Research Proposal

Forecasting Pakistan Wheat Production (2022–2027)

Submitted by:

Saif Ullah

Department of Statistics

Bachelor's Thesis Project

October 14, 2025

Contents

Chapter 1

Introduction

Agriculture plays a vital role in Pakistan's economy, with wheat being the staple crop and the backbone of food security. Forecasting wheat production allows policymakers to plan import/export strategies, stabilize markets, and ensure food availability. This research applies statistical forecasting models—Linear Regression, Quadratic Regression, and Exponential Smoothing—to historical wheat production data from 1948–2021 to predict output for 2022–2027.

Chapter 2

Objectives

- To analyze the historical trends of wheat production in Pakistan.
- To apply time series forecasting models (linear, quadratic, exponential).
- To compare model performance using RMSE (Root Mean Square Error).
- To forecast production for the years 2025–2027.

Chapter 3

Data Description

The dataset consists of 74 annual observations of wheat production (in metric tons) from 1948 to 2021, obtained from official agricultural records. The variable used is:

- **Year:** 1948–2021
- **Production:** Annual wheat production (tons)

Chapter 4

Methodology

The forecasting process was implemented in Python using the following libraries: `pandas`, `numpy`, `matplotlib`, `statsmodels`, `sklearn`. Three models were trained and compared:

1. **Linear Trend Model:** Fits a straight line to the centered data.
2. **Quadratic Trend Model:** Captures curvature and non-linear growth.
3. **Exponential Smoothing (Holt's Method):** Captures trend and level dynamics adaptively.

Python Implementation

The following code was executed in Google Colab:

```
# Paste this into a Google Colab cell and run.
# Requirements: pandas, numpy, matplotlib, statsmodels, sklearn

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from sklearn.metrics import mean_squared_error
import math
plt.rcParams['figure.figsize'] = (10,6)

# Dataset
data_values = [
3301, 3974, 3862, 3930, 2367, 3587, 3136, 3317, 3581, 3508, 3845, 3847,
    3754,
```

```
3963, 4104, 4096, 4518, 3854, 4266, 6317, 6513, 7179, 6476.3, 6890.4,
    7442.3,
7628.9, 7673.5, 8690.7, 9193.7, 8367.2, 9950, 10856.5, 11474.6,
    11304.2,
12414.4, 10881.9, 11703, 13923, 12015.9, 12675.1, 14419.2, 14315.5,
    14565,
15684.2, 16156.5, 15213, 17002.4, 16907.4, 16650.5, 18694, 17857.6,
    21078.6,
19023.7, 18226.5, 19183.3, 19499.8, 21612.3, 21276.8, 23294.7, 20958.8,
     24032.9,
23310.8, 25213.8, 23473.4, 24211.4, 25979.4, 25086.1, 25633.1, 26673.6,
     25076.1,
24358.1, 25249.5, 27464.5, 26393.65
]

# Automatic year detection
n = len(data_values)
end_year = 2021
start_year = end_year - n + 1
years = list(range(start_year, end_year+1))
df = pd.DataFrame({'year': years, 'production': data_values}).set_index
    ('year')

# Model fitting
x = np.array(df.index)
y = np.array(df['production'])
x_centered = x - x.mean()

coeffs_lin = np.polyfit(x_centered, y, 1)
poly_lin = np.poly1d(coeffs_lin)
coeffs_quad = np.polyfit(x_centered, y, 2)
poly_quad = np.poly1d(coeffs_quad)

forecast_years = np.array([2022,2023,2024,2025,2026,2027])
fx_centered = forecast_years - x.mean()
lin_forecasts = poly_lin(fx_centered)
quad_forecasts = poly_quad(fx_centered)

es_model = ExponentialSmoothing(y, trend='add', seasonal=None,
                                initialization_method="estimated")
es_fit = es_model.fit(optimized=True)
es_forecast = es_fit.forecast(len(forecast_years))

# RMSE Evaluation
from sklearn.metrics import mean_squared_error
```

```
rmse_lin = math.sqrt(mean_squared_error(y, poly_lin(x_centered)))
rmse_quad = math.sqrt(mean_squared_error(y, poly_quad(x_centered)))
rmse_es = math.sqrt(mean_squared_error(y, es_fit.fittedvalues))

print("RMSE Values:")
print(f"Linear: {rmse_lin:.2f}, Quadratic: {rmse_quad:.2f}, Exponential
    : {rmse_es:.2f}")
```

Chapter 5

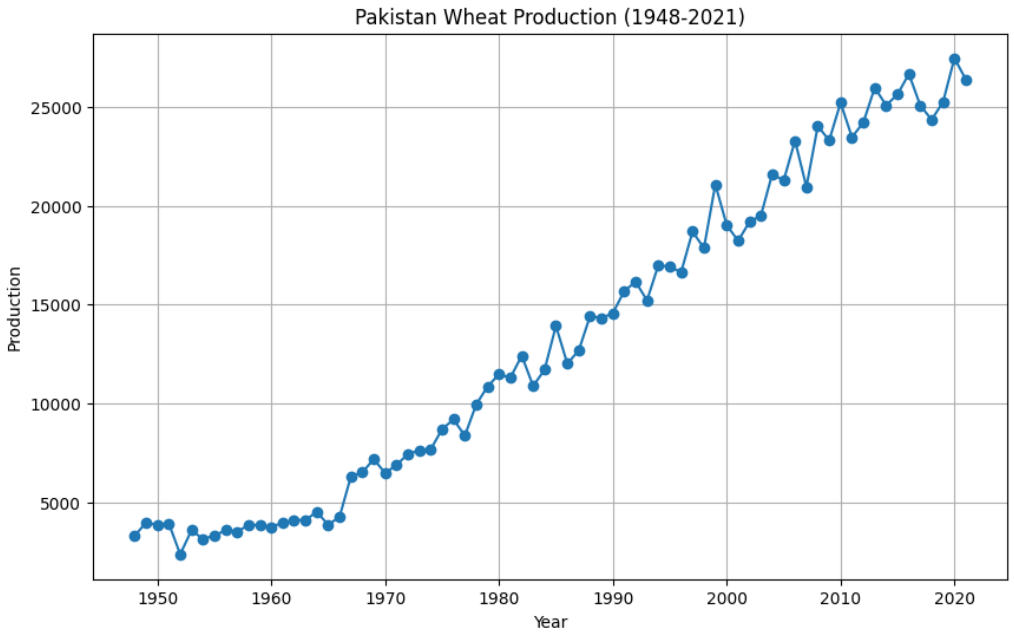
Results and Discussion

The following table summarizes the forecasted values for 2025–2027:

Year	Linear Forecast	Quadratic Forecast	Exponential Smoothing
2025	28241.5	30676.6	27270.7
2026	28610.4	31214.1	27433.5
2027	28979.3	31755.8	27596.3

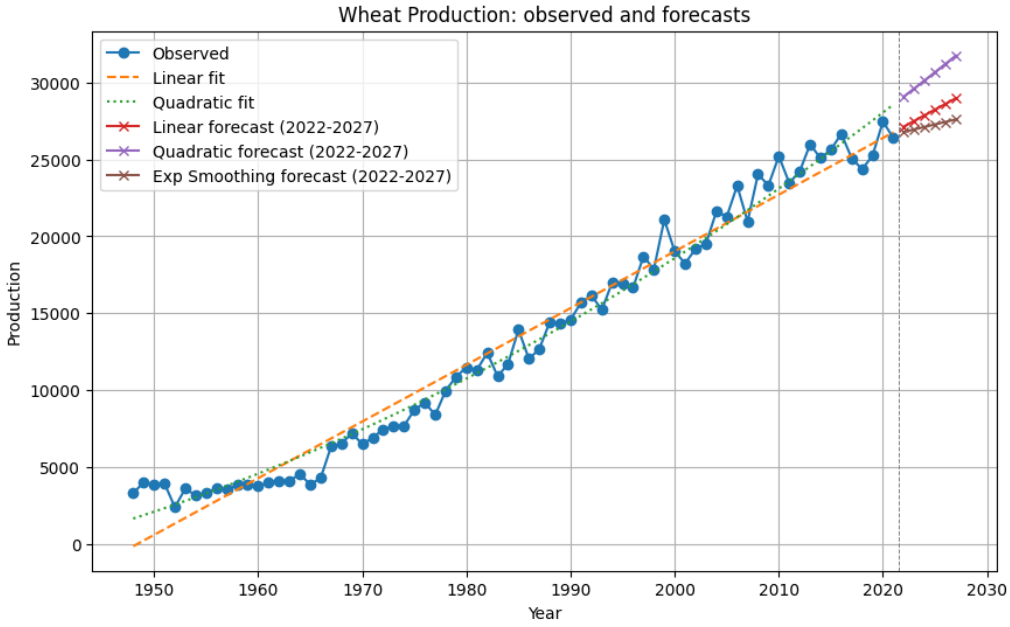
The exponential smoothing model had the lowest RMSE (**950.36**), indicating it best captures the production pattern. Quadratic trends show an upward curvature, predicting higher yields, while the exponential model is more conservative.

Figure 1: Historical Wheat Production



The above figure shows the historical production trend of wheat from 1948 to 2021.

Figure 2: Forecasted Wheat Production (2022–2027)



Forecasted values from three models showing different growth trajectories for future years.

Chapter 6

Conclusion

The results suggest that Pakistan's wheat production will continue a gradual upward trend through 2027. Exponential Smoothing proved the most accurate based on RMSE, balancing trend-following and adaptability. These findings can guide agricultural policy and resource allocation decisions.

Chapter 7

Future Work

Future studies could:

- Include climate, rainfall, and temperature data as external regressors.
- Compare ARIMA or LSTM neural models.
- Incorporate uncertainty intervals around forecasts.