A Hybrid Pork Price Forecast Model Based on LightGBM and Segment-wise Attention

Yan Wang*, Fang Deng, Xudong Yang

Beijing University of Posts and Telecommunications, Beijing, China

Abstract

Long-term forecasting of pork prices is important for production investment and price regulation. However, the pork price series exhibits the characteristics of non-stationary, non-linear, and pseudo-periodic, which makes the forecasting task challenging. To deal with the intricate temporal patterns, in this paper, we propose a novel hybrid model based on the decomposition-ensemble framework and provide tailored prediction methods for different components. We begin by decomposing the original series into trend and cyclical components using the HP filter. To address the problem of pseudo-cycle, we design a segment-wise neural network, which introduces the attention mechanism to model correlations at the segment level. Then LightGBM is used to implement a dynamic regression model for the trend component. Finally, the predicted series are aggregated. Predictions are carried out in a rolling forward manner to avoid the problem of information leakage. Experiments show that our model outperforms other single or hybrid models.

Keywords

Time series, decomposition-ensemble framework, dynamic regression model, segment-wise correlation

1. Introduction

Pork production has always been topping the list of meat production in China. According to the National Bureau of Statistics, the average domestic pork production in the past five years reached 49.30 million tons, accounting for 59.52% of the total domestic meat production. However, pork prices have been shrouded in frequent fluctuations. Especially in recent years, against the background of transformation and upgrading of the hog industry chain, pork prices have shown great volatility, impacting the nation's daily life, the operation of farmers' enterprises and the smooth operation of the economy. Therefore, studying the pattern of pork price fluctuations and establishing an effective long-term forecasting model are of great significance for the pig industry.

Decomposition-ensemble model is a standard method in time series analysis, and it has been widely used in dealing with the intricate temporal patterns[1]. Choosing a proper method to separate the series and designing a suitable algorithm for each component are two key problems when constructing a forecasting model under the decomposition-ensemble framework[2]. Inspired by the value theory, we use the HP filter to decompose the original price series into the trend and cyclical components. Dynamic regression models combine the features of explanatory and time series models and are able to model the causes of price changes and the dependent information in historical prices. Given the good performance of the tree model in generalization, we choose the LightGBM model to implement the dynamic regression model and to forecast the trend series. The final prediction is the combination of the predicted series.

This paper is organized as follows: Section 2 introduces the data and the model we proposed. Section 3 describes the experimental settings, results and analysis. Conclusions are drawn in Section 4.

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control
EMAIL: *wangyan_cs@bupt.edu.cn (Yan Wang)

2. Materials and Methods

2.1. Data Source and Pre-processing

Weekly pork and hog pricing as well as monthly breeding sow stock are collected from January 2006 to May 2022 from the National Development and Reform Commission, the China Pig website, and the Ministry of Agriculture. For these datasets, a few pertinent statistics indices are computed and displayed in Table 1.

Table 1. Description of datasets.

Description	Sample Num	Mean	Max	Min	Standard deviation
Pork price	856	24.29	59.58	9.81	9.86
Hog price	856	15.76	40.97	5.88	6.65
Sow livestock	197	4297.21	5078.20	2960.73	585.43

As can be seen, there is a positive association between price volatility and price level. The series also contains several missing values. In the preprocessing stage, the initial sequence is transformed using the box-cox method to guarantee the consistency of the data distribution. The Box-cox transformation is a family of transformations that include both logarithmic and power transformations depending on the parameter λ, which is defined as follows:

$$w_t = \begin{cases} \log(y_t), & if\ \lambda = 0; \\ ({y_t}^{\lambda} - 1)/\lambda, & otherwise. \end{cases} \tag{1}$$

Figure 1 Time series of pork and hog prices from 2006 to 2022 before and after data pre-processing.

For the predicted results, we need to reverse the transformation to obtain forecasts on the original scale. The reverse Box-Cox transformation is given by:

$$y_t = \begin{cases} \exp(w_t), & if\ \lambda = 0; \\ (\lambda w_t + 1)^{\frac{1}{\lambda}}, & otherwise. \end{cases} \tag{2}$$

We use the 3σ rule to identify outliers in the sequence. Outliers and missing values are filled using the weighted average of nearby points. The hog and pork price series before and after preprocessing are shown in Figure 1.

2.2. Overall Process of the Proposed Hybrid Model

Figure 2 shows the entire process of the prediction model, including the following steps:

Step 1: Separate the cyclical component and the trend component of the transformed hog price series using the HP filter.

Step 2: Extract features on price series and other variable series by time series feature engineering and build the LightGBM model to forecast the trend component.

Step 3: Predict the cyclical component using encoder-decoder model with segment-wise attention.

Step 4: Add the two components and reverse the Box-cox transformation.

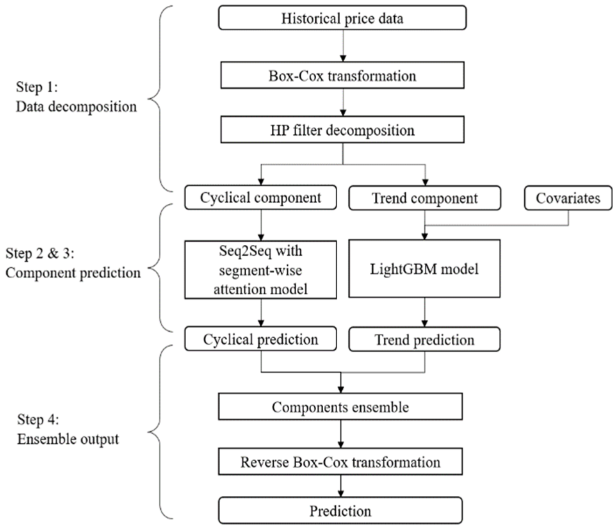

Figure 2 Structure of the proposed hybrid model.

2.3. HP Filter

It is generally agreed that separating the temporal patterns of entanglement by time series decomposition can highlight the intrinsic properties of the sequence of components. Inspired by the law of market value, we choose HP filter to fit the original sequence to obtain a trend sequence that is easier to predict. The trend series is the solution to the following optimization problem:

$$\min_{\{x_t\}_{t=1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - x_t)^2 + \lambda \sum_{t=1}^{T} [(x_t - x_{t-1}) - (x_{t-1} - x_{t-2})]^2 \right\} \tag{3}$$

where y_t is the original series and x_t is the trend series. This optimization problem can be solved using the method of least squares.

The periodic series is obtained by the difference between the original series and the trend series, as shown in the following formula:

$$C_t = P_t - T_t \tag{4}$$

$$P_t = \{x_i\}_{t=1}^{T}, T_t = \{y_i\}_{t=1}^{T} \tag{5}$$

2.4. Prediction of Trend Component Based on LightGBM

2.4.1. Constructing the Inputs

Our experiments show that autoregressive time series models relying only on historical information cannot effectively deal with the huge fluctuations in pork and hog prices, which will be discussed in Section 3.3, so we do extensive feature engineering to extract effective features from price series and covariate series. According to the research of relevant literature[3,4] and the exploratory analysis of the data, the following features are selected under the framework of the dynamic regression model. On the one hand, timestamps, holidays and breeding sow stocks are chosen as three main covariates because capacity, seasonality and holiday effects play an important role in the formation of pork and hog prices. A set of dummy variables are used to represent the month of the year. It should be noted that the number of dummy variables needs to be less than the number of categories to avoid the "dummy variable trap". For holiday features, we customize holiday variables and consider the window effect of holidays. In addition, considering that there is a time lag in the correlation between the number of reproductive sows and the number of live pigs, and the number of live pigs is an important determinant of pork and pig prices, we use the following formula to calculate the number of lag periods. Then the shifted sow stock series is included as a covariate.

$$TLCC_k(x,y) = \frac{\sum_{i=0}^{N-1-k}(x_i - \bar{x})(y_{i+k} - \bar{y})}{\sqrt{(x_i - \bar{x})^2}\sqrt{(y_{i+k} - \bar{y})^2}} \tag{6}$$

On the other hand, the price series within the lookback window and its statistical characteristics (variation and variance) are used as the input of the time series model to reflect the short-term dependence of the price series.

2.4.2. LightGBM Model

Gradient Boosting Decision Tree (GBDT) is a member of the Boosting family in the field of ensemble learning. Once it was proposed, it attracted widespread attention for its outstanding effects. Although the application of GBDT to time series prediction is rarely found, it does not mean that the algorithm is uncompetitive in this problem. Shereen2021's research shows that a properly configured Recent research shows that a properly configured GBDT model can outperform the current SOTA DNN model in the time series prediction field[5]. LightGBM is a variant of GBDT which proposes a gradient-based one-side sampling algorithm and an exclusive feature bundling algorithm, which solves the poor computational efficiency problem of the XGBoost model.

2.5. Prediction of Cyclical Component Based on Encoder-Decoder Model

2.5.1. Encoder-decoder Model with Attention Mechanism

The encoder-decoder model is widely used in the field of NLP, and attention mechanisms are frequently applied to improve it[6]. The general structure of the encoder-decoder model with attention is shown in Figure 3. In this model, the hidden state of the final RNN layer at all time steps is the key and value of the attention layer. At each time step of the decoding process, the final RNN layer hidden state of the decoder at the previous time step is used as a query. With the attention score we can amplify the attention-worthy parts of the sequence and reduce the influence of irrelevant parts. In this model, we choose 1D-CNN as the encoder and LSTM as the decoder.

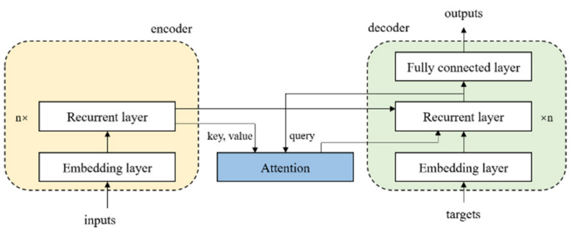

Figure 3 Structure of encoder-decoder model with attention mechanism

2.5.2. Segment-wise Attention Model

The similar sub-series search algorithm has a better effect on the time series prediction problem, which shows that the correlation between time segments can reflect the correlation of time series better than point-level sequence correlation. Therefore, we design a segment-wise attention model. We use the sliding window to construct the segment list in an overlapping way as the input. Considering that the similarity of the clips is too small if the time interval is too small, we take the step size of the window movement as a hyperparameter. The input sequence is constructed as follows.:

$$y_{1:T} \rightarrow \{y_{1:1+s} y_{1+step:1+step+s}, \cdots\} \quad (7)$$

where $y_{1:T}$ represents $\{y_1, y_2, \ldots, y_T\}$, s represents the segment length, and step represents the window moving step. Figure 4 shows the difference between our attention model and the traditional attention mechanism. Our motivation is to distinguish the difference in reasoning logic between machine translation and time series prediction. In the machine translation scenario, attention is used to align the vocabulary of the source and target languages, but in the time series prediction scenario, the historical segment similar to the current segment is not our target, but the subsequent segments. During the training phase, the input to the decoder is the ground-truth value at the previous time step, and during the prediction phase, it is the predicted value. The prediction ends when the total segment length reaches the prediction step size.

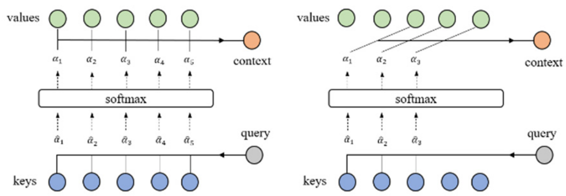

Figure 4 Comparison between traditional and segment wise attention predictive paradigm

3. Materials and Methods

3.1. Experimental Settings

The disparity between the supply and demand for pig products is largely due to information lag. We chose the prediction step as 25 weeks, which is the breeding cycle from piglets to killed pigs, in order to make the prediction results support farmers' predictions of future income before investing. To eliminate the possibility of experimental findings, we ran 10 experiments on each dataset. The test set and training set were often dissected simultaneously in earlier studies, although this practice can result in data leaks and unnaturally high accuracy[7]. We carry out experiments using rolling decomposition to get around this issue. Only the sequences that fall inside the training set window are broken down for training in each trial. Figure 5 shows the schematic diagram.

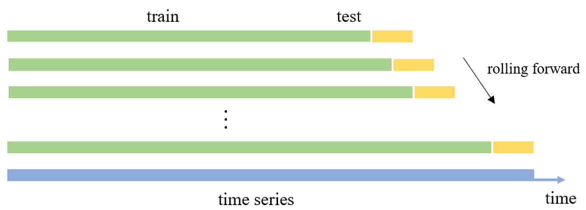

Figure 5 Rolling decomposition experiment

All experiments are implemented in Python 3.6 and 3.8. We use a few popular Python packages. The Statsmodels package provides support for statistical correlation methods, including HP filters, STL decomposition, ARMA, and SARIMA models. SVR modelling is carried out using the LinearSVR class that sklearn package provides. We use darts to implement LightGBM because the library naturally supports historically known covariates and future known covariates. The PyTorch package is used to implement neural network models like the 1D-CNN and LSTM models.

3.2. Performance Criteria

Three statistical indicators, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), are used to evaluate the performance of the model, which are the most commonly used evaluation metrics for regression models. The formulas are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widetilde{y_i}| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widetilde{y_i})^2} \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widetilde{y_i}}{y_i} \times 100\%\right| \tag{10}$$

where y and $\tilde{y}$ represent the ground truth and the predicted sequence respectively, $\bar{y}$ represent the average of y, and N represents the length of the predicted sequence, which is 25 in this experiment.

3.3. Result and Discussion

To verify the effect of our model on the predictive ability and generalization ability, we conducted two sets of experiments. The time range of the data used in the first set of experiments is 2006.1-2018.4, which is the same as the data time range used in researches of Zhu et al. and Liu et al.[2,8]. The pig cycle in this time range is more significant and relatively stable. The time range of the data used in the second set of experiments is extended to 2022.5, including situations of extreme price volatility brought on by market imbalances, in order to test the generalization ability and stability of the model. Our method is compared with other methods for pork price series forecast, including single model (ARIMA, Prophet[9] and LSTM[10]) and combined model[2,8]. Out of the 20 trials conducted in the two groups, we only chose the results of 4 for presentation due to space restrictions (Figure 6 and 7, Table 2 and 3).

Table 2 and Figures 6 show the comparison of forecast results in a relatively stable price period. It can be seen from the table that the prediction error for hog prices is typically higher than the forecast error for pork prices. Compared with the single model, the combined model under the decomposition ensemble framework generally performs better. On both datasets, our approach produces results that

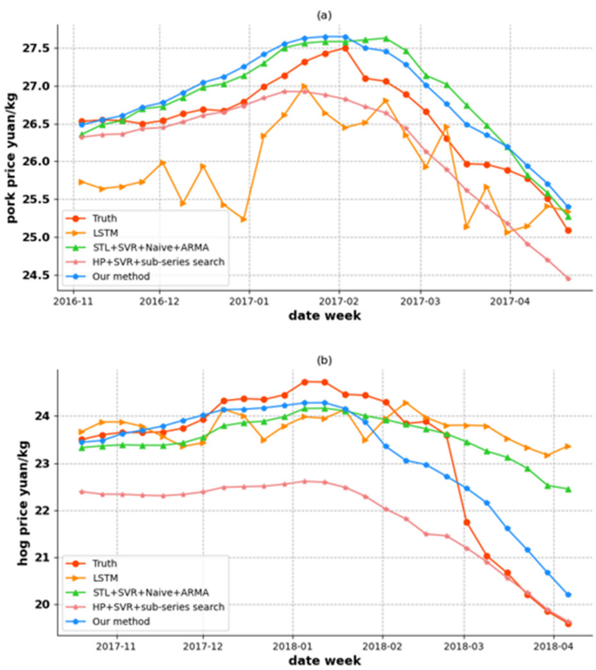

Figure 6 Result of pork and hog prices prediction during 2016 to 2018

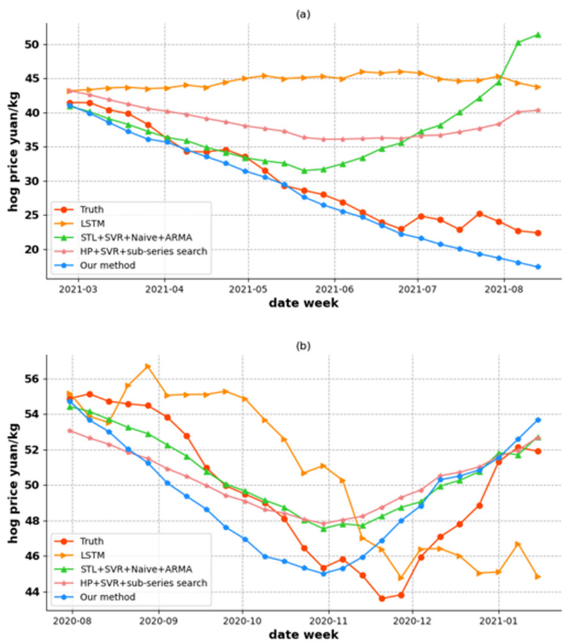

Figure 7 Result of pork and hog prices prediction during 2020 to 2022

Table 2. Statistical results of the predictions for pork price and hog price during 2016 to 2018

Type	Methods	MAE	RMSE	MAPE
Pork price	ARIMA	1.4699	2.0174	6.4212
	prophet	1.4429	1.9819	5.9394
	LSTM	1.2361	1.6401	5.1259
	STL+SVR+naïve+ARMA	1.0018	1.3662	3.9916
	HP+SVR+ similar sub-series search	0.9183	1.1517	3.9266
	Our Method	0.8992	0.9989	3.6816
Hog price	ARIMA	1.5657	2.1764	6.8211
	prophet	1.5396	2.0191	6.7319
	LSTM	1.4081	1.9410	6.1569
	STL+SVR+naïve+ARMA	1.1325	1.4616	4.9772
	HP+SVR+ similar sub-series search	1.0689	1.1979	4.6276
	Our Method	1.0717	1.2011	4.6827

Table 3. Statistical results of the predictions for pork price and hog price during 2020 to 2022

		MAE	RMSE	MAPE
Pork price	ARIMA	2.9299	4. 0685	11.5935
	prophet	4.1081	5.7524	16.0042
	LSTM	3.4793	4.6971	12.6139
	STL+SVR+naïve+ARMA	2.6201	3.6686	8.9772
	HP+SVR+ similar sub-series search	2.3917	3.1781	8.8917
	Our Method	1.6862	2.1828	6.0251
Hog price	ARIMA	3.4292	4.5205	12.5935
	prophet	5.1747	7.0191	19.0042
	LSTM	3.9793	5.2810	14.6139
	STL+SVR+naïve+ARMA	2.9890	4.2456	10.9772
	HP+SVR+ similar sub-series search	2.6128	3.3865	9.6276
	Our Method	1.7557	2.2524	6.4480

are competitive. The results of the predictions made during times of high price volatility are displayed in Table 3 and Figures 7. In the case of a market imbalance, all models' accuracy falls, but our model performs the best. The mean values of MAE, RMSE, and MAPE of our model are lowered in the prediction of pork prices when compared to the suboptimal model by 29.49%, 29.16%, and 32.23%, respectively, while the mean errors of pig price predictions are reduced by 32.80%, 33.48%, and 33.02%. Through the above analysis, we can draw the following conclusions: firstly, the price prediction model proposed in this paper is effective in predicting the pork price and hog price, and secondly, the prediction model considering the causes and historical price information of live pig and pork prices has greater generalization ability.

4. Conclusion

This paper proposes a hybrid model based on a decomposition ensemble framework for predicting hog prices and pork prices in the next hog breeding cycle. Considering the historical dependence information of the time series and the impact of covariates that affect the price of live pigs, the dynamic regression model based on LightGBM is chosen to forecast the trend component of the price. This is done in order to ensure the stability and generalization ability of the model. In order to deal with pseudo-periodic components, we extend the recognition ability of similar subsequence search algorithm with the potent embedding capability of neural networks and propose an encoder-decoder model based on segmental attention for periodic components. Finally, we combine the trend and cyclical components to provide the forecast outcomes. The advantages of the proposed method are reinforced by experiments on real datasets.

Considering the distinct characteristics of Chinese provinces in the pork industry and the transmission of pork price fluctuations between provinces, we will consider introducing geospatial information into the pig price forecast model to achieve more fine-grained price forecasting in the future.

5. Reference

[1] HYNDMAN R J, ATHANASOPOULOS G. Forecasting: principles and practice[M]. 2nd Edition. Melbourne, Australia, 2018.

[2] ZHU H, XU R, DENG H. A novel STL-based hybrid model for forecasting hog price in China[J/OL]. Computers and Electronics in Agriculture, 2022, 198.

[3] LUO Q, ZHANG L. Fluctuation Characteristics of Hog Price in China: Evidence Based on the Method of B-N Decomposition[J/OL]. Journal of Agrotechnical Economics, 2018, 7: 93-106.

[4] DU YANG. Analysis, prospect and countermeasures of periodic fluctuation of pig price in China[J/OL]. Prices Monthly, 2020, 515: 1-5.

[5] ELSAYED S, THYSSENS D, RASHED A, et al. Do We Really Need Deep Learning Models for Time Series Forecasting?[J/OL]. arXiv preprint arXiv, 2101.02118, 2021.

[6] KONDO K, ISHIKAWA A, KIMURA M. Sequence to sequence with attention for influenza prevalence prediction using Google Trends[J/OL]. ACM International Conference Proceeding Series, 2019: 1-7

[7] YU L, MA Y, MA M. An effective rolling decomposition-ensemble model for gasoline consumption forecasting[J/OL]. Energy, 2021, 222: 119869.

[8] LIU Y, DUAN Q, WANG D, et al. Prediction for hog prices based on similar sub-series search and support vector regression[J/OL]. Computers and Electronics in Agriculture, 2019, 157: 581-588.

[9] TAYLOR S J, LETHAM B. Forecasting at Scale[J/OL].

[10] SABU K M, KUMAR T K M. Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala[J/OL]. Procedia Computer Science, 2020, 171: 699-708.