Annals of Bangladesh Agriculture
Journal homepage: bsmrau.edu.bd/aba

ORIGINAL ARTICLES

Trend model for forecasting sustainable potato production in Bangladesh

Zarrin Sovha Khan[1], Mst. Noorunnahar[2*], Keya Rani Das[2], Kaynath Akhi[3]

[1] Department of Agrometeorology, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh
[2] Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh
[3] Department of Agricultural Economics, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh

ARTICLE INFO.

Keywords:
Trend models, Mean absolute percentage error (MAPE), Mean absolute deviation, Box jenkins approach, Prediction.

Received : 08 April 2024
Revised : 06 June 2024
Accepted : 30 June 2024
Published : 15 July 2024

Citation:
Khan, Z. S., M. Noorunnahar, K. R. Das and K. Akhi. 2024. Trend model for forecasting sustainable potato production in Bangladesh. *Ann. Bangladesh Agric.* 28(1): 111-125

ABSTRACT

Bangladesh, predominantly an agricultural nation, deeply used its culture with farming practices. Among its staple crops, potatoes hold significant importance, following rice and wheat. Every year there is a notable increase in potato cultivation and consumption, reflecting the vegetable's growing popularity. The increasing popularity of potatoes presents a significant opportunity to enhance Bangladesh's socioeconomic development with minimal investment required. This research aims to assess various trend models to understand how effectively they capture changes in potato production over time. Secondary data from FAOSTAT spanning from 1971 to 2021, the coefficient of determination (R^2), and adjusted R^2 were employed to evaluate the performance of seven trend models. Notably, the quadratic, cubic, and S-curve models yielded the highest R^2 and adjusted R^2 values, indicating their superior performance in tracking potato production trends. Furthermore, to determine the accuracy of fitted models, we utilized metrics like Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD). The quadratic model, with a MAPE of 20.93% and MAD value of 514185.63, emerges as the most suitable option after rigorous statistical diagnostics. The ARIMA (0,2,2) model exhibits the most optimal correlation for forecasting potato production within the framework of Bangladesh. Through a comparative analysis between the projected dataset and the original data, it is evident that the fitted model ARIMA (0,2,2) outperforms statistically others. This implies that the model possesses the capacity to precisely forecast potato production in Bangladesh for the upcoming decade. These findings will provide critical insights for stakeholders and policymakers in developing sustainable potato production strategies.

Introduction

Potatoes (*Solanum tuberosum*), a member of the Solanaceae family, have been used as an important food crop since ancient times (Aksoy *et al.*, 2021). According to the FAO (Food and Agriculture Organization) statistics for 2017, China is the largest producer of potatoes, with a production of 92,205,580 tons. India follows with a production of 48,605,000 tons of potatoes. Eventually, in the late 19th century, potato farming started in Bangladesh; but commercial

*Corresponding Author: Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh. Email: noorunnahar.stt@bsmrau.edu.bd

https://doi.org/10.3329/aba.v28i1.73300

production did not take hold until the 1920s (Singha and Maezawa, 2019). Potatoes, known for their considerable output when compared to other food and vegetable crops, have received significant amounts of attention in Bangladesh's agricultural sector, with area allocations prevailing. Potatoes flourish in moderate environments, but they are vulnerable to frost, which can cause serious harm when temperatures fall below 0°C (Hijmans, 2003). Particularly during winter when they are abundantly available, potatoes significantly supplement overall food intake, especially in the off-season for rice, the staple crop (Zaheer and Akhtar, 2016). Despite the country's reputation as a rice-consuming nation, potato consumption and production continue to rise (Kadiri *et al.*, 2021). As Bangladesh's population expands, it's essential to broaden food options and decrease dependence on rice. Emphasizing high-nutrient production per unit area and time is crucial, with potatoes offering substantial nutritional value, rich in vitamins B and C, and various minerals (Çalışkan *et al.*, 2023). Compared to rice and wheat, potatoes supply twice as many calories per unit of land and mature faster. With the aim of promoting nutritional diversity and reducing reliance on rice, the government of Bangladesh has lately launched initiatives to increase the use of potatoes (Chowdhury and Chowdhury, 2015). Potato production plays a vital role in Bangladesh's agricultural sector, offering essential sustenance, income, and job opportunities. With the rising demand for potatoes driven by population growth and evolving dietary habits, implementing reliable forecasting techniques is essential to maintain sustainable production levels (Patel *et al.*, 2021). Trend models and forecasting techniques are invaluable assets that provide a glimpse into future production trends, empowering policymakers and farmers to make well-informed decisions (Jha *et al.*, 2020).

Previous studies about different agricultural crops in Bangladesh used ARIMA model to analyze and forecast the production of different agricultural crops like jute, tea, sugarcane, tomatoes etc. (Hossain and Abdulla, 2015; Hossain and Abdulla, 2015a; Hossain and Abdulla, 2015b; Hossain and Abdulla, 2015c; Abdulla and Hossain, 2015; Hossain and Abdulla, 2016). Research has shown that ARIMA (Autoregressive Integrated Moving Average) models showing promising results in predicting crop yields across various regions. So, Accurate forecasting is essential for sustaining and enhancing agricultural productivity in an increasingly uncertain environment (Jha *et al.*, 2020). However, there is a noticeable gap in the literature when it comes to comparing trend models and evaluating their accuracy for potato production in Bangladesh. Our research aims to fill this gap by exploring various trend models from existing literature on other crops and applying rigorous statistical diagnostics to ensure the precision of these models. Additionally, while some research has employed the ARIMA model for forecasting potato production in Bangladesh, up-to-date forecasted values remain scarce. Our study seeks to address this shortfall by providing current and accurate forecasts. Keeping these considerations in mind, the major objective of this study is to evaluate and analyze the performance of various trend models in order to determine which one is the most effective. To meet the nutritional needs of a growing population and boost the economy through potato exports, policymakers need to prioritize forecasting output. This study examines the present potato production pattern and forecasts it for the next ten years. The forecasting approach combines with trend analysis with an Autoregressive Integrated Moving Average (ARIMA). Ultimately, the research aims to bridge the gap between agricultural trends and policy formulation, fostering informed decision-making for the future of potato production in the country.

Methodology

Data Source

This study considered the published secondary data of yearly potato production in Bangladesh which was collected over the period 1971 to 2021 from the website of Food and Agricultural Organization (FAO) https://www.fao.org/faostat/en/#data.

Trend Model

The trend is a general systematic linear or (most often) nonlinear component that changes over time and does not repeat. A trend model represents the evolving pattern of a condition, output, or process, depicting the average or prevailing direction of a series of data points as time progresses. Trend analysis helps us understand and anticipate how a phenomenon behaves over time by looking at both past and present data Top of Form The seven trend models are as follows:

1) Logarithmic trend model: $Y_t = \alpha + \beta \log t_i + e_i$
2) Inverse trend model: $Y_t = \alpha + \beta / t_i + e_i$
3) Quadratic trend model: $Y_t = \alpha + \beta t_i + \beta t_i^2 + e_i$
4) Cubic trend model: $Y_t = \alpha + \beta t_i + \beta t_i^{\,2} + \beta t_i^3 + e_i$
5) Compound trend model: $Y_t = \alpha \beta^t_i + e_i$
6) Power trend model: $Y_t = \alpha t_i^{\,\beta} + e_i$
7) S-curve model: $Y_t = e^{n+\beta/t}{}_i + e_i$

Here, t represents the time (in years) and e_i denotes the error term. (Das *et al.* 2023)

Model Selection

During our model selection process, we employed the criterion of maximizing both the R^2 (coefficient of determination) and adjusted R^2 values (Raza *et al.*, 2015). Mean absolute percentage error (MAPE) and mean absolute deviation (MAD) are the measurement of accuracy of the estimated models (Raza *et al.*, 2015). Lower values of MAPE and MAD indicate a more suitably fitted model (Karim *et al.*, 2010).

$$\text{MAPE} = \frac{\sum|(yt-\hat{y}t)/yt|}{T} \times 100$$

$$\text{MAD} = \frac{\sum|(yt-\hat{y}t)/yt|}{T}$$

Where, $_t$ = actual observation,

$\hat{y}_t$ = fitted observation,

T = number of observations (Das *et al.*, 2023).

ARIMA Model

The combination of Autoregressive (AR) models with Moving Average (MA) models constitutes a flexible category of time series models referred to as Autoregressive Moving Average (ARMA) models. However, their applicability is limited to stationary data. To address non-stationary data, this class of models can be extended by incorporating differencing, resulting in Autoregressive Integrated Moving Average (ARIMA) models (Makridakis, 1998). Therefore, an ARIMA model is a synthesis of AR and MA processes designed specifically for non-stationary data series. Top of Form The general non-seasonal formulation is denoted as ARIMA (p, d, q):

Where, AR: p signifies the order of the autoregressive component,

I: d represents the degree of differencing applied,

MA: q indicates the order of the moving average component.

The equation representing the simplest ARIMA (p, d, q) model is as follows:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots\dots\dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots\dots\dots\dots - \theta_p e_{t-p}$$

Where, C = constant term,

$\varphi_i = i^{th}$ autoregressive parameter,

$\theta_j = j^{th}$ moving average parameter,

e_t = the error term at time t,

$B^k = k^{th}$ order backward shift operator (Makridakis, 1998).

To pinpoint the optimal ARIMA model for a specific data series, Box and Jenkins proposed a methodology that consists of three phases are known as Box-Jenkins's methodology, or in short BJ methodology. The total process of selecting a model is nothing but an iteration procedure that contains the following phases, namely-Phase I: Identification, Phase II: Estimation and Diagnostic Checking, Phase III: Application.

Identification phase

In this phase, a tentative model is usually obtained by using the following four steps:

Step 1: Stability in variance

At first, the observed data series are plotted to identify any unusual observations and a decision is taken on whether any transformation is necessary to stabilize the variance or not. Might be the need arise; transformations may be applied to the data series to attain variance stationarity.

Step 2: Checking the stationary

Before identification of the model, the Augmented Dickey-Fuller (ADF) test can be used to check the data to see if there is any unit root or not (Mila *et al.*, 2022). Different ADF tests can be performed depending on the data generating process such as pure random walk, random walk with drift and random walk with drift plus linear trend. The equations are as follows:

$Y_t = Y_{t-1} + \varepsilon_t$; Pure random walk

$Y_t = \theta_0 + Y_{t-1} + \varepsilon_t$; Random Walk with drift

$Y_t = \theta_0 + \theta_1 t + Y_{t-1} + \varepsilon_t$; Random Walk with drift and deterministic trend (Dickey and Fuller, 1979).

Here, pure random walk predicts the value of Y at time t is equal to the last year value (t − 1) plus a white noise error term (ε_t). Likewise, random walk with drift combines pure random walk plus a drift parameter (θ_0) whereas random walk with drift and deterministic trend combines a random walk with a drift component (θθ0) and a deterministic trend ($\theta_1 t$).

For the ADF unit root test, the null hypothesis is that the data has a unit root whereas for an alternative hypothesis, there is no unit root in the data.

> Null hypothesis: H_0 = There is a unit root in the data
>
> Alternative hypothesis: H_1 = There is no unit root in the data (Dickey and Fuller, 1979).

If p-value of observed t-statistic is less than the nominated significance level then the null hypothesis will be rejected, i.e. there is no unit root. The KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test can be performed to assess whether the data series is stationary or not (Kwiatkowski *et al.*, 1992). For the KPSS test, if the p-value is less than 0.05, then the null hypothesis will be rejected, i.e., the data is non stationary (Mila *et al.*, 2022).

Step 3: Obtaining stationary

If the data appear non-stationary, it can be made stationary through differencing. For non-seasonal data, the usual differencing procedure is enough.

Step 4: Model selection

When stationarity has been achieved, the ACF and PACF of the stationary series and the resulting correlogram are observed to see if any pattern remains. Top of Form Significant autocorrelations and/or partial autocorrelations observed at seasonal lags indicate the presence of seasonality. To find out the appropriate values of the parameters and the pattern in the ACF and PACF are considered which are summarized in the table (Makridakis, 1998).

Table 1. Anticipated patterns in the ACF and PACF for simple AR and MA models

Model	ACFs	PACFs
AR(p)	Decay to zero with exponential pattern (Dies down)	Cuts off after lag p
MA(q)	Cuts off after lag q	Decay to zero with exponential pattern (Dies down)

Estimation and diagnostic checking

The estimation of the parameters, their significance test, and if necessary, identification is revisited in this phase that involves the following three steps:

Step 1: Estimating the parameters

When a tentative model is identified, we want the best estimate of the AR and MA parameters to fit the time series that is being modeled. When utilizing the ARIMA model, the least squares method can be employed, initially selecting an estimate that is subsequently fine-tuned by a computer program (such as auto. Arima). In this point the sum of squared errors is minimum.

Step 2: Selection of the best model

To select the optimal model from among the feasible options, we require an evaluation method, and the AIC can be found approximately using the formula:

$$AI\,C \approx n\,(1 + \log(2\pi)) + n\log\sigma^2 + 2m$$

Where, σ^2 = is the residual variance and

n= is the number of observations in the series.

We choose the best model that has the minimum AIC value. The AIC lacks intrinsic significance; its utility lies in its comparison with another model fitted to the same dataset.

Step 3: Diagnostic checking

The Portmanteau test offers an additional means of assessing the fit of residuals. If this test yields significance, it suggests that the model may lack significance overall. In such instances, reconsideration of alternative ARIMA models becomes necessary. Any new models must undergo parameter estimation, followed by the comparison of their AIC values with those of other models. Usually, choosing the model with the lowest AIC value leads to residuals that closely resemble white noise. Yet, it's wise to choose a model with a slightly higher AIC value if it means getting more manageable residuals (Makridakis,1998).

The Ljung-Box test statistics: Again, for checking the absence of serial autocorrelation of the fitted model, the Ljung-Box (1978) statistics can be used. It is a modification of the Box Pierce test. The Ljung-Box test can be written as

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{p}_K^2}{n-k}$$

Where, $\hat{p}_K^2$= the residual autocorrelation at lag k,

n= the number of residuals,

h = the number of time lags being tested (Ljung and Box, 1978).

If the calculated p-value is higher than 0.05, the model suggests that there is strong autocorrelation of the residuals. The study then develops a new model until the model proves that there is no autocorrelation among the residuals, i.e. the calculated p-value is less than 0.05 (Mila *et al.*, 2022).

Results

To ascertain the most suitable trend model for this dataset, descriptive statistics have been analyzed alongside the performance metrics of eight trend models. These models are evaluated based on their R^2, adjusted R^2, and F-statistics values concerning potato production data over time. Furthermore, the accuracy of each fitted trend model is assessed using MAPE and MAD values. A scatter diagram depicted in Fig. 1. illustrates the relationship between time (years) and potato production (tons). Notably, the scatter plot indicates a lack of linearity between the two variables, suggesting that alternatives to the linear model may be more appropriate for modeling.

As per the dataset, the highest recorded potato production in Bangladesh spanning from 1971 to 2021 reaches 10,215,957 tons (as indicated in Table 2). Furthermore, the average production stands at 3,672,825 tons, with the lowest recorded output hitting 730,067 tons.

Table 3 shows that, for all trend models, the F-statistic is significant at the 1% level. Among these models, the inverse model exhibits the lowest R^2 and adjusted R^2 values, while the quadratic, cubic, and S-curve models display the highest R^2 and adjusted R^2 values. Consequently, the quadratic, cubic, and S-curve trend models emerge as the most appropriate options, given their superior R^2 and adjusted R^2 values, when compared to the logarithmic, inverse, compound, and power models, for analyzing potato production data in Bangladesh from 1971 to 2021.

Table 4 was utilized as the foundation for assessing the accuracy of the fitted model. Assessing the accuracy of the estimated model across the dataset, we observe a MAPE of 20.93% for the quadratic model, which stands out with the lowest MAPE among the models considered. Conversely, the inverse model displays the highest MAPE at 100%.

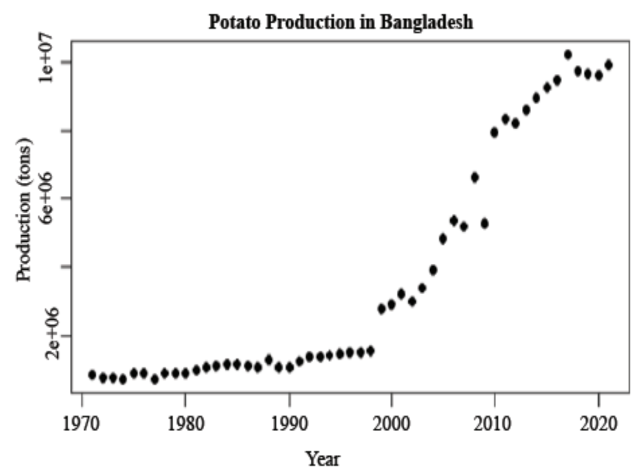


Fig. 1. Scatter diagram for potato production in Bangladesh against time

Additionally, examining the MAD values for the entire fitted model, we noted that the quadratic model exhibits the lowest MAD value of 514185.63, while the inverse model recorded the highest MAD value of 3672824.84. Finally, based on the MAPE and MAD values derived from the fitted model, the Quadratic model emerged as the most suitable choice for analyzing potato data.

In this study, we used potato production data from FAOSTAT over the period 1971 to 2021. The lower

Table 2. Various descriptive statistics of potato production in Bangladesh

Min value (tons)	Max value (tons)	Mean	SD
730067	10215957	3672825	3385115

Table 3. Estimated values for seven trend models

Model/ Statistics					R^2	Adj R^2	F
Logarithmic	-3.12e+09	4.11e+08***			0.817	0.813	219.11***
Inverse	5.04e-04	-9.95e-13***			0.815	0.811	215.81***
Quadratic	2.56e+10	-2.58e+07***	6.53e+03***		0.962	0.961	610.01***
Cubic	3.97e+11	-5.85e+08***	2.86e+05***	4.68e+01***	0.963	0.961	412.01***
Compound	-1.08e+02	6.14e-02***			0.927	0.926	630.20***
Power	-915.606	122.42***			0.926	0.925	620.11***
S-curve	1.26e+07	2.01e+03***			0.969	0.966	1451.28***

*** denotes for significant value at the 1% level of significance.

MAPE and MAD indicate higher reliability of the fitted quadratic model. Furthermore, in contrast to alternative fitted models, the quadratic, cubic, and S-curve models exhibited superior R^2 and adjusted R^2 values. Specifically, all fitted models have maximum R^2 and adjusted R^2 values of 0.963 and 0.960, respectively. Following an assessment of seven trend models, it was observed that the quadratic trend model delivers a more effective fit compared to the remaining six trend models. Now, for forecasting we have applied ARIMA model. There are some analyzing procedure which we should follow before forecasting.

Stationarity checking

At first, the Augmented Dickey-Fuller (ADF) test was used to check the unit root. Again, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was used for checking the stationarity of the time-series data. The ADF and KPSS tests have opposite null hypothesis statement, i.e. the null hypothesis is non-stationary for ADF but it is stationary for KPSS.

The ADF test result showed that for potato production data, the p-value (0.002179) was smaller than 0.05 at 5% level of significance indicating no unit root, as shown in Table 5. Again, the KPSS test showed that the p-value was less than 0.05 indicating that the data became stationary at 5% level of significance.

Obtaining stationarity

Numerous statistical tests have been devised to ascertain the stationarity of a series. Among these, the Augmented Dickey-Fuller test stands out as the most prevalent and widely-utilized method. Null hypothesis is that the series is non-stationary. The p-value obtained in this test is 0.002179. So, the parameter is significant because we can reject null hypothesis here. So, the series is stationary which is alternative hypothesis.

Fig 2 depicted the time plot of Potato production in Bangladesh. It is clear that the potato production data series show an increasing trend over the year.

Model selection

In Figure 3, the Auto Correlation Function (ACF) model and Partial Auto Correlation Function (PACF) model for potato production in Bangladesh are likely used to analyze the correlation and partial correlation between the potato production data at different time lags.

As mentioned earlier, the plot of ACF and PACF can gives a primary guess about the order of parameters p and q for ARIMA model. The ACF plot (Fig.3.a) shows the correlation between the potato production values at different time lags. A spike or peak in the ACF plot at a specific lag indicates a strong correlation between the potato production values at that lag. This information can help in identifying any seasonal patterns or trends in the data. The PACF plot (Fig.3.b) shows the partial correlation between the potato production values at different time lags, while controlling for the effects of the intermediate lags. Similar to the ACF plot, spikes or peaks in the PACF plot indicate significant partial

Table 4. Accuracy of fitted trend models of potato production

	Logarithmic	Inverse	Quadratic	Cubic	Compound	Power	S- curve
MAPE	75.90	100	20.93	24.59	99.99	99.99	32.47
MAD	1294576	3672824.84	514185.63	532288.80	3672810.16	3672810.16	514940.89

Table 5. ADF and KPSS tests of potato production

Unit Root and Stationarity test	Test Statistics	p-value
ADF Unit Root Test	3.2841	0.002179
KPSS Test	1.1846	0.01

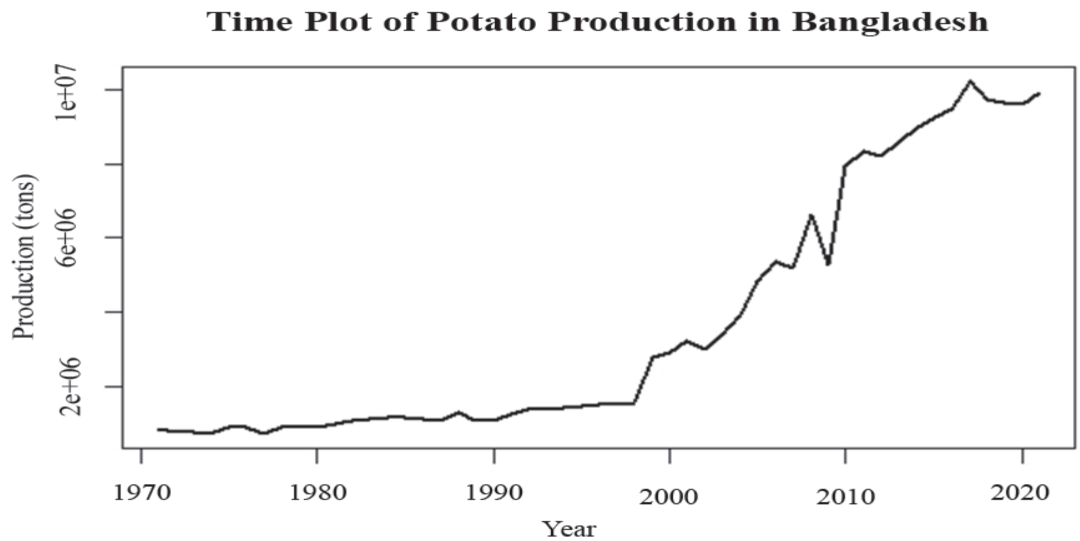
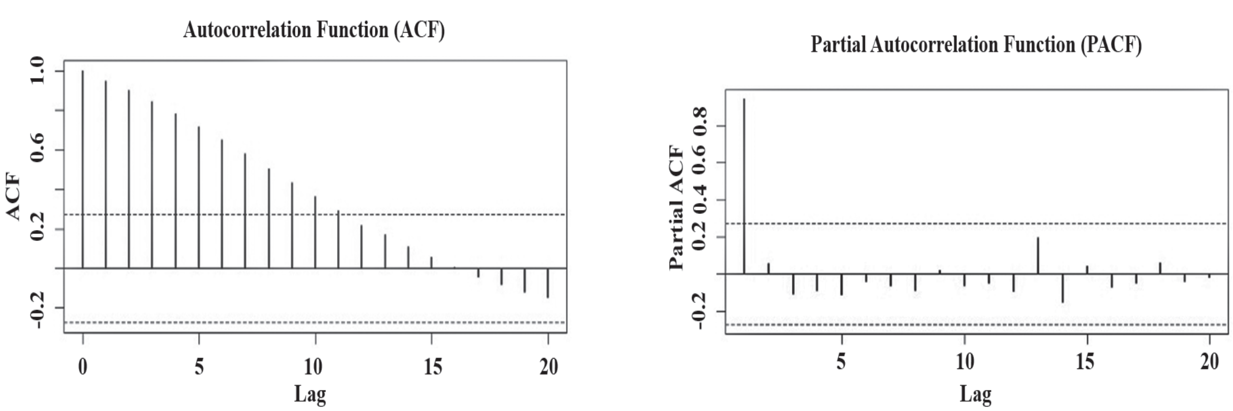


Fig. 2. The time plot of potato production in Bangladesh

correlations at those lags. But from the ACF and PACF plot in fig 3, we cannot make any guess about the parameter p and q, since few spike fall outside the limit at lag 20. Now primarily we say that the series does not follow AR or MA process. Now if we want to see which ARIMA model is appropriate for this series by using computer programming software R. There is no seasonality in data of potato production in Bangladesh. So, there is no need to SARIMA model. After thorough analysis, we have identified the ARIMA (0,2,2) model as the most suitable for modeling potato production in Bangladesh.

The white noise test

For the Auto correlation values $_k$ the Box- Ljung test is applied**.** For autocorrelation values, the Ljung-Box Q* statistics is computed. In addition, the Ljung-Box test was also used to examine the presence or absence of autocorrelations in the residuals. From Table 6, it is observed that the p-value (0.2725) is higher than 0.05, which points to a white noise error in the residuals of the fitted model. It also indicates that there is no autocorrelation among the residuals. Therefore, ARIMA (0, 2, 2) was the best-fitted model for forecasting potato production at 5% level of significance.



a) ACF model

b) PACF model

Fig. 3. (a) ACF model and (b) PACF model for potato production in Bangladesh

ARIMA model encompasses nonseasonal AR and MA coefficients with AIC (Akaike Information Criterion) at 1426.01, AICC (Akaike's Information Corrected Criterion) at 1426.54, and BIC (Bayesian Information Criterion) at 1431.68, it stands out as the optimal choice for forecasting. The parameter estimates for the fitted ARIMA (0,2,2) model are detailed in Table 7.

It is clear from the mean absolute percent error (MAPE) that potato production is predicted with roughly 92.716% accuracy.

Diagnostic checking

For Diagnostic checking, we have to look at Fig. 4 which shows the behavior of the residuals left over after fitting the ARIMA (0,2,2) model. The plot of ACF and PACF of residuals are given. In both cases, all the spikes are in the 55% limits and near to zero. The plot of p-values of the Ljung box test 198.64 with 5 percent level of significance indicates that residuals left over after fitting the model are white noise. All these diagnostic checks support that our selected model is not only have the smallest AIC value but also the better-behaved residuals. Using the most recent model selection criteria (e.g., AIC, BIC, etc.), this study was to determine the most suitable model for predicting potato production in Bangladesh. As we obtain an appropriate ARIMA model for forecasting, we use it to predict the future values in the data set.

Figure 5 presents a graphical comparison between the actual series and the forecast or fitted time series. It is evident that the actual time series (depicted in sea-green color) initially exhibits consistent production, followed by a gradual upward trend, and then experiences a notable acceleration in potato production in Bangladesh after 1999. But in 2010 there's a sudden downturn for a short time and after some time again gradually demonstrates an upward trend in the potato production in Bangladesh. The fitted time series (red color) closely follows the original series, exhibiting fluctuations with minimal variance. As a result, the forecasted series offers a more precise representation of the actual potato production trends in Bangladesh.

Figure 6 displays the forecast with drift for potato production in Bangladesh using the ARIMA (0,2,2) model. The graph illustrates the predicted values of potato production over time, incorporating a drift component in the forecast. The forecasted values, represented by the line in the graph, show the expected trend of potato production in Bangladesh based on the ARIMA (0,2,2) model. The inclusion of a drift component accounts for a systematic increase or decrease in the series over time, capturing any long-term trends or shifts in production levels. By incorporating the drift component, the forecasted values provide a more comprehensive outlook on the future trajectory of potato production in Bangladesh, considering not only short-term fluctuations but also gradual changes in production levels over the forecasted period. Overall, Figure 6 offers a predictive visualization of how potato production in Bangladesh is expected to evolve over

Table 6. Residual test of potato production

Variable	Ljung-Box Test	Value
Potato Production	χ-squared Statistic	23.339
	p-value	0.2725

Table 7. Summary statistics of the parameter estimates of the fitted ARIMA (0,2,2) model

Parameter	Estimate	Standard error	MPE	MAPE
ma1	-1.5541	0.1093	1.078718	9.271067
ma2	0.7873	0.1105		

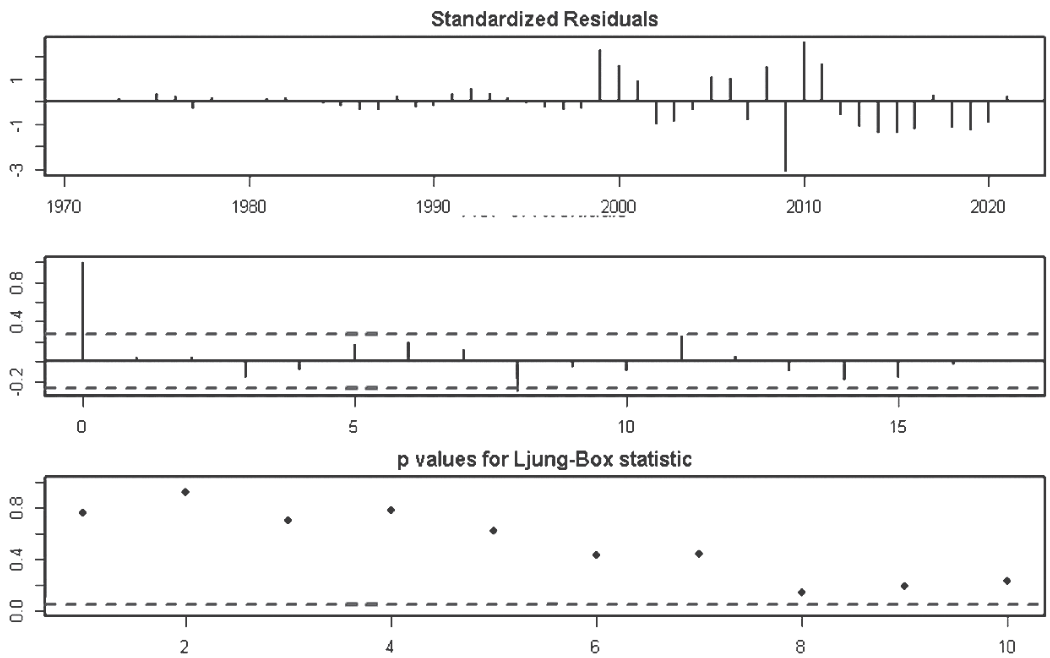


Fig. 4. The diagnostic checking of ARIMA (0, 2,2) of potato production in Bangladesh

time, taking into account both short-term variations and long-term trends in the agricultural sector. So, the forecasted series gives really better representation of the original potato production series in Bangladesh.

Using the best fitted model **ARIMA (0,2,2)** the forecasted values of next 10 years of Potato production were presented in Table 8.

Discussion

The study evaluates seven trend models to understand their effectiveness in tracking changes in potato production over time. The quadratic, cubic, and S-curve models identify as yielding the highest R^2 and adjusted R^2 values among the models considered. Similar study was incorporated in the recent year 2022, by Das *et*

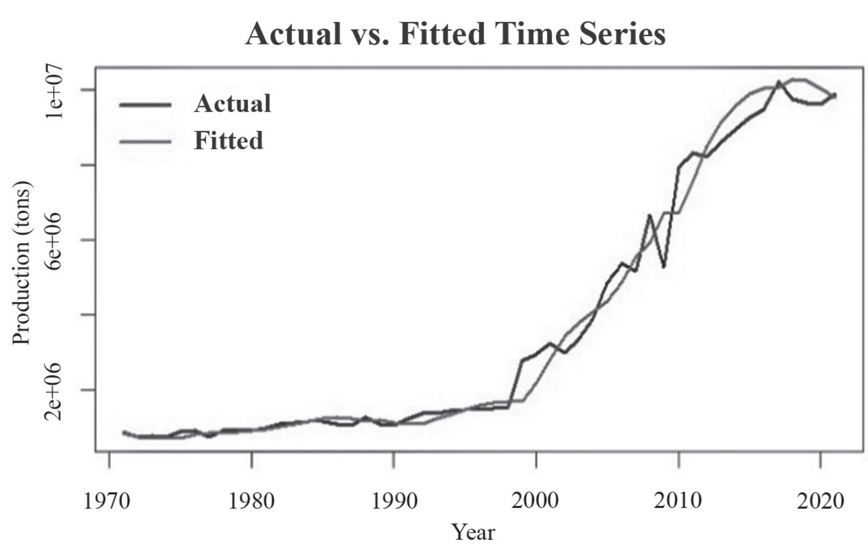


Fig. 5. The contrast between the actual and fitted values of potato production in Bangladesh

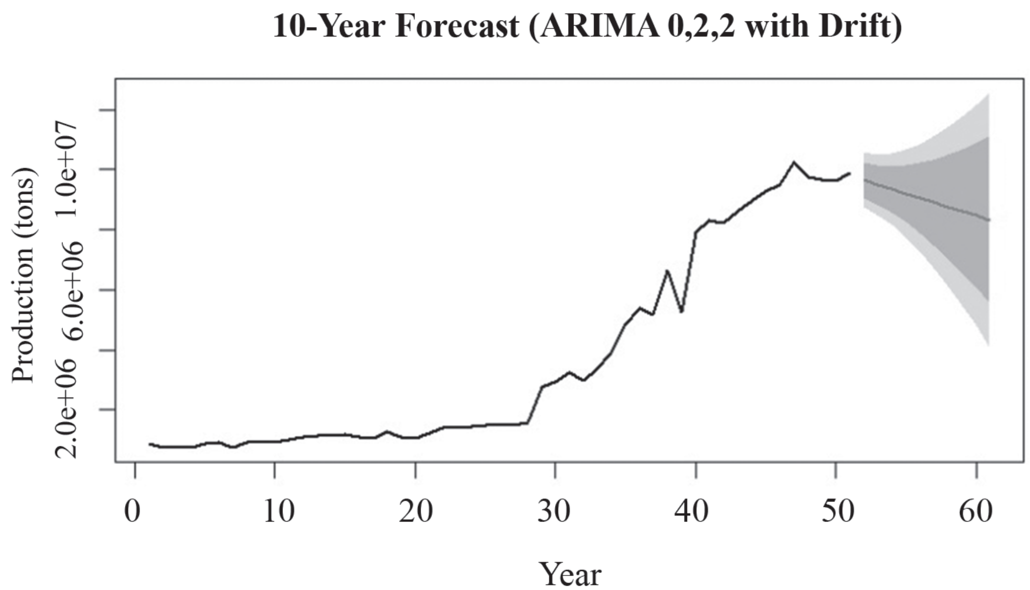


Fig. 6. Plot of forecasted value of ARIMA (0,2,2) model with drift for potato production

al., where six trend models were used for properly checking the performance of the trend models for tea production in Bangladesh (Das *et al.*, 2022). Again, another study explained that compound and growth trend model are suitable for the ginger production in Bangladesh because of their highest R^2 and adjusted R^2 values (Das *et al.*, 2023). The quadratic model emerges as the most suitable choice based on its low MAPE of 20.93% and MAD value of 514185.63, whereas Das showed in her research that the fitted compound and growth models had lower MAPE and MAD values of 12.14% and 6057.54, respectively, for ginger production in Bangladesh. In our study, ARIMA (0,2,2) model emerges as the most advantageous selection for forecasting potato production in Bangladesh. Previous research has found ARIMA (0,1,1) (AMIS,

Table 8. The estimated forecasted values of potato production for the next 10 years (2022 -2031) at 95% confidence interval

Year of Prediction	Forecasted Values of Potato Production (tons)		
	Forecasted Values	Lower Confidence Limit	Upper Confidence Limit
2022	9647089	8728517	10565661
2023	9499784	8494025	10505543
2024	9352479	8168970	10535988
2025	9205175	7755012	10655337
2026	9057870	7266179	10849561
2027	8910565	6716479	11104651
2028	8763260	6116315	11410205
2029	8615956	5472894	11759017
2030	8468651	4791242	12146060
2031	8321346	4074984	12567708

The forecasted values of next 10 years of Potato production was presented in Table 8.

2017) and ARIMA (0,2,1) (Hossain *et al.*, 2016) as suitable for Potato production. The disparity may be attributed to structural changes over time. On the other hand, ARIMA (0,1,1) model was suitable for potato production estimated by Celik in Turkey based on highest R^2 and Lowest MAPE values (Celik, S. 2019). Apart from this, ARIMA (2,1,3), ARIMA (3,1,2) and ARIMA (0,2,1) were the best model to forecast the Mango, Banana and Potato productions respectively in Bangladesh (Hamjah *et al.*, 2014; Hossain, M. M., and Abdulla, F., 2016). Our suggested model contains nonseasonal AR and MA coefficients, with AIC (Akaike Information Criterion) at 1426.01, AICC (Akaike's Information Corrected Criterion) at 1426.54, and BIC (Bayesian Information Criterion) at 1431.68. According to Amir and colleague's study, the AIC for potato production in Bangladesh is 299.46 and the BIC is 305.1, using the fitted model ARIMA (0,2,2) (Rahman *et al.*, 2022). The variation is mainly due to the length of time series and structural changes examined for analysis. The rigorous statistical diagnostics conduct in the study to ensure the accuracy and reliability of the forecasting models. By validating the forecasted series against actual data, the study reinforces the credibility of the selected models and their utility in predicting long-term trends in potato production.

Conclusion

The findings of this study have broader implications for agriculture and the economy of Bangladesh. Accurate forecasting of potato production contributes to better agricultural management, ensures a stable food supply, and drives economic development by improving yields and maintaining market balance. Predicting production trends enables stakeholders to make well-informed decisions that promote productivity and sustainability in agriculture. The statistical performance of the fitted model in estimating potato output in Bangladesh is commendable, demonstrated by its capability to accurately predict both within and beyond the estimation period. Hence, these models can be utilized for forecasting potato production in Bangladesh, serving as valuable tools to inform policy-making decisions. Consequently, there's a pressing need for researchers and policymakers to innovate and devise novel approaches to potato cultivation to meet the burgeoning demand effectively. More investments should be done in infrastructure development such as cold storage facilities, irrigation system and transportation networks to ensure the growing trend of potato. Capacity building for farmers through training program, access to credit and extension services, fostering collaboration among government agencies, research organizations, industry stakeholders and farmers groups to withstand potential challenges and promoting sustainable growth in the area of potato production in Bangladesh.

Acknowledgments

The authors acknowledge FAOSTAT for the availability of secondary data used in this study. Also, the authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

Conflict of interest

The authors affirm that no financial or commercial relationships that might be construed as a potential conflict of interest existed during the course of the research.

Author contributions

Conceptualization, Mst. Noorunnahar and Zarrin Sovha Khan.; methodology, Mst. Noorunnahar and Zarrin Sovha Khan.; software, Zarrin Sovha Khan and Mst. Noorunnahar.; validation, Mst. Noorunnahar and Keya Rani Das.; resources, Zarrin Sovha Khan, Mst. Noorunnahar and Keya Rani Das; data curation, Mst. Noorunnahar and Keya Rani Das.; writing—preparation of the initial draft, Mst. Noorunnahar and Zarrin Sovha Khan; writing, review and editing, Kaynath Akhi and Mst. Noorunnahar; visualization, Kaynath Akhi and Mst. Noorunnahar.; supervision, Mst. Noorunnahar and Keya Rani Das.; project administration, Mst. Noorunnahar, Kaynath Akhi and Keya Rani Das.; All authors have reviewed the manuscript in its current form and given their approval".

References

Abdulla, F and M. M. Hossain. 2015. Forecasting of Wheat Production in Kushtia District & Bangladesh by ARIMA Model: An Application of Box-Jenkin's Method. *Journal of Statistics Applications & Probability*. 4(3): 465-474.

Aksoy, E., U. Demirel, A. Bakhsh, M. A. B. Zia, M. Naeem, F. Saeed and M. E. Çalışkan. 2021. Recent advances in potato (Solanum tuberosum L.) breeding. *Advances in Plant Breeding Strategies: Vegetable Crops*. 8: 409-487

AMIS. 2017. Chapter VI. Model based forecasting. Strengthening Agriculture Market Information System (AMIS) in Bangladesh project. Analytical report on methodologies of crop estimation and forecast and private stock of food grain survey 2016-17. Statistics and Informatics Division (SID), Bangladesh Bureau of Statistics (BBS). Ministry of Planning. 38-59.

BBS. 2019. Statistical Yearbook of Bangladesh. Bangladesh Bureau of Statistics. Statistical Division, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.

Box, G. E. P and G. C. Tiao. 1978. Applications of Time Series Analysis. *Contributions to Survey Sampling and Applied Statistics*. Academic Press. 203-219.

Box, G and G. Jenkins. 1970. Time series analysis: Forecasting and Control, San Francisco: Holden-Day.

Çalışkan, M. E., M. F. Yousaf, C. Yavuz, M. A. B. Zia and S. Çalışkan. 2023. History, production, current trends, and future prospects. In *Potato production worldwide*. Academic Press. 1-18. https://doi.org/10.1016/B978-0-12-822925-5.00016-5.

Celik, S. 2019. Modeling and estimation of potato production in Turkey with time series analysis. *International Journal of Trend in Research and Development*. *6*(5): 111-116.

Chowdhury, M. D and A. H. Chowdhury. 2015. Problems and prospects of potato cultivation in Bangladesh. *Asian Business Review*. *5*(1):28-32. http://dx.doi.org/10.17613/a4bs-t827

Das, K. R., M. Jahan and M. Noorunnahar. 2023. Performance of trend models of ginger production. *Journal of Bioscience and Agriculture Research*. 30(1): 2508-2512. https://doi.org/10.18801/jbar.300123.302.

Das, K. R., Sultana, N., Karmokar, P. K and Hasan, M. N. 2022. A comparison of trend models for predicting tea production in Bangladesh. *Nepalese Journal of Statistics*. 6, 51-62. https://doi.org/10.3126/njs.v6i01.50804.

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

FAO. 2017. Food and Agriculture Organization of the United States. Production statistics, crops. http://www.fao.org/faostat/en/#data/QC

FAOSTAT online database. 2023. Crops and livestock products (Potato), Concatenated, 1971 to 2021, Food and Agricultural Organization, 00153 Rome, Italy.

Hamjah, M. A. 2014. Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. *Journal of Economics and Sustainable Development*. *5*(7).

Hijmans, R. J. 2003. The effect of climate change on global potato production. *American journal of potato research*. *80*: 271-279.

Hossain, M. M and F. Abdulla. 2015. Jute Production in Bangladesh: A Time Series Analysis. *Journal of Mathematics and Statistics*. 11(3): 93-98. https://doi.org/10.3844/jmssp.2015.93.98.

Hossain, M. M and F. Abdulla. 2015a. Forecasting the Tea Production of Bangladesh: Application of ARIMA Model. *Jordan Journal of Mathematics & Statistics*. 8(3): 257-270.

Hossain, M. M and F. Abdulla. 2015b. Forecasting the Sugarcane Production in Bangladesh by ARIMA Model. *Journal of Statistics Applications & Probability*. 4(2): 297-303.

Hossain, M. M and F. Abdulla. 2015c. Forecasting the Garlic Production in Bangladesh by ARIMA Model. *Asian Journal of Crop Science*. 7(2): 147-153.

Hossain, M. M and F. Abdulla. 2016. Forecasting potato production in Bangladesh by ARIMA model. *Journal of Advanced Statistics*. 1(4): 191-198.

Hyndman, R. J and G. Athanasopoulos. 2018. Forecasting: principles and practice. 2nd edition.

Jha, G. K., K. Sinha and S. Pal. 2020. Time series and econometric modelling for crop yield forecasting: A review. *Agricultural Economics Research Review*. 33(1): 91-105.

Kadiri, O., B. Olawoye and D. Oluwajuyitan. 2021. Potatoes: Processing, Properties, and Application. In *Handbook of Cereals, Pulses, Roots, and Tubers*. 551-568. CRC Press.

Karim, R. A. Awala and M. Akhter. 2010. Forecasting of wheat production in Bangladesh. *Bangladesh Journal of Agricultural Research*. 35(1): 17-28.

Kwiatkowski, D. P. C. Phillips, P. Schmidt and Y. Shin. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*. *54*(1-3): 159-178.

Ljung, G. M and G. E. Box. 1978. On a measure of lack of fit in time series models. *Biometrika*. *65*(2): 297-303.

Makridakis, S. 1988. Meta forecasting: ways of improving forecasting accuracy and usefulness. *International Journal of Forecasting*. 4(3): 467-491. https://doi.org/10.1016/0169-2070(88)90112-4.

Mila, F. A., M. Noorunnahar, A. Nahar, D. C. Acharjee, M. T. Parvin and R. J. Culas. 2022. Modelling and forecasting of tea production, consumption and export in Bangladesh. *Current Applied Science and Technology*. 10-55003.

Moazzem, K. G. and K. FUJITA. 2004. The potato marketing system and its changes in Bangladesh: From the perspective of a village study in Comilla District. *The Developing Economies*. *42*(1): 63-94. https://doi.org/10.1111/j.1746-1049.2004.tb01016.x.

Patel, H., S. Thakar and P. Patel. 2021. Machine learning techniques for crop yield prediction: A comprehensive review. *Computers and Electronics in Agriculture*. 187: 106260.

Rahman, M. M., M. A. Islam, M. G. Mahboob, N. Mohammad and I. Ahmed. 2022. Forecasting of potato production in Bangladesh using ARIMA and mixed model approach. *Scholars Journal of Agriculture and Veterinary Sciences*. 10: 136-145.

Sarker, K. R. 2004. Marketing System and Price Behavior of Potato areas of Munshigonj District, Department of Agricultural Economics, Bangladesh Agricultural University, Mymensingh, Bangladesh, M.S. thesis.

Singha, U and S. Maezawa. 2019. Production, marketing system, storage and future aspect of potato in Bangladesh. *Reviews in Agricultural Science*. 7: 29-40. https://doi.org/10.7831/ras.7.29.

Zaheer, K and M. H. Akhtar. 2016. Potato Production, Usage, and Nutrition—A Review. *Critical Reviews in Food Science and Nutrition*. 56(5): 711-721. https://doi.org/10.1080/10408398.2012.724479.

Appendix

Year	Production (tons)	Year	Production (tons)
1971	862600	1997	1507860
1972	752797	1998	1553180
1973	758710	1999	2762000
1974	730067	2000	2933000
1975	880372	2001	3216000
1976	903025	2002	2994000
1977	735336	2003	3385910
1978	913846	2004	3907120
1979	909319	2005	4855377
1980	917120	2006	5368400
1981	998909	2007	5167000
1982	1095151	2008	6648000
1983	1131098	2009	5268000
1984	1147686	2010	7930000
1985	1159385	2011	8326389
1986	1102790	2012	8205470
1987	1069295	2013	8603000
1988	1275650	2014	8950000
1989	1089298	2015	9254285
1990	1065680	2016	9474099
1991	1236805	2017	10215957
1992	1379320	2018	9744412
1993	1384010	2019	9655082
1994	1438055	2020	9606000
1995	1468400	2021	9887242
1996	1491560		