



## Article

# Monthly Pork Price Prediction Applying Projection Pursuit Regression: Modeling, Empirical Research, Comparison, and Sustainability Implications

Xiaohong Yu <sup>1</sup>, Bin Liu <sup>2,\*</sup> and Yongzeng Lai <sup>3</sup><sup>1</sup> College of Humanities and Law, Shanghai Business School, Shanghai 200235, China; 21090015@sbs.edu.cn<sup>2</sup> Shanghai Institute of Commercial Development, Shanghai Business School, Shanghai 200235, China<sup>3</sup> Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada; ylai@wlu.ca

\* Correspondence: liubin@sbs.edu.cn

**Abstract:** The drastic fluctuations in pork prices directly affect the sustainable development of pig farming, agriculture, and feed processing industries, reducing people's happiness and sense of gain. Although there have been extensive studies on pork price prediction and early warning in the literature, some problems still need further study. Based on the monthly time series data of pork prices and other 11 influencing prices (variables) such as beef, hog, piglet, etc., in China from January 2000 to November 2023, we have established a projection pursuit auto-regression (PPAR) and a hybrid PPAR (H-PPAR) model. The results of the PPAR model study show that the monthly pork prices in the lagged periods one to three have an important impact on the current monthly pork price. The first lagged period has the largest and most positive impact. The second lagged period has the second and a negative impact. We built the H-PPAR model using the 11 independent variables (prices), including the prices of corn, hog, mutton, hen's egg, and beef in lagged period one, the piglet's price in lagged period six, and by deleting non-important variables. The results of the H-PPAR model show that the hog price in lagged period one is the most critical factor, and beef price and the other six influencing variables are essential factors. The model's performance metrics show that the PPAR and H-PPAR models outperform approaches such as support vector regression, error backpropagation neural network, dynamic model average, etc., and possess better suitability, applicability, and reliability. Our results forecast the changing trend of the monthly pork price and provide policy insights for administrators and pig farmers to control and adjust the monthly pork price and further enhance the health and sustainable development of the hog farming industry.

**Keywords:** monthly pork price prediction; projection pursuit regressive model; parasitism–predation algorithm; sustainability implications; empirical research



**Citation:** Yu, X.; Liu, B.; Lai, Y. Monthly Pork Price Prediction Applying Projection Pursuit Regression: Modeling, Empirical Research, Comparison, and Sustainability Implications. *Sustainability* **2024**, *16*, 1466. <https://doi.org/10.3390/su16041466>

Academic Editor: Hossein Azadi

Received: 16 October 2023

Revised: 21 January 2024

Accepted: 22 January 2024

Published: 9 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pork is the primary source of animal protein for residents in China. Pork production has consistently topped the list of domestic meat production in China. According to the National Bureau of Statistics (<http://data.stats.gov.cn/easyquery.htm?cn=C01>, accessed on 15 October 2023), hog yields in China reached 55.41 million tons in the year 2022, accounting for 59.4% of the domestic livestock yields, which accounted for about 50% of the world's total amount. The pork supply chain is composed of a wide range of links, including its upstream industry, such as feed processing and transportation, the farming of soybean and corn, slaughter, and the downstream sector, such as package, storage, transportation, and sales while satisfying consumers' needs, etc. It can be seen that the pig farming industry has a core impact on the national economy and people's livelihood and also affects the changes in international and domestic pork futures index to a certain extent. Keeping pork prices stable, avoiding big ups and downs, and accurately and reliably predicting the

law of pork price change are of great value to ensure the safety, stability, and sustainable development of the pork meat supply chain, as well as the pig farming industry.

The pork production sector makes a vital contribution to the agricultural industry. However, due to the rapid development of pork production and poor management, incomplete regulation, and the decoupling of crop and pork production systems, pork production and its related feed production have significantly increased environmental pollution, especially through the improper disposal of manures and slurries and waste of feed resources, as well as the associated greenhouse gas emissions and non-renewable energy and resource use. Annual pig manure production exceeded 60 Mt in 2017, accounting for about 30% of total pollutants sourced from the animal husbandry industry. Therefore, strengthening the prediction and monitoring of pork prices is the foundation for achieving stable prices and production in the pig industry. It plays a vital role in promoting the sustainable development of the pig industry, affecting the sustainable development of upstream sectors such as pig feed processing and slaughter and ultimately affecting the sustainable development of the breeding industry and agriculture industry [1–3].

Moreover, the pork meat supply chain significantly impacts the farming of the soybean and corn used as feed for pigs, increasing farmers' income, sustainable agricultural development, and rural revitalization. In other words, the fluctuation, soaring, or continuous decline of pork prices not only affects the nerves of the general public but also involves the interests of the pig industry practitioners and even affects the stability and harmony of the economy and society. The increase in pork prices will increase the number of pigs raised, thereby driving the sustainable development of the planting industry of raw materials for pig feed and agricultural products and the sustainable expansion of the pig feed-processing industry. Conversely, the development of related sectors cannot be sustained and will lead to shrinkage. In other words, studying the changes in pork prices, predicting price changes, and warning against fluctuation risks are the foundation for the sustainable development of the pig industry and its upstream and downstream related industries. They are also fundamental requirements for achieving sustainable agricultural development. Therefore, the reliable, reasonable, and accurate prediction of the pork price for the pig industry chain, pork supply chain management, and practical arrangement of production, sales front production activities, commodity price departments, and control pork prices, as well as the consumer, has important theoretical significance and impacts decision making, has long been critical in government supervision and academic circles that carry out the hot and challenging issues of price forecasting and early warning research [1–16].

Sarle [17] and Ezekiel [18] established multiple linear regression (MLR) models to predict the price of pork (or hogs) in the 1920s. Since then, a large number of studies on the prediction and early warning of pork prices have been published. In summary, there are three main categories of prediction models. The first type is based on the price fluctuation mechanism and influencing factors (or independent variables, hereafter referred to as based on the price fluctuation mechanism). In building multivariable models with different lagged periods, the number of influencing factors can be as small as four to five and as large as twenty or more. It mainly involves the cost of piglets, pig feeding, the price of pork substitutes, consumer demand, the feeding environment related to African swine fever, the catering industry, logistics, the international environment, the money supply, pork imports, futures index, and so on [1–18]. The second model type starts from the result of price fluctuation results (hereafter referred to as the result of price fluctuation results). The time series univariate auto-regression (i.e., multiple periods) model is established based on the time series data of pork price (daily, weekly, half-monthly, monthly, quarterly, semi-yearly, and yearly). The third category is a hybrid multivariate model, which includes both the autoregressive time series data on pork price and the data on influencing factors (the number of lagged periods can be different). Theoretically, the first model type is conducive to studying the pork price transmission mechanism, supply chain management, pig industry chain, upstream and downstream enterprise management, etc. The reliability and effectiveness of the model prediction results directly depend on the comprehensive,

scientific, and systematic nature of the influencing factors determined. According to the existing literature, different authors often use different independent variables (influencing factors), which shows that the factors affecting pork prices are difficult to reliably and reasonably determine, which reduces the reliability and rationality of the model and the accuracy of the prediction results. Moreover, according to such models, the price of pork can generally only be predicted in phase 1. All the influencing factors of pork prices, and even the result of the interaction of multiple influencing factors, are finally reflected in the time series data of pork prices. The time series data of pork prices contain the combined influence results of all the influencing factors. The model established based on time series data of pork prices also has good rationality and reliability. According to this model, it can be very convenient to realize the multi-step prediction of the changing trend of pork price and to predict when or whether the pork price will appear at an inflection point in a certain period to make up for the deficiency of the first type of model. Therefore, the establishment of the two different models above is to study the fluctuation of pork prices from two perspectives, both of which have important theoretical and practical significance. Each has its characteristics, advantages, and disadvantages, and the two models should not be ignored. In principle, both models are valid and reasonable. The third model type is a synthesis of the first two types of models, which integrates the advantages of the above two models. Of course, there are also some deficiencies, such as only predicting the price of the one ahead. As for the established pork price prediction model, almost all the modeling methods for time series data have been applied to pork price prediction, mainly including traditional prediction models (TPMs, including time series in single independent variable or multi-independent variables) and modern data-mining technologies (MDTs). The TPMs include MLR [17–19], grey prediction model [5], vector auto-regression (VAR) [14], auto-regressive integrated moving model (ARIMA) [8,18,19], etc. The MDTs (or machine learning methods) have artificial neural networks (mainly based on error backpropagation neural network (BPNN)), radial basis function neural networks, generalized regression neural networks, extreme learning machine neural networks [5,9,11,15,19], support vector machine/regression (SVM/SVR) [6], multi-regime smooth transition autoregressive model [16], and dynamic model average (DMA) [4]. Meanwhile, the data decomposition methods or independent variable compression methods, such as empirical mode decomposition (EMD, including EEMD, CEEMD, and CEEMDAN, etc.) [6], filter algorithm [10], principal component analysis [5], are applied to decompose the original data into intrinsic mode functions (IMFs). The IMFs are modeled by the TPMs or MDTs abovementioned. Furthermore, the combination models are established using two or three models above [5,13]. The existing studies and the above literature have achieved certain results in predicting pork prices. Most of the literature has shown good performance for training and validation datasets. At the same time, there are also some problems in the modeling process, such as generality, applicability, and reliability.

On the other hand, the research shows that the projection pursuit regression models have a good generalization ability and outperform SVR/SVM, BPNN, RF, etc., in suitability, applicability, and reliability [20–22] for small samples as well as large samples. The primary purpose of this article is split into four aspects. Firstly, we apply the PPR model, which is particularly suitable for modeling high-dimensional, nonlinear, and non-normal distribution data, to the study of monthly pork price prediction for the first time. Secondly, we introduce the principles and precautions of PPR modeling. We compare the modeling performance metrics between the PPR models and other models in terms of reliability, prediction accuracy, generalization ability, etc., and analyze the main problems in the literature. Thirdly, based on the established PPR models, we put forward some measures and suggestions for regulating pork prices to avoid sudden increases and decreases, better promote the stable and efficient development of pig farming and upstream and downstream industries, promote sustainable agricultural development, and enhance the people's happiness and sense of gain. Finally, we apply the PPR models to actually predict the monthly pork price more accurately and reliably using the latest available data.

The structure of this paper is as follows: Section 1 is the introduction; Section 2 is a review of the literature on price prediction of livestock, pork, and crop and PPR models; Section 3 describes the data resources of the time series of pork price and its influencing factors; Section 4 discusses the principles and precautions of PPR modeling based on univariate time series data and multivariate time series data such as hog–corn ratio, piglet’s price, etc.; Section 5 is positive research and the results of establishing the PPR model; Section 6 analyses the particular procedure and results of the H-PPR model; Section 7 is the results and the discussion; Section 8 includes the main conclusions, policy recommendations, limitations, and future research.

## 2. Literature Review

In order to conduct better research, it is necessary to review the existing literature on price forecasting comprehensively, absorb valuable achievements, identify problems, and make improvements. The literature provides multiple techniques to forecast livestock, pork, and crop products. The proposed solutions included mathematical and statistical models (MLR, GM, ARIMA, etc.) and machine learning approaches combining statistical and artificial intelligence models to provide better predictions. Among these models, ARIMA, error backpropagation neural networks (BPNN), support vector machines for regression (SVR), random forests, and LSTM are the most popular, but other models have also been used. Refs. [23,24] comprehensively reviewed the literature on the price prediction of pork, livestock, and agriculture products. Since the 1920s, scholars have widely researched pork price prediction. These previous studies added meaningful value to this article, and we only provide a brief review and summary in Table 1.

**Table 1.** The detailed information of the articles for predicting prices of pork, hog, etc.

Refs.	Model Used	Samples	Univariate/Multivariate Time Series	Simple/Challenging to Implement	Price
[2]	ARIMA, SARIMA	2010–2018, 100 obs.	Univariate	Simple &	Monthly, Sugar, Russian
[4]	DMA, TVP, AR	January 2000–March 2019, 321 obs.	Multivariate, 11 factors	Challenging	Monthly, Pork, China
[5]	PCA-GM-BPNN	January 2010–December 2018, 108(96/12) *	Multivariate, 12 factors	Challenging	Monthly, Pork, China
[6]	RBFNN, GA-SVR, EMD-GA-SVR, EMD-GA-SVR, EEMD-GA-SVR, CEEMD-GA-SVR, CEEMD-PEFFT-GCD-SVR	January 2006–June 2018, 150(120/30)	Univariate	Challenging	Monthly, Pork, China
[11]	GRNN, BPNN	1 March 2011–25 March 2014, 732(502/110/110)	Univariate	Challenging	Daily, pork, China
[13]	GM, TS, BPNN, TS-GM, TS-BPNN, BPNN-BPNN, GM-GM, BPNN-GM, GM-BPNN	January 2000–June 2008, 102(90/12)	Multivariate, 4 factors	Challenging	Monthly, hog, China

Table 1. Cont.

Refs.	Model Used	Samples	Univariate/Multivariate Time Series	Simple/Challenging to Implement	Price
[17]	MLR	January 1897–December 1916	Multivariate, 4 factors	Simple	Monthly, hog, USA
[18]	Empirical formula, Demand-curve method	January 1903–December 1914	Multivariate	Simple	Monthly, hog, USA
[19]	ARIMA, BPNN, Econometric model	January 1974–December 1996	Multivariate, Univariate	Challenging	Quarterly, monthly, hog, USA
[23]	ARIMA, SARIMAX, RF, SVR, Ridge, LGBM, XGBoost, RNN, LSTM, CatBOOST	January 2016–February 2022, 322(80:20) **	Univariate	Challenging	Weekly, pork, Spain
[25]	LDA, Deep learning (LSTM), RF, BPNN, CNN, Gradient, Boosting, Ridge	January 2010–December 2019, 1175(987/188)	Univariate	Challenging	Daily, pork, South Korea
[26]	CAR, CART, SA-CART, SSA-CART, WSO-CART	January 2011–December 2015, 257(80:10:10)	Multivariate	Challenging	Weekly, pork, China
[27]	LSTM, STL-ATTLLSTM, BERTLSTM, GCNLSTM, HGLSTM	January 2013–December 2020	Multivariate	Challenging	Weekly, hog, China
[28]	ARIMA, EMD-ARIMA, VMD-ARIMA, SVR, EMD-SVR, VMD-SVR, RNN, EMD-RNN, VMD-RNN, LSTM, EMD-LSTM, VMD-LSTM	January 1974–December 2017, 11,085(80:10:10)	Univariate	Challenging	Daily, corn and soybean, USA
[29]	ARIMA, ETS, SVM, LSTM, BPNN, Other 12 combined models	January 2013–December 2018	Univariate	Challenging	Monthly, onion and potato, India
[30]	SARIMA, TDNN, ELM, STL-ELM	January 2010–December 2020, January 2005–December 2020	Univariate	Challenging	Monthly, potato, India
[31]	ARMA, GM, BPNN, GRNN, RBFNN, FOA-GRNN, FOA-RBFNN	January 2010–April 2020, 112(80/30)	Univariate	Challenging	Monthly, vegetables, China

Table 1. Cont.

Refs.	Model Used	Samples	Univariate/Multivariate Time Series	Simple/Challenging to Implement	Price
[32]	Lasso, SARIMA, STL-SVR-ARMA, SVR, RF, LSTM, VMD-LR-ARMA, STL-SVR-SNN-ARMA	January 2006–December 2018, 678(658/20)	Univariate	Challenging	Weekly, pork and hog, China
[33]	RF, XGB, LGBM, BPNN, RNN, LSTM, EEMD-BPNN, EEMD-RNN, EMD-MiLSTM, EEMD-MiLSTM	240(90:10)	Univariate	Challenging	Monthly, hog, China
[34]	SVR-cyclical component, BPNN, Wavelet-SVR	January 2011–March 2017	Univariate	Challenging	Monthly, pork and piglet, China
[35]	SVR, SVR-CIs, SVR-CIs-CIp, SVR-CIs+p, SVR-WD, SVR-EMD, SVR-EEMD, SVR-SSA	January 2011–December 2017, 84(80:20)	Univariate	Challenging	Monthly, pork, China

Notes: \* 180(96/12) means that the total number of samples is 108, and the number of training and testing samples are 96 and 12, respectively. 732(502/110/110) means that the total number of samples is 732, and the split for training, validation, and testing is 502, 110, and 100. \*\* 322(80:20) means that the total number of samples is 322, which have been divided into training and testing in an 80:20 ratio. 257(80:10:10) means that the total samples have been split into training, validation, and testing in an 80:10:10 ratio. & Simple means that it is simple or easy to implement the model. Challenging implies that it is challenging or complex to implement the model and train it with over-training or overfitting easily, which possesses good flexibility and nonlinear approximation ability.

It can be seen from Table 1 that there are various models, including statistical methods (e.g., ARIMA, SARIMA, GM, etc.) and machine learning models such as BPNN, SVR, RF, and LSTM. Furthermore, swarm intelligent optimizations such as FOA, WSO, and SSA are used to optimize the parameters of the models. Decomposition techniques such as EMD, EEMD, CEEMD, STL, and VMD are applied to decompose the time series data into several independent components for each component to establish a model and finally to combine them to build the model for prediction prices. We can conclude that more and more machine learnings and their combined models, as well as more and more complicated models, are used to predict prices. In fact, the more complicated the models are, the higher the fitting accuracy for the training samples the models, the poorer the generalization for the models' validation, and the greater the challenge to establish the models. Meanwhile, the conclusions of these articles are usually very vague or ambiguous. According to their study results, some scholars thought the SVR, as well as its combined models, outperformed other models [34,35], some scholars thought the LSTM and its combined model outperformed other models [25,27,34], some scholars thought BPNN had better generalization ability than other models [5,11,13], Ref. [4] thought the DMA had better performance than other models, Ref. [18] found that the BPNN performed considerably worse than the econometric model, etc. Theoretically, the traditional statistical model is a "white box" model with a clear working mechanism, and its flexibility is relatively insufficient. It can achieve better results only when the pork price changes conform to the model's function. Usually, the fluctuation of the monthly/weekly/daily prices of pork, hog, vegetables, and other agricultural products and the future is far more complex than the function of the traditional



model. Although modern data mining technologies or machine learning models such as BPNN, SVM/SVR, RF, and LSTM have good nonlinear approximation ability, they are not the panacea for price prediction [36]. The machine learning models are well-designed, trained under the monitoring of validation samples in the training or optimization process, meaning they avoid over-training or over-fitting, and have good generalization ability, reliability, and applicability. Otherwise, “overtraining” and “overfitting” can easily occur when modeling. To avoid “overtraining” and “overfitting,” certain modeling principles must be followed. For example, the BPNN modeling process must follow basic principles and steps [37–40]: (1) The sample data must be divided into training and verification subsets with similar properties. Monitoring the root-mean-square error (RMSE) for verification cases in the training process is necessary. If the RMSE on the verification cases does not improve and begins to rise, the training process will cease (called the early stop training method). Characterize the model performance metrics with the error of the test samples; (2) Meet the accuracy requirements. We take the neural network structure topology as compact as possible (with hidden layer nodes as few as possible). The number of training samples must be at least 3–5 times greater than the number of network connection weights, and it is better to reach 5–10 times and above; (3) Use the regularization method to determine the reasonable number of hidden layer nodes. Unfortunately, much of the existing BPNN modeling literature does not follow the above principles. Although the SVR model can be applied to moderate samples, it is not easy to choose reasonable model parameters. In addition, BPNN, SVR, and others belong to the data-driven “black box” and “recessive” models [40]. It is not convenient to analyze the working mechanism and study the transmission mechanism of pork prices. The follow-up application is not convenient, which is not conducive to formulating measures to control the pork price and strengthening the macro management of the pork supply chain and upstream and downstream enterprises. However, the DMA model involves more than 2000 prediction models, each with 4–5 independent variables. The model is very complex and significant in theory; its practicability is insufficient, and the prediction accuracy is not very high (referring to Section 7.3). Therefore, two problems exist in the existing research literature on pork price prediction. First, the process of establishing the machine learning models (including BPNN, RBFNN, SVR, LSTM, DMA, and various combination models) is too complex to have good applicability. Second, for most of the literature on establishing machine learning models, the basic principles of modeling are not followed, which makes it difficult to ensure the generalization and prediction ability. For SVR models (including various combination models), the results are directly related to the model’s parameters search range, making it difficult to ensure its robustness and stability.

For the pork price prediction problem, under the condition of meeting the prediction accuracy requirements or prediction accuracy, we should choose to use a simple dominant model as far as possible. The model contains independent variables that should be as few as possible to facilitate data collection and reduce costs, making it more convenient, according to the prediction model, to take effective measures to control and adjust pork prices, analyze the pork price transmission mechanism, strengthen the pork supply chain management, and improve the pig industry chain. Therefore, the existing literature cannot meet the above requirements for pork price prediction. On the other hand, projection pursuit regression (PPR) technology is also a nonlinear data mining technology. Research has shown that it has the same nonlinear approximation ability as BPNN. Still, it is especially suitable for small and medium sample data modeling that does not obey the ordinary distribution law [20–22,41–46]. Due to PPR, the model of independent variable weight sum is equal to 1 for multiple independent variables with collaborative constraints. The PPR model has been widely used in agriculture, water conservancy, earthquake, and experimental optimization design with less data for complicated changes and fluctuations, but has not been used in pork price prediction research.

This paper has the following features and contributions compared with the existing literature. First, from the perspective of theoretical model selection, modeling, and prediction

ability, we innovatively established the PPAR model for the time series data of monthly pork prices, using the H-PPR model for the monthly data, including 12 influencing factors (lag period or sliding window data). The predation–parasitic algorithm [20,47] is adopted to obtain the real global optimal solution. Since the constraint of the PPR model is the sum of squares of the best variables' best weights equal to 1, "overtraining" and "overfitting" can be effectively avoided. At the same time, through comparison, non-significant independent variables (influencing factors) are deleted one by one to establish more concise and practical PPAR and H-PPR models and a more straightforward and valuable model. Comparative studies show that the data-fitting ability of several machine learning algorithms (models) is equal. Still, the prediction ability of PPAR and H-PPR models is better than SVR, BPNN, DMA, and other models, and the model's results are more robust, reliable, and reasonable. Moreover, PPAR and H-PPR are dominant models. Because of this, given pork price and multiple influencing factors (independent variables), this paper constructs the PPAR and H-PPR pork price prediction model and applies various performance metrics to evaluate the prediction ability of the model, avoid the subjectivity of the model and its parameter selection, improve the effectiveness, robustness, and effectiveness of the model, expand the new method of pork price prediction research, and provides a guiding research framework for the subsequent pork price prediction modeling.

Second, from the perspective of the practical application of the model, the PPAR model established in this paper only uses the pork price data lagging behind 1–3 periods. The established H-PPR model, which removed non-significant influencing factors and included only six independent variables, greatly simplified the prediction model, making PPAR and H-PPR models more practical and obtain higher prediction accuracy. Third, we formulate the basic principles of the regulation and control of pork price according to the best weight size and ranking of the influencing factors obtained, reveal the main factors affecting the fluctuation of pork price and their transmission mechanism, and put forward the principles of strengthening the management of the pork supply chain. The research methods and conclusions in this paper make up for the deficiency of the existing literature and also provide an essential basis for decision making for the relevant government departments to take appropriate measures to stabilize the pork price.

### 3. Materials: Data Resource

#### 3.1. Collecting the Monthly Pork Price

The China Animal Husbandry Information Network (<http://www.caaa.cn/market/zs/article.php?zsid=3/>, accessed on 15 October 2023) published the complete data on the price of pork (from now on referred to as pork) and the hog–corn ratio from January 2000 to September 2020 (detailed data omitted). The data are similar to Xiong et al. [4]. Therefore, we establish the PPAR model by applying the above sample data in this paper.

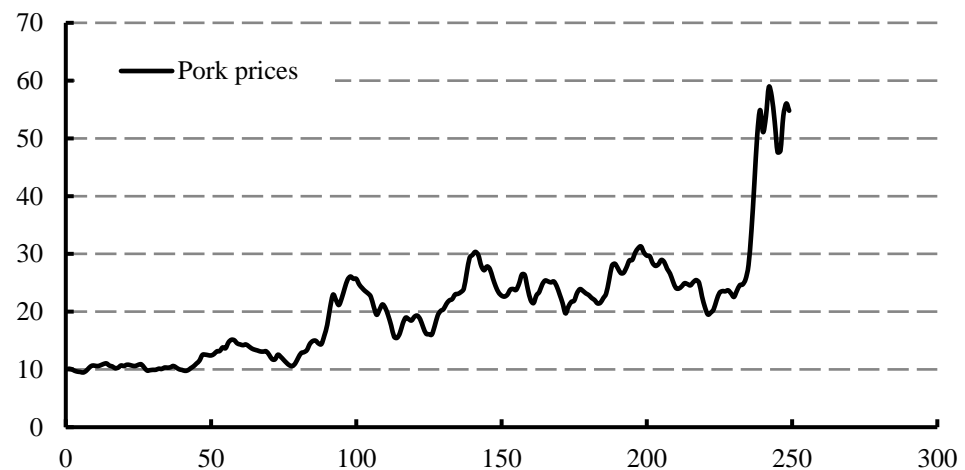
#### 3.2. Preliminary Determination of the Main Variables Affecting the Fluctuation of Pork Prices and Collecting the Data

According to the literature review and theoretical analysis, many factors affect the monthly price of pork, such as piglet cost, corn, beef, and mutton substitutes, consumer demand side and African swine fever, restaurants, logistics, the international environment, and M2 and futures prices related to pork supply chain and pig industry chain factors. However, African swine fever has dramatically reduced the swine population in China. In some other Asian and European countries, this has directly lead to an increased retail price for pork and is a main factor impacting sustainable pork production as well as its related industry, but the detailed data are difficult to obtain accurately. Considering that the price prediction model should not be too complex and that the data on environmental factors are more challenging to obtain, the first three factors are mainly considered. Due to the lack of soybean meal and wheat bran data, this paper collected the hog–corn ratio, piglets, slaughtered hog, boneless beef, with-bone mutton, eggs, chicken, commercial eggs



with chickens, corn, pigs, and chicken feed data, for a total of 12 meal-to-belt bone pork (hereafter referred to as pork) monthly price factors.

Figure 1 shows a schematic diagram of the monthly pork price changing over time (starting from January 2000). It can be seen that the monthly cost of pork shows an upward trend, with the typical characteristics of different sizes of cycles, but it is not easy to judge directly. Moreover, from August to October 2019, the price of pork soared significantly and then stabilized at about CNY 47~60. Still, the number of samples in these parts is relatively small, making prediction and modeling difficult.



**Figure 1.** Schematic diagram of pork price change over time. (The ordinate is pork price, unit CNY; the horizontal coordinate is time, starting from January 2000, the same below).

#### 4. Principles of PPR Modeling

This paper mainly establishes the PPAR and the hybrid multivariate prediction pursuit regression (H-PPR) models based on the time series data of pork prices and the other independent influencing factors.

##### 4.1. Principle of Establishing the PPAR Model

Two basic assumptions exist for establishing a PPAR prediction model for monthly pork prices based on time series historical data. Firstly, multiple factors affect the monthly pork prices, and the relationship between these factors is very complex, making it difficult to have a mathematical model to represent them. However, the results of these factors are reflected in the changes in monthly pork prices. Secondly, the changes in monthly pork prices have a certain regularity, which autoregressive time series data can represent.

According to the research results of the existing literature, there are several short and large cycles in the monthly price of pork. Some of the literature asserts that the extensive process (low-frequency fluctuation) should be around 36~48 months (3~5 years), which is too long for the monthly price forecast modeling. Therefore, establishing PPAR modeling is generally dominated by small and medium cycles. To this end, this paper analyzes whether the 12-month autoregression delay  $x(i-1), x(i-2), \dots, x(i-12)$  is significantly associated with monthly pork prices  $x(i)$ . The modeling principle is as follows [20]:

Step 1: The autocorrelation coefficient  $R(k)$  of the delay  $k$  step of the time series data  $\{x(i)\}$  is

$$R(k) = \frac{\sum_{i=k+1}^n [x(i) - Ex][x(i-k) - Ex]}{\sum_{i=1}^n [x(i) - Ex]^2} \quad (1)$$

where  $Ex = \frac{1}{n} \sum_{i=1}^n x(i)$ ,  $k = 1, 2, 3, \dots, m$ , in general,  $m < \frac{n}{4}$ ,  $n$  is the number of time series data. With the  $k$  increasing, the variance of  $R(k)$  increases, and the estimation accuracy decreases. Therefore, it is usually required to take a smaller value for  $m$ . According to

the sampling distribution theory, the confidence level is  $(1 - \alpha)$  (generally being 70~80%). When the autoregressive correlation coefficient value meets

$$R(k) \notin [R_L(k), R_U(k)] = \left[ \frac{-1 - \mu_{\frac{\alpha}{2}} \cdot \sqrt{(n-k-1)}}{(n-k)}, \frac{-1 + \mu_{\frac{\alpha}{2}} \cdot \sqrt{(n-k-1)}}{(n-k)} \right] \quad (2)$$

it can be inferred that delay steps  $x(i-k)$  are significantly correlated with  $\{x(i)\}$ , and  $x(i-k)$  are used as predictors. The quantile values  $\mu_{\frac{\alpha}{2}}$  can be found in the standard normal distribution table.

Step 2: According to the delay step  $k$ , we obtain the predictors  $x(i-k)$  ( $k = 1, 2, 3, \dots, p$ ;  $i = k+1, k+2, \dots, n$ ), and  $p$  is the number of autoregressive predictors. Because it is difficult to judge the maximum pork price (because it may continue to rise) and the minimum, standardization (normalization) preprocessing is generally adopted, and the prediction model of  $x(i)$  with  $x(i-k)$  is established. According to the principle of PPAR modeling, the normalized data of the  $p$  dimensional predictors  $x(i-k)$  are projected to obtain one-dimensional projection values

$$z(i) = \sum_{j=1}^p a(j)x(i-p-1+j), \quad (i = p+1, p+2, \dots, n) \quad (3)$$

where  $a(j)$  is the best projection vector coefficient or weight of the  $p$ -dimensional autoregressive predictor.

Step 3: Build the PPAR model between  $x(i)$  and  $x(i-k)$ . To study the fitting effect and predictive ability of the model more intuitively, the monthly data of the dependent variable pork price  $x(i)$  is not normalized. A PPAR model based on the power index polynomial ridge function is established between the one-dimensional projection value  $z(i)$  and the pork price  $x(i)$  (dependent variable). To set the objective function as the minimum sum of error squares (least squares), that is,

$$Q(a, C) = \min \left\{ \sum_{i=1+p}^n [x(i) - f(i)]^2 \right\}, \quad (4)$$

where  $f(i)$  is the predicted value of the PPAR model. The formula based on the cubic polynomial ridge function (PRF) is

$$\begin{aligned} f(i) &= f[z(i)] = c_0 + c_1 z(i) + c_2 [z(i)]^2 + c_3 [z(i)]^3 \\ &= c_0 + c_1 \sum_{j=1}^p a(j)x(i, j) + c_2 \left[ \sum_{j=1}^p a(j)x(i, j) \right]^2 + c_3 \left[ \sum_{j=1}^p a(j)x(i, j) \right]^3 \end{aligned} \quad (5)$$

where  $c_0 \sim c_3$  are the coefficients of the PRF.

In practice, to prevent "overtraining" and "overfitting", we try the linear ridge function first. The quadratic and cubic polynomial ridge functions are established if the accuracy requirements are unmet.

Step 4: Optimize the objective function (4) to obtain the optimal global solution and obtain the fitting error of the PPAR model based on the first ridge function  $e(i) = [x(i) - f(i)]$ . If the appropriate error meets the prediction accuracy requirements, stop building more PRFs and output the model parameters and the performance indicators such as RMSE and MAPE. Otherwise, follow Step 5 to create more dimensional ridge functions.

Step 5: Replace  $e(i)$  with  $y(i)$ , return to Step 2, repeat Steps 3 and 4, and establish a PPAR model based on the second and third ridge functions until the prediction accuracy requirements are satisfied.

Generally, the higher the order of PRFs or the more the number of PRFs, the more likely it is to have "overtraining" and "overfitting." Therefore, the verification (test) sample should be set in modeling. The verification sample error decreases gradually and then increases, which indicates that "overtraining" and "overfitting" have occurred; the number of polynomials and the ridge function before "overtraining" and "overfitting" must be taken.

To verify the predictive and generalization capabilities of the PPAR model, we used the monthly data of pork prices in the last 12 months as a validation sample.

#### 4.2. The Principle of Establishing the H-PPR Model of Monthly Pork Price Prediction Based on Multivariate Time Series

There are two basic assumptions for establishing an H-PPR prediction model for monthly pork prices based on multivariate time series historical data. Firstly, the prices of live pigs, beef, piglets, etc., are the main factors affecting the monthly pork prices, and the effects of other factors can be ignored. Secondly, there is a specific quantitative relationship between the monthly prices of live pigs, piglets, pork, etc., that lags 1–6 periods and the current monthly pork prices.

The PPAR model generally has relatively high fitting accuracy, generalization, and prediction ability. Still, the PPAR model only contains the monthly pork price data so that it can perform multi-period and inflection point price predictions. Still, it cannot forecast the pork prices that soared rapidly according to the PPAR model. Providing strategic decisions for pig industry development is challenging, and we cannot study the influence mechanism of pork price fluctuation, etc. To achieve these goals, it is generally necessary to establish a nonlinear model between the monthly pork price and its influencing factors. The correlation analysis between the 12 collected factors affecting the monthly pork prices (referred to as the independent variables) and the monthly pork prices show that all the independent variables were significantly correlated with the pork price, and the pork prices lagging 1 to 6 periods were also significantly associated with the current pork price.

It is of no practical significance to study the relationship between the prices of the independent variables and the pork prices in the same period because these independent variables also need to be predicted. Therefore, building a prediction model between pork price and the independent variable lagging several periods is standard practice. According to the current research results, the monthly price of piglets generally lags behind by six periods (months), while other independent variables are assumed to lag behind by one period (sometimes there are specific differences between different scholars, see [4,5,14,15,20,48]). The results of the correlation analysis of the monthly price of pork and the data of other independent variables lagging 1 to 6 periods show that (1) the longer the lag period, the lower the correlation; (2) there is a high correlation between the price of piglets and the monthly price of hogs to slaughter, as well as the cost of all feeding (monthly prices); (3) the pig ratio has a certain independence, but is highly correlated with the monthly price of pork. Therefore, considering the model's practicality and meeting the need to study the transmission mechanism of the monthly price of pork, we should first establish a PPR model of the monthly price of pork and all 12 indicators. The modeling principle consists of the following two steps:

Step 1: The monthly piglet price with lagging six periods and the data of the other 11 independent variables lagging one period (from now on referred to as predictors or independent variables  $v(j)$ ), and the monthly price of pork in the current period  $y(j)$  is not normalized.

Step 2: Build the construction data and make a one-dimensional projection of the  $p$ -dimensional independent variable predictor data  $v(j)$  to obtain the one-dimensional projection value of the sample

$$z(i) = \sum_{j=1}^p a(j)v(i, j)$$

Steps three to five are the same as those for establishing the PPAR model.

We established two models to predict monthly pork prices: the first is a PPAR model based on the time series data of pork prices, and the second is an H-PPR model based on time series data of multiple factors with lagged periods. We compare the performance metrics of two models, BPNN, SVR, LSTM, and other models, and study the applicability, advantages, and disadvantages of the models.

## 5. An Empirical Study on Establishing a PPAR Model for Monthly Pork Price Prediction

### 5.1. Determination of the Reasonable Number of Time Series Lagged Periods

We assume the delay period  $k = 1, 2, 3, \dots, 12$ . Then,  $R(k)$  is calculated according to Equation (1). The autoregressive correlation coefficients  $R(k) = (0.9628, 0.9116, 0.8619, 0.8215, 0.7802, 0.7285, 0.6708, 0.6145, 0.5676, 0.5238, 0.4715, 0.4213)$ , and the lower and upper bound  $[R_L(k), R_U(k)] = [-0.07, 0.06]$ . From the values of the correlation coefficient  $R(k) \notin [-0.07, 0.06]$ , we can see that the smaller the delay period  $k$  is, the more significant the correlation is, and the more significant the recent price's influence on the current period's price is, which is entirely in line with the general price fluctuation law.

The stationarity and unit root test of the time series data show that the time series data are unstable, the first order difference is stable and follows the ordinary distribution law, and the original time series data have an increasingly more significant overall trend, indicating that the data are suitable for predictive modeling by applying PPAR. At the same time, according to the original data of the monthly time series of pork prices, they do not obey the normal distribution. It is reasonable to apply regression analysis modeling directly with the traditional statistical model, which is normally inappropriate in principle.

The above monthly pork price normalization data and the current monthly price data of pork lagging 1–12 periods are imported into the PPA-based PPAR program compiled by Lou [20–22] and Mohamed et al. [47]. In the PPAR program, the PPAR model based on the first linear PRF is established, and the actual global optimal solution is obtained. The best weights and the optimal coefficients  $c_0$  and  $c_1$  are shown in the “1-0-PPAR” row in Table 2. At the same time, the PPAR model with the second linear PRF is established, and the best weights and coefficients are shown in the “1-1-PPAR” row in Table 1. It can be seen that the PPAR model with two linear PRFs performs better than the model with one linear PRFs (that is, the fitting errors of the training samples and the verification samples of the PPAR model with two linear PRFs are smaller than the errors of that with one linear PRF).

**Table 2.** Comparison of the optimal weights, polynomial coefficients, and objective function values of different PPAR models for monthly pork prices.

Model	The Best Weight $a(1) \sim a(12)$	Polynomial Coefficients $c_0, c_1, c_2$	$Q_T(a), Q_V(a)$ #
1-0-PPAR	−0.273, 0.226, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.935	20.984, 11.002	253.9, 234.3
1-1-PPAR	0.314, −0.845, 0.401, 0, 0, 0, 0, 0, 0, 0.1631, 0, 0	0.008, −3.739	239.3, 220.3
2-0-PPAR	−0.339, 0.488, −0.200, 0, 0, 0, 0, 0, 0, 0, 0, 0.779	20.988, 13.397, −0.048	246.3, 234.9
2-2-PPAR	−0.321, 0.687, −0.430, 0, 0, 0, 0, 0, 0, 0.490, 0, 0	0.276, −1.002, −4.117	225.6, 973.9
2-1-PPAR	0.267, −0.761, 0, 0, 0, 0, 0, 0, 0, 0.591, 0, 0	0.010, −1.135	235.7, 224.4

Notes: #  $Q_T(a)$  and  $Q_V(a)$  represent the sum of squared error (SSE) (objective function value) of the training and validation samples, respectively. The smaller the values are, the better the model's fitting accuracy and prediction ability (generalization ability) are. The autoregressive term with the best weight of “0” can be deleted, as shown below.

For comparison, we further established three PPAR models. The first one is the 2-0-PPAR model with one quadratic PRF. The second model is a 2-2-PPAR model with two quadratic PRFs. The third PPAR model is a 2-1-PPAR model with two PRFs; the first PRF is quadratic, and the second PRF is linear. The best weights and optimal coefficients of PRFs are shown in the “2-0-PPAR”, “2-2-PPAR”, and “2-1-PPAR” rows in Table 2.

According to the results shown in Table 1, the error of the verification samples of the PPAR model with two quadratic PRFs  $Q_V(a) = 973.9$ , which is greater than the model with one quadratic PRF  $Q_V(a) = 234.9$ , which shows that “overfitting” occurs. The performance of the 1-0-PPAR model was comparable to that of the 2-0-PPAR model, and the performance of the 1-1-PPAR model was similar to that of the 2-1-PPAR model.

For this example's data, the authors also tried to establish a PPAR model with one cubic PRF, and “overtraining” and “overfitting” occurred in each training.

According to the results shown in Table 2, all the PPAR models have multiple predictors (autoregressive terms) with the best weight equal to “0”, indicating that these autoregressive

items are “invalid” and can be removed from the reduced model. To this end, we try to establish the PPAR model only with the first six predictors  $x(i-1), x(i-2), \dots, x(i-6)$ , composed of linear and quadratic PRFs. In the PPAR model with one linear PRF, only the weights of the first three predictors do not equal “0”. In the PPAR model with one quadratic PRF, only the weights of the first four predictors are not equal to “0” (the specific results are no longer listed due to space limit). Further studies showed that the PPAR model with three predictors showed better generalization. Therefore, the reasonable number of time series lagged periods is three, and we establish the PPAR model with three predictors  $x(i-1), x(i-2), x(i-3)$ .

### 5.2. Establishment of the Optimal PPAR Model

The sample data of three predictors were imported into the PPA-based PPAR program compiled by Lou [20–22] to obtain the real global optimal solutions. The 1-1-PPAR model comprises two linear PRFs, and the 2-2-PPAR model comprises two quadratic PRFs. The 2-1-PPAR model is composed of two PRFs; the first PRF is quadratic, and the second is linear. The 1-2-PPAR model is composed of two PRFs; the first PRF is linear, and the second is quadratic. The best weights, the PRF’s coefficients, and the objective function values of the training and verification samples are shown in Table 3.

**Table 3.** Optimization results and performance comparisons of different PPAR models with three predictors (autoregressive items).

Model	The Optimal Weights $a(1)–a(3)$	Polynomial Coefficients $c_0, c_1, c_2$	$Q_T(a), Q_V(a)$
1-0-PPAR *	0.174, −0.575, 0.800	20.973, 24.627	131.58, 178.32
1-1-PPAR	−0.770, 0.597, 0.227	0.0000, 0.0001	131.58, 178.32
1-2-PPAR	−0.208, 0.775, −0.597	−0.0977, 2.7771, 35.0328	119.02, 289.15
2-0-PPAR	0.176, −0.574, 0.800	20.916, 24.563, 0.8681	138.84, 244.96
2-2-PPAR	0.209, −0.778, 0.592	−0.083, −2.607, 32.529	120.78, 392.86
2-1-PPAR	0.193, −0.790, 0.582	0.0006, −0.0804	130.83, 244.02
1-PPAR	0.173, −0.545, 0.820	20.913, 21.674	93.591, 5.549

Notes: \* for 1-2-, the first digit represents the order of the first PRF, and the second digit the order of the second PRF; that is, the 2-1-PPAR model is composed of two PRFs, where the first PRF is quadratic, and the second PRF is linear.

It can be seen from the results shown in Table 3 that the 2-2-PPAR model composed of two quadratic PRFs has an excellent data-fitting ability for training samples. The  $Q_T(a)$  is small and  $Q_V(a)$  is large indicates that “overfitting” has occurred, the prediction ability is low, and the model’s generalization ability is difficult to guarantee. The SSE of the 1-1-PPAR model, composed of one linear PRF,  $Q_V(a)$  is the smallest of all models, which shows that the model has the best generalization ability. The fitting ability of the training samples is slightly lower than that of the 1-2-PPAR model composed of the first linear PRF and the second quadratic PRF. In fact, from the actual demand for monthly pork price forecasts and early warning, as well as for the government price management, regulatory authorities, consumers, pig industry practitioners, and upstream and downstream enterprises and their employees, we pay more attention to the generalization ability and prediction ability of the PPAR model. Therefore, we take the 1-0-PPAR model as the optimal monthly pork price prediction model. The model’s performance metrics of the training and verification samples are shown in Table 4.

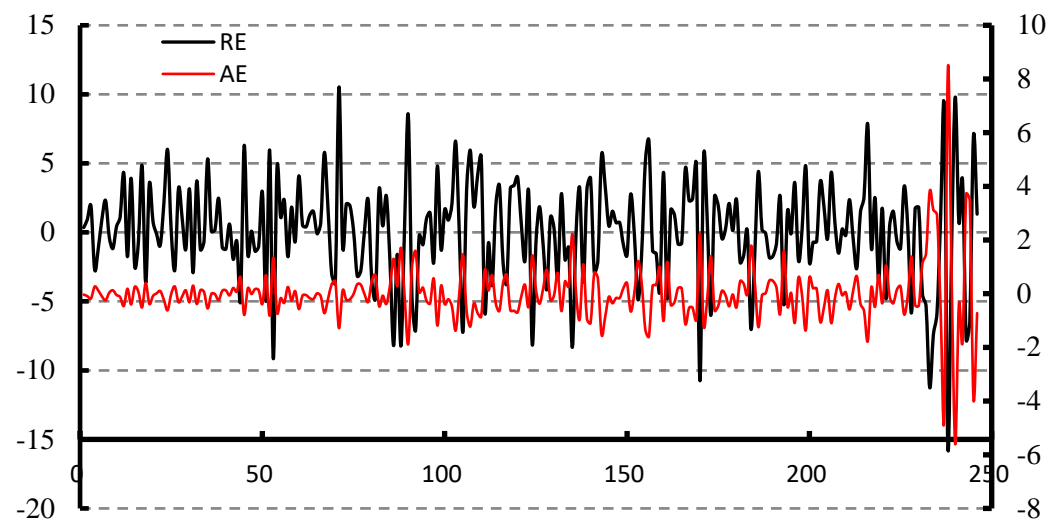
As seen from Table 4, the mean absolute error (MAE) of the prediction value is CNY 3.08, and the MAPE and Max\_RE are 5.80% and 15.7%, respectively. For soaring monthly pork prices, the model’s generalization ability is outstanding. Generally, when the price changes dramatically, no matter what model is used, the prediction (especially as a verification sample) error will be relatively large. The absolute errors (AE) and the relative errors (RE) of the training and verification samples are shown in Figure 2.



**Table 4.** Comparison of performance metrics between the training and verification samples of the 1-0-PPAR model with one linear PRF.

Model	Sample Subset	RMSE *	MAE	Max_AE	MAPE (%)	Max_RE (%)
1-0-PPAR	training	0.750	0.525	3.82	2.63	11.25
	verification	3.855	3.076	8.44	5.80	15.70
1-PPAR	training	0.654	0.479	2.408	2.52	9.78
	verification	0.680	0.496	1.371	2.11	5.81

Note: \* RMSE, MAE, Max\_AE, MAPE, Max\_RE are the root mean squared error, mean absolute error, maximum absolute error, mean absolute percentage error and maximum relative error; the smaller the performance metrics are, the better the model performance is [20,48], the same as below. The deadline date for the samples was May 2019 in the “1-PPAR” model.

**Figure 2.** Schematic diagram of the AE and RE changes of the PPAR model with one linear PRF for the monthly pork price.

As can be seen from Figure 2, for the relative error (RE), the difference between the training samples and the verification samples is not significant; for the absolute error (AE), the verification samples are all the surged pork prices, and so the error is relatively high. For the verification samples, the model performance metrics (MAPE, Max\_RE) of the relative error is 1.5–2 times as large as that of the training samples, indicating that the PPAR model has good generalization ability and practical value.

For comparison, we take the monthly pork prices from January 2000 to May 2019 (before the pork prices surged) as the studied data, take the last 12 months for validation samples, and establish a 1-PPAR model with one linear PRF. The best weights, the PRF’s coefficients, the objective function value, and the model’s performance metrics are shown in Tables 2 and 3. It can be seen that the model’s performance metrics of the verification samples of the 1-PPAR model are equal to (even smaller than) that of the training samples, which indicates that the 1-PPAR model has good generalization ability and practical value, and the prediction error is very small.

## 6. Establishment of the H-PPR Model of Monthly Pork Price Prediction with Mixed Multivariate Time Series Data

Although the PPAR model based on time series data has relatively high fitting accuracy and prediction ability, according to the PPAR model, it is challenging to propose measures to control the drastic changes in pork price and the development strategy of the pig industry, and it is impossible to study the transmission mechanism and effect of monthly pork price. Therefore, it is necessary to establish a mixed monthly pork price prediction model based on independent variables, such as hog–corn ratio, piglet price, and pork price lag in 1–2 periods, etc. In other words, by establishing an H-PPR pork price prediction model, the

transmission mechanism of pork price fluctuations can be studied, and targeted measures can be taken to stabilize pork prices and ensure pork market supply, helping to build a long-term mechanism for the sustainable development of China’s pig industry [49].

6.1. Selection of the Critical Variables

According to Section 3.2, the normalized and monthly pork price data of the 12 independent variables above were imported into the PPA-based H-PPR program compiled by Lou [20–22]. First, the H-PPR model based on the linear ridge function was established, and the actual global optimal solution was obtained. The “1-PPR” row of Table 5 is the optimal result of the PPR model with one linear PRF. The “2-PPR” row of Table 5 is the optimal weights and other parameters of the H-PPR model with one quadratic PRF. It can be seen that in the PPR model, the weight of the variable  $v(3)$  (hoghog price) is close to “1”, and the weights of the other variables are all minimal. The monthly hog price significantly impacts the monthly pork price. We deleted the variable  $v(3)$  and re-established a PPR model with one linear and quadratic PRF, respectively. The best weights in the “1-PPR-11” and “2-PPR-11” rows are in Table 5. The weights of the two variables  $v(1)$  and  $v(9)$  are greater than 0.60, and the weights of other variables are minimal. Moreover, the performance metrics of the verification samples based on the PPR model with one quadratic PRF are worse than those with one linear PRF, which shows that “overtraining” and “overfitting” of the PPR model with one quadratic PRF have occurred.

**Table 5.** Comparison of the optimal weights, polynomial coefficients, and objective function values of different PPR models for predicting pork prices.

Model	The Best Weights $a(1) - a(12)$	Coefficients $c_0, c_1, c_2$	$Q_T(a), Q_V(a)$
1-PPR	0.020, -0.056, <u>0.955</u> *, 0.195, -0.075, -0.040, 0.050, 0.072, 0.122, -0.113, -0.062, 0.007	20.901, 9.142	189.66, 156.23
2-PPR	-0.024, -0.048, <u>0.961</u> , 0.191, -0.068, -0.037, 0.039, 0.066, 0.088, -0.126, -0.049, 0.005	21.008, 9.738, -0.180	189.22, 198.75
1-PPR-11	<u>0.642</u> , 0.116, 0.157, -0.012, -0.013, 0.153, -0.001, <u>0.673</u> , -0.111, -0.245, -0.030	20.820, 8.5692	232.07, 252.84
2-PPR-11	<u>0.688</u> , 0.031, 0.148, -0.058, -0.026, 0.173, 0.030, 0.650, -0.072, -0.204, 0.015	20.357, 8.3245, 0.5764	219.68, 309.95
1-PPR-9	<u>0.656</u> , -0.372, 0.219, 0.188, 0.335, -0.152, 0.245, 0.092, -0.379	20.653, 11.428	423.70, 810.91
2-PPR-9	<u>0.651</u> , -0.367, 0.201, 0.189, 0.346, -0.154, 0.253, 0.103, -0.384	20.542, 11.327, 0.2637	423.04, 717.64
H-PPR-13	0.015, -0.024, <u>0.968</u> , 0.163, -0.055, -0.033, 0.058, 0.071, 0.077, -0.113, -0.007, 0.012, -0.060	21.168, 9.067	188.00, 162.70
H-PPR-8	0.003, -0.023, <u>0.983</u> , 0.129, <u>0</u> , <u>0</u> , 0.075, 0.036, -0.088, <u>0</u> , <u>0</u> , -0.043	21.188, 9.093	189.80, 155.70
H-PPR-7	0.049, -0.031, <u>0.939</u> , 0.195, <u>0</u> , <u>0</u> , <u>0</u> , 0.127, -0.189, <u>0</u> , <u>0</u> , 0.158	21.215, 8.031	196.70, 162.00
H-PPR-6	-0.0474, -0.0549, <u>0.9787</u> , 0.1305, <u>0</u> , <u>0</u> , <u>0</u> , -0.0482, <u>0</u> , <u>0</u> , <u>0</u> , 0.132	21.170, -8.796	201.50, 165.00
H-PPR-5	-0.105, -0.045, <u>0.633</u> , <u>0</u> , <u>0</u> , <u>0</u> , <u>0</u> , -0.044, <u>0</u> , <u>0</u> , <u>0</u> , <u>0.764</u>	21.131, 7.982	237.40, 198.30
H-PPR-6a	0.028, 0.001, <u>0.992</u> , 0.077, <u>0</u> , <u>0</u> , <u>0</u> , 0.050, <u>0</u> , <u>0</u> , <u>0</u> , 0.076	20.990, 7.789	118.20, 10.69

Table 5. Cont.

Model	The Best Weights $a(1)–a(12)$	Coefficients $c_0, c_1, c_2$	$Q_T(a), Q_V(a)$
H-PPR-12b	0.417, $-0.064$ , 0, 0.247, $-0.099$ , $-0.039$ , 0.050, 0.050, 0.451, $-0.201$ , $-0.111$ , 0.025, <u>0.698</u>	21.105, 7.553	209.88, 194.19
H-PPR-6b	0.377, $-0.059$ , 0, 0.191, <u>0</u> , <u>0</u> , <u>0</u> , 0.394, $-0.292$ , <u>0</u> , <u>0</u> , <u>0.760</u>	21.147, 7.921	214.49, 191.39

Notes: \* A double underline represents the first (or second) largest best weight. “0” represents the weight after deleting the variable. Additionally, “11” in “1-PPR-11” indicates 11 predictors in the model, and the digits “1” and “2” indicate the model with one linear or quadratic PRF, the same below.

We once deleted the two variables  $v(1)$  and  $v(9)$  and built the PPR models with one linear and one quadratic PRF, respectively. The parameters, such as the best weights, etc., are shown in the “1-PPR-9” and “2-PPR-9” rows in Table 4. The weights of other variables are relatively small, except that the weight of variable  $v(2)$  is greater than 0.65. Till now, the objective function values  $Q_T(a)$  and  $Q_V(a)$  of the PPR model of the training and verification samples are greater than the PPR model with 11 variables. The MAPE of the training and the verification samples are greater than 4.26% and 13%, respectively. The model’s prediction accuracy can no longer meet the requirements of the actual prediction.

From the above process of establishing the PPR model, the most significant variables impacting pork price are the hog price  $v(3)$ , hog–corn ratio  $v(1)$ , corn price  $v(9)$ , and piglet price  $v(2)$ . Therefore, we can study the transmission mechanism of pork price and the effects of the four variables above. The impact of other variables is implied in the above variables.

According to the results shown in Table 5, the established PPR model with one linear PRF can meet the prediction accuracy requirements. The value  $Q_V(a)$  of the PPR model with one quadratic PRF is large and indicates that “overtraining” and “overfitting” have occurred.

## 6.2. Establishment of an H-PPR Model Based on the Monthly Pork Price and Multivariate Price Time Series

According to the PPAR model established above, the other factors will be less critical if the pork prices with lag periods of 1–3 are included. Therefore, only the monthly pork price with a lag period of one was included to establish an H-PPR model with pork price and the other 12 independent variables.

First, the data of those abovementioned 12 independent variables and the lag period one pork price were normalized with a mean of zero and a variation of one. Then, the normalized data and the monthly pork price (the dependent variable) are input into the H-PPR program. We obtain the global optimal solutions. Because the H-PPR model with one quadratic PRF has occurred “overtraining” and “overfitting”, we only list the results of the H-PPR model with one linear PRF in Table 5. Without deleting the above four critical variables, we delete only one variable with the least weight each time and establish the model step by step. The parameters, such as the best weights, etc., are shown in Table 5. It is clear that the sum of errors  $Q_T(a)$  and  $Q_V(a)$  (objective function value) of the training and verification samples changes a little when the number of independent variables reduces from 13 to 8 (H-PPR-13 and H-PPR-8). If we once again delete the variables  $v(5)$  (mutton price) and  $v(10)$  (finishing pig feed price), the objective function values of the training and verification samples are raised (referring to “H-PPR-7” and “H-PPR-6” rows in Table 5). The accuracy still meets the actual requirement. If we continuously delete the variable  $v(4)$  (beef price), the objective function values of the training and verification samples are too great to meet the practical requirement (referring to the “H-PPR-5” row in Table 5). Therefore, we built the final H-PPR model with six independent variables and a lag period one of pork price. The H-PPR model only contains the hog–corn ratio, the hog price, beef, corn, finishing pig feed, pork with a lag period of one, and the piglet price with a lag

period of six. The performance metrics of the training and verification samples are shown in Table 6.

**Table 6.** Comparison of the performance metrics of the H-PPR model with different data deadline.

Model	Sample	RMSE	MAE	Max_AE	MAPE (%)	Max_RE (%)	Bias
H-PPR-6	Training	0.936	0.628	6.323	3.03	15.05	0
	Verification	3.873	3.462	7.535	6.62	14.92	0.872
H-PPR-6a	Training	0.742	0.559	2.640	2.88	12.71	0
	Verification	0.986	0.788	2.159	3.40	9.24	0.706
Equation (7)	Training	0.910	0.628	6.214	3.13	14.79	0
	Verification	3.701	3.064	7.568	5.85	14.99	1.256
Equation (8)	Training	0.706	0.537	2.121	2.84	12.69	0
	Verification	1.000	0.751	2.502	3.21	10.60	0.741

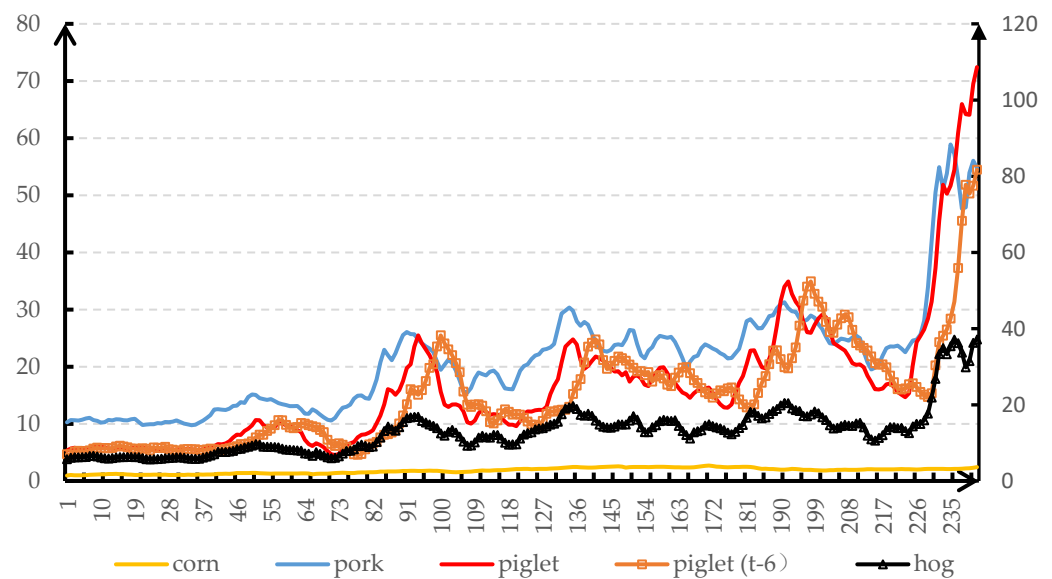
Notes: The date deadline for the samples for the H-PPR-6 and H-PPR-6a models are September 2020 and May 2019, respectively.

Similarly, from August to October 2019, the monthly pork price soared rapidly. As a verification sample, its performance metrics are 1–2 times those of the training samples. If we establish an H-PPR model using the time series data from January 2000 to May 2019 before the price surge, its performance metrics of the H-PPR-6a are shown in the “H-PPR-6a” row of Table 6, and the optimal parameters are shown in the “H-PPR-6a” row of Table 5. It can be seen that the performance metrics of the verification samples are equal to those of the training samples. Of course, the predictive ability of the H-PPR-6a model, with a deadline of September 2020, is also strong enough.

The H-PPR-6 model seems to differ from general cognition. It is usually believed that piglet and corn prices are the most critical factors affecting pork prices. The modeling results confirm these conclusions, but this is only partially true. The relationship between the monthly pork price and the hog–corn ratio, the price of corn, hog with lag period one, and the piglet price with lag period six is shown in Figure 3. It can be seen that the monthly pork price trend is consistent with that of the above independent variables. Still, there are also some differences, and the correlation is lower than the monthly pork price with a lag period of one. The best weight of the PPR model shows that except for the monthly pork price with a lag period of one, the hog–corn ratio with a lag period of one has the most significant impact on the pork price, followed by the corn price, indicating that the feeding cost is the most critical factor determining the monthly pork price. The prices of piglet, beef, and other variables have a particular impact on the pork price but are less important than the feeding cost.

Due to the large number of verification samples, the “overtraining” and “overfitting” occurred quickly in building the PPR model with Hermite orthogonal PRFs. At the same time, we cannot establish a reliable and reasonable neural network-based projection pursuit regression model (PPBP) [20].

From the objective function value of the PPAR model and the H-PPR model (referring to Tables 3 and 5) and the performance metrics (referring to Tables 4 and 6), we can conclude that the performance metrics of the training and verification samples of the PPAR model are both higher than those of the H-PPR model.



**Figure 3.** The schematic diagram of the changing relationship between the pork price and the price of the hog, corn, the piglet with a lag period of one, and the piglet price with a lag period of six (the piglet's price with lag periods one and six are on the right ordinate).

## 7. Results and Discussion

### 7.1. Comparison of the PPAR, H-PPR, and MLR Models

(1) The PPAR model has very high fitting and prediction accuracy. According to the results shown in Table 2, it can be seen that the fitting accuracy of the training samples and the prediction ability of the verification samples of the PPAR model are both higher than that of the H-PPR model. This once again shows that the established models, such as ARIMA, BPNN, PPAR, and SVR with univariate time series data, are feasible and meaningful for monthly pork price prediction, which confirmed again that the monthly pork price time series data contains a variety of factors. The monthly pork price volatility and changing trends include a certain regularity. Compared with ARIMA and BPNN models, the PPAR model is more concise, has a clear mathematical meaning, and has a relatively simple topology. According to the best weight, the PPAR model can be very convenient to determine the monthly pork price fluctuations and change trends based on the pork price with lag periods. The best weight of the pork price with lag period one is 0.800, and the price with lag period two is  $-0.575$ , which indicates that the pork price with lag period one has the most significant impact on the pork price and the pork price with lag period two has a reverse harmonic effect and the second significant impact on the monthly pork prices. The weight of the price with lag period three is only 0.174, and the impact is significantly lower than the lag periods one and two. Conversely, we cannot draw similar conclusions from ARIMA, BPNN, etc.

(2) The H-PPR model established has an excellent ability to fit the data, test the prediction and generalization of the samples, and reveal the transmission mechanism and effect of pork price, which can effectively regulate the monthly pork price. Although the H-PPR model's data fitting accuracy and prediction ability are slightly lower than the PPAR model, according to the optimal weight of multiple factors, we can analyze the transmission mechanism of pork price change and judge the pork price fluctuation and changing trend, put forward more targeted measurement, and control pork price fluctuations or soaring, etc. Therefore, establishing the H-PPR model is essential for strengthening the pork supply chain management and promoting the healthy development of the pig industry chain. The H-PPR model also provides the basis for decision making.

It can be seen from the best weight of influencing factors in the H-PPR model that the hog price with a lag period of one has the most significant impact on the monthly pork price, followed by beef price, the pork price with a lag period of one, the hog-corn ratio, and



the piglet price with a lag period of six. Therefore, if the departments for price monitoring and management find that the hog price has risen significantly, they must provide more pork supply to the market. Otherwise, the pork price will increase significantly in the next month. Similarly, if beef prices rise significantly, the departments must take corresponding measurements to provide more pork or beef supply to the market. Otherwise, the pork prices in the following months will certainly rise. If we delete the hog price to establish the H-PPR model, its objective function value is 209.88 (referring to the “H-PPR-12b” row in Table 4), which is significantly greater than 188.00. If we delete the seven non-important independent variables, such as mutton, and establish the H-PPAR-6b model, the pork price with a lag period of one has the most significant impact on the pork price, followed by the hog–corn ratio, then the beef price, the finishing pig feed price, etc. Therefore, if we establish the models with various variables, we may obtain different prediction results. Thus, we must carefully select proper and reliable influencing factors for modeling.

(3) The reliability of the MLR model is difficult to guarantee. We establish the MLR model with the same data and 12 variables. This is because most of the variables do not obey the normal distribution, even if the fitting accuracy of the MLR model is not low. Theoretically, its reliability and robustness are challenging to guarantee. The MLR model is obtained as follows

$$y(i+1) = 21.166 + 0.161v(1,i) - 0.246v(2,i-6) + 8.316v(3,i) + 1.412v(4,i) - 0.442v(5,i) - 0.292v(6,i) + 0.510v(7,i) + 0.606v(8,i) + 0.711v(9,i) - 0.911v(10,i) - 0.119v(11,i) + 0.031v(12,i) \quad (6)$$

Because there is collinearity between the variables, only the variable  $v(3,i)$  is significant at level 0.01, the variables  $v(4,i)$  and  $v(8,i)$  are significant at level 0.1, and Equation (6) has a little sense. Using the stepwise regression method, we establish Equation (7) with the significant variables at level 0.10,

$$y(i+1) = 21.166 - 0.322v(2,i-6) + 8.604v(3,i) + 1.014v(4,i) + 0.719v(8,i) + 0.711v(9,i) - 0.341v(11,i) \quad (7)$$

The model performance metrics of Equation (7) are shown in Table 5.

Similarly, if the deadline date for the monthly pork price is May 2019, the MLR is established as

$$y(i+1) = 20.920 + 7.901v(3,i) + 2.496v(4,i) - 2.493v(5,i) + 0.639v(8,i) + 0.789v(9,i) \quad (8)$$

The model performance metrics of Equation (8) are shown in Table 5. Compared with the performance metrics of various models in Table 5, it can be seen that although the performance metrics of the MLR are almost the same as H-PPR-6 and H-PPR-6a, the bias of Equation (7) is much greater than that of H-PPR, indicating that the predicted value of Equation (7) is skewed. Its robustness and reliability could be better.

Comparing Equations (7) and (8), we found that some significant variables differ. The coefficients of the piglet, the mutton, and the compound feed of broiler chickens were less than 0, indicating that these variables adversely affect the pork price, which is difficult to explain in theory. At the same time, the hog–corn ratio has nothing to do with pork prices and is inconsistent with common sense and truth. Therefore, although the fitting and prediction accuracy of the MLR is not low, its results are challenging to explain reasonably, and its practicability is poor.

## 7.2. Comparison with Xiong et al. [4]

Xiong et al. [4] used 11 variables, including monthly pork price, piglet price, lean pork futures prices, west Texas light (West Texas Intermediate) crude oil prices, etc., from January 2000 to March 2019, and established a dynamic model average (DMA) consisting of 2000 models (each model has four–five variables). Its results are compared with the

Bayesian model, time-varying parameter model, etc. The deadline date for this paper is May 2019, which is almost the same as that of Xiong et al. [4]. The RMSE and SMAPE of the three training sample models (DMA, dynamic model selection, Bayesian model average) are shown in Table 6. The RMSE and SMAPE of the PPAR and H-PPR models established in this paper are also shown in Table 6. It can be seen that the SMAPE of PPAR and H-PPR models for training and verification samples are both smaller than those in Xiong et al. [4]. Meanwhile, the RMSE is slightly larger than that of Xiong et al. [4]. The leading cause is the large prediction error in February 2019 (AE = CNY 2.27, RE = 9.63%). In fact, from November 2018 to January 2019, the pork price was CNY 23.69, 23.16, and 22.55, respectively, which was gradually reduced. It suddenly turned upward in February, rose by more than one to CNY 23.61, and resulted in a large prediction error. The prediction error in March returned to normal, indicating that the PPR model has good robustness.

The DMA model in Xiong et al. [4] is too complex to practice and only suitable for theoretical research. The applicability needs to be higher. We need to solve 8000~10,000 models' parameters to establish DMA, and its prediction accuracy is similar to that of PPAR and PPAR models. In contrast, we need to solve a few parameters, such as the best weights of variables and the coefficients of PRF applying PPA, which is convenient.

Moreover, we cannot analyze the transmission mechanism of affecting pork prices through the DMA model.

### 7.3. Comparison of PPR with SVR, BPNN, etc.

The SVR/SVM and BPNN models have their characteristics and advantages. Although the results in many articles show that SVR's fitting accuracy and prediction ability are better than that of BPNN, BPNN is still more used for price prediction and early warning. We apply data process system software [50] and the STATISTICA Neural Network [40] to establish the SVR and BPNN; the results are shown in Table 6. The results of the SVR are closely related to the specified ranges of the parameters to optimize. For univariate pork price time series data, the BPNN network topology is 3-2-1 (the number of neurons on the input, hidden, and output layer are 3, 2, and 1), and the number of its connection weights is 11. The network topology 6-2-1 is used for multivariate times series data of 12 variables, and its connection weights are 17. The 24 verification samples (about accounting for 10%) are randomly selected. During the training process, we monitor the RMSE of the verification samples, stop training when the REMS of verification samples begin to rise, and take the network weights before "overtraining". The number of the training samples is ten times greater than that of connection weights, which meets with the principle of modeling BPNN. The following can be seen from Table 7: (1) For the univariate pork price time series data with lag periods of 1-3, the SVR has the smallest RMSE and SMAPE of the training sample and the largest values of the verification samples, which indicates that the generalization ability is poor. The PPAR and BPNN, their RMSE, and SMAPE of the training and verification samples are almost the same, which indicates that PPAR and BPNN without "overtraining" have good generalization ability. (2) For the multivariate pork price time series data, the RMSE and SMAPE of the SVR are the smallest, but those of the verification samples are large, which indicates that the generalization ability of the SVR is poor; the RMSE and SMAPE of the training samples of H-PPR are good agreement with that of verification samples, which indicates that the H-PPR model has good generalization ability and is also better than the BPNN model. The H-PPR, SVR, and BPNN outperform DMA, dynamic model selection, and Bayesian model averaging for training samples. Therefore, compared with the BPNN and SVR, the PPAR and H-PPR have similar fitting abilities but generally do not occur as "overtraining" and "overfitting" during modeling and have a better predictive ability and generalization ability.

**Table 7.** Comparison of RMSE and SMAPE of data for various models and different cases.

Model	Univariate Time Series		Multivariate Pork Price Times Series	
	RMSE	SMAPE	RMSE	SMAPE
DMA	/	/	0.541 */	3.387 /
Dynamic model selection	/	/	0.557 /	3.391 /
Bayes model averaging	/	/	0.664 /	3.906 /
PPAR/H-PPR	0.650/0.680	2.509/2.493	0.743/0.986	2.887/3.212
SVR *	0.503/1.195	1.752/3.415	0.517/1.0216	2.044/3.155
BPNN	0.655/0.657	2.561/2.127	0.684/1.420	2.819/5.125

Notes: \* The values before and after “/” denote the RMSE and SMAPE of the training and verification sample, respectively.

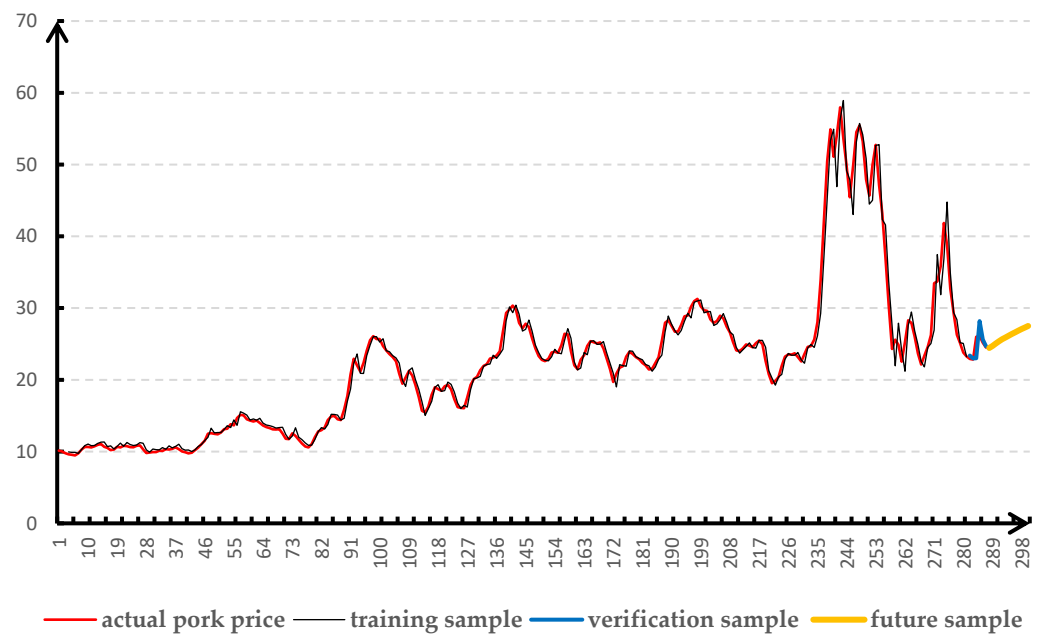
Through the above comparison, we can conclude the following. Firstly, the PPAR and H-PPR models not only have simple structures but also have explicit structures with precise mathematical meanings, and their prediction accuracy is higher than other machine learning models such as SVR and BPNN. Secondly, the PPAR and H-PPR models are semi-parametric models. When establishing the models, only the coefficients (weights) of multiple factors or autoregressive terms and the coefficients of the ridge function need to be optimized, which is not easy to cause “over-training”. Thirdly, based on the established models, the importance of multiple influencing factors or autoregressive terms can be directly judged, making it easier to analyze the transmission mechanism of pork price, build a pork price control mechanism, and strengthen pork supply chain management. This promotes the sustainable development of pig farming, as well as upstream and downstream industries such as cattle, sheep, and chickens, agricultural product production, feed processing, and sales, and lastly, promotes sustainable agricultural and regional development.

#### 7.4. To Predict the Pork Price Using the Latest Data Available

We collect the latest pork price data from January 2020 to November 2023 from the National Bureau of Statistics of PRC (<http://www.stats.gov.cn> accessed on 5 January 2024) and the Ministry of Agriculture and Rural Affairs of PRC (<http://www.moa.gov.cn> accessed on 5 January 2024). The websites do not provide multivariate time series data. So, we only establish the PPAR model using the data from January 2000 to November 2023 to predict pork prices in the following 13 months.

We input the normalized data into the PPA-based PPAR program, build a PPAR model with one quadratic PRF, and obtain the global optimal solution. The best weights are 0.1236,  $-0.4823$ , and  $0.8672$ , and the best coefficients of the PRF are  $22.3053$ ,  $19.7423$ , and  $-0.45775$ . We obtain the sample projection values  $z(i) = 0.8672x(i-1) - 0.4523x(i-2) + 0.1236x(i-3)$ , and the predicted pork price  $y(i) = 22.3053 + 19.7243z(i) - 0.45775z^2(i)$ . The performance metrics of MAE, RMSE, MAPE, Max\_AE, and Max\_RE of the training samples are 0.8684, 1.5089, 3.37%, 7.708, 18.89%, and those of the verification samples are 0.8368, 1.2516, 3.23%, 2.145, 8.21%. The predicted values of the training and verification samples, as well as the forecasted samples in the following months, are shown in Figure 4.

From Figure 4, we can conclude that the pork price will gradually increase in the following months, and the departments of price management and business administration should pay more attention to the market and provide more pork, beef, etc., to the market.



**Figure 4.** The comparison of the actual and predicted pork price and the forecasted values.

## 8. Conclusions, Policy Recommendations, Limitations, and Future Research

### 8.1. Conclusions

(1) The sustainable development of the pig industry is an important component of animal husbandry, feed processing industry, and agriculture, which significantly impacts achieving sustainable economic, social, and environmental development. The reliable and accurate prediction and risk warning of pork price fluctuations are the foundation and guarantee for achieving the sustainable development of the pig industry, playing a leading role. Establishing PPAR and H-PPR models and accurately and reliably predicting the pork price changing trend help the Chinese government to establish a long-term mechanism to promote the sustainable development of the pig industry, improve and strengthen the system for pork (pig) price prediction and warning mechanisms, collect the information about feed prices such as corn and finishing pig feed as well as piglet prices in a timely manner, strengthen monitoring of African swine fever and other diseases, strengthen the management of the pig industry chain, ensure controllable price fluctuations and stable production, and achieve the sustainable development of the pig industry (animal husbandry) and its related industries, laying a solid foundation for sustainable agricultural development.

(2) There is important theoretical significance and practical value in establishing the PPAR and H-PPR models to forecast the monthly pork price and expand the method. We collect the time series data of the monthly pork prices from January 2000 to September 2020 as well as the other 12 influencing factors (variables), such as the piglet and corn prices. For the monthly pork price, the studied results of the PPAR model with one linear or quadratic PRF show that the pork price lagged by 1–3 periods has a significant influence, and the lagged period of one has the most and positive impact, while the lagged period of two plays is of secondary significance and has a reverse and harmonic impact. The PPAR model possesses high fitting accuracy and good generalization ability. According to the time series data of the piglet price with a lagged period of six, the other variables, and the pork price with lagged period one, we established an H-PPR model with one linear PRF. We found that seven variables, including the hog price, beef price, pork price, finishing pig feed price, piglet price, hog–corn ratio, and corn price, are important influencing factors. Among them, the hog price had the most significant impact, playing a decisive and positive role, followed by the beef and pork prices with a lagged period of one. The influence impacts of other

variables are almost the same. Therefore, we established the PPAR and H-PPR models to expand a method for monthly pork price prediction.

(3) The generalization ability and applicability of the established PPAR and H-PPR models are better than SVR, BPNN, DMA, and other methods. Compared with SVR, BPNN, and DMA models, the PPAR and H-PPR models are semi-parametric and “white box” models. We established the PPAR and H-PPR models with a few parameters, which are more straightforward, more explicit in mathematical meaning, and more convenient for applications than the other models. According to the best weights of the established PPR models, we can directly judge the importance of the lagged periods of the pork price, the importance of each variable, and its ranking, put forward the practical measurement of adjusting the pork price, and study the transmission mechanism and effectiveness of the pork price. According to market surveys or collected data, if the hog price has risen significantly in a month, we should increase the pork and beef supply to stabilize the pork price. Otherwise, the pork price will dramatically increase in the next month. Similarly, if the corn and beef prices in a month have increased significantly, it indicates that the pork prices in the next month will also rise significantly. If the monitoring finds that the piglet price increases significantly, the monthly pork price will rise considerably in the sixth month.

(4) According to the PPAR model, we can forecast the monthly pork price in multi-periods with higher accuracy, and the government departments can conveniently judge the changing trend of the pork price. With the H-PPR model, we can forecast the monthly pork price with a lagged period of one and study the transmission mechanism and effectiveness of the pork price. The related government departments take adequate measures to strengthen pork supply chain management and take steps to control the pork price. The studied results of the PPAR model show that only the periods lagged by 1–3 of the monthly pork price have an important impact on the current pork price; it is not necessary to introduce more lagged periods into models, and it is beneficial to simplify the model, improving its practicability. The prediction accuracy of the PPAR model is even higher than the H-PPR model. Still, its shortcomings are not suitable for studying the pork price transmission mechanism and the measures and suggestions to control the pork price. According to the results of the H-PPR model, we can analyze the transmission mechanism and effectiveness of the monthly pork price, and the government authorities can strengthen the management of the pork supply chain and promote the healthy development of the pig industry chain. We established the H-PPR model to delete seven factors with lower influence, although this does not mean that these seven factors are unrelated to the monthly pork price. Their influence impact has been reflected by factors such as hog–corn ratio, corn price, etc. The transmission mechanism of the monthly pork price is very complex and needs to be studied further.

(5) We establish a PPAR model using the latest pork price data from January 2000 to November 2023 to forecast the trend of pork prices changing in the following months. The results show that the pork price will rise in the future. The departments of price management and business administration should closely monitor the changes in pork prices and take timely measurements to adjust pork, hog, beef, etc., supply to ensure stable prices and increased efficiency in the pig farming industry.

## 8.2. Policy Recommendations

- (1) To improve the monitoring of the monthly pork price, piglet price, other information, and the timeliness of monthly pork price prediction.

The pork price is the center of the whole price system of the pig industry chain. There is a lagged effect in the price transmission of pig breeding, and the transmission effectiveness of slaughtering and sales links also has information asymmetry, as well as sudden situations such as swine fever, which is highly likely to lead to drastic price fluctuations. Therefore, if the monthly pork price is to be controlled within a reasonable range, the relevant government departments must further improve the daily monitoring



of the monthly pork price, piglet price, and other information and timely feedback on the drastic changes in relevant prices, to improve the timeliness and reliability of the monthly pork price forecast.

Many factors influence the pork price. According to the results of this paper, the pork price monitoring system mainly involves primary data collection, management, processing, etc. It should focus on monitoring the baby cost (piglets price), feeding cost (corn, pig ratio, pig, and chicken feed prices, etc.), alternative production prices (such as beef, mutton, live chicken, etc.), and hog price index etc. We must apply the timely data to establish the PPAR and H-PPR models to ensure the timeliness of the monthly pork price prediction. Based on timely pork price predictions, the market subject can make good decisions and take corresponding measures to keep the pork price fluctuation within a reasonable range, ensuring the orderly operation of the market mechanism.

- (2) To standardize the release of the pork price information and to realize real-time information sharing.

Information asymmetry is a fundamental reason for the risk of the pork market. The regulatory information department should promptly release the pork price forecast results and the price information of related products, simplify the information query process, and realize information sharing. In this way, the market administrators, producers, and operators in the pig industry chain can, in a timely and accurate manner, grasp the market development trend and reliably guide the market administrators, producers, and operators to adjust the production and operation decisions according to the forecasting information, and actively adapt to the changes in the market situation.

- (3) To improve the risk early warning system of the monthly pork price and the government's coordinating ability.

Relevant government departments should establish an emergency control mechanism for pork prices to ensure market supply and price stability. Sudden outbreaks such as African swine fever are unpredictable and quickly lead to drastic changes in pork prices in the short term. Therefore, in addition to monitoring the price information, the relevant government departments must also closely monitor the epidemic situation of pigs, coordinate the release and storage of frozen pork meat from the central reserve in a timely fashion, and ensure the essential balance between the supply and demand of pork, to reduce the adverse impact of the pig epidemics.

### 8.3. Limitations and Future Research

Theoretically, the relationship between pork supply and demand should be one of the essential factors in determining the monthly price change of pork. Data composition techniques, such as VMD, EEMD, etc., have been widely applied in modeling time series data, and there are still some differences in their effectiveness. So, there are two limitations in this paper. First, without complete data on pork's supply and demand, similar to the other literature, we do not consider the monthly supply and demand of pork in our modeling. Furthermore, infectious and sow reproductive diseases have always threatened the sustainable development of the pig farming industry; there is a shortage of related information, so we do not consider these factors. Second, we establish PPAR and H-PPR models using the original data, do not decompose the pork price time series data into independent components, and do not compare whether the data decomposition will improve the generalization ability, applicability, and reliability. In future research, we should collect and consider the pork supply, demand, and disease factors to establish H-PPR models. Secondly, we will decompose the pork price into independent components using VMD and EEMD, etc., and study whether data decomposition techniques will improve the model performance or not.

**Author Contributions:** Conceptualization, X.Y. and B.L.; methodology, X.Y. and Y.L.; software, X.Y.; validation, X.Y.; formal analysis, X.Y.; investigation, B.L.; resources, B.L.; data curation, X.Y.; writing—original draft preparation, X.Y.; writing—review and editing, B.L. and Y.L.; visualization, B.L.; supervision, B.L.; project administration, B.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original data and results are available upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Zhang, S.; Wu, X.; Han, D.; Hou, Y.; Tan, J.; Kim, S.; Li, D.; Yin, Y.; Wang, J. Pork production systems in China: A review of their development, challenges and prospects in green production. *Front. Agric. Sci. Eng.* **2021**, *8*, 15–24. [\[CrossRef\]](#)
- Tatarintsev, M.; Korchagin, S.; Nikitin, P.; Gorokhova, R.; Bystrenina, I.; Serdechnyy, D. Analysis of the Forecast Price as a Factor of Sustainable Development of Agriculture. *Agronomy* **2021**, *11*, 1235. [\[CrossRef\]](#)
- Šrédli, K.; Prášilová, M.; Severová, L.; Svoboda, R.; Štěbeták, M. Social and Economic Aspects of Sustainable Development of Livestock Production and Meat Consumption in the Czech Republic. *Agriculture* **2021**, *11*, 102. [\[CrossRef\]](#)
- Xiong, T. Do the Influencing Factors of Pork Price Change Over Time—The Analysis and Forecasting Based on Dynamic Model Averaging. *J. Huazhong Agric. Univ.* **2021**, *3*, 63–73+186.
- Li, Y.; Wang, X.G. Analysis of Pork Price Prediction Based on PCA-GM-BP Neural Network. *Math. Pract. Theory* **2021**, *51*, 56–63.
- Zhang, D.B.; Cai, C.M.; Ling, L.W.; Chen, S.Y. Pork Price Ensemble Prediction Model Based on CEEMD and GA-SVR. *J. Syst. Sci. Math. Sci.* **2020**, *40*, 1061–1073.
- Kim, H.; Choi, I. The Economic Impact of Government Policy on Market Prices of Low-Fat Pork in South Korea: A Quasi-Experimental Hedonic Price Approach. *Sustainability* **2018**, *10*, 892. [\[CrossRef\]](#)
- Ndwandwe, S.; Weng, R. Competitive Analyses of the Pig Industry in Swaziland. *Sustainability* **2018**, *10*, 4402. [\[CrossRef\]](#)
- Pröll, S.; Grüneis, H.; Sinabell, F. Market Concentration, Producer Organizations, and Policy Measures to Strengthen the Opportunities of Farmers for Value Addition—Empirical Findings from the Austrian Meat Supply Chain Using a Multi-method Approach. *Sustainability* **2022**, *14*, 2256. [\[CrossRef\]](#)
- Ye, F.; Xie, J.; Ma, J.G. Periodic Study of Pork Price Fluctuation in China based on HP Filter. *Prices Mon.* **2017**, *10*, 27–30.
- Lou, W.G.; Chen, F.; Zhang, B.; Liu, L.J.; Fan, X. Forecasting the Market Daily Price of the Fresh Pork based on a General Regression Neural Network and Positive Research. *J. Syst. Sci. Math. Sci.* **2016**, *36*, 1986–1996.
- Han, M.; Yu, W.; Clora, F. Boom and Bust in China's Pig Sector during 2018–2021: Recent Recovery from the ASF Shocks and Longer-Term Sustainability Considerations. *Sustainability* **2022**, *14*, 6784. [\[CrossRef\]](#)
- Ping, P.; Liu, D.Y.; Yang, B.; Jin, D.; Fang, F.; Ma, S.J.; Ye, T.; Wang, Y. Research on the Combinational Model for Predicting the Pork Price. *Comput. Eng. Sci.* **2010**, *32*, 109–112.
- Sun, D.Y.; Chen, L. Analysis of the Dynamic Relationship between Pork Prices and Influencing Factors. *Stat. Decis.* **2020**, *36*, 74–77.
- Feng, S.J.; Chen, F. Pork Price Trend Fluctuation and Its Forecast in China-Analysis Based on Elman Neural Network Model. *Price: Theory Pract.* **2018**, *6*, 90–93.
- Zhang, M.; Yu, L.A.; Liu, F.G. Study on Regime Transition and Nonlinear Dynamic Adjustment Behavior of Hog Industry Chain Price in China. *Chin. J. Manag. Sci.* **2020**, *28*, 45–56.
- Sarle, C.F. The forecasting of the price of hogs. *Am. Econ. Rev.* **1925**, *15*, 1–22.
- Ezekiel, M. Two Methods of Forecasting Hog Prices. *J. Am. Stat. Assoc.* **1927**, *22*, 22–30. [\[CrossRef\]](#)
- Hamm, L.; Brorsen, B.W. Forecasting Hog Prices with a Neural Network. *J. Agribus.* **1997**, *15*, 37–54.
- Lou, W. *The Projection Pursuit Theory Based on Swarm Intelligence Optimization Algorithms—New Developments, Applications, and Software*; Fudan University Press: Shanghai, China, 2021.
- Yu, X.; Lou, W. An Exploration of Prediction Performance Based on Projection Pursuit Regression in Conjunction with Data Envelopment Analysis: A Comparison with Artificial Neural Networks and Support Vector Regression. *Mathematics* **2023**, *11*, 4775. [\[CrossRef\]](#)
- Yu, X.H.; Xu, H.Y.; Lou, W.G.; Xu, X.; Shi, V. Examining energy eco-efficiency in China's logistics industry. *Int. J. Prod. Econ.* **2023**, *258*, 108797. [\[CrossRef\]](#)
- Suaza-Medina, M.E.; Zarazaga-Soria, F.J.; Pinilla-Lopez, J.; Lopez-Pellicer, F.J.; Lacasta, J. Effects of data time lag in a decision-making system using machine learning for pork price prediction. *Neural Comput. Appl.* **2023**, *35*, 19221–19233. [\[CrossRef\]](#)
- Sun, F.; Meng, X.; Zhang, Y.; Wang, Y.; Jiang, H.; Liu, P. Agricultural Product Price Forecasting Methods: A Review. *Agriculture* **2023**, *13*, 1671. [\[CrossRef\]](#)

25. Chuluunsaikhan, T.; Ryu, G.; Kwan-Hee Yoo, K.; Rah, H.; Nasridinov, A. Incorporating Deep Learning and News Topic Modeling for Forecasting Pork Prices: The Case of South Korea. *Agriculture* **2020**, *10*, 513. [[CrossRef](#)]
26. Qin, J.; Yang, D.; Zhang, W. A Pork Price Prediction Model Based on a Combined Sparrow Search Algorithm and Classification and Regression Trees Model. *Appl. Sci.* **2023**, *13*, 12697. [[CrossRef](#)]
27. Ye, K.; Piao, Y.; Zhao, K.; Cui, X. A Heterogeneous Graph Enhanced LSTM Network for Hog Price Prediction Using Online Discussion. *Agriculture* **2021**, *11*, 359. [[CrossRef](#)]
28. Wang, J.; Wang, Z.; Li, X.; Zhou, H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *Int. J. Forecast.* **2022**, *38*, 21–34. [[CrossRef](#)]
29. Purohit, S.K.; Panigrahi, S.; Sethy, P.K.; Behera, S.K. Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. *Appl. Artif. Intell.* **2021**, *35*, 1388–1406. [[CrossRef](#)]
30. Jaiswal, R.; Choudhary, K.; Kumar, R.R. STL-ELM: A Decomposition-Based Hybrid Model for Price Forecasting of Agricultural Commodities. *Natl. Acad. Sci. Lett.* **2022**, *45*, 477–480. [[CrossRef](#)]
31. Li, B.; Ding, J.; Yin, Z.; Li, K.; Zhao, X.; Zhang, L. Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting. *Expert Syst. Appl.* **2021**, *168*, 114232. [[CrossRef](#)]
32. Zhu, H.; Xu, R.; Deng, H. A novel STL-based hybrid model for forecasting hog price in China. *Comput. Electron. Agric.* **2022**, *198*, 107068. [[CrossRef](#)]
33. Fu, L.; Ding, X.; Ding, Y. Ensemble empirical mode decomposition-based preprocessing method with Multi-LSTM for time series forecasting: A case study for hog prices. *Connect. Sci.* **2022**, *34*, 2177–2200. [[CrossRef](#)]
34. Liu, Y.; Duan, Q.; Wang, D.; Zhang, Z.; Liu, C. Prediction for hog prices based on similar sub-series search and support vector regression. *Comput. Electron. Agric.* **2019**, *157*, 581–588. [[CrossRef](#)]
35. Ling, L.; Zhang, D.; Chen, S.; Muger, M. Can online search data improve the forecast accuracy of pork price in China? *J. Forecast.* **2020**, *39*, 671–686. [[CrossRef](#)]
36. Krambia-Kapardis, M. Neural Networks: The Panacea in Fraud Detection? *Manag. Audit. J.* **2010**, *25*, 659–678. [[CrossRef](#)]
37. Lou, W.G. Early Warning Model of Financial Risks and Empirical Study Based on Neural Network. *Financ. Forum* **2011**, *11*, 52–61.
38. Lou, W.G. Evaluation and Prediction of Soil Quality based on Artificial Neural Network in the Sanjiang Plain. *Chin. J. Manag. Sci.* **2002**, *10*, 79–83.
39. Zhang, G.; Patuwo, E.; Hu, M. Forecasting with Artificial Neural Networks: The State of the Art. *Int. J. Forecast.* **1998**, *14*, 35–62. [[CrossRef](#)]
40. Statsoft, Inc. *STATISTICA Neural Networks*; Statsoft, Inc.: Tulsa, OK, USA, 2007.
41. Bhasin, H.; Khanna, E. Neural Network based Black Box Testing. *ACM SIGSOFT Softw. Eng. Notes* **2014**, *39*, 1–6. [[CrossRef](#)]
42. Friedman, J.H.; Stuetzle, W. Projection Pursuit Regression. *J. Am. Stat. Assoc.* **1981**, *76*, 817–823. [[CrossRef](#)]
43. Hall, P. On Polynomial-based Projection Indices for Exploratory Projection Pursuit. *Ann. Stat.* **1989**, *17*, 589–605. [[CrossRef](#)]
44. Zhan, H.R.; Zhang, M.K.; Xia, Y.C. Ensemble Projection Pursuit for General Nonparametric Regression. *arXiv* **2022**, arXiv:2210.14467. [[CrossRef](#)]
45. Durocher, M.; Chebana, F.; Ouarda, T.B.M.J. Delineation of Homogenous Regions using Hydrological Variables Predicted by Projection Pursuit Regression. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 4717–4729. [[CrossRef](#)]
46. Chen, C.; Tuo, R. Projection Pursuit Gaussian Process Regression. *IISE Trans.* **2023**, *55*, 901–911. [[CrossRef](#)]
47. Mohamed, A.A.; Hassan, S.A.; Hemeida, A.M.; Alkhalaf, S.; Mahmoud, M.M.M.; Eldin, A.M.B. Parasitism-Predation Algorithm (PPA): A Novel Approach for Feature Selection. *Ain Shams Eng. J.* **2020**, *11*, 293–308. [[CrossRef](#)]
48. Chengdu Development and Reform Commission; Sichuan University School of Economics. A Study on Optimizing Quantitative Prediction and Application of Agricultural Products Price: Take Chengdu's Main Agricultural Products as an Example. *Macroeconomics* **2021**, *6*, 107–116.
49. Zhang, L.; Luo, Q.; Han, L. Research on the construction of long-term mechanism for sustainable development of China's pig industry. *Issues Agric. Econ.* **2020**, *12*, 50–60.
50. Tang, Q.Y.; Zhang, C.X. Data Processing System (DPS) Software with Experimental Design, Statistical Analysis and Data Mining Developed for Use in Entomological Research. *Insect Sci.* **2013**, *20*, 254–260. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.