Using ETS And SARIMA To Fit the Prices of Ice Cream

Shijun Shen *

School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, China

* Corresponding Author Email: 122090460@link.cuhk.edu.cn

Abstract. The accurate prediction of economic variables is important in economics, as expectation plays a crucial role in decision-making. The prediction of the prices of ice cream, which is a common dessert, is an interesting and little investigated topic. The paper uses both ETS and SARIMA models to fit the monthly average prices of ice cream in the United States from 1980-2020 and build a forecasting model. Then a comparison between the two models using MAE and RMSE on the test set as measures of the goodness of fit is made. The paper shows that the ETS model has a lower RMSE and MAE on the test set than the SARIMA model. The paper concludes that the ETS model may provide more accurate results when predicting the prices of ice cream. The paper discusses an interesting topic and provides a new perspective on the prediction of the prices of ice cream.

Keywords: ETS; SARIMA; ice cream.

1. Introduction

The accurate prediction of economic variables is of vital importance in the process of economic activities. When a firm is making a decision on when and how much to produce a certain good, and on how to manage the inventory of the particular good, the firm needs to estimate the future demand for that good. In this way, the firm can plan its production based on the estimated demand and reduce the waste of overproduction. When a household is deciding how much to save in a month, it needs to estimate the future income (for example, wages) and price levels in the future. Based on an expectation of future income and prices, the household can allocate resources between different time periods and optimize their total life utility.

Due to the critical role of expectations and predictions, a substantial body of economic literature has been devoted to developing accurate forecasts for key economic indicators. A food price index prediction system using time series methods for household food items like cereals, millets, and pulses was developed in 2023 [1]. Such a system can help retailers and suppliers to manage their production and inventories. The prediction of prices of other agricultural commodities like grains and corn is also a topic that has been discussed frequently [2-3]. The prediction of macroeconomic variables like inflation has always been of vital importance for the central banks and other macroeconomic agents. Studies have been devoted to how to increase the robustness and accuracy of prediction of inflation rates in the short term and in the medium to long term in various ways, like exploiting extra information, using machine learning methods [4-8]. The forecasting of asset prices is also an important topic in finance. For example, the impact of variables like macroeconomic shocks, consumer sentiment on asset prices has been studied in the past [9-10].

As ice cream is a common kind of dessert, the prediction of its prices is an interesting and important topic. However, existing literature has rarely discussed the topic. This paper uses the ETS and SARIMA models to fit the prices of ice cream. The two methods are then compared based on two measures. The ETS method is shown to have lower MAE and RMSE on the test set, which may be useful for future predictions. The forecasting of the prices of ice cream can provide valuable insights for the production decisions of the firms that produce and sell ice cream. Besides, the price of ice cream is affected by the prices of energy, sugar and dairy. Thus, the prediction of ice cream prices may have further implications in the prices of other goods.

2. Data

This paper focuses on predicting the average prices of ice cream in U.S. cities. The ice cream discussed in this article is pre-packaged, bulk ice cream that is sold in cardboard containers. The average price is computed based on both organic and non-organic ice cream, regardless of package size, flavor, or dietary features.

The ice cream price data is retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/APU0000710411. The data is collected from U.S. Bureau of Labor Statistics. Monthly average prices from 1980, Jan to 2025, Aug are used in this research. Additionally, the research uses linear interpolation to assign values to months when the prices are missing to improve the continuity of the data. The average prices of ice cream from 1980, Jan to 2025, Aug are shown in Fig. 1 below.

Fig. 1 Average Price of Ice Cream

Fig. 1 shows the trend of changes in the data. In general, the data shows an upward trend and does not have obvious seasonal characteristics. In 1990-1995 and 2015-2020, the prices did not experience dramatic increases and remain rather flat. Table 1 contains basic descriptive statistical data of the prices during the sample period.

Table 1. Descriptive Statistics

	mean	Sd	min	max	skew	kurtosis
Ice Cream	3.69	1.2	1.72	6.5	0.29	2.00

According to Table 1, the mean value is 3.69, the standard deviation is 1.2, the minimum value is 1.72, the maximum value is 6.5, the skew is 0.29, and the kurtosis is 2.00. Almost all data lies within 2 standard deviations, which indicates that the data fluctuations are relatively small. The kurtosis is 2, which is less than 3, indicating that the data distribution is flatter than the normal distribution and has fewer extreme outliers than the normal distribution. A skewness of 0.29 (which is above 0) indicates that the data distribution is more right-skewed in shape relative to a normal distribution.

3. Method

This paper firstly uses the ETS and SARIMA models to fit the data from 1980, Jan to 2019, Dec, which is the training set. Then the model will use the actual data from 2020, Jan to 2025, Aug as the test set to compute the MAE and RMSE to measure the model's prediction accuracy.

3.1. SARIMA

The SARIMA model with drift (Seasonal Autoregressive Integrated Moving Average with Drift) uses seasonal and non-seasonal autoregressive, integrated, moving average components plus an additional drift term to capture the dynamics of ice cream prices. The formula for the model is given below:

$$W_t = c + \delta \cdot t + \sum_{i=1}^{p} \varphi_i W_{t-i} + \sum_{i=1}^{P} \Phi_i W_{t-s*i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} - \sum_{i=1}^{Q} \Theta_i \varepsilon_{t-s*i} + \varepsilon_t \quad (1)$$

After performing d-order non-seasonal differencing and D-order seasonal differencing on the original time series, the stationary time series W_t is obtained. The differencing process removed the long-term trend and fixed-cycle seasonal fluctuations of the original series. C is a constant term, representing the base level of W_t. $\delta \cdot t$ captures slow, long-term linear trends in the stationary series W_t. $\sum_{i=1}^{p} \varphi_i W_{t-i}$ corresponds to the non-seasonal autoregressive (AR) component of the SARIMA model. φ_i describes the relationship between the current value W_t and its p lagged value. $\sum_{i=1}^{P} \Phi_i W_{t-s*i}$ corresponds to the seasonal autoregressive (SAR) component of the SARIMA model. s represents the seasonal cycle length; W_{t-s*i} is the lagged value from the past i seasonal cycles. Φ_i describes the relationship between the current value W_t and W_{t-s*i}. $\sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$ corresponds to the non-seasonal moving average (MA) component of the SARIMA model. ε_{t-i} represents the error of the stationary series at the i-th lagged ordinary time period. θ_i is used to describe the relationship between the current value W_t and these non-seasonal lagged prediction errors. $\sum_{i=1}^{Q} \Theta_i \varepsilon_{t-s*i}$ corresponds to the seasonal moving average (SMA) component of the SARIMA model. ε_{t-s*i} refers to the error of the stationary series at the i-th lagged seasonal cycle. Θ_i is used to describe the relationship between the current value W_t and ε_{t-s*i}. Finally, ε_t is the random error term of the stationary series at time period t. It represents the unpredictable fluctuations in W_t that cannot be explained by the autoregressive and moving average (both non-seasonal and seasonal) components.

3.2. ETS

The ETS model can be used to analyze time series data. It contains three parts that are trend, seasonality and error. The ETS model allows for different types of trends (no trend, linear trend and damped trend), seasonality (no seasonality, additive seasonality and multiplicative seasonality) and error (additive and multiplicative). Different combinations of error, trend and seasonality in the ETS model offers great flexibility and can adapt to data with different characteristics.

The trend component of the ETS model describes the long-term trend in the time series. The seasonal component describes the regular fluctuations in the time series during a fixed period. The error component represents the random fluctuations that cannot be explained by trend or seasonality.

The ets() function in R fits the time series data using the ETS model with different specifications. Secondly, the ETS model with lowest AICc will be identified by the ets() function. Thirdly, the ets() function will output the appropriate type of error, trend and seasonality and the estimated model parameters.

4. Results

Using the test set, the SARIMA (2,1,1) (1,0,1) (12) with drift and ETS (M, A, N) models are chosen. Table 2 shows the estimated parameters of the SARIMA model.

Table 2. Parameters of the SARIMA model

	Ar1	Ar2	Ma1	Sar1	Sma1	drift
coefficient	0.2405	0.1623	-0.7266	0.6829	-0.5420	0.0090
s.e.	0.1117	0.0713	0.0992	0.1079	0.1219	0.0028

Fig.2 shows the fitted values given by the SARIMA model and the actual values in the test set. The grey line represents the actual values and the blue line represents the fitted values.

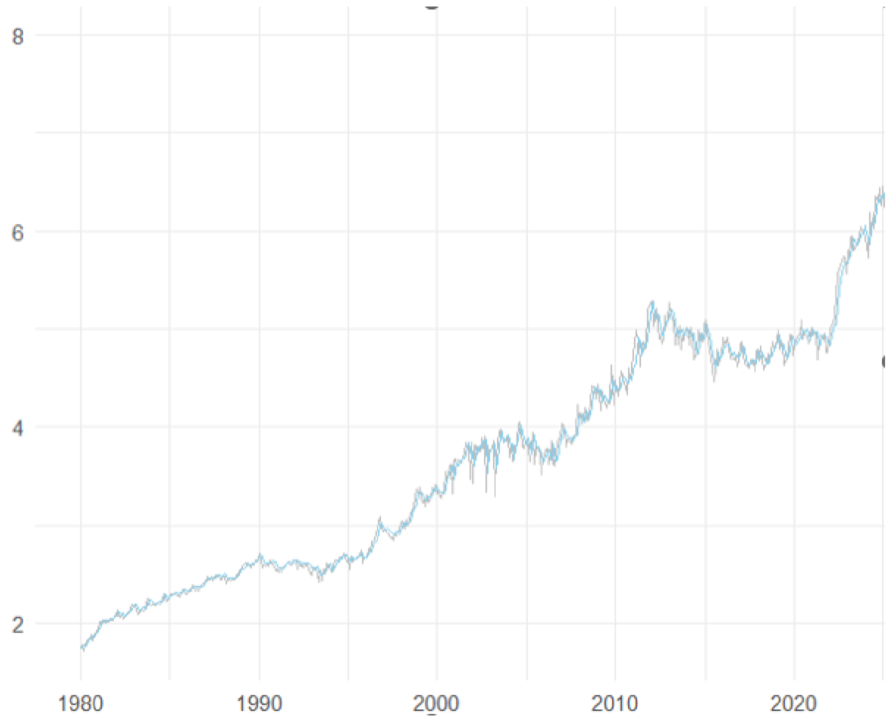

Fig. 2 SARIMA Results

The ETS (M, A, N) model has multiplicative errors, additive trend and no seasonal terms. Its observation equation is $y_t = l_{t-1} \cdot (1+\varepsilon_t)$. Its state update equations are $l_t = l_{t-1} + b_{t-1} + \alpha \cdot (y_t - l_{t-1})$ and $b_t = b_{t-1} + \beta \cdot \alpha \cdot (y_t - l_{t-1})$. $\alpha = 0.5218$. $\beta = 1e\text{-}04$. $l_0 = 1.7436$. $b_0 = 0.0077$.

Fig. 3 shows the fitted values given by the ETS model and the actual values in the test set. The grey line represents the actual values, and the blue line represents the fitted values.

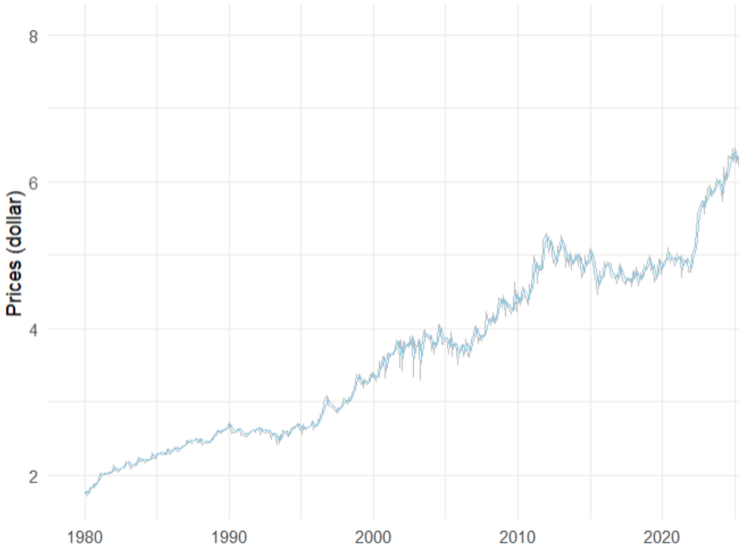

Fig. 3 ETS Result

After the parameters of the ETS and SARIMA model are estimated, the paper further evaluates their model performance. The performance of the two models on the test set is shown in Table 3.

Table 3. Performance of different models on the test set

	MAE	RMSE
SARIMA	0.5741	0.7306
ETS	0.5268	0.6730

The ETS model, which has lower MAE and RMSE than the SARIMA model, is the superior model under two measures of prediction accuracy. It is interesting that although the SARIMA model includes seasonality and the ETS model does not include seasonality, the ETS model still performs better in predicting the prices of ice cream.

5. Conclusion

Predicting economic variables accurately is of great importance for economic agents. The prediction of prices of ice cream is an interesting and important topic, which has not been discussed by others extensively. Using the data in the United States from 1980-2020, the paper fits both ETS and SARIMA models and compares their performance on the test set. Under two measures, the ETS model always performs better than the SARIMA model. Thus, the ETS model may be a better choice when predicting the prices of ice cream compared to the SARIMA model.

The current research's limitation is that it only exploits the information in the time series of prices of the ice cream. In the future research, researchers can include variables that influences the supply and demand of ice cream like temperature, the prices of energy, sugar and dairy. As these variables tend to be mutually determined, the VAR model could be established to analyze the relationships between different variables. Future research could further analyze the causal relationships between different variables and the impact of shocks in other variables on the price of ice cream.

References

[1] Majhi S K, Bano R, Pradhan R, et al. Food price index prediction using time series models: A study of Cereals, Millets and Pulses. International Conference on Technology Advances for Green Solutions and Sustainable Development. Cham: Springer Nature Switzerland, 2024: 216-238.

[2] Brignoli P L, Varacca A, Gardebroek C, et al. Machine learning to predict grains futures prices. Agricultural Economics, 2024, 55(3): 479-497.

[3] Xu X, Zhang Y. Yellow corn wholesale price forecasts via the neural network. EconomiA, 2023, 24(1): 44-67.

[4] Forte F D. Forecasting short-term inflation in Argentina with Random Forest Models. arXiv, 2024. arXiv:2410.01175.

[5] Zaman S. Improving inflation forecasts in the medium to long term. Economic Commentary, 2013(2013-16).

[6] Fulton C, Hubrich K. Forecasting US inflation in real time. Econometrics, 2021, 9(4): 36.

[7] Ang A, Bekaert G, Wei M. Do macro variables, asset markets, or surveys forecast inflation better. Journal of monetary Economics, 2007, 54(4): 1163-1212.

[8] Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects. Science, 2015, 349(6245): 255-260.

[9] Qian Y, Zhang Y. Long-term forecasting in asset pricing: Machine learning models' sensitivity to macroeconomic shifts and firm-specific factors. The North American Journal of Economics and Finance, 2025, 78: 102423.

[10] Lemmon M, Portniaguina E. Consumer confidence and asset prices: Some empirical evidence. The Review of Financial Studies, 2006, 19(4): 1499-1529.