A Multi-Model, Ensemble Approach to Forecasting United States Food Prices

Weifang Liang

Department of Agricultural Economics, Texas A&M University

wf.liang@tamu.edu

Yong Liu

Department of Agricultural Economics, Texas A&M University.

yong.liu@ag.tamu.edu

Simon Somogyi

Department of Agricultural Economics, Texas A&M University

simon.somogyi@ag.tamu.edu

David Anderson

Department of Agricultural Economics, Texas A&M University

david.anderson@ag.tamu.edu

Selected paper prepared for the presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

A Multi-Model, Ensemble Approach to Forecasting United States Food Prices

Weifang Liang, Yong Liu, Simon Somogyi, and David Anderson

Abstract

It is important for stakeholders in agricultural economics to accurately predict food prices to make better decisions. We propose an ensemble approach that combines multiple models to predict food prices more accurately. We choose autoregressive integrated moving average (ARIMA), exponential smoothing, local linear regression, and Gaussian Process as candidate models. The final prediction is the weighted average of the predictions from the candidate models. We use ARIMA as the baseline model. It outperforms most candidate models when fitting individually. By leveraging strengths from diverse models, all ensemble models demonstrate superior forecast accuracy. They achieve lower mean absolute percentage error and root mean square error than ARIMA when adding all selected food categories together.

Keywords: food price, forecast, ensemble learning

Introduction

Food price is an essential component in food and agricultural economies. It is often influenced by various factors and impacts a wide range of individuals. The inherent uncertainty in food prices increases the risks for stakeholders involved in agricultural products. Accurately predicting food prices enables informed decision-making and strategic planning for farmers, consumers, producers, and policymakers. The prediction process provides invaluable insights into understanding the factors driving food price fluctuations and makes significant impacts on the food supply chain. It can also help consumers make food budgeting decisions that can influence their food choices and subsequently their nutritional intakes. For primary producers, processors, wholesalers, and retailers, it can influence their production and marketing strategies and help them anticipate market trends and adjust their operations accordingly. With more accurate food price forecasts, policymakers gain better insights into market dynamics and consumer behavior which can assist them in making timely policies that ensure food security and market stability. In essence, forecasting food prices is not just about predicting numbers but about understanding the complexity of the food supply chain and its impact on various members of society. It can help guide decisions that impact all processes from production to consumption.

Food prices can exhibit linear and nonlinear trends, with significant fluctuations and seasonality. Previous research has shown the difficulty of capturing the linearity and non-linearity simultaneously in a single model and has demonstrated the superiority of hybrid models over the single models in its components (Wang et al., 2020). To capture more relevant features in time series and enhance the accuracy of food price predictions, we employ a robust ensemble approach. It integrates multiple forecasting methods to build the food price predictor. By averaging various models, the ensemble approach reduces the risk of over-fitting and the likelihood of encountering a local minimum. When the true model is not one of the candidates, averaging multiple models can increase the possibility of getting closer to the true model and provide a better match for the data (Sagi and Rokach, 2018). Additionally, the candidate models need to vary significantly to achieve the desired accuracy in predictions (Breiman, 1996). The ensemble approach has demonstrated its efficacy in enhancing predictive capabilities in the agriculture field (Huang et al., 2017; Ribeiro and dos Santos Coelho, 2020). As an extension of previous studies, our research incorporates a

more diverse set of candidate models and uses four methods to determine optimal weights for each model, aiming to improve the forecasting performance.

Ensemble learning encompasses a series of techniques, including bagging, boosting, and stacking (Brown, 2010). Stacking treats a set of candidate models as the level-0 models, and a meta-model as the level-1 generalizer. Outputs from the level-0 models are the inputs of the level-1 model, thereby combining the strengths of all candidate models (Wolpert, 1992). We select four models as the level-0 candidate models, including autoregressive integrated moving average (ARIMA), exponential smoothing, local linear regression, and Gaussian Process (GP). Each candidate is chosen because they specialize in one of the conditions when dealing with time series data. ARIMA and exponential smoothing are effective when the data does not display complex patterns (Zhang, 2003). Non-parametric approaches are more flexible and excel in capturing nonlinear trends (Vilar-Fernández and Cao, 2007). GP is a probabilistic machine learning approach that can flexibly model complex datasets with uncertainty captured (Williams and Rasmussen, 2006). Each model is independently trained on historical data and then combined to leverage the strengths of these methods so that the model can acquire the best collective performance. A non-negative least-squares linear regression is employed as the level-1 model (Breiman, 1996). As a comparison, a linear regression model without the non-negative constraint is fitted (Wu et al., 2017). We also conduct a weighted ensemble approach that assigns weights proportional to the inverse of mean squared errors (Pawlikowski and Chorowska, 2020). Models that demonstrate better predictive performance will receive higher weights and the final model is the weighted average of the candidate models.

Methodology

Level-0 Models

High accuracy and the diversity of the candidate models are crucial in building the ensemble model. However, there is a subtle trade-off between the two features (Li et al., 2021). By balancing accuracy and diversity, we can improve the ensemble performance and lower the risk of over-fitting. We select four candidate models specialized in different conditions to incorporate diversity while ensuring their capability to capture the temporal relationships to mitigate the risk of low accuracy.

ARIMA

ARIMA is introduced to deal with the potential non-stationary and autocorrelation problems in time series data (Box and Jenkins, 1976). The model contains three parts. The autoregressive (AR) process examines how the current value is related to the past values by regressing it with the lagged observations. The integrated (I) part solves the non-stationary problem by differencing the data. The moving average (MA) process focuses on the relationship between the current value and the past residual terms. The combination of the three processes makes it easy to implement and interpret. An ARIMA(p, d, q) model is defined as:

$$(1 - L)^d y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}, \tag{1}$$

where L is the lag operator. p and q define the number of lagged terms that explain the current value in the AR and MA process respectively. d defines the times the differencing process is performed to achieve stationarity in the time series.

Exponential Smoothing

In a time series, recent values are more indicative of predicting future behavior. Exponential smoothing addresses this by predicting future values based on the weighted average of past observations with exponentially decreasing weights assigned to older values (Brown, 1956). The process can be described as

$$\begin{aligned}
\bar{y}_t &= b y_t + (1 - b)\bar{y}_{t-1} && (2)\\
&= b y_t + b(1 - b) y_{t-1} + (1 - b)^2 \bar{y}_{t-2} && (3)\\
&= b[y_t + (1 - b) y_{t-1} + (1 - b)^2 y_{t-2} + ... + (1 - b)^{t-1} y_1] + (1 - b)^t y_0, && (4)
\end{aligned}$$

where $\bar{y}_t$ is the expected value at time t while y are the observed values. By adjusting the smoothing parameter b, we can control the balance between capturing short-term fluctuations and smoothing out noise. With a large b, the impact from historical data fades away quickly. Exponential smoothing can also handle time series with seasonality, additive, or multiplicative trends (Holt,

1957; Winters, 1960). The Holt linear trend model includes a trend adjustment increment r:

$$\bar{y}_t = ay_t + (1-a)(\bar{y}_{t-1} + r_{t-1}). \tag{5}$$

The trend adjustment is estimated with a parameter c, which controls the rate of change in the trend increment:

$$r_t = c(\bar{y}_t - \bar{y}_{t-1}) + (1-c)r_{t-1}. \tag{6}$$

A T-step-ahead prediction is estimated as:

$$\bar{y}_{t+T} = \bar{y}_t + r_t T. \tag{7}$$

Local Linear Regression

Local linear regression is a non-parametric approach that does not impose a specific functional form. Relaxing the parametric assumption lets the data reveal the appropriate model. Hence, it is capable of capturing relationships not detected by specific functional forms. Instead of fitting a single global model, local linear regression fits separate linear models to local neighborhoods around the data (Cleveland, 1979). When modeling a time series, it can capture the time-varying relationship of the data. Estimating $g(t)$ is equivalent to estimating the intercept a of a local linear regression in the neighborhood of time t (Racine et al., 2008):

$$g(t_0) \approx g(t) + (\partial g(t)/\partial t)(t_0 - t) = a + b(t_0 - t). \tag{8}$$

The marginal effect can be captured by slope b. a and b are estimated by minimizing

$$S = \sum_{i=1}^{n} (Y_i - a - b(T_i - t))^2 K(\frac{T_i - x}{h}), \tag{9}$$

where K is a kernel function that assigns weights to the neighborhood T_i based on their distance to the target time t. Altering the kernel function does not significantly change the later estimate (Racine et al., 2008). Hence we employ a Gaussian kernel, a proper choice for continuous data (Li and Yuan, 2019).

The bandwidth h decides the size of the local neighborhood, thereby also named the smoothing parameter. Bandwidth selection is essential in the non-parametric approach since incorrectly specifying the bandwidth will cause either over-smoothness or over-roughness on the estimated trend (Hansen, 2004). Instead of using a fixed bandwidth, we employ the generalized nearest neighbors bandwidth and the adaptive nearest neighbors bandwidth that vary the bandwidth size across the data (Dette and Gefeller, 1995; Shi, 2010). By adjusting the bandwidth locally, the model becomes more flexible and robust.

Gaussian Process

Gaussian Process is a probabilistic machine learning model for flexible non-parametric regression. Instead of using a parametric function, GP uses time index t as input and takes the following form:

$$f(\mathbf{t}) \sim GP(m(\mathbf{t}), K(\mathbf{t}, \mathbf{t}')). \tag{10}$$

It can be intuitively understood as a multivariate Gaussian distribution with infinite variables. A mean function and a covariance function can fully define it. The mean function reveals the average behavior of the model, while the covariance function captures the relationship between data points. Data points that are closer on the x-axis indicate a shorter time gap between them, implying a stronger relationship. Hence, it is important to select a suitable covariance function, as it determines how effectively the model can capture the relationships exhibited by the observed data points. GP with a linear kernel can model a linear relationship between the input and the output variables. The squared exponential (SE) kernel is effective for smooth non-linear trends, while the Matérn class of kernels excels at capturing rougher trends (Williams and Rasmussen, 2006). Adjusting the hyperparameters for the kernels allows us to generate trends with different degrees of smoothness and length scales. It enables us to determine the maximum fluctuation levels or the range of slopes in the case of a linear trend. When dealing with trends exhibiting complex relationships, combining multiple kernels can effectively capture more relevant features.

GP provides probabilistic estimation rather than point estimates. Each estimation is a Gaussian distribution. Hence, the joint distribution of observed data and future values is needed to predict

values at time $\mathbf{t}^*$. Denote $\mathbf{f}^* = f(\mathbf{t}^*)$, the joint distribution is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim N\left(m(\mathbf{t}), \begin{bmatrix} K(\mathbf{t},\mathbf{t}) + \sigma_n^2 I & K(\mathbf{t},\mathbf{t}^*) \\ K(\mathbf{t},\mathbf{t}^*)' & K(\mathbf{t}^*,\mathbf{t}^*) \end{bmatrix} \right). \tag{11}$$

The covariance matrix $K(\mathbf{t},\mathbf{t}^*)$ captures the degree of correlation between observed data and future values. By the Bayes' theorem $p(\mathbf{y}_1|\mathbf{y}_2) = p(\mathbf{y}_1,\mathbf{y}_2)/p(\mathbf{y}_2)$, the predictive distribution of $\mathbf{f}^*$ given the data is

$$\mathbf{f}^*|\mathbf{t},\mathbf{y},\mathbf{t}^* \sim N(\bar{\mathbf{f}}^*, cov(\mathbf{f}^*)), \tag{12}$$

with predictive mean and predictive covariance:

$$\bar{\mathbf{f}}^* = K(\mathbf{t}^*,\mathbf{t})[K(\mathbf{t},\mathbf{t}) + \sigma_y^2 I]^{-1}\mathbf{y}; \tag{13}$$

$$cov(\mathbf{f}^*) = K(\mathbf{t}^*,\mathbf{t}^*) - K(\mathbf{t}^*,\mathbf{t})[K(\mathbf{t},\mathbf{t}) + \sigma_y^2 I]^{-1}K(\mathbf{t},\mathbf{t}^*). \tag{14}$$

The predictive mean at $\mathbf{t}^*$ is linearly related to $\mathbf{y}$. The predictive covariance is the prior uncertainty minus the reduction in uncertainty after observing the data. By enlarging the dataset, more information is incorporated and we can be more certain about our prediction compared to the prior.

Level-1 Model

We compare multiple methods to search for a set of optimal weights assigned to the level-0 models. First, we use a non-negative least-squares (NNLS) linear regression as the level-1 model. It almost always outperforms the candidate models in the simulations (Breiman, 1996). The leave-one-out cross-validation (LOOCV) is applied to the level-0 models to generate predictors that fit into the level-1 model. Then, the optimal weight assigned to candidate model M_k in level-0 is:

$$\hat{w}_k = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \sum_k w_k \hat{f}^{-i}(x_i) \right), \tag{15}$$

where $\hat{w}_k \geq 0, \sum_{k=1}^{K} = 1$, K is the number of candidate models. The final prediction is the weighted average of all candidate models:

$$\hat{y} = \sum_{k=1}^{K} \hat{w}_k f_k(x). \tag{16}$$

We implement the NNLS model through two approaches: a) directly deriving non-negative weights that minimize the mean squared error (MSE) through disciplined convex optimization (Fu et al., 2020), and b) using the Lawson-Hanson algorithm for NNLS (Lawson and Hanson, 1995). As a comparison, we conduct a traditional linear regression without imposing non-negativity (Wu et al., 2017). Considering that the non-negative constraint is necessary to improve the prediction accuracy (Breiman, 1996; LeBlanc and Tibshirani, 1996), we apply a coefficient transformation after the linear regression. We set negative values to zero and normalize the remaining coefficients so that they sum up to one. Additionally, we conduct a weighted average ensemble (WAE) by setting the weights as:

$$\hat{w}_k = \frac{g_k(S)}{\sum_{k=1}^{K} g_k(S)}, \tag{17}$$

where S is the mean squared error (MSE) from model M_k, and $g_k(S) = 1/S$ (Pawlikowski and Chorowska, 2020). The models with lower MSE exhibit higher forecast accuracy and are given more weight. We denote these approaches as "Ensemble MSE", "Ensemble NNLS", "Ensemble Linear", and "Ensemble WAE" respectively.

Application

To analyze changes in food prices across various categories in the United States (U.S.), we collect Consumer Price Index (CPI) data from the U.S. Bureau of Labor Statistics. Our dataset contains seven food categories: food, food at home, food away from home, meats, pork, fish and seafood, and fresh fruits. They are monthly data with seasonal adjustments. All categories have 588 observations, spanning from January 1975 to December 2023.

In January 2023, the U.S. Department of Agriculture (USDA), Economic Research Service's Food Price Outlook revised its forecasting approach. They adopted ARIMA to predict the food prices for the next 18 months (MacLachlan et al., 2022). Hence, we set ARIMA as the benchmark. Instead of 18 months, we plan to forecast the food price index for 12 months at the beginning of

each year. We simulate this process for the recent 10 years (2014-2023) and evaluate the forecast accuracy using the mean absolute percentage error (MAPE) and the root mean square error (RMSE) between the predicted values and the actual values:

$$\text{MAPE} = 100\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{18}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{19}$$

where y_t are the actual values, $\hat{y}_t$ are the fitted values.

In the LOOCV process, each data point is left out once as a validation set, and the model is trained on all other data points. Applying the LOOCV on time series may cause using recent data to predict past data. To implement the method for time series data, we adopt a time series cross-validation (TSCV), which adjusts the traditional cross-validation (CV) process to the temporal structure. It splits data into training and testing sets while preserving the temporal order. An exact LOOCV can be computationally expensive, and a smaller fold CV is more effective than LOOCV in the simulation (Breiman, 1996). Therefore we expand the training set by 12 months for each CV iteration, rather than incrementing it by 1 month. Each test dataset contains a 12-month period that immediately follows the training period.

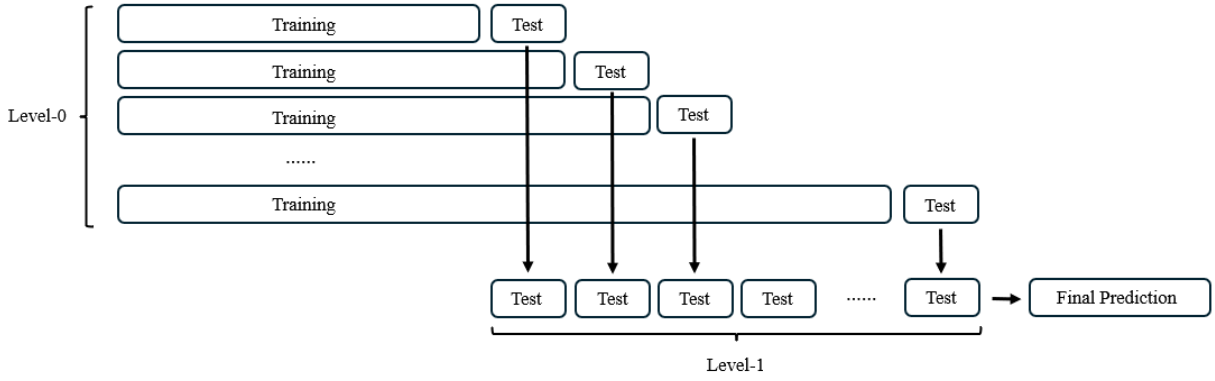


Figure 1: TSCV in Ensemble

We train the level-0 models individually using data preceding each year from 2007 to 2023 to forecast the index for that respective year. Each candidate model is fitted separately with different approaches to tune parameters. The order of the ARIMA model is selected by choosing the

one with the lowest Akaike information criterion (AIC). Exponential Smoothing uses the latest 30% training data as the validation set to identify the optimal smoothing parameter. The non-parametric estimation uses least-squares cross-validation to select appropriate bandwidth (Racine and Li, 2004). GP estimates the hyperparameters through Maximum a Posteriori (MAP) using the Limited-memory quasi-Newton (L-BFGS) optimization algorithm (Wright, 2006). To simultaneously capture the long-term upward trend and the short-term fluctuations of the food price index, we combine the linear kernel, SE kernel, and Matérn 5/2 kernel as our covariance matrix to build the GP.

Table 1: Level-0 Models 2007 - 2023

Item	Baseline ARIMA	Gaussian Process	Exponential Smoothing	Local Linear Regression Generlized	Local Linear Regression Additive
Average MAPE					
Food	1.19	**0.99**	1.15	1.48	1.06
Food at home	1.66	**1.34**	1.69	2.24	1.62
Food away from home	**0.47**	0.52	0.70	0.59	0.50
Meats	**2.79**	2.99	3.12	3.84	4.07
Pork	3.33	**3.22**	3.51	4.76	3.31
Fish and seafood	**1.60**	1.75	2.17	2.15	2.28
Fresh fruits	**2.32**	2.44	2.54	2.37	6.81
Total	13.36	**13.25**	14.88	17.43	19.65
Average RMSE					
Food	3.73	3.20	3.47	4.54	**3.12**
Food at home	4.79	**4.04**	4.78	6.44	4.54
Food away from home	**1.58**	1.80	2.42	2.06	**1.58**
Meats	**8.26**	8.93	9.27	11.18	11.90
Pork	**8.26**	**8.26**	8.61	11.61	8.29
Fish and seafood	**5.52**	5.79	7.53	7.41	7.63
Fresh fruits	**9.11**	9.79	10.11	9.79	27.88
Total	**41.25**	41.81	46.19	53.03	64.94

Note: The lowest score in each category is highlighted in bold.

Table 1 presents the average MAPE and RMSE for all candidate models in predicting food CPI from 2007 to 2023. Most models in our results can not outperform the ARIMA model when evaluating candidate models individually. Among the seven categories, ARIMA achieves the lowest average MAPE in four categories, and the lowest average RMSE in five categories. ARIMA and GP achieve the lowest total RMSE and MAPE respectively by adding all seven categories.

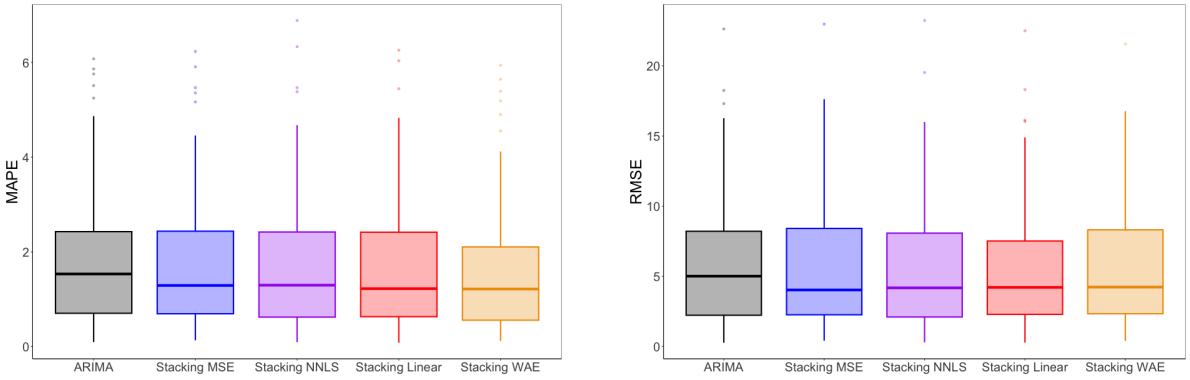
In the next step, the outputs from the level-0 models are used as the inputs for the level-1 model. It creates a dataset from 2007 to 2023 with 5 variables, each representing the prediction from a candidate model. The ensemble predictors are created for the last 10 years (2014-2023). For each year, we use the forecasts from the level-0 models up to the end of the preceding year to construct the level-1 models and make forecasts for the subsequent year. Table 2 demonstrates the average MAPE and RMSE of ARIMA, GP, and the ensemble models over the last ten years.

Table 2: Ensemble Models 2014 - 2023

Item	Baseline ARIMA	Gaussian Process	Ensemble MSE	Ensemble NNLS	Ensemble Linear	Ensemble WAE
Average MAPE						
Food	1.27	1.15	1.41	**1.08**	1.10	1.10
Food at home	1.69	1.50	**1.41**	1.45	1.51	1.55
Food away from home	0.45	0.53	0.55	0.47	0.51	**0.43**
Meats	2.62	**2.48**	2.85	2.74	2.70	2.51
Pork	2.83	3.03	2.49	2.79	2.53	**2.48**
Fish and seafood	1.69	**1.50**	1.62	1.66	1.56	1.67
Fresh fruits	2.22	2.31	1.93	1.94	1.94	**1.92**
Total	12.77	12.50	12.26	12.13	11.85	**11.66**
Average RMSE						
Food	4.30	3.93	4.77	**3.65**	3.73	**3.65**
Food at home	5.23	4.80	**4.38**	4.56	4.71	4.76
Food away from home	1.70	2.03	2.20	1.78	1.95	**1.66**
Meats	8.71	8.58	9.33	9.11	8.91	**8.33**
Pork	7.68	8.46	6.98	7.66	6.99	**6.77**
Fish and seafood	6.10	**5.36**	5.81	5.97	5.62	6.04
Fresh fruits	9.07	9.57	**7.83**	8.04	8.00	8.06
Total	42.79	42.73	41.30	40.77	39.91	**39.27**

Note: The lowest score in each category is highlighted in bold.

By assigning more weights to the models that effectively capture time series features, ensemble models exhibit a lower average of MAPE and RMSE in most categories than ARIMA. All ensemble models exhibit lower total average MAPE and average RMSE than ARIMA and GP models. The median of MAPE and RMSE from the ensemble models are also lower than ARIMA as demonstrated in Figure 2. Ensemble WAE achieves the lowest total average MAPE and total average RMSE among all the models.



(a) MAPE (b) RMSE

Figure 2: MAPE and RMSE Boxplots

Conclusion

In this paper, we adopt an ensemble model for predicting U.S. food prices, which trains multiple models and combines them in a way that maximizes their collective predictive performance. We simulate a 12-month forecast at the beginning of each year over the past decade. The ARIMA model, employed by the current USDA Food Price Outlook to forecast the food price index, is selected as our baseline model. When fitting each model individually, the ARIMA model outperforms most candidate models. After leveraging the strengths of diverse models, the ensemble models outperform the ARIMA model with both lower MAPE and RMSE.

The ensemble model is promising in achieving more accurate food price prediction. Instead of fitting existing data, it focuses on enhancing the forecast ability by assigning weights to models with higher predictive accuracy in the validation set. It effectively handles the situation when the true model might not be among the candidates. It also provides a general framework that can be extended to forecast other categories and can be adapted to a larger data setup. Its adaptability allows for the modification of candidate models, either by adding or removing them, to achieve optimal predictions. Further study should consider incorporating a more diverse set of candidate models, adding covariates to the candidate models, and exploring alternative methods for the weight assignment.

References

Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day series in time series analysis and digital processing. Holden-Day.

Breiman, L. (1996). Stacked regressions. *Machine learning*, 24:49–64.

Brown, G. (2010). Ensemble learning. *Encyclopedia of machine learning*, 312:15–19.

Brown, R. (1956). *Exponential Smoothing for Predicting Demand*. Little.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Dette, H. and Gefeller, O. (1995). The impact of different definitions of nearest neighbour distances for censored data on nearest neighbour kernel estimators of the hazard rate. *Journal of Nonparametric Statistics*, 4(3):271–278.

Fu, A., Narasimhan, B., and Boyd, S. (2020). CVXR: An R package for disciplined convex optimization. *Journal of Statistical Software*, 94(14):1–34.

Hansen, B. E. (2004). Bandwidth selection for nonparametric distribution estimation. *manuscript, University of Wisconsin*.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *ONR Memorandum*, 52.

Huang, X., Huang, G., Yu, C., Ni, S., and Yu, L. (2017). A multiple crop model ensemble for improving broad-scale yield prediction using bayesian model averaging. *Field Crops Research*, 211:114–124.

Lawson, C. L. and Hanson, R. J. (1995). *Solving least squares problems*. SIAM.

LeBlanc, M. and Tibshirani, R. (1996). Combining estiamates in regression and classification. *Journal of the American Statistical Association*, 91(436):1641–1650.

Li, T. and Yuan, M. (2019). On the optimality of gaussian kernel based nonparametric tests against smooth alternatives. *arXiv preprint arXiv:1909.03302*.

Li, W., Paffenroth, R. C., and Berthiaume, D. (2021). Neural network ensembles: Theory, training, and the importance of explicit diversity. *arXiv preprint arXiv:2109.14117*.

MacLachlan, M., Chelius, C., and Short, G. (2022). Time-series methods for forecasting and modeling uncertainty in the food price outlook. Technical report.

Pawlikowski, M. and Chorowska, A. (2020). Weighted ensemble of statistical models. *International Journal of Forecasting*, 36(1):93–97.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130.

Racine, J. S. et al. (2008). Nonparametric econometrics: A primer. *Foundations and Trends® in Econometrics*, 3(1):1–88.

Ribeiro, M. H. D. M. and dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied soft computing*, 86:105837.

Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1249.

Shi, X. (2010). Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *International Journal of Geographical Information Science*, 24(5):643–660.

Vilar-Fernández, J. M. and Cao, R. (2007). Nonparametric forecasting in time series—a comparative study. *Communications in Statistics—Simulation and Computation®*, 36(2):311–334.

Wang, L., Feng, J., Sui, X., Chu, X., and Mu, W. (2020). Agricultural product price forecasting methods: research advances and trend. *British Food Journal*, 122(7):2121–2138.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

Wright, S. J. (2006). *Numerical optimization*.

Wu, H., Wu, H., Zhu, M., Chen, W., and Chen, W. (2017). A new method of large-scale short-term forecasting of agricultural commodity prices: illustrated by the case of agricultural markets in beijing. *Journal of Big Data*, 4(1):1–22.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.