Aryan Sharma

Forecasting Wheat Price in India (2012–2024) with MSP Influence

Context and Purpose of Analysis

The impetus for this analysis stems from the ongoing debate surrounding the legal guarantee of the Minimum Support Price (MSP) in India. Amidst prolonged farmer protests from 2020 to 2025, this issue has taken center stage in agricultural policy discussions. The core objective of this study is to evaluate whether MSP, as it currently functions, aligns with market realities and offers tangible protection to farmers.

To achieve this, we analyze 10 years of price data for wheat—one of the 20 crops covered under the MSP regime. This analysis uses trend modeling, seasonality adjustment, lag structures, and ARIMA forecasting to compare market wheat prices against MSP over time. The findings serve as an empirical basis for assessing the credibility of farmer claims advocating for MSP as a legally binding guarantee.

The urgency of this analysis is amplified by the human and economic toll of the farmers' movement. According to the National Crime Records Bureau (NCRB), a total of 11,290 individuals involved in farming died by suicide in 2022, representing 6.6% of all suicide victims in India. The state of Maharashtra alone recorded 822 farmer suicides in its Marathwada region in 2024, yet compensation was granted in only 303 cases. These statistics underscore the distress and systemic challenges that farmers continue to face.

The economic impact of the farmers' protests has also been substantial. National highways, railways, and toll plazas were repeatedly blocked, leading to major disruptions. Between October 2020 and November 2024, toll revenue losses in Punjab totaled approximately ₹1,639 crore. This includes ₹1,348.77 crore lost from October 2020 to December 2021, ₹41.83 crore in 2022-23, ₹179.10 crore from January to July 2024, and ₹69.15 crore between October and November 2024.

Additionally, the protests led to the cancellation of over 80 trains in Punjab during December 2024 alone, creating widespread transportation and supply chain challenges. Blockades on the Haryana–Punjab border further disrupted trade and local industries, resulting in significant economic losses. These events collectively highlight the scale of disruption and the critical need for structural agricultural reforms in India.

1. Introduction

This report analyzes the relationship between the monthly market price of wheat per quintal and the Minimum Support Price (MSP) announced by the Indian government. The objective is to understand the time-based behavior of wheat prices, identify seasonality and autocorrelation, and select the most accurate forecasting model for future price prediction. The models considered include:

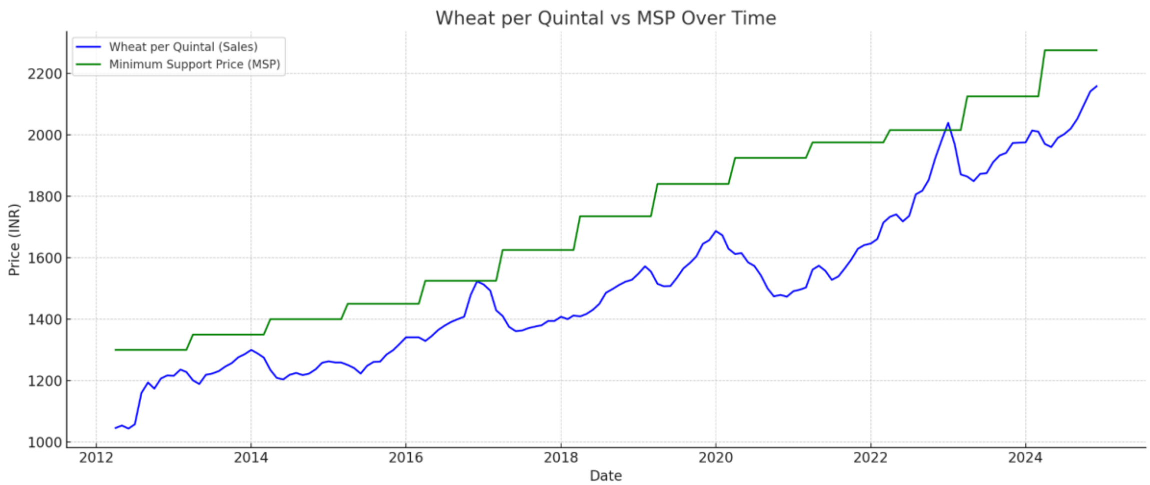
- Linear time trend

Aryan Sharma

- Seasonal adjustment using quarter dummies
- Lagged autoregressive components
- Performance-based model comparison using RMSE
- Durbin-Watson test for autocorrelation
- ACF and PACF diagnostics

2. Visual Analysis: MSP vs Market Price

The figure below compares the time-series evolution of market wheat prices and MSP between 2012 and 2024:



Observation:
- MSP increases in a staircase pattern (reflecting policy revisions).
- Market price fluctuates, showing seasonal variation, spikes, and autocorrelated behavior.
- The gap between MSP and market price narrows during certain periods, suggesting that MSP might act as a floor or anchor price for the market.

3. Dataset Summary

Variable	Description
Duration	Monthly time index from April 2012–2024
Sales	Wheat price per quintal (market price)
MSP	Minimum Support Price (Govt announced)
Qtr	Quarter (1–4)
Wheat Atta per quintal	Parallel retail metric (not modeled)

4. Methodology and Model Specifications

Why Use a Training Dataset?

To evaluate predictive performance, data is split into:
- Training set (first 40 observations): Used to estimate model coefficients.
- Test set (remaining 113 observations): Used to compute Root Mean Squared Error (RMSE)

between actual and predicted values.
This prevents overfitting and provides an unbiased estimate of a model's real-world forecasting ability.

5. Model Comparison

Model 1: Linear Trend Model

Time coefficient: +5.74 INR/month
Adjusted R^2: 0.897
RMSE (Test): 187.43
Durbin-Watson: 0.095 → Severe positive autocorrelation

Model 2: Trend + Seasonal (Quarterly Dummies)

Adjusted R^2: 0.901
RMSE (Test): 198.05
Durbin-Watson: 0.117 → Positive autocorrelation persists

Model 3: Trend + Seasonality + Lag 1

SalesLag1 coefficient: ~0.71
Adjusted R^2: 0.835
RMSE (Test): 109.33

Model 4: Trend + Seasonality + Lag 1–3

Adjusted R^2: 0.854
RMSE (Test): 113.59

Model Summary Table:

Model	Adjusted R^2	RMSE (Test)	Durbin-Watson	Notes
M1	0.897	187.43	0.095	High autocorrelation
M2	0.901	198.05	0.117	Overfits slightly
M3	0.835	109.33	Improved	Best forecasting performance
M4	0.854	113.59	Improved	Complex; marginally worse RMSE

6. Autocorrelation Diagnostics

Durbin-Watson statistic < 1 for both M1 and M2 indicates strong positive serial correlation.
ACF and PACF plots confirm lag structures.
Including lagged sales significantly reduces residual variance and improves predictive accuracy.

7. Recommended Modeling Approach

Based on RMSE and interpretability:
Model 3 (Trend + Qtr + Lag 1) is the most effective model for short-term wheat price prediction.
It captures:
- Upward trend over time
- Seasonal effects (Q3, Q4 uplift)
- Price inertia (previous month impacts current price)

8. ARIMA Extension

This includes two non-seasonal autoregressive terms (AR1, AR2), a seasonal AR term (SAR1), and a drift term representing consistent growth.

Parameter	Estimate (Std. Error)
AR1	0.4966 (0.0807)
AR2	-0.1718 (0.0800)
SAR1	0.0151 (0.0887)
Drift	7.3336 (2.8664)

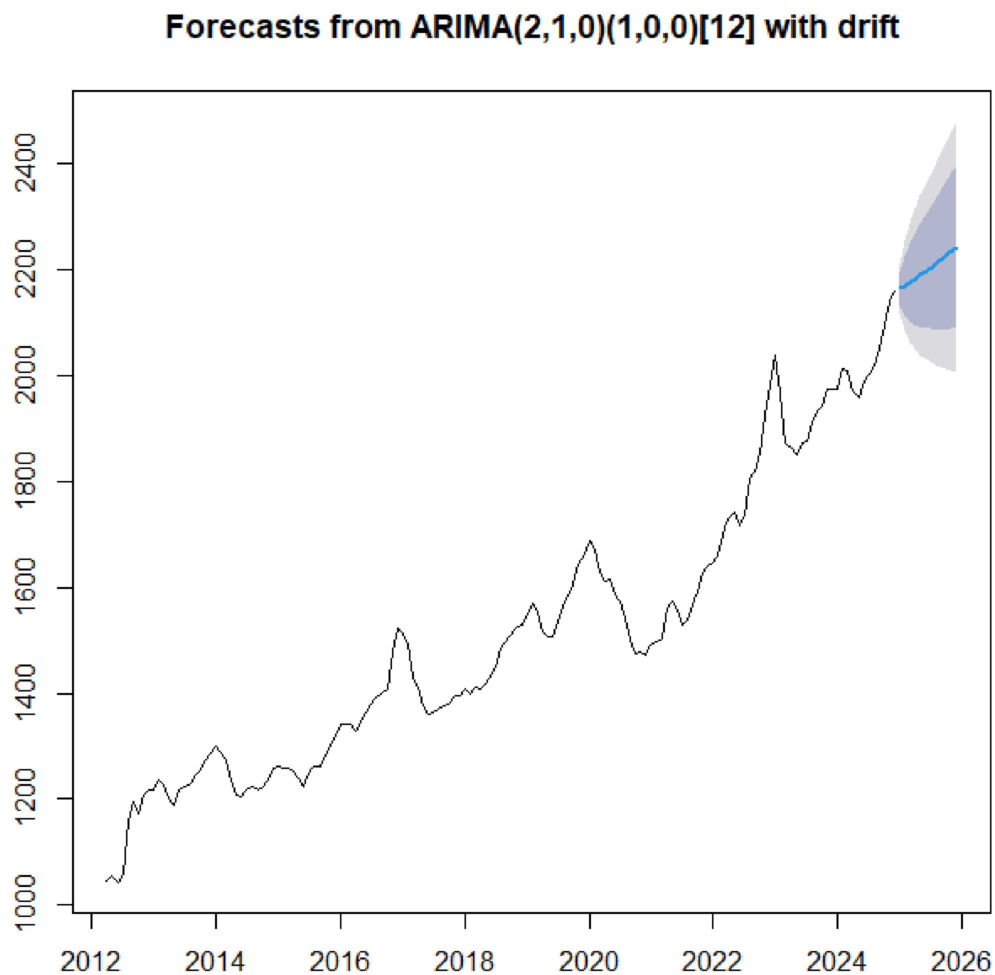
Model Diagnostics:

- $Sigma^2$ (residual variance): 569.3
- Log-likelihood: -695.96
- AIC: 1401.92, BIC: 1417.03
- Residual ACF1: 0.0109 → indicates minimal autocorrelation in residuals
- RMSE (Train): 23.47, MAPE: 1.12%

The model shows excellent residual diagnostics and minimal autocorrelation, confirming its robustness for short-term forecasting. Forecasts for the next 12 months were generated and visually analyzed. ARIMA provides slightly better predictive stability and adaptability to unforeseen shifts.

9. ARIMA Modeling and Forecasting

ARIMA (AutoRegressive Integrated Moving Average) is a powerful time-series modeling technique used to forecast data with trends, seasonality, and autocorrelation. Unlike regression-based models, ARIMA operates directly on time-ordered observations and requires the series to be stationary, which is often achieved by differencing. It allows both autoregressive terms (AR), moving average terms (MA), and differencing (I). Seasonal ARIMA models also incorporate seasonal lags.



Why ARIMA?

- It accommodates non-stationarity through differencing.
- It captures both short-term autocorrelation and long-term trend.
- It enables direct forecasting, unlike regression which estimates relationships.
- ARIMA models can be extended to ARIMAX when exogenous variables like MSP are included.

9. Conclusion

This report demonstrates that:

- MSP is a slow-moving, policy-driven anchor for wheat prices.
- Market prices follow a monthly upward trend with clear quarterly seasonality.
- Autoregression (lag models) significantly improve forecasting accuracy.
- Training-test splitting is essential to evaluate real-world performance.