d,q)$ processes having polynomials with the d unit roots. The above equation (4) can be generalized to account for the drift term as follows.

$$\left(1-\sum_1^{p}\phi_i L^i\right)(1-L)^d X_t = \sigma + \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\epsilon_t \quad (5)$$

Where the drift term $\frac{\sigma}{(1-\sum\phi_i)}$ defines an $ARIMA\,(p,d,q)$ process with the drift. Where the differencing in equation (4) means the differencing between the consecutive values of the time series. Mathematically, $y' = y_t - y_{t-1}$ shows the first difference of the series and, so on for the higher order differencing if necessary.

The impulse indicator saturation (IIS) approach to account for the abrupt changes in the modelled series can be adjusted as follows. Let X_t be the series of the interest, then the impulse indicators I_{it} are defined as.

$$I_{it} = \left\{ \begin{array}{l} 1\ if\ t = t_i \\ 0\ otherwise \end{array} \right\}$$

Where t_i is the period when the i^{th} exogenous variable has an impact on the endogenous variable. In the case of wheat yield these may be the joint effect of the weather, temperature, fertilizer, heat waves and any other event that may have occurred to affect the wheat yield. While integrating these impulse indicators into the ARIMA model, generates a hybrid model.

3.2 Hybrid Auto Regressive Integrated Moving Average (Hyb-ARIMA) ARIMA-IIS approach

Generalizing the impulse indicator saturation (IIS) approach to include *N* indicators for *N* exogenous variables ca be done as follows.

$$X_t = \alpha + \sum_{i=1}^{N}\beta_i\, I_{it} + \epsilon_t \quad (6)$$

α is the intercept term, β_i is the coefficient representing the impact of the i^{th} exogenous indicator on X_t and ϵ_t is the stochastic error term. Now, integrating equation (6) into the equation (5) generates the hybrid ARIMA-IIS model. This family of ARIMA models is also known as the ARIMAX approach. The logic of the hybridization contained "the way the vector of Exogenous variables $\{X_t\}$ are identified."

$$\left(1-\sum_1^{p}\phi_i L^i\right)(1-L)^d X_t = \alpha + \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\epsilon_t + \alpha + \sum_{i=1}^{N}\beta_i\, I_{it} + \epsilon_t \quad (7)$$

An alternate way to write the equation (7) is as follows.

$$\Delta^d X_t = \alpha + \sum_{i=1}^{N}\beta_i\, I_{it} + \phi_1\Delta X_{t-1} + \phi_2\Delta X_{t-2} + \cdots + \phi_p\Delta X_{t-p} + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} + \epsilon_t \quad (8)$$

This integrated model allows for the incorporation of both the ARIMA components (autoregressive, moving average, and differencing) and the IIS approach with exogenous variables and their impulse indicators to model and forecast the endogenous variable.

3.3. Estimation of ARIMA-IIS Model

The Box et al. (2015), also known as the ARIMA modelling approach is historically available to model the time series following the ARIMA processes. This is the systematic procedure for identifying, estimating, and diagnosing auto regressive integrated moving average models for time series. While for the IIS process there is procedure known as the Auto-metrics Hendry and Doornik (2014), implemented in the famous software Ox-metrics (Nikinmaa et al., 2023). The estimation steps involved in the ARIMA modelling are further described below.

3.3.1 Identification

Identification means to recognize the appropriate order of the integration that Δ^d and the possible order of AR and MA terms in the ARIMA model. For the identification of the possible order of the integration is identified using the Muhammad, Xu, and Karim (2015) test for unit root. The general form of the model of the ADF is given by the equation below.

$$\Delta Y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_1^p \delta_i \Delta\, y_{t-i} + \epsilon_t \quad (9)$$

Imposing restriction $\alpha = 0$ corresponds to the random walk with the trend and imposing both $\alpha = 0\,, \beta = 0$ constitutes a model with the random walk only. The equation (9) presents a model with drift and deterministic trend. The unit root test is conducted under the null hypothesis that $\gamma = 0$ in the equation (9) against the alternative that $\gamma < 0$. The test statistics of the ADF regression is given by.

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

This test statistics is sensitive to the number of lags of the auto-regressive terms. Therefore, care must be taken when including the p lags into the model above. One approach to select the lags is to evaluate down from higher orders and examine the t- values of coefficients. The other approach is to examine the information criterion such as Akaike information criterion, Bayesian information criterion or the Hanna-Quin information criterion. These criteria are also useful to select the appropriate ARIMA model from the many iterations.

To identify appropriate order of the ARIMA model, we use the auto correlation function (ACF) and partial auto correlation function (PACF). The PACF of an AR process i.e $\phi_{kk} = \frac{\gamma_k}{\gamma_0}$ breaks off after a lag say 'L.' Therefore, the given series follows order of an 'L' number and thus, PACF can be used to identify the possible AR process. Similarly, ACF of an MA process is zero after a certain lag thus, it can be used to identify the possible order of an MA process. but the ARIMA process cannot be identify when taking together. AS the PACF does not breaks after a certain number of lags due to the presence of the MA terms and vice versa. Therefore, to exactly find the ARIMA process. one should move in data driven way and select the appropriate model based on the certain criterion, known as the Akaike information criterion, Schwartz information criterion and Schwartz Bayesian information criterion. The model which gives the minimum value of these criterion and highest value of the adjusted R^2 is considered as the final model. The model selection criterions are given as:

$$Akaike\ information\ Criterion\ (AIC) = \log\left(\frac{RSS}{n}\right) + \frac{2k}{n} \quad (10)$$

$$Schwartz\ Bayesian\ information\ criterion\ (SBC) = \ln\left(\frac{RSS}{n}\right) + \frac{k}{n}\ln(n) \quad (11)$$

$$adj\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{12}$$

When there are larger data points available then AIC is preferred choice.

3.3.2 Estimation

The estimation of ARIMA models is often preferred by the Exact Maximum Likelihood assuming the Gaussian Innovations. However, this can also be performed by the Generalized least squares and Constrained Least squares functions Akpan and Moffat (2016); Hamilton (2020), The maximum likelihood function for estimating ARIMA models involves calculating the likelihood of observing the data given the model parameters. For an ARIMA(p, d, q) model, the likelihood function is based on the normal distribution assumption for the errors. Here is the general form of the maximum likelihood function.

$$L(\theta|y_1, y_2, \dots y_t) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_t - \mu_t}{2\sigma^2}\right) \tag{13}$$

Where, θ represents the parameters of the ARIMA model (AR and MA coefficients, variance of the errors). y_t is the observed value at time *t.* u_t is the predicted value at time *t* based on the ARIMA model and σ^2 is the variance of the errors. For the computational convenience, the log-likelihood function is given by.

$$L(\theta|y_1, y_2, \dots y_t) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - u_t)^2 \tag{14}$$

The goal of estimation is to find the values of the parameters θ that maximize the likelihood function. This is typically done using numerical optimization techniques, such as the Newton-Raphson method or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.

3.3.3 Diagnostics (Error Metrics)

Diagnostic metrics are used to assess the goodness-of-fit of a model and to diagnose any potential issues or violations of model assumptions. Some common diagnostic metrics used in ARIMA modelling are the auto correlation (ACF) and partial auto correlation (PACF) plot of the residuals from the estimated model. In addition, there are test statistics used to ensure the fit of the model which are as.

1. Box-Ljung test (Q-Test)

This test is performed to check the correlation in the residuals of the estimated at some specified number of lags. This ensures that at or above the given number of lags the residuals are not auto correlated. The test statistics is given by.

$$Q = T(T\} + 2)\sum_{i=1}^{l}\frac{\rho_i^2}{t-k} \quad \sim \quad \chi_\alpha^2\ , (l - p - q)df$$

If $Q_c > \chi^2$, reject the H_0.

2. Breusch-Godfrey Test

The BP test is Lagrange Multiplier based test. Given by the following Auxiliary regression model. Run the possible ARIMA model and predict residuals. That is,

$$\hat{\epsilon_t} = \Delta^d X_t - \alpha + \left(1 + \sum_{i=1}^{q}\theta_i L^i\right)\epsilon_t + \alpha + \sum_{i=1}^{N}\beta_i I_{it} + \left(1 - \sum_{1}^{p}\phi_i L^i\right)$$

Then, regress the estimated residuals on the auto regression of the residuals and include the ARIMA process along with the auto- regression of the residuals. This is given by the following formula.

$$\hat{\epsilon_t} = \sum_{i=1}^{k} \vartheta\_i\, \epsilon_{t-i} + \Delta^d X_t - \alpha + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\epsilon_t + \alpha + \sum_{i=1}^{N} \beta_i\, I_{it} + \left(1 - \sum_{1}^{p} \phi_i L^i\right) + \epsilon_t$$

The test statistics for the BP test is thus given by,

$$LM = nR^2 \quad \sim \quad \chi^2_{\alpha},\ (n-p)df$$

In addition to these tests Normality test of the residuals, actual versus fitted plots and the plots of the ARMA structures with the impulse responses are also performed side by side.

3.3.4 Forecast

Once a satisfactory model is identified, it is submitted for the forecasting. The general equation for the forecasting in an $ARIMA(p.d.q)$ can be expressed as follows:

$$\hat{Y}_{(t+h)|t} = u + \sum_{i=1}^{p} \phi_i\,(Y_{t+i} - u\,) + \sum_{j=1}^{q} \theta_j\,(e_{t+j} - \hat{e}\,)$$

The forecast accuracy depends on the selection of most appropriate model and the power of a model to forecast out of sample. For this, the data is split up into training and validation. The forecast is made for the out sample (which is in-sample out data) for example if we have a data starting from 1947 to 2023, then last 5 years set for the validation and remaining data points are made to use for the model estimation. It is quite possible that a good model selected in the phase 1 may not have good forecast in the validation period but this does not happen in general the out of sample forecast is compared based on certain errors metrics like, MSE, MAPE, MSE or the theil-inequality coefficient etc., after out of sample testing or validation. The best model is used for the ex-ante forecasting and implementation. It is to reiterate here that The Box-Jenkins method is an iterative process, and steps may need to be revisited and revised based on the results of diagnostic checks and model selection criteria. Therefore-estimate the model periodically if needed and check its performance periodically.

4. Results and Discussion

This part of the article discusses the main findings of the data analysis. This divided into three main parts. One is related to pre-estimation, other is based on the estimation and finally post estimation analysis.

4.1 Exploratory Analysis

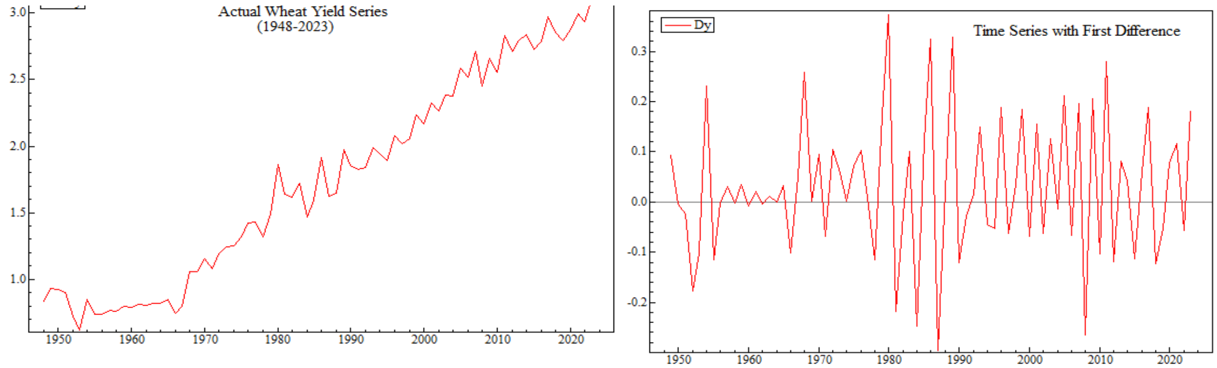

Figure 1: Time Series Plot of the Actual and First Difference of the Wheat yield (t/h)

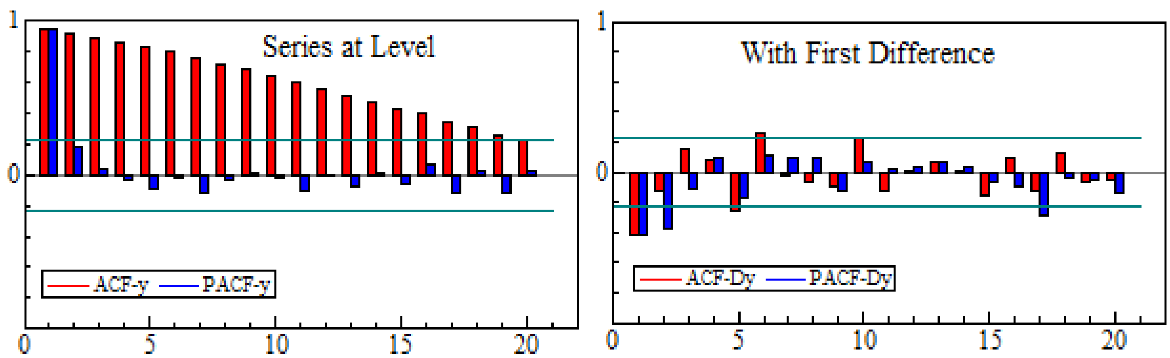

Figure 2: ACF and PACF of the Time Series with Level and First Difference.

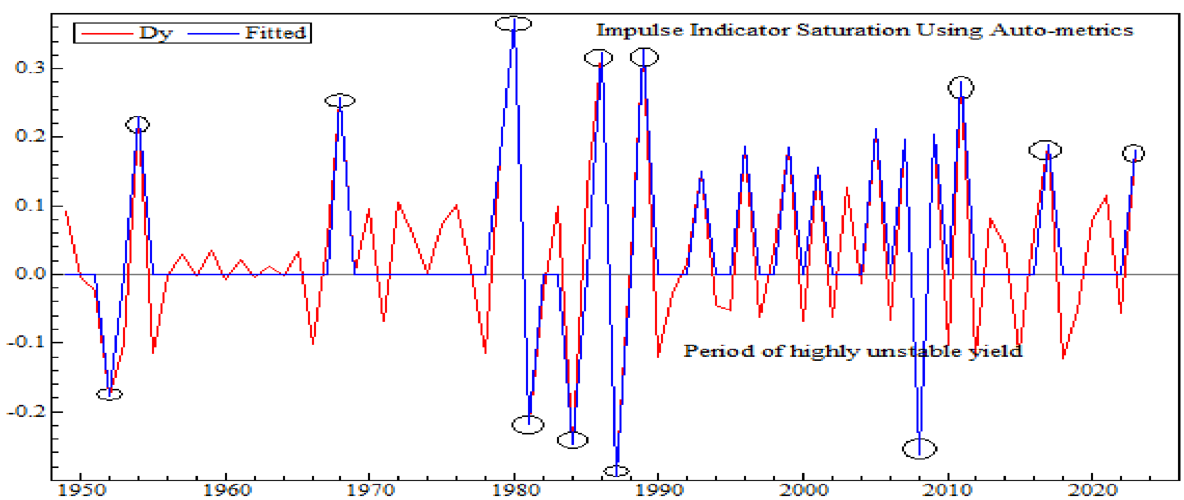

Figure 3: Impulse Indicators at Various Point in Tine Obtained from the Auto-metrics.

Figure (1) shows the time series plot of the accrual and differenced series. The actual series is trendy with many abrupt changes in the data generating process of the series. Therefore, it indicates that the series is non-stationary. Where, the difference plot shows that the series is stable but periods with many higher outliers. Similarly, ACF and PACF plot of the time series given in the figure (2) shows that ACF does not break-off at any period, therefore, showing that the series is clearly an integrated process. Similarly, the ACF and PACF of the differenced series shows that the series has become stationary showing that there are two AR process and one with at higher lag (15). Similarly, there is an identification of few moving average (MA) processes. Visually inspection shows some intuitive idea of the presence of the ARIMA process. the possible order of ARIMA model that can be inspected from the graphs above is intuitively ARIMA (3,1,3). But the true order of ARMA process is identified statistically based on certain criterion as discussed in the methodology section. The figure (3) shows the possible Impulse Indicators or the abrupt changes in the data generating process of the wheat yield series. This graph is the output of the Auto-metrics approach used to identify the abrupt changes in the wheat yield. There are several indicators associated with data generating process of the wheat yield, indicated by the points with the circles in the figure (3). These Impulse indicators are used as covariates with the ARIMA process in case of the hybrid ARIMA model.

4.2 Statistical Analysis

The AD(F) result suggests that the series y_t is trend stationary. The first difference of the series shows that $\Delta y_t \sim I(0)$ which shows that $y_t \sim I(1)$.

Table 1
Augmented Dickey-Fuller Unit Root Test on Wheat Yield

Yield	Level Stationary			1st Difference Stationary		
	Drift	Drift and Trend	None	Drift	Drift and Trend	None
	0.9961	-4.039***	3.917	-10.4312***	-10.6701***	-8.8716***

*** *p<0.01*

Table 2
Augmented Dickey-Fuller Regression Equations and Model Selection

D(y)	coefficient	Adj. R^2	AIC	SIC	HQIC
At level	-4.039	0.4186	-1.584	-1.4280	-1.5224
1st diff	-8.8716	0.7424	-1.3877	-1.2936	-1.3502

AIC = Akaike information criterion, SIC = Schwartz information criterion, HQIC = Hannan Quin Information criterion

The series at first difference is preferred using the information metrics of the model selections. Although the series becomes stationary at level when including the trend and drift term, yet differencing is necessary to make the stationary to avoid the trend.

Table 3
IIS Selection Based on the Auto-metrics

D(y)	Coefficient	t-values	t-prob.	Part. R^2
I: 1953	0.3728	3.30	0.0015	0.1366
I: 1985	0.3244	2.88	0.0005	0.1070
I: 1987	-0.2951	-2.62	0.0109	0.0902
I: 1988	0.3281	2.91	0.0004	0.1092
I: 2008	-0.2641	-2.34	0.0221	0.0736
I: 2023	0.2800	2.48	0.0155	0.0820
Sigma (σ^2)		0.1128	Mean(Dy)	0.0904
Likelihood		60.346	Se(Dy)	0.1373
No. of observations		75	F(2, 67)	3.4497(0.0375)

Note: These indicators are retained because of the auto-metrics application on wheat yield data

Table 4 describes the ARIMA models of the wheat yield. The results are based on the most appropriate model because of the multiple iterations. Using the appropriate model in case of ARIMA (2,1,2) and integrating with the IIS models. The hyb-ARIMA(x) is also estimated, and results are presented in the same table above. The hyb-ARIM(x) model has best fit. The sum of squared residuals is much less than the simple ARIMA (2,1,2). The impulse indicators I:1953, I:1985, I:1986, I:2008 are all significant and has positive effect in the case of 1986 where the negative and significant is observed for the remaining impulse indicators. Further, model selection criterion for the hyb-ARIMA(x) are also suggesting that the model has best goodness of fit.

Table 4
Fitted Models of the ARIMA (p.d.q) and hyb-ARIMA(x)

	ARIMA (2,1,2)			HYB-ARIMA(X)		
Variables	**Coefficient**	**t-statistic**	**Prob.**	**Coefficient**	**t-statistic**	**Prob.**
C	0.0264	2.1363	0.0362	0.0355	6.1243	0.0000
AR(1)	0.9120	6.2872	0.0000	0.4776	0.4776	0.0120
AR(2)	-0.4477	-2.6852	0.0108	-0.3638	-0.3638	0.0102
MA(1)	-0.6792	-5.9400	0.0000	-1.3353	-1.3353	0.0000
MA(2)	0.8052	8.3923	0.0000	0.7320	0.7320	0.0000
I:1953				-0.2875	-3.7907	0.0003
I:1985				-0.1083	-2.5507	0.0132
I:1986				0.1698	2.2049	0.0311
I:2008				-0.1872	-2.9886	0.0040
SIGMA	0.0151		0.0000	0.0109	6.1157	0.0000

R-SQUARED	0.3784	AIC	-1.08174	R-SQUARED	0.5504	AIC	-1.3612
ADJ. R^2	0.3333	SIC	-0.93603	ADJ. R^2	0.4801	SIC	-1.0213
F-STAT	8.4002	HQIC	-1.02320	F-STAT	7.8361	HQIC	-1.2255
PROB(F-STAT)	0.0000			PROB(F-STAT)	0.0000		
SS(RESIDS)	1.1377			SS(RESIDS)	0.8212		

SS = Sum of Squared Residuals

The models above have been estimated using the time series from 1948-2018, the remaining data points have been reserved for the testing or the validation of the models. As the ARIMA models' estimation is an iterative process for the selection of the best model. Therefore, different models with the different orders have been estimated, but only top 10 models and their model selection criterion have been presented in the table 5 below.

The frost row of the table indicates the minimum criterion for the selection of the most optimal model from the top 10 models. As per model selection criterion opted the optimal model for the wheat yield prediction is the $ARIMA(2,1,2)$. Similarly, this optimal model was integrated with the IIS. Thus, the estimated model produced the most reliable results as reported by the MSE and relative lower criterion values in the estimation table 4. When the hyb-ARIMA(x) model was estimated few of the Impulse indicators became insignificant like, I:1988 and I:1987, therefore, they dropped from the final calculation of the model.

Table 5
Model selection criterion for the top 10 Models

Model	Log.L	AIC*	BIC	HQ
(2,2)(0,0)	49.8922	-1.0817	-0.9360	-1.0232
(0,3)(0,0)	49.3283	-1.0682	-0.9224	-1.0096
(0,1)(0,0)	47.0403	-1.0612	-0.9738	-1.0261
(3,0)(0,0)	48.9992	-1.0602	-0.9145	-1.0017
(0,2)(0,0)	47.8388	-1.0564	-0.9398	-1.0095
(2,0)(0,0)	47.6110	-1.0509	-0.9343	-1.0040
(2,1)(0,0)	48.5084	-1.0484	-0.9027	-0.9899
(3,1)(0,0)	49.4996	-1.0482	-0.8733	-0.9779
(2,2)(0,0)	49.4878	-1.0479	-0.8730	-0.9777

Note: The under lined model at the top is the selected model for analysis

The post estimation analysis using various criterion as given below also supported that $ARIMA(2,1,2)$ is an optimal model for wheat yield.

Figure 4 shows the inverse roots of the AR and MA process for an ARIMA process to be stationary. The inverse roots must lie inside the unit root circle as indicated above. The figure 5 shows the impulse response to the one standard deviation innovation. The impulse response must asymptote to the zero values as indicated in the upper part of the graph. Therefore, it is concluded that the ARMA process is stable and stationary. The stationary ARMA process is necessary for the forecasting.

The Box Ljung test statistic (Q-test) calculated for fifteen lags has been given in the table 6 which shows the absence of the serial correlation. The calculated value for the Breusch Pagan (BP) langrage Multiplier (LM) test is 0.0975 which is less than the critical value 8.84 at 5%. It also indicates that the estimated model is free from any serial correlation. The final stage of the Box-Jenkins method is to use the estimated model for the prediction and forecasting. For this the data points are divided into the training and validation sets.

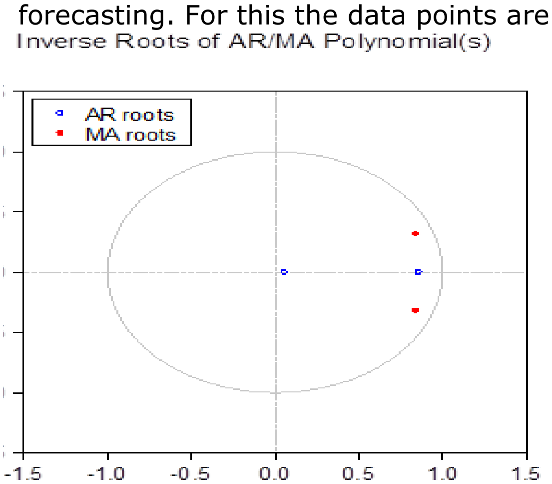

Figure 4: Inverse roots of the ARIMA polynomials

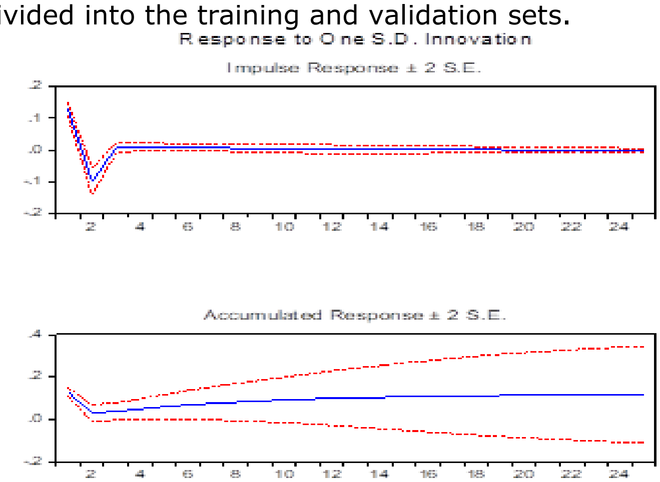

Figure 5: Impulse Response to one SD innovation.

Table 6
Q-Statistic Probabilities Adjusted for ARMA Terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.008	-0.008	0.0046	
		2	-0.029	-0.029	0.0731	
		3	0.059	0.058	0.3506	
		4	0.027	0.027	0.4117	
		5	-0.181	-0.178	3.1233	0.177
		6	0.009	0.005	3.1295	0.209
		7	-0.071	-0.086	3.5621	0.313
		8	-0.110	-0.095	4.6018	0.331
		9	-0.140	-0.144	6.3265	0.276
		10	0.145	0.118	8.1819	0.225
		11	-0.008	0.001	8.1877	0.316
		12	0.000	-0.001	8.1877	0.415
		13	0.080	0.043	8.7833	0.458
		14	-0.015	-0.077	8.8037	0.551
		15	-0.179	-0.162	11.888	0.372

The estimated sample is from 1948 to 2018 the rest five data points that is from 2019-2023 are used to validate the models. This is known as the in sample out or the ex-post forecasting. But for the ex-ante forecasting the production and yield forecasting for the year 2024 is estimated using both the models and their relative performance is compared based on the validation metrics.

Table 7
Forecast Comparison of the ARIMA and hyb-ARIMA(x) Models

Years	Yield actual	Predicted		Validation ARIMA			Hyb-ARIMA(x)		
		ARIMA	Hyb-ARIMA(x)	$(y_t - \hat{y}_t)^2$	$\sqrt{(y_t - \hat{y}_t)^2}$	$\lvert(y_t - \hat{y}_t)\rvert$	$(y_t - \hat{y}_t)^2$	$\sqrt{(y_t - \hat{y}_t)^2}$	$\lvert(y_t - \hat{y}_t)\rvert$
2018-19	2.7980	2.9055	2.8267	0.0017	0.0408	0.0409	0.00006	0.20221	0.0081
2019-20	2.8763	2.9572	2.8535	0.0016	0.1075	0.1075	0.00007	0.32792	0.0087
2020-21	2.9921	2.9438	2.9718	0.0065	0.0808	0.0809	0.00044	0.28442	0.0228
2021-22	2.9347	2.9726	3.0125	0.0016	0.0403	0.0403	0.00014	0.20079	0.0203
2022-23	3.1156	3.0714	3.1397	0.0001	0.0091	0.0091	0.00079	0.09545	0.0121
2023-24		3.0714	3.0980						
	MSE				0.0039			0.0003	
	RMSE				0.0632			0.0183	
	MAPE				1.854			0.565	

The table 7 presents a comparison of actual wheat yield values with predictions from two different models: ARIMA and a hybrid ARIMA model with exogenous variables (Hyb-ARIMA(x)). For each year from 2018-19 to 2022-23, the actual yield is listed alongside the predictions from both models. The table also includes validation metrics such as squared error, RMSE, absolute error, MSE, RMSE, and MAPE for each prediction.

The comparison indicates that the hybrid ARIMA model with exogenous variables outperforms the standard ARIMA model in terms of prediction accuracy. This is evident from the lower values of RMSE and MAPE for the hybrid model across most years. For example, in the last row of the table, the RMSE for the ARIMA model is 0.0632, while for the hybrid model it is 0.0183, indicating that the hybrid model has a smaller average error in predicting wheat yield. Similarly, the MAPE for the ARIMA model is 1.854, whereas for the hybrid model it is 0.565, further supporting the superior performance of the hybrid model. Overall, based on these metrics, the hybrid ARIMA model with exogenous variables appears to be the better model for predicting wheat yield in this context.

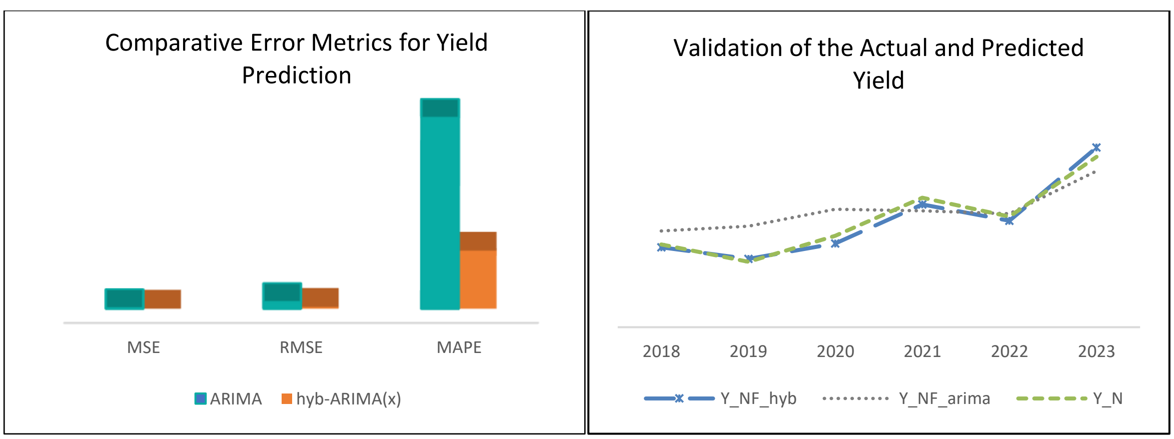

Figure 6: Comparison of the Evaluation Metrics vs Actual and Predicted Yield from 2018-19 to 2022-23

Figure 6 describes the comparative findings of the ARIMA and hyb-ARIMA(x). it is concluded that hyb-ARIMA(x) has the good error metrics for the comparative analysis. The second part of the figure portrays the actual versus predicted values of the wheat yield from both the models. The ex-post predicted values of the wheat yield have close match to the actual values. The next section describes the ex-ante forecasting of the wheat yield.

Table 7 also shows the ex-ante forecasted values for the year 2023-24. The forecasted yield values based on the ARIMA model is 3.0714 (t/h) while the forecasted value from the hyb-ARIMA(x) is 3.0980 (t/h). The Figure 7 shows the actual versus predicted values of the wheat yield with the uncertainty bands at 95% CI and 99% CI. The predicted values lie inside the uncertainty bands indicating that the forecast is stable.

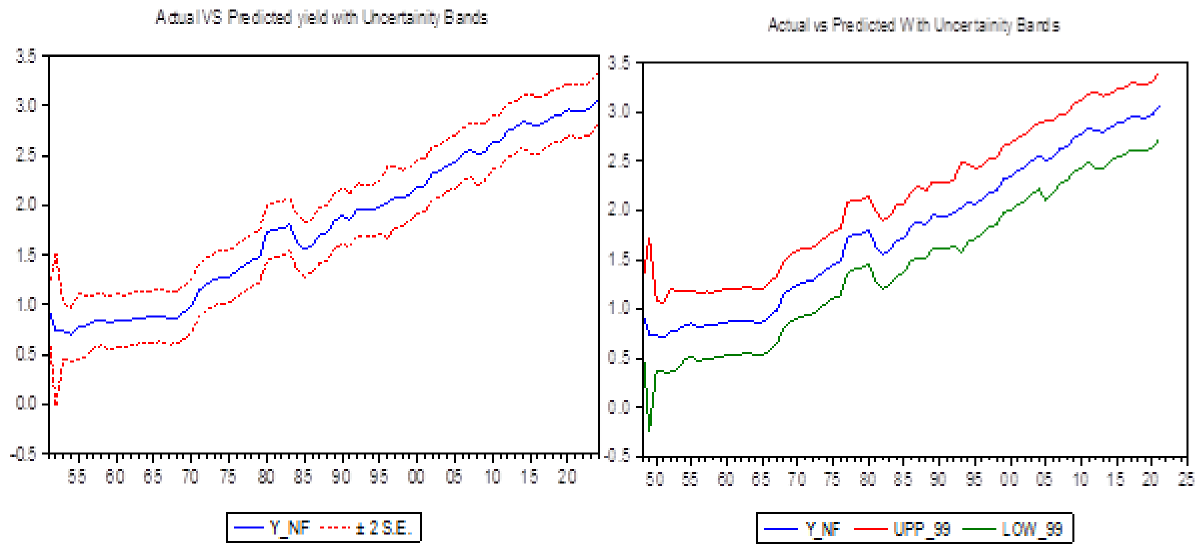

Figure 7: Actual versus Forecasted with the uncertainty bands at 95 % CI and 99% CI for the wheat yield series.

The essence of ex-ante forecast is to utilize this into the policy making. Because this is an intuition that the nature is going to act like determined way. Therefore, wheat yield forecast together with the area cultivated can be used to predict the production of wheat in 2023-24. Production with the annual requirement justifies the potential gap between the requirement and availability.

4.2.1 Policy analysis of the Ex-ante Forecast

Predictions of wheat yield can inform food security planning by anticipating potential shortages or surpluses. This information can help policymakers develop strategies to ensure an adequate food supply for the population. As per the food security policy of the Ministry of National Food Security and Research (MNFS&R). The Ministry is mandated to ensure the food security in the country. The Federal Committee on the agriculture constitute on Food Security commissioners and other agriculture planning entities set the targets for the cropping area and production. So, the province reports to the Ministry over the sowing position and final estimates of the area and production.

The Final estimates of the production are reported generally in August then the planning for import or export is started which is too late in the sense that the harvesting season normally ends in April and in case when there is a potential gap between the requirement and availability there starts another planning that is to get permission from federal cabinet to import for wheat which is another delay in the import if any.

Table 8

National Level Production Forecast Using the Information on Area Cultivated

	year	**Area Cultivated (000)ha**	**Yield forecast. (t/h)**	**Production Forecast (MMT)**	**Low 95%**	**Upper 95%**
ARIMA	2023-24	9,656	3.098	29.916	28.363	30.401
ARIMA(x)	2023-24	9,656	3.071	29.656	27.210	31.630

So, these forecasting techniques have deep implications for the policy planning and risk management. The following section defines the wheat security issue for the year 2023-24, using the information on historical consumption and area cultivated.

The forecasted values given in the
Table 8 are based on the static forecast. The same can also be done for dynamic forecasts. But for the specific case here, the production is based on the area cultivated reported by the Federal Committee on the Agriculture, which is decided at least two months before the sowing season. Therefore, this cannot be anticipated for longer period but can be done by directly modelling and forecasting the production series.

5. Conclusion

The box-Jenkins methodology was applied to the ARMA and Hyb-ARMA(x) models for wheat yield forecasting. The impulse indicator saturation approach was used to identify the joint effects of weather, temperature and/or other related variables that may cause changes in the data generating process of the wheat yield, accordingly five potential impulses were identified. These potential impulse indicators were then used as a covariate together with the ARIMA model. The results suggest that the ARIMA(x) model is more appropriate for modeling wheat yield. The estimated models and the comparison of different error metrics indicate that the ARIMA(x) model provides a better fit to the data and more accurate forecasts compared to the ARIMA model. Specifically, for the forecasted year 2023-24, the ARIMA(x) model predicts a wheat yield of 3.098, while the ARIMA model predicts a slightly lower yield of 3.0741. This difference in forecasted values suggests that the hyb-ARIMA(x) model captures the underlying patterns and variability in the wheat yield data more effectively, leading to more reliable forecasts.

In addition to enhancing the accuracy of wheat yield forecasting, the study's findings hold significant economic implications. Accurate forecasting plays a vital role in agricultural planning and management, enabling policymakers, farmers, and other stakeholders to make informed decisions regarding planting, harvesting, and resource allocation. By providing more reliable forecasts, the Hyb-ARIMA(x) model can contribute to improved crop management practices, increased agricultural productivity, and ultimately, greater economic returns for the wheat sector.

5.1 Policy Recommendations

Based on these findings, policymakers and stakeholders involved in agricultural planning and food security can consider several policy suggestions and recommendations. Firstly, the adoption of the ARIMA(x) model for wheat yield forecasting can enhance the accuracy of predictions, enabling better planning for agricultural production and food distribution. Secondly, policymakers can use the forecasted wheat yield values to assess potential food security risks and develop strategies to address them. Additionally, the use of advanced forecasting models like ARIMA(x) can improve decision-making regarding agricultural policies, resource allocation, and risk management.

5.2. Limitations

Since the Box-Jenkins method is an iterative process to find the best model for the prediction. Therefore, care must be taken to identify the optimal model for the prediction based on the different criterions. The iterative approach used here in this study is to implement, monitor the forecast performance and re-estimate. Additionally, the IIS indicators are integrated as a covariate in the data generating process of the wheat yield series. This data driven way may be different for different researchers and they may have different potential year based on different model selection criterion. So, their findings may be bit different. The crux of this study is to utilize the power of IIS approach with the univariate framework. In the end, the method has used an indirect approach to forecast the production, that is given the final area estimates of the wheat sown, forecasted wheat yield is multiplied by this area to have the final production of wheat. This analysis can also be done using the production series on some assumptions on the area sown. Lastly, this model can be extended to incorporate remote sensing-based area estimations techniques to forecast the area sown and then for the prediction of yield and production. In technical terminology, the hybrid ARIMA-IIS-RS approach may also be opted. Producing some forecast uncertainty when using diverse variety of models.

Author's Contribution:
Hanzala Zulfiqar: Manuscript Preparation, Literature Review, Policy Suggestions
Rizwan Ahmad: Methodology and Data Analysis
Umar Shehzad: Thorough review, Data Validation, and Introduction Preparation

Conflict of interest/ Disclosures:
The authors declared no potential conflicts of interest w.r.t the research, authorship and/or publication of this article.

References

Akpan, E. A., & Moffat, I. U. (2016). Dynamic Time Series Regression: A Panacea for Spurious Correlations. *International Journal of Scientific and Research Publications, 6*(10), 337-342.

Amin, M., Amanullah, M., & Akbar, A. (2014). Time Series Modeling for Forecasting Wheat Production of Pakistan. *JAPS: Journal of Animal & Plant Sciences, 24*(5), 1441-1451.

Azhar, B., Chaudhry, M. G., & Shafique, M. (1973). A Model for Forecasting Wheat Production in the Punjab. *The Pakistan Development Review, 12*(4), 407-415.

Badmus, M., & Ariyo, O. (2011). Forecasting Cultivated Areas and Production of Maize in Nigerian Using Arima Model. *Asian Journal of Agricultural Sciences, 3*(3), 171-176.

Baier, W. (1975). Crop-Weather Models and Their Use in Yield Assessments.

Basso, B., & Liu, L. (2019). Seasonal Crop Yield Forecast: Methods, Applications, and Accuracies. *advances in agronomy, 154*, 201-255. doi:https://doi.org/10.1016/bs.agron.2018.11.002

Bian, C., Shi, H., Wu, S., Zhang, K., Wei, M., Zhao, Y., . . . Chen, S. (2022). Prediction of Field-Scale Wheat Yield Using Machine Learning Method and Multi-Spectral Uav Data. *Remote Sensing, 14*(6), 1474. doi:https://doi.org/10.3390/rs14061474

Boruah, B. B., Roy, P., Mahanta, K. K., & Barhoi, K. (2020). Forecasting Rice and Wheat Production of India Using Arima Models. *International Journal of Management (IJM), 11*(11), 3117-3128. doi:https://doi.org/10.34218/IJM.11.11.2020.297

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*: John Wiley & Sons.

Erciulescu, A. L., Cruze, N. B., & Nandram, B. (2019). Model-Based County Level Crop Estimates Incorporating Auxiliary Sources of Information. *Journal of the Royal Statistical Society Series A: Statistics in Society, 182*(1), 283-303. doi:https://doi.org/10.1111/rssa.12390

Feng, P., Wang, B., Li Liu, D., Waters, C., Xiao, D., Shi, L., & Yu, Q. (2020). Dynamic Wheat Yield Forecasts Are Improved by a Hybrid Approach Using a Biophysical Model and Machine Learning Technique. *Agricultural and Forest Meteorology, 285*(5), 107922. doi:https://doi.org/10.1016/j.agrformet.2020.107922

Feng, Q., An, C., Chen, Z., & Wang, Z. (2020). Can Deep Tillage Enhance Carbon Sequestration in Soils? A Meta-Analysis Towards Ghg Mitigation and Sustainable Agricultural Management. *Renewable and Sustainable Energy Reviews, 133*(11), 110293. doi:https://doi.org/10.1016/j.rser.2020.110293

Fisher, R. A. (1925). Iii. The Influence of Rainfall on the Yield of Wheat at Rothamsted. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, 213*(402-410), 89-142. doi:https://doi.org/10.1098/rstb.1925.0003

Hamid, N., Pinckney, T. C., Gnaegy, S., & Valdes, A. (2015). The Wheat Economy of Pakistan: Setting and Prospects.

Hamilton, J. D. (2020). *Time Series Analysis*: Princeton university press.

Hendry, A. P., & Kinnison, M. T. (1999). Perspective: The Pace of Modern Life: Measuring Rates of Contemporary Microevolution. *Evolution, 53*(6), 1637-1653. doi:https://doi.org/10.1111/j.1558-5646.1999.tb04550.x

Hendry, D. F., & Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*: Mit Press.

Hossain, Z., Abdus Samad, Q., & Ali, Z. (2006). Arima Model and Forecasting with Three Types of Pulse Prices in Bangladesh: A Case Study. *International Journal of Social Economics, 33*(4), 344-353. doi:https://doi.org/10.1108/03068290610651652

Iqbal, N., Bakhsh, K., Maqbool, A., & Ahmad, A. S. (2005). Use of the Arima Model for Forecasting Wheat Area and Production in Pakistan. *Journal of Agriculture and Social Sciences, 1*(2), 120-122.

Karim, M. R., Awal, M. A., & Akter, M. (2010). Forecasting of Wheat Production in Bangladesh.

Khaki, S., Pham, H., & Wang, L. (2021). Simultaneous Corn and Soybean Yield Prediction from Remote Sensing Data Using Deep Transfer Learning. *Scientific Reports, 11*(1), 11132. doi:https://doi.org/10.1038/s41598-021-89779-z

Khin, A. A., Chong, E., Shamsudin, M. N., & Mohamed, Z. A. (2008). Natural Rubber Price Forecasting in the World Market. *Paper in agriculture sustainability through participative global extension, Putrajaya*, 15-19.

Li, Y., Zhang, H., & Shen, Q. (2017). Spectral–Spatial Classification of Hyperspectral Imagery with 3d Convolutional Neural Network. *Remote Sensing, 9*(1), 67. doi: https://doi.org/10.3390/rs9010067

Mandal, B. (2005). Forecasting Sugarcane Production in India with Arima Model. *Inter Stat, October*.

Masood, M. A., Raza, I., & Abid, S. (2018). Forecasting Wheat Production Using Time Series Models in Pakistan. *Asian Journal of Agriculture and Rural Development, 8*(2), 172-177. doi:Https://doi.org/10.18488/journal.1005/2018.8.2/1005.2.172.177

Melekşen, A., & EYDURAN, S. P. (2017). Forecasting Harvest Area and Production of Strawberry Using Time Series Analyses. *Journal of Agricultural Faculty of Gaziosmanpaşa University (JAFAG), 34*(3), 18-26. doi:https://doi.org/10.13002/jafag4298

Mishra, P., Sahu, P., Dhekale, B., & Vishwajith, K. (2015). Modeling and Forecasting of Wheat in India and Their Yield Sustainability. *Indian Journal of Economics and Development, 11*(3), 637-647. doi:https://doi.org/10.5958/2322-0430.2015.00072.4

Muhammad, F., Xu, T., & Karim, R. (2015). Impact of Expenditure on Economic Growth in Pakistan. *International Journal of Academic Research in Business and Social Sciences, 5*(2), 231-236. doi:https://doi.org/10.6007/IJARBSS/v5-i2/1480

Muhammad, K. (1989). Description of the Historical Background of Wheat Improvement in Baluchistan, Pakistan. *Agriculture Research Institute (Sariab, Quetta, Baluchistan, Pakistan)*.

Nikinmaa, L., Lindner, M., Cantarello, E., Gardiner, B., Jacobsen, J. B., Jump, A. S., . . . Seidl, R. (2023). A Balancing Act: Principles, Criteria and Indicator Framework to Operationalize Social-Ecological Resilience of Forests. *Journal of Environmental Management, 331*(4), 117039. doi:https://doi.org/10.1016/j.jenvman.2022.117039

Pandit, A., Sawant, S., Mohite, J., Rajpoot, N., & Pappula, S. (2023). Can Radar Backscatter Ratio Be Used to Estimate Ndvi and Act as Its Substitute?–a Study over Winter Wheat Crop in India. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10*(12), 857-862. doi:https://doi.org/10.5194/isprs-annals-X-1-W1-2023-857-2023

Paswan, S., Paul, A., Paul, A., & Noel, A. S. (2022). Time Series Prediction for Sugarcane Production in Bihar Using Arima & Ann Model. *The Pharma Innovation J, 11*(4), 1947-1956.

Peng, Z., Wang, L., Xie, J., Li, L., Coulter, J. A., Zhang, R., . . . Whitbread, A. (2020). Conservation Tillage Increases Yield and Precipitation Use Efficiency of Wheat on the Semi-Arid Loess Plateau of China. *Agricultural Water Management, 231*(3), 106024. doi:https://doi.org/10.1016/j.agwat.2020.106024

Qureshi, S. K. (1963). Rainfall, Acreage and Wheat Production in West Pakistan: A Statistical Analysis. *The Pakistan Development Review, 3*(4), 566-593.

Rachana, W., Suvarna, M., & Sonal, G. (2010). Use of Arima Model for Forecasting Pigeon Pea Production in India.

Ray, M., Rai, A., Ramasubramanian, V., & Singh, K. (2016). Arima-Wnn Hybrid Model for Forecasting Wheat Yield Time-Series Data. *J. Ind. Soc. Agric. Stat, 70*(1), 63-70.

Saeed, N., Saeed, A., Zakria, M., & Bajwa, T. M. (2000). Forecasting of Wheat Production in Pakistan Using Arima Models. *International Journal of Agricultural Biology, 2*(4), 352-353.

Saha, A. K., Chakraborti, C., Barua, N., Gyatsho, J., & Ghosh, P. (2023). Factors Associated with Visual Outcome after Traumatic Hyphema Following Closed Globe Injury Due to Blunt Trauma: A Tertiary Care Center Study. *National Journal of Physiology, Pharmacy and Pharmacology, 13*(11), 2347-2347. doi:https://doi.org/10.5455/njppp.2023.13.01019202325042023

Salam, A. (1981). Farm Tractorization, Fertilizer Use and Productivity of Mexican Wheat in Pakistan. *The Pakistan Development Review, 20*(3), 323-345.

Salem, M. (1989). Farmers Attitudes Towards Risk in Dryland Wheat Production Areas in Jordan. *Dirasat, 16*.

Sapkota, B., Singh, V., Neely, C., Rajan, N., & Bagavathiannan, M. (2020). Detection of Italian Ryegrass in Wheat and Prediction of Competitive Interactions Using Remote-Sensing and Machine-Learning Techniques. *Remote Sensing, 12*(18), 2977. doi:https://doi.org/10.3390/rs12182977

Shtewy, N., Hamzah, I. J., & Alwan Alsharifi, S. K. (2020). Effect of Mechanical Properties on Some Growth Characteristics for Wheat Crop. *Plant Archives (09725210), 20*(1).

Singh, L. N., Darji, V., & Parmar, D. (2015). Forecasting of Wheat Production and Productivity of Ahmedabad Region of Gujarat State by Using Arima Models. *Indian Journal of Economics and Development, 3*(6), 1-7.

Sudheer, S., Bai, R. G., Muthoosamy, K., Tuvikene, R., Gupta, V. K., & Manickam, S. (2022). Biosustainable Production of Nanoparticles Via Mycogenesis for Biotechnological Applications: A Critical Review. *Environmental research, 204*(3), 111963. doi:https://doi.org/10.1016/j.envres.2021.111963

Yücel, C., & Erkan, M. E. (2020). Evaluation of Forage Yield and Silage Quality of Sweet Sorghum in the Eastern Mediterranean Region. *JOURNAL OF ANIMAL AND PLANT SCIENCES-JAPS, 30*(4). doi:https://doi.org10.36899/JAPS.2020.4.0108