



# The Study of Prediction of Hog Price in China

YiKang, Cui\*

The Rural Survey Department of National Bureau of Statistics of China

No. 57, Yuetan Nanjie, Sanlihe, Xicheng District

Beijing, China

[Cuiyk@stats.gov.cn](mailto:Cuiyk@stats.gov.cn)

DOI: 10.1481/icasVII.2016.g43

## ABSTRACT

Pork is the main meat for Chinese consumer, the price of hogs has a significant impact on the price of agricultural products and the Consumer Price Index(CPI), and it has always been the focus of the government and the academic community to stabilize the price of hogs. In this paper, we use the intermediate consumption survey data from National Bureau of Statistics of China, by using the machine learning algorithm, to predict the price of hogs in each provinces in the first half of year 2016. The result of this research would be used to lay the foundation for more accurate and efficient policy control and market regulation.

**Keywords:** hog prices, intermediate consumption survey, machine learning algorithm

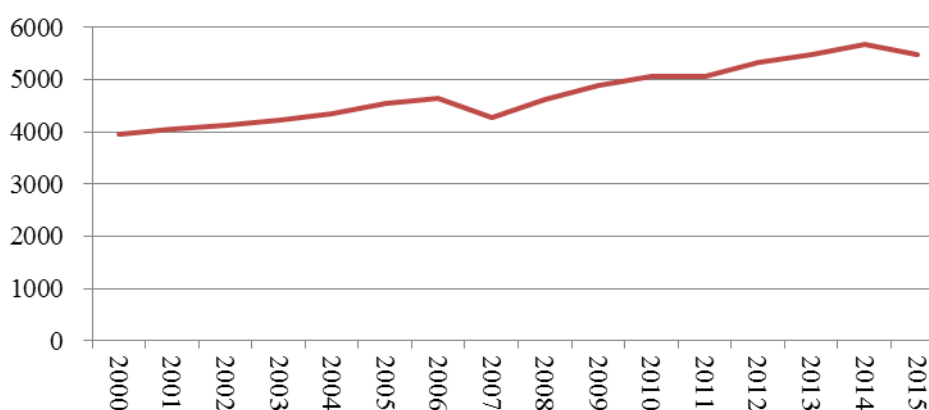
## 1. INTRODUCTION

Pork is the main meat for Chinese consumers. In 2015, The production of pork reached 54.87 million tons, about 41 kg per person, accounting for about 65% of the total output of meat. In recent years, The hog price fluctuates frequently. On the one hand, the producers may face more risks in the market, on the other hand, the quality of people's lives would be affected. Therefore, The hog

price controlling mechanism for stabilizing the hog prices and reducing the volatility of the hog prices become a hot topic for government, the society and the scholars.

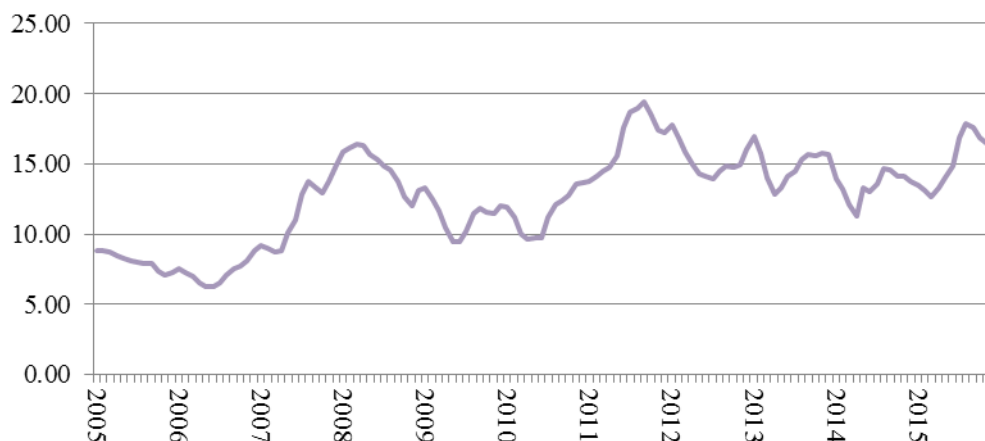
Most studies on the mechanism for producing hog price at home and abroad focus on: Effects between supply and demand, Production costs, the level of money supply, the international market price transmission, the scale of production and epidemic disease etc. The main research methods include Cobweb Theorem, Vector Autoregression (VAR), Neural Networks (NN), Grey System etc., based on the analysis of the seasonal, cyclical and trend fluctuations in the price of Hogs. For example, Ying Qi, Jie Li (2007) considered that the relationship between supply and demand, the deviation of production decisions, the cycle of hog production, Consumption habits of the residents are the main causes of hog price volatility. By using the method of Co integration analysis, with the monthly data from 2002 to 2006, Fang Wang (2009) found out that the long term influence factors of pork price, sorted by impact strength, are the price of corn, piglet, hogs and pork. Feng Luo et al. (2009) indicated the long-term equilibrium relationship between the Chinese and the international prices of agricultural products by using the Co-integration Analysis and Impulse Response Function, with the data from 2003 to 2008. LiXiang Zhang et al. (2011) found the high degree of integration in the hog chain market by using the Co integration Analysis to study the price data of corn, feed, piglets, pig and pork, from January 2003 to July 2010, and the prices of Piglets and hogs are the core of the whole chain by the Granger Causal Relation Test. YongBo Ma (2011) considered the price volatility of agricultural products mainly came from the supply, cost-push, the money supply and international agricultural price transmission and so on. GuoChang Xin (2013) found the supply of cattle products can provide a stronger affect on the prices than the demand did, and the supply, demand, production costs, the price of alternatives, macro-control can cause significant different affect on the hog price by path, intensity and period.

Since the 1990s, the scale production of poultry breeding industry have developed rapidly in China, the marketable fattened stock from the scale-producing units (units that have the size of the marketable fattened stock more than 1000) accounts for about 17% of the total amount of the marketable fattened stock. In recent years, the yield of Chinese pork goes up periodically, there were inflection points once every 5-8 years. In the last 15 years, the inflections occurred in the year of 2007, 2011 and 2015.



**Figure 1:** The yield of Chinese pork, 2000-2015 (Unit: 10,000 tons)

The hog price presents the obvious periodicity and seasonality. From a monthly point of view, the price is usually lower from April to May and higher from January to February. Viewed annually the hog market price has a fluctuation cycle of 3-5 years, and recently the cycle became shorter.



**Figure 2:** *The Chinese Hog market price, 2005-2015 (unit: yuan / kg)*

There are four surveys related to hog in the Rural Division of the National Bureau of Statistics of China including the animal husbandry production survey (quarterly), the agriculture, forestry, animal husbandry and fishery output survey (quarterly), the agricultural products price survey (quarterly or monthly) and the intermediate consumption survey (semiannually). The relevant statistical indicators include the amount of output, slaughter and feeding; the output value and added value; the producer price and the market prices; the feeding costs, costs of piglets, labor costs, operating expenses and so on. In order to adapt to the industrial development of the hog production, the National Bureau of Statistics continually reforms and improves the statistical investigation system and sampling methods to better reflect the production and development of hog.

## 2. EMPIRICAL ANALYSIS

### 2.1 Model Selection

In the study of forming mechanism of the agricultural products prices, the methods of Cobweb Theorem, the VAR model, the neural networks and the grey system all have been widely used and achieved fruitful results. In the empirical study, the quality of data has a significant impact on the results. The study of the hog price has its special difficulties: firstly, with the rapid development of industrialization and the necessity of environmental protection, the picture of hog producing changes rapidly, which makes the analysis of the long-term hog prices become difficult. Secondly, since the producing and consuming of hogs have the attribute of periodic and seasonal characteristics, the relationships between the hog price and its influencing factors are relatively complex, meanwhile, the function between the hog price and its influencing factors are different on path, intensity and cycle, there are both the positive and reverse conduction between the hog price

and its influencing factors and the transmission intensity is also different. Thirdly, the quality of the data itself affects the results. With the different sources of data and differently processing methods, the scholars may draw different conclusions from the same object. There are a large number of high quality data in NBS. To develop and utilize these data has always been an important issue of our work. There are about 2000 sample units in the intermediate consumption survey of hogs nationwide in China, Each year about 1/5 of the sample units are rotated. There are two surveys for one year, one is held in the first half of the year and the other held in the second half. The interview of the survey covers the main information of each sample unit, including address, the hog slaughter, the sale price ,the piglets cost ,the feeding cost, the fuel cost, the water cost, the electricity cost, the labor costs and so on. It could be theoretically and practically instructive to predict hog price movements by using these data.

On the basis of the represent above, this study based on samples which have many variables and been observed in a short-term period. Considering the characteristics of the data and the applicability of the model ,we decide the research method of this paper is using the data of the intermediate consumption survey, applying the machine learning method (the four models of Decision Tree, Support Vector Machine ,Neural Networks and Random Forest) to construct the forecast model to forecast the price movements of hogs, and then put forward some specific suggestions.

## 2.2 Variables Selection and Data Processing

To form the index system from the intermediate consumption survey data we preliminarily select the price of hog as the explained variables, and the explanatory variables include the price of purchased piglets, the price of since numerous bred piglet, the price of concentrated feed, the price of crude feed, the price of water fees, the price of electricity fees, the price of gasoline, the price of diesel, the price of coal, the price of hired worker, the price of family labor and the cost of the epidemic prevention.

**Table 1:** *preliminary selection of variables*

Name	Shortcut	Type
the price of hog	y	explained
the price of purchased piglets	gzx	explanatory
the price of since numerous bred piglet	zfzx	explanatory
the price of concentrated feed	jsl	explanatory
the price of crude feed	csl	explanatory
the price of water fees	sf	explanatory
the price of electricity fees	df	explanatory
the price of hired worker	gg	explanatory
the price of family labor	jg	explanatory
the cost of the epidemic prevention	Fy	explanatory

the price of gasoline	qy	explanatory
the price of diesel	cy	explanatory
the price of coal	m	explanatory

The intermediate consumption survey began in 2002 in the Rural Survey of the National Bureau of Statistics. Taking into account the sample rotation and the characteristics of the hog production in different periods, we selected the data from 2011 to 2015 into the analysis. The objects of the intermediate consumption survey are households or units, considering the sample rotation and the different characteristic of hog production in different regions, we selected the provincial average prices into analysis after comparison. Therefore we get the final analysis data of a total of 10 cycles from 2011 to 2015 and the provincial average prices.

Before carrying out the machine learning method, we need to translate the time series data into the panel data. We tested for the correlation coefficients between the hog prices and all the influential factors in the cycles (T-4,T+4) by applying the Pearson Correlation Coefficient Test. Then we get to a conclusion that the price of purchased piglets, the price of concentrated feed, the price of water fees, the price of electricity fees, the price of hired worker in the period of T-1 have the biggest correlation with the price of hog, while the correlation between the price of self breeding piglets, the price of crude feed, the price of gasoline, the price of diesel, the price of coal, the price of family labor in period T-1 did not correlate significantly with the price of hog. Hence we put a one-cycle-advance on each influencing factor then matched with the hog price to make up the training and testing samples for the machine learning model, and impact factor data of the last cycle was used to predict the hog prices of the coming cycle.

**Table 2:** *The correlation test between the price of hogs and the T-1 influencing factors*

Name	Coefficient	P Value
the price of purchased piglets	0.837	0.0000
the price of concentrated feed	0.867	0.0002
the price of water fees	0.513	0.0307
the price of electricity fees	0.551	0.0312
the price of hired worker	0.732	0.0237
the price of since numerous bred piglet	0.637	0.1017
the price of crude feed	0.523	0.0749
the price of gasoline	0.314	0.2022
the price of diesel	0.350	0.2812
the price of coal	0.211	0.3714
the price of family labor	0.273	0.0747

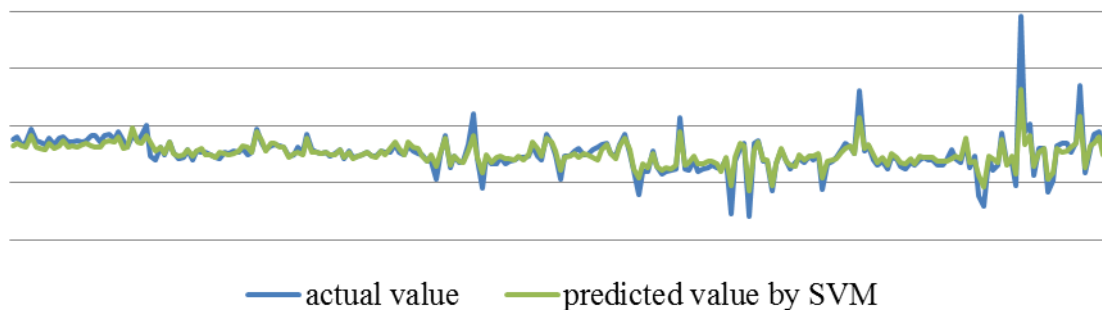
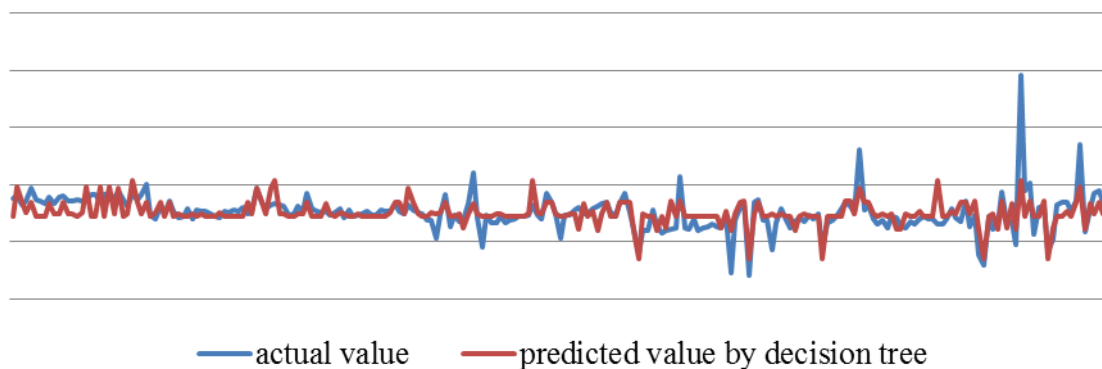
---

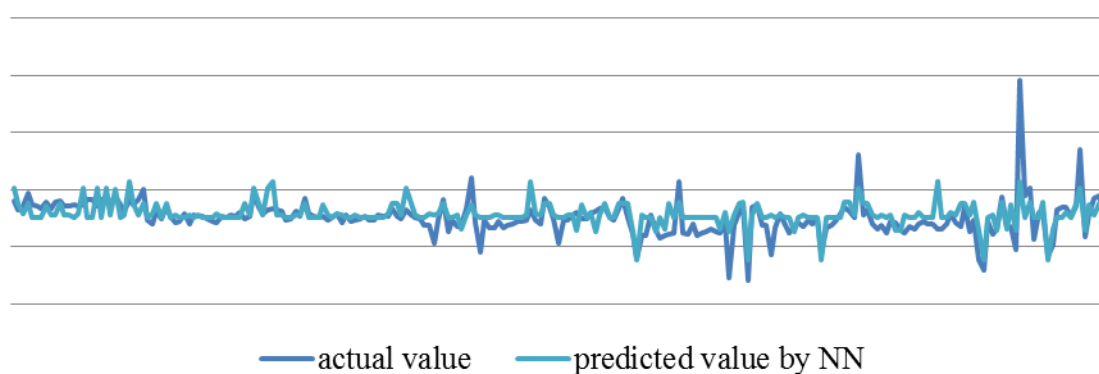
the cost of the epidemic prevention	0.416	0.0814
-------------------------------------	-------	--------

---

### 2.3 Model Implementation

The training and test data are used to model in R3.2.4 by the models of the Decision Tree, Random Forest, Support Vector Machine and BP Neural Networks. The program code is shown in the Appendices. The fitting curves of the four models are shown below.





**Figure 3:** *The fitting curves of the four models*

By applying the 10-fold cross validation, we get the Root Mean Square Errors (RMSE) of the four models decision tree, random forest, support vector machine and BP neural network are respectively 0.2214, 0.0125, 0.3756 and 0.3392. This shows that the four models have good fitting effects and it is feasible by using these models to predict the hog prices, By comparison of the values of RMSE, in these four models, the Random Forest model is relatively better. Finally, we used the 10th period data of the influential factors to predict provincial hog prices in the first half of 2016, then calculated the national price index of the hog which was about 113.7.

### 3.CONCLUSIONS AND RECOMMENDATIONS

Firstly, this paper shows that it is feasible to use the intermediate consumption survey data such as the cost of piglets, feed, operating and labor etc. to predict the price of live pigs by applying the machine learning algorithm.

Secondly, this paper provides a reference of building a price forecasting model in multi regions and shorter periods with the exploration of forecasting the price of hogs in many provinces and the semiannual.

Thirdly, since the obtainable data is limited, this article has not yet taken the effect of market demand, money supply level, international market price transmission, alternative goods prices, the macro-control policy etc. into scope of study. we will continue to explore the impact of these factors on the price of hog in the next stage.

Finally, by looking into the leading indicators we can forecast the price movements of hogs in advance of one or several cycles. In addition, analysis on the price movements of the different regions with the leading indicators could help to make more accurate and sensitive policy and market regulation.

**REFERENCES**

- Lv Jie, Qi Ying. (2007)The economic analysis of the cyclical fluctuations of the hog market price, *Issues in Agricultural Economy* ,7,89-92.
- Guochang Xin. (2013)A study on the price regulation mechanism of the livestock products ,*Ph. D thesis of Inner Mongolia Agricultural University*.
- Junan Chen, Fang Wang. (2009)The analysis of the transmission mechanism of the hog price fluctuation, *China Rural Survey* ,7,31-41.
- Yinchu Ceng, Xuefeng Mao.(2008) The identification of hog price cycle based on time series decomposition, *Chinese Rural Economy*, 12,4-13.