Predicting the Price of a Food Basket in India

Simon Quach

February 2024

Abstract

This project investigates the trends and forecasting of food basket prices in Mumbai, India, from January 2000 to July 2023. Utilizing data from data.humdata.org, the project focuses on a comprehensive range of commodities. Methodologies include data cleaning, descriptive statistical analysis, and advanced forecasting techniques like exponential smoothing and Holt's Linear Trend Method. The findings reveal significant trends in food prices and provide future price forecasts. These insights are critical for understanding the economic aspects of food security in urban India.

1 Introduction

Food prices are a critical aspect of urban economies, impacting household budgets and food security. In rapidly developing cities like Mumbai, understanding these price dynamics is essential for policymakers, businesses, and consumers. This study aims to analyze the trends and forecast the future prices of a comprehensive food basket in Mumbai from January 2000 to July 2023. The research leverages historical data to identify patterns and apply forecasting methods. By examining these trends, the study provides insights into the economic challenges and opportunities within Mumbai's food market, contributing to a deeper understanding of urban food security in India.

2 Methodology

2.1 Data Collection

Data was sourced from data.humdata.org [1], focusing on a diverse range of commodities relevant to Mumbai's food market. This dataset included essential items such as wheat, sugar, rice, oil, potatoes, chickpeas, onions, milk, wheat flour, lentils, tomatoes, salt, tea, and ghee, among others. The dataset ranges from January 1994 to July 2023. The prices are listed in USD ($) for each commodity during each time frame.

2.2 Data Cleaning

The data cleaning process was vital to prepare the dataset for analysis. Initially, we removed the first row to eliminate the example descriptive sample, ensuring that only actual data was retained.

Subsequent to the initial clean-up, we discarded columns that were not essential to our analysis, such as administrative divisions and geographic coordinates, along with certain pricing details that were not relevant.

The date column was converted to the pandas datetime type to facilitate time-series analysis, while the price columns were transformed into numeric values.

Lastly, a crucial step was to standardize the unit of measurement for price to 'per kg' to maintain consistency across the dataset. This involved applying a custom function that adjusted prices based on the unit specified, ensuring uniformity and accuracy in price representation.

2.3 Data Preparation

The focus of this study is exclusively on the food basket prices of Mumbai from January 2000 onwards. Our data preparation involved consolidating the various commodities that constitute the food basket. This comprehensive list included staples like wheat, sugar, and rice, along with a variety of oils, legumes, and other essential food items, totaling 23 distinct commodities.[2][3]

We filtered the dataset to include only those records from the Mumbai market, ensuring a targeted analysis of the regional price trends. Additionally, we refined the dataset to consider dates post-January 2000, to align with the study's timeframe.

For each commodity on each date, we calculated the average price. These averages were then aggregated to ascertain the total food basket price per date, providing a singular figure that represented the overall cost of the food basket over time.

The culmination of this preparation was a time-series analysis, where we plotted the food basket price trend. The visualization utilized a line graph, displaying the evolution of prices across the years, with a focus on major temporal shifts and patterns.

2.4 Model Selection

In the initial phase of model consideration, three types of exponential smoothing were evaluated. However, the choice was narrowed down to **Double Exponential Smoothing**, also known as **Holt's Linear Trend Method**. This method was selected for its ability to capture both the level and trend of the data—essential for the time-series analysis of food basket prices, which are subject to both seasonal variability and long-term trends. Unlike single exponential smoothing, which only accounts for the level of the series, or Triple Exponential Smoothing, which also incorporates seasonal changes, Holt's method provided a balance of complexity and computational efficiency, making it well-suited for the prediction of food prices where seasonality is less pronounced.

2.5 Model Implementation

To implement Double Exponential Smoothing (Holt's Linear Trend Method), we must specify the mathematical formulas used to compute the model.

Forecast equation:

$$\hat{y}_{t+h|t} = l_t + hb_t$$

Level equation:

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$$

Trend equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$$

For the implementation of **Holt's Linear Trend Method** in this study, we utilized the forecast, level, and trend equations to establish a predictive model for food basket prices in Mumbai. The model parameters α and β were calibrated through an iterative process by Statsmodels, balancing responsiveness to recent changes against smoothing out random fluctuations. After applying this library's model, we attempted to create our own model from scratch, with our own functions used to compute the prices of the food baskets.

Listing 1: Python code for Holt's Linear Trend Method

```python
def holt_linear_trend(series, dates, alpha, beta, n_forecast):
    n = len(series)
    level = np.zeros(n)
    trend = np.zeros(n)
    forecast = np.zeros(n_forecast)

    # initializing level and trend
    level[0] = series[0]
    trend[0] = series[1] - series[0]

    # Holt's Linear Trend equations
    for t in range(1, n):
        level[t] = alpha * series[t] + (1 - alpha) * (level[t-1] + trend[t-1])
        trend[t] = beta * (level[t] - level[t-1]) + (1 - beta) * trend[t-1]

    # forecasting future values
    last_level = level[-1]
    last_trend = trend[-1]
    for t in range(n_forecast):
        forecast[t] = last_level + (t+1) * last_trend

    # setting dates for forecasted values
    last_date = dates[-1]
    forecast_dates = pd.date_range(start=last_date, periods=n_forecast+1, freq='M')[1:]

    return forecast, forecast_dates
```

We set α to 0.3 and β to 0.1. We also set our forecast period to 60, which is equivalent to 5 years.

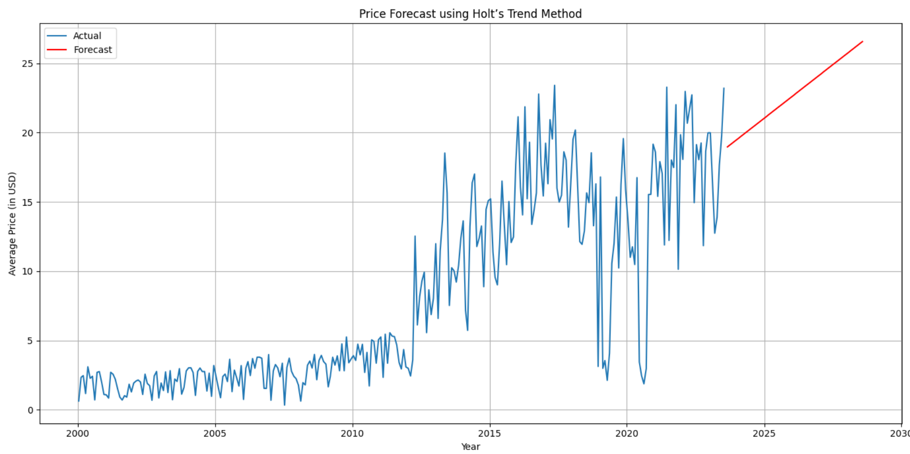

Figure 1: Our forecast for the next 60 periods (5 years)

References

[1] India - food prices, Dec 2023. URL: https://data.humdata.org/dataset/wfp-food-prices-for-india.

[2] NISHA GUPTA, MAHUA BHATTACHARJEE, and ANINDITA ROY SAHA. Sustainability and dietary change: An analysis of indian food consumption patterns. *INDIAN JOURNAL OF APPLIED ECONOMICS AND BUSINESS*, 5, 2022. URL: https://www.arfjournals.com/image/catalog/Journals

[3] Manika Sharma, Avinash Kishore, Devesh Roy, and Kuhu Joshi. A comparison of the indian diet with the eat-lancet reference diet. *BMC Public Health*, 20(1), May 2020. https://doi.org/10.1186/s12889-020-08951-8 doi:10.1186/s12889-020-08951-8.