Dynamic Influence of Network Public Opinions on Price Fluctuation of Small Agricultural Products Based on NLP-TVP-VAR Model—Taking Garlic as an Example

Xingchen Lv 1, Jun Meng 2,* and Qiufeng Wu 2

1 College of Engineering, Northeast Agricultural University, Harbin 150030, China; lvxingchen@neau.edu.cn
2 College of Arts and Sciences, Northeast Agricultural University, Harbin 150030, China; qfwu@neau.edu.cn
* Correspondence: merd@neau.edu.cn

Abstract: In recent years, the price of small agricultural products has both plummeted and skyrocketed, which has a great impact on people’s lives. Studying the factors affecting the price fluctuation of small agricultural products is of great significance for stabilizing their price. With the development and application of social media, farmers and consumers are more greatly influenced by online public opinion, resulting in irrational planting behavior or purchasing behavior, which has a complex impact on the price of small agricultural products. Taking garlic as an example, we crawled through network public opinions about garlic price from January 2015 to December 2020 using web crawler technology. Then, the network public opinions were quantified using a natural language processing and time-varying parameter vector autoregression (NLP-TVP-VAR) model to empirically analyze their dynamic influence on garlic price fluctuation. It was found that both public attitude and public attention have a short-term influence on garlic price fluctuation, and the influences of each differ according to direction, intensity and timing. The influence of public attitude on garlic price fluctuation is positive, while the influence of public attention on garlic price fluctuation is largely negative. The influence intensity of public attitude is stronger than that of public attention on garlic price fluctuation. The influence of public attitude on garlic price fluctuation shows a trend of intensifying, while that of public attention has been weaker than in previous years. In addition, based on the results of our study, we present some recommendations for improving the comprehensive information platform and price fluctuation early warning system for the whole industry chain of small agricultural products.

Keywords: small agricultural products; network public opinions; natural language processing; TVP-VAR model; dynamic influence

1. Introduction

Agriculture is the basic industry of social development, and agricultural products are the primary guarantee and prerequisite for human life and production; thus, the price of agricultural products has always been a key concern of state and society [1]. In China, with the upgrading of residents’ consumption structure and the continuous adjustment of the agricultural cultivation structure, the economic value of small agricultural products is increasing. The prices of small agricultural products directly impact the national economy and quality of residents’ lives [2]. For example, garlic price plummeted to ¥1.31/kg in May 2008 and skyrocketed to ¥14.98/kg in May 2017. Similarly, the price of ginger plummeted to ¥2.47/kg in June 2007 and skyrocketed to ¥16.64/kg in October 2014. Such drastic price fluctuations seriously affect the normal consumption patterns of residents and is prone to systemic risks [3]. There are many reasons resulting in the aforementioned phenomenon. Firstly, small agricultural products are characterized by significant seasonality, aggregation and asymmetry, which has a certain impact on price fluctuation [4,5]. Secondly, the volume of yield, market demand, trade and distribution are small, and the production areas are concentrated, which makes the price vulnerable to disruptions from external factors [6].
Finally, when the existing management cannot keep pace with developments in the current era, the scale, standardization, branding and informatization of production are relatively low, and the price is unable to respond to emergencies in a timely manner [7]. It can be seen that, in addition to being influenced by their own factors, there is a greater likelihood of interference in the price of small agricultural products due to external factors, especially with the progress of times and the development of technology, and the influencing factors are thus becoming increasingly diversified, and their contribution to small agricultural product prices is increasing. Therefore, it is important to analyze the factors influencing small agricultural product prices toward stabilizing small agricultural product prices and increasing farmers’ income.

At present, social media platforms, such as Sina Weibo, WeChat and online forums, play an important role in people’s lives. On the one hand, people can express their opinions or ideas through social media anytime, anywhere. On the other hand, our own opinions or decisions can be influenced by other’s sentiments [8]. According to social identity theory [9], when an opinion or a behavior is accepted by most people, individual behavior can be influenced by group behavior within the same social group, or the so-called “herd effect”. Furthermore, according to the theory of planned behavior [10], when someone publishes opinions with strong sentiments on the social media, others seem more easily influenced, resulting in irrational decisions being made. Conversely, the herd and strong sentiment effects are potentially useful for farmers when making decisions related to the price of small agricultural product and the planting area for next season. On the one hand, since farmers may be influenced by other people’s opinions on fixing small agricultural product price and the planting area for next season expressed on the social media, they make decisions that are aligned with other people’s opinions. On the other hand, since farmers easily trust opinions reflecting obviously polarized attitudes on social media, they make decisions aligned with these sentiments. Decisions made by farms in the case of the herd and strong sentiment effects result in price fluctuation of small agricultural products. For example, the mung bean price skyrocketed in 2010 because most people believed that mung bean could cure all diseases after a health expert published a book about the advantages of mung bean [11]. Furthermore, the extreme public panic caused by the African swine fever in 2018 led to plummeting pork prices [12]. Therefore, by analyzing the public attitude and attention to the fluctuation of small agricultural product prices, we can understand the farmers’ planting willingness and consumers’ purchase intention, which can be used in predicting trends in small agricultural product prices.

China is the largest garlic grower, consumer and exporter in the world, and the garlic industry plays an important role in China’s agricultural industry [13]. However, in the process of development of the garlic industry, a serious event occurred due to the lack of attention to network public opinions, called “Cangshan Garlic” [14], which resulted in serious damage to the entire garlic industry as well as the image of government and social security. With the continuous improvement of agricultural informatization construction in China, the Ministry of Agriculture and Rural Affairs issued the Agricultural and Rural Big Data Pilot Program in 2016 and actively carried out a big data pilot study of eight important agricultural products, including garlic [15]. However, since the existing management cannot keep pace with developments in the current era and the scale, standardization, branding and informatization of production are relatively low, garlic prices are still plummeting and skyrocketing. Therefore, this paper taking garlic as an example, explored the influences of network public opinions on garlic prices.

Equilibrium price theory states that the interactions between basic demand and supply market forces are the main driving forces determining small agricultural market equilibrium and agricultural price fluctuations [16]. In the garlic market, the influences of public attitude and attention on garlic price fluctuation act mainly through changing supply and demand. The network public opinions contain a lot of information, among which the most important is public attitude and public attention [17]. Public attitude and attention, to an extent, influence farmers’ willingness to plant garlic, causing the relationship between supply
and demand in the garlic market to change and thus affect garlic price. When the public attitude toward garlic prices shows positive tendencies, farmers will choose to raise the price of garlic and tend to plant more garlic. The more attention this positive sentiment receives will, on the one hand, make farmers more willing to plant garlic, which increases the garlic supply; on the other hand, consumers will choose to buy other alternative small agricultural products because the garlic price is too high. When the supply of garlic in the market exceeds the demand, the price of garlic will decline. When the public attitude about garlic prices turns negative, farmers would prefer to reduce the garlic price to sell out their stock and be less willing to plant garlic. The more attention this negative sentiment receives, the more unwilling farmers will be to plant garlic, and the garlic supply will decline. When the market supply of garlic is less than demand, garlic prices will rise. Therefore, this paper proposes the following hypotheses:

H1. Public attitudes have a short-term promoting effect on price fluctuations.

H2. Public attentions have a short-term inhibitory effect on garlic price fluctuation.

This study aims to examine the influence of network public opinions on the price fluctuation of small agricultural products. Taking garlic as an example, the influences of public attention and public attitude on garlic price fluctuation were studied and discussed. However, how to quantify the public attitude and public attention and construct relevant indicators reasonably is the main challenge of this study, which directly affects the estimation and analysis of subsequent model.

The main contribution of this paper is to propose a two-stage method, including a natural language processing and time-varying parameter vector autoregression (NLP-TVP-VAR) model to quantitative the public attention and public attitude, and analyze the dynamic impact of network public opinions on garlic price fluctuation. Firstly, this paper crawls the network public opinions, including examining the content of posts about garlic (such as price, international trade volume and weather, etc.), number of views and post time, etc., from garlic forums using crawler technology to construct a dataset on public opinions about garlic. In the first stage of NLP-TVP-VAR, this paper classifies polarity sentiment of network public opinions on garlic using sentiment analysis, and the result on polarity sentiment labels is then used to construct the sentiment indicators. The procedure can be divided into three steps: (a) segmentation of network public opinion on garlic into several words using a natural language processing and information retrieval (NLPIR) big data analysis platform; (b) sentiment classification based on new sentiment lexicon and SVM; and (c) construction of sentiment indicators for public attention and public attitude. Public attention is constructed according to the number of views, which is called Read, and public attitude is constructed according to the number of positive sentiment posts and the number of negative sentiment posts, which is called BI. In the second stage of NLP-TVP-VAR, this paper uses TVP-VAR model, together with the garlic price and sentiment indicators, to analyze the dynamic influence of network public opinions (including public attention and public attitude) on garlic price fluctuation. The procedure mainly includes two steps: (a) A Markov chain Monte Carlo (MCMC) algorithm is used to estimate the time-varying parameters in terms of unobserved latent variables. (b) Impulse response analysis is used to examine the dynamic relationship between public attention, public attitude and garlic price fluctuation. Finally, the influences of public attention and attitude on the garlic price are assessed. These results can assist the government and relevant departments in guiding and developing online public opinion in a positive direction toward controlling the prices of small agricultural products and stabilizing small agricultural product markets.

The rest of the paper is structured as follows: Section 2 reviews the literature on the main factors influencing agricultural product price fluctuation, methods of sentiment analysis, and the relationship between network public opinions and market prices. Section 3 delineates the research variables and methodology issues, which include the consideration
of web crawler technology, natural language processing techniques and autoregressive models. Section 4 introduces the NLP results and discusses the dynamic influence of garlic network public opinions on garlic price based on TVP-VAR. Conclusions and suggestions to keep the price of small agricultural products stable are provided in Section 5. Section 6 details the research limitations and future research directions.

2. Literature Review

2.1. Factors Influencing Agricultural Product Price Fluctuation

As small agricultural products have obvious characteristics, such as being seasonal, having a concentrated origin, easy storage, etc., their prices are vulnerable to influence by the external environment, leading to frequently appearance of skyrocketing and plummeting phenomena [18,19]. The price fluctuations of small agricultural products are mainly influenced by several factors, such as supply and demand [20,21], climate [22] and international macroeconomic factors [23], etc. Considering supply and demand, Hassan [24] found that supply logistic was the main factor influencing vegetable price fluctuation. Liu [25] pointed out that inventory has a certain impact on the ginger price. From the viewpoint of climate, Deressal [26] suggest there is a certain relationship between climate and agricultural product output, which causes agricultural product prices to be influenced. Yuan [27] demonstrated that climate is the main factor influencing the price of perishable agricultural products. In the view of international macroeconomic factors, Benavides [28] found that the exchange rate is the main factor influencing the price of agricultural products. The research of Westcoot [29] and Wu [30] showed that the international energy market is highly correlated with the agricultural product market in China. In addition, many studies have shown that inflation, domestic future price and money supply exacerbate abnormal price fluctuations of agricultural products [1,31]. The results from studies on the influencing factors of small agricultural products price are in accordance with those from previous social economy studies, and the collection of historic data has a lagging nature. However, these influences were found based on the environment of social economy at the particular time, and with the change in economic environment, the main influences also change. Furthermore, since collecting historic data of conventional influencing factors has a certain lagging nature, results on the analysis of price fluctuations of small agricultural product based on the use of conventional influencing factors are also relatively lagging and do not reflect real-time performance. When emergency situations occur, the analysis of prices based on the historic data of previous conventional influencing factors cannot produce accurate results; thus, the market is not flexible in dealing with emergencies, and prices will skyrocket and plummet.

2.2. Methods of Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a fundamental task in natural language processing (NLP) that involves the process of analyzing, processing, generalizing and reasoning network public opinions. Sentiment analysis has a wide range of application in domains such as the financial [32], e-commerce [33,34] and assessment teaching quality [35] domains, among others. The main methods of sentiment analysis are based on a sentiment lexicon, which in turn is based on deep learning, a subset of machine learning. Sentiment analysis methods based on sentiment lexicon refer to the division of sentiment polarity at different granularities based on the sentiment polarity of sentiment words provided by different sentiment lexicons. Yekrangi [36] and Beigi [37] constructed domain sentiment lexicons based on building upon the basic sentiment lexicon, which could effectively identify and extend on sentiment words in the corpus and improve the classification effect of the domain dataset. Machine-learning-based sentiment analysis methods refer to feature extraction and sentiment classification using machine learning methods with a large amount of annotated corpus information. Go and Bhayani [38] proposed using a supervised learning method to classify sentiment reviews as positive or negative when extracted from Twitter using naive Bayes, maximum entropy and SVM analysis algorithms,
and the experiment showed 80% analysis accuracy. Deep-learning-based sentiment analysis methods are able to actively learn text features and learn semantic information in text based on deep network models to achieve sentiment classification [39]. Dashtipour et al. [40] automatically mined relevant features in text information for sentiment classification by convolutional neural network and LSTM model. Gopalakrishnan et al. [41] performed sentiment analysis on Twitter datasets by constructing an LSTM model with optimal parameters, and satisfactory sentiment classification performance was achieved.

However, it can be seen from the above literature analysis that most of the existing sentiment lexicons are manually constructed, which requires a lot of human and material resources, so machine-learning-based methods do not make full use of the contextual information of the contextual text, and deep-learning-based methods require large amounts of data to support them and are not suitable for small datasets.

2.3. Relationship between Network Public Opinions and Market Price Behavior

The majority of research on the relationship between network public opinions and market prices has focused on financial markets. Strau et al. [42] used the Granger causality test to study the relationships between the emotions in Dutch newspaper articles and stock market prices and found that negative emotions better reflected stock market trends. Derakhshan et al. [43] improved the accuracy of stock price predictions by building a new stock prediction system that combined technical stock price indicators and the sentiments expressed in news articles. In addition, there are some studies on the sentiment influences on agricultural market price behaviors. Hassouneh et al. [44] developed an avian influenza food panic information index to analyze the impact of the avian influenza epidemic on vertical poultry prices in Egypt. Chen et al. [45] calculated the positive and negative emotional tendency values on social networks based on CNN and tested the Granger causality between emotions and vegetable prices. Hsl et al. [46] analyzed the relationships between African swine fever (ASF) and meat prices based on a TVP-VAR model, and the results showed that there were some differences in the impact size, direction and duration of ASF on the prices of pork, chicken, beef and mutton. Liu et al. [47] quantified online negative sentiment using microblog text mining and the results of SVM as the sentiment value, which was then used to empirically analyze the dynamic impact of negative public emotions on agricultural product prices during the COVID-19 pandemic in China.

However, some common problems are encountered in the aforementioned studies. First of all, the public sentiment index in the abovementioned literature was not constructed according to public speech, so it cannot adequately reflect the real sentiment of public attitude. Second, only public attitude, and not public attention, was considered as an influencing factor of price fluctuation. Finally, many scholars only analyzed the influence of network public opinion on market price for a specific event and did not discuss the influence of network public opinions as a separate influencing factor.

Therefore, this paper tried to combine the natural language processing technology (NLP) and time-varying parameter vector autoregression (TVP-VAR) model to reasonably and objectively construct the indicators of public attention and public attitude, and investigate the time-varying relationship of public attitude and public attention on garlic price fluctuation. This study analyzes the dynamic influence of public attention and public attitude on the price fluctuation of small agricultural products from the perspective of network public opinions, which enriches the influencing factors of small agricultural product price fluctuation, fills the gap of influence of public attention on price fluctuation and provides new ideas for the prediction and early warning of price fluctuation of small agricultural products.

3. Materials and Methods

3.1. Overview

This paper takes garlic as an example; the research approach is grounded on web crawler technology and a natural language processing technology (NLP) and time-varying...
parameter vector autoregression (TVP-VAR) model. Firstly, web crawler technology was used to obtain the data for network public opinions about garlic price fluctuation from garlic forums, and the content of posts and number of views were extracted as the source for construction of public attitude and public attention indices. Then, on the one hand, sentiment analysis was conducted, and the number of monthly positive and negative sentiment content of posts calculated, and the time series of BI was developed as Formula (3). On the other hand, the monthly number of views was calculated, and the time series of Read was developed as Formula (2). At the same time, the garlic market price data of the whole country were processed using Formula (1) to develop a basic time series of garlic price. Finally, before model estimation, a unit root test was conducted on the time series of BI, Read and R. To ensure time series stability, a model optimal lag order was selected, and the TVP-VAR model was used to analyze the time-varying influence of network public opinions on garlic price fluctuation. Figure 1 shows a flowchart of the experimental methods, and each method will be introduced in following section.

![Flowchart of the experimental methods used in this study.](image)

**Figure 1.** Flowchart of the experimental methods used in this study.

### 3.2. Research Variables

#### 3.2.1. Garlic Price Dataset

Data on garlic prices were obtained from the Business Forecast website ([http://cif.mofcom.gov.cn](http://cif.mofcom.gov.cn)) (accessed on 27 August 2021). Business Forecast is a comprehensive information release platform of the Ministry of Commerce for domestic trade statistics monitoring and industry management. Thus, there is no doubt regarding the accuracy and comprehensiveness of the data. Therefore, we obtained the weekly garlic prices from the Business Forecast website and determined the time range to obtain the garlic price data for the six years from 2015 to 2020. Since the garlic price is to be standardized and averaged afterwards, the monthly average price was used as the garlic price for the month. In order to calculate this conveniently, the rate of garlic price volatility was defined, as shown in Formula (1).

\[ R = \ln P_t - \ln P_{t-1} \]  

(1)

Here, \( P_t \) and \( P_{t-1} \) represent the prices in months \( t \) and \( t - 1 \), respectively.

#### 3.2.2. Network Public Opinion Datasets

##### (1) Data acquisition

Although there are many carriers related to the topic of garlic prices on the Internet, in order to ensure the authority and authenticity of corpus, the selected website must meet the following requirements: (1) The website should focus on agriculture. (2) The amount of information should be relatively large and reflect the real fundamentals more...
comprehensively. (3) The time series should be complete to allow for prediction research with a wide range of time periods, so as to obtain a general conclusion. (4) The website should be influential enough to have some impact on price fluctuations. Based on the above four factors, this paper chose International Garlic Forum (http://bbs.51garlic.com) (accessed on 27 August 2021) as the source for public sentiment information. The forum of international garlic forum net is a large platform for everyone to discuss the factors of and trends in garlic price, with a large amount of information and complete time series.

With the popularity and rapid development of Internet, the network is flooded with a large amount of information. Web crawler technology emerged to facilitate effectively using the network resources and to quickly find useful information [48]. Web crawler technology is a program or script that automatically crawls web pages for information according to certain rules, simulating a browser, and is, thus, a way to obtain network resources [49]. Octopus is a universal Internet data collector for the whole network that can obtain standardized data from different websites or web pages. This paper collected information based on Octopus v8.2.0, that developed by Shenzhen Vision Information Technology Co., Ltd., Shenzhen, China, and the founder, Mr. Liu, is dedicated to the collection and analysis of big data information through the software of Octopus. The detailed principle is shown in Figure 2.

$$R = \ln P_t - \ln P_{t-1}$$

Here, $P_t$ and $P_{t-1}$ represent the prices in months $t$ and $t-1$, respectively.

![Figure 2. Principle of data collection.](image-url)

With regard to the type of information collected on posts, we obtained the username, post content, number of views, number of comments, and post time. After the content was collected, all information was exported and stored as an Excel file.

(2) Data Processing

The dataset we used in the experiment is the garlic forum data. Using crawler technology, the relevant posts in the forum on garlic price from 2015 to 2020 were extracted. A total of 3562 pieces of posts were obtained. Then, the raw posts were cleaned up by excluding duplicates, advertising, intentional abuse and other useless information. Finally, 3357 posts that met the requirements remained. These posts were divided into two groups. One
was for the posts with obvious sentiment, which were manually classified into 800 posts with positive sentiment and 800 posts with negative sentiment. The remaining posts were classified using machine learning methods. The classified posts were used to train and test the classifier. Some examples of the classified posts are shown in Table 1.

### Table 1. Example posts in the forum dataset.

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major good news came, immediately ushered in a wave of rising peak.</td>
<td>There is no lowest, only lower.</td>
</tr>
<tr>
<td>Prices continue to rise today, the upward trend is clear, the market is healthy.</td>
<td>The downside has been established, and a deep decline is coming.</td>
</tr>
<tr>
<td>Garlic prices will continue to rise after the Spring Festival. From most of the domestic production areas at present, there are not too many supplies left.</td>
<td>Urgent tips, the plunge is coming, the new garlic will be listed in large quantities soon, and the plunge of garlic is in sight.</td>
</tr>
</tbody>
</table>

#### (3) Construct Sentiment Index

In order to reflect the public attitude and public attention to garlic price from multiple angles, this paper divided the sentiment indicators into two categories. One category is the state of public attention based on the number of readings and posts, and the other category is the state of public attitude based on the number of positive sentiment posts and negative sentiment posts.

(a) The change rate of monthly reading ($Read_t$), which mainly reflects the change in public attention to the garlic price. The larger the reading volume, the more intense the public emotion regarding the price of garlic [50]. It reflects the intensity of price fluctuation from the side. The calculation formula is shown in Formula (2).

(b) The bullish index ($BI_t$) reflects the public’s positive sentiment toward prices. The sentiment index is computed as the spread between the percentage of positive and negative [51–53]. $BI_t > 0$ indicates that the public sentiment is positive, whereas $BI_t < 0$ means that the sentiment public is negative. The calculation formula is shown in Formula (3).

\[
Read_t = \ln \frac{x_t}{x_{t-1}} \quad (2)
\]

\[
BI_t = \ln \frac{1 + pos_t}{1 + neg_t} \quad (3)
\]

Here, $x_t$ is the number of readings in month $t$, $pos_t$ is the number of positive posts in month $t$, and $neg_t$ is the number of negative posts in month $t$.

#### 3.3. Sentiment Analysis Based on NLP

Sentiment analysis is a type of data mining measuring the inclination of people’s opinions through natural language processing (NLP) [54]. It has two components of application, namely, feature extraction and valence. The structure of sentiment analysis is shown in Figure 3.

#### 3.3.1. Feature Extraction Based on NLPIR

Feature extraction consists of three main steps: word segment, stop word removal and keyword extraction. It is mainly used to extract useful information, such as topics or keywords, from a large amount of text through word segment techniques. Feature extraction is heavily reliant on dictionary analysis and aims to perform word segmentation from sentences and to clearly distinguish and classify each word form. Word segmentation in dictionary analysis for Chinese; however, is difficult due to the specific language characteristics. This paper uses the Chinese Academy of Science NLPIR Bigdata Semantic Intelligence Analytics Platform to segment the textual information of garlic network public
opinions. NLPIR is a comprehensive software based on the needs of Chinese data mining that integrates the research results of web precision collection, natural language processing, text mining and semantic search.

![Figure 3. The structure of sentiment analysis.](image)

### 3.3.2. Valence Based on Sentiment Lexicon and SVM

Since the basic sentiment lexicon is still incomplete, the generalization of sentiment words is limited [55], so it is necessary to identify some unique sentiment new words in the garlic market posts in constructing a sentiment lexicon for garlic. The support vector machine (SVM) model is mainly used for classification problems and can solve problems such as small sample size and nonlinearity, especially in solving text classification problems. Therefore, this paper proposes a sentiment classification method based on sentiment lexicon and SVM. The steps of valence based on sentiment lexicon and SVM are as follows:

1. **Step 1: Construct basic sentiment lexicon.**

   The basic sentiment lexicon was constructed based on the Chinese Sentiment Dictionary Database of Dalian University of Technology [56] through integration and optimization by removing repetitive and useless words.

2. **Step 2: Expanding sentiment words based on Word2Vec.**

   Word2Vec is a tool of training word vectors that was introduced by Google in 2013 [57]. The calculation formula is shown in Formula (4):

   \[
   \cos(w_1, w_2) = \frac{\sum_{i=1}^{n} w_{1i} w_{2i}}{\sqrt{\sum_{i=1}^{n} w_{1i}^2} \sqrt{\sum_{i=1}^{n} w_{2i}^2}}
   \]

   where \(w_1, w_2\) denote two words or phrases, and \(w_{1i}, w_{2i}\) denote their respective values taken in dimension \(i\). The \(\cos\) denotes the cosine value between the two words, where a higher value indicates a higher degree of association between the two words.

   The polarity strength of the added sentiment words was determined by the semantic similarity between the candidate words and the baseline words. The calculation formula is shown in Formula (5):

   \[
   SentiScore(\text{word}) = \max \left[ \frac{1}{N} \sum \cos(\text{word}, \text{set}) \right]
   \]

   where \(\text{word}\) denotes the added sentiment word and \(\text{set}\) denotes the set of basic sentiment words. \(SentiScore\) is the polar intensity of new words. New words with sentiment are taken
into the basic sentiment lexicon as the new sentiment lexicon, which is the domain lexicon for the garlic market.

Step 3: Sentiment classification based on SVM

The basic model of SVM is the linear classifier defined as the one with largest interval on feature space. The addition of kernel tricks can turn basic SVM into a substantially nonlinear classifier with the learning goal of finding a separating hyperplane in a feature space: \( w \cdot x + b = 0 \), where \( w \) is the normal vector and \( b \) is the intercept. The radial basis function (RBF) kernel function was used to classify the sentiment polarity of garlic network public opinions.

3.4. TVP-VAR Model

The TVP-VAR model was employed in Primiceri [58] and Nakajima [59] to detect the time-varying nature of the effects of public opinions on garlic price fluctuation. Compared with the VAR model, the TVP-VAR model can not only characterize the time-varying characteristics of the variables but also deal with abnormal changes in the variables, which enhances the stability of the estimation results.

The TVP-VAR model is derived from the same algebraic formula as the SVAR model. The SVAR model with a \( p \)-order lag term can be written in Formula (6):

\[
A_t y_t = F_1 y_{t-1} + \cdots + F_p y_{t-p} + u_t
\]

where \( y_t \) is a \((n \times 1)\) vector of observed dependent variables and \( F_1, \ldots, F_p \) are \((n \times n)\) matrices of time-varying coefficients. \( u_t \) is a structural shock with zero mean and \( u_t \sim N(0, \Sigma) \), \( A_t \) is a lower triangular matrix, and the \( \Sigma_t \) is a diagonal matrix.

Thus, both sides of Formula (6) are simultaneously multiplied by the inverse matrix of matrix \( A_t \) to obtain Formula (9):

\[
y_t = B_1 y_{t-1} + \cdots + B_s y_{t-1} + A_t^{-1} \sum_t \epsilon_t
\]

where \( \epsilon_t \sim N(0, I_k) \), \( B_i = A_t^{-1} F_i \), \( i = 1, \ldots, s \). In order to simplify the calculation, each element of the row in \( B_{ij} \) is collapsed into a \( k^2 \times 1 \) dimensional vector \( \beta \). Furthermore, \( X_t = I_n \otimes [y_{t-1}', \ldots, y_{t-p}'] \) is set, where the Kronecker product is shown by \( \otimes \) symbols. Therefore, Formula (5) can be further simplified as \( y_t = X_t \beta + A_t^{-1} \sum_t \epsilon_t \), and all parameters in the model are time-varying. Thus, the general form of TVP-VAR model is shown in Formula (10):

\[
y_t = \beta_t + A_t^{-1} \sum_t \epsilon_t
\]

Therefore, Formula (10) can be further simplified as \( y_t = X_t \beta + A_t^{-1} \sum_t \epsilon_t \), and all parameters in the model are time-varying. Thus, the general form of TVP-VAR model is shown in Formula (10):
\[ V = \text{Var} \begin{pmatrix} u_t \\ v_t \\ \xi_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} I_k & 0 & 0 & 0 \\ 0 & \Sigma_\beta & 0 & 0 \\ 0 & 0 & \Sigma_a & 0 \\ 0 & 0 & 0 & \Sigma_h \end{pmatrix} \]  

(11)

where the n-dimensional matrix \( I_k \) is an identity matrix, and \( \Sigma_\beta, \Sigma_a \) and \( \Sigma_h \) are definite positive matrices. Estimation using this method is difficult because of the large number of parameters. The MCMC algorithm and TVP-VAR model based on the Bayesian inference method can be used to overcome this problem.

4. Results and Discussions

4.1. Sentiment Analysis

Using the NLPIR big data platform for word separation and discovery, new words are added to the base dictionary. Examples of new words added to the domain sentiment lexicon are shown in Table 2.

Table 2. Example of new words added to the domain sentiment lexicon.

<table>
<thead>
<tr>
<th>Sentiment Polarity</th>
<th>Sentiment Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>rising, boom, price increase, rebounding, stability, bullish, bullish market</td>
</tr>
<tr>
<td>Negative</td>
<td>decline, plummet, downward slide, downward, crag, weakness, price pressure, trough, risk, shocks, hidden risks, stagnant sales, alerts, bear market, lackluster, plague</td>
</tr>
</tbody>
</table>

In order to demonstrate the validity of the algorithm, we compared the results with SVM, logistic regression, KNN, decision tree and random forest. In the experiment, we use the classified dataset to divide into the training set and test set in proportion to \( w \). The result with different \( w \) is presented in Figure 4.

Figure 4. Accuracy of different \( w \).

It can be seen from Figure 4 that when the proportion of training set to test set is 8:2, the classifiers of SVM, logistic regression and decision tree have the highest accuracy. When the proportion of training set to test set is 9:1, random forest has the highest accuracy. After comprehensive comparison, the training and test sets are randomly divided in the
It can be seen from Figure 4 that when the proportion of training set to test set is 8:2, the classification performance of the five classifiers was compared according to accuracy and time. The results are shown in Figure 5.

![Figure 5. Comparison of experimental results.](image)

As can be seen from Figure 4, the SVM has the highest classification accuracy and takes the least amount of time. This is mainly because SVM can classify linear and nonlinear data by adding a dimension to the feature space, and it shows better performance than other classifiers for small samples. Logistic regression models are better for classifying linear data. KNN, decision tree and random forest do not classify small sample data well and are prone to overfitting. Therefore, this paper used an SVM classifier to classify the remaining forum data.

The relationship between BI, Read and R is shown in Figure 6.

![Figure 6. Relationship between BI, Read and R.](image)
4.2. Pre-Tests for the Time Series Data

4.2.1. Stability Test of Variables

To avoid the possibility of a “false regression”, a stability test was conducted on each time series before the model estimation. The augmented Dicky–Fuller (ADF) test [60] was used to assess the stability of the public attention, public attitude and garlic price fluctuation time series, the results of which are shown in Appendix B Table A1, which shows that the public attitude, public attention and garlic price fluctuation time series were stable. As the time series were stationary, it was necessary to determine the optimal lag order. To determine the number of lags in the VAR, the model is estimated from one to six lags, and the appropriate lag with the lowest values of Akaike information criterion and Schwarz criterion is selected [61]. The result is shown in Appendix B Table A2, which shows that the optimal lag order is 1.

4.2.2. Stability Test of Model

As the time series of public opinions and garlic price fluctuation may have some structural breaks, the traditional linear model may not be stable, which may not only influence the estimation result, but also tend to cause systematic errors [62]. Therefore, before proceeding to TVP-VAR estimation, we needed to check the stability of the model. For this, we first estimated the linear version of the VAR model and then checked the stability based on the plots of recursive residuals. The recursive residuals of the linear VAR model are presented in Appendix A Figure A1, which suggest the presence of serious parameter instabilities in the linear VAR model. Therefore, the nonlinear TVP-VAR model was employed to analyze the shock effect of public attention and public attitude on garlic price fluctuation.

4.3. The Model of Garlic Price Fluctuation with Network Public Opinions

4.3.1. Estimate the Time-Varying Parameters Based on MCMC Algorithm

The MCMC algorithm is used to estimate the time-varying parameters in terms of unobserved latent variables. Following the method of Nakajuma, the MCMC algorithm is used to draw 10,000 samples and discard the initial 1000 samples to obtain the model of valid samples. Table 3 reports the standard deviations, lower and upper 95% confidence intervals and the posterior means of the selected parameters based on the MCMC estimation of TVP-VAR model. The results based on Geweke [63] show that the null hypothesis of the convergence to the posterior distribution is not rejected for the parameters at the 5% level of significance. The sample autocorrelation functions, sample paths and posterior densities for selected parameters show that the simulation produced stable and uncorrelated samples, as highlighted in Appendix A Figure A2. Convergence of the time-varying parameters was successful, as demonstrated from the diagnostic tests. Most of the inefficiency factors were also found to be low, which shows that the number of iterations is sufficient for stable estimation using the TVP-VAR model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>95%L</th>
<th>95%U</th>
<th>CD</th>
<th>Inefficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_1$</td>
<td>0.0023</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0029</td>
<td>0.539</td>
<td>8.42</td>
</tr>
<tr>
<td>$\sum_2$</td>
<td>0.0023</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0028</td>
<td>0.759</td>
<td>8.04</td>
</tr>
<tr>
<td>$\sum_1$</td>
<td>0.0057</td>
<td>0.0016</td>
<td>0.0034</td>
<td>0.0096</td>
<td>0.443</td>
<td>40.96</td>
</tr>
<tr>
<td>$\sum_2$</td>
<td>0.0053</td>
<td>0.0013</td>
<td>0.0033</td>
<td>0.0084</td>
<td>0.146</td>
<td>25.94</td>
</tr>
<tr>
<td>$\sum_1$</td>
<td>0.0059</td>
<td>0.0021</td>
<td>0.0033</td>
<td>0.0115</td>
<td>0.698</td>
<td>56.63</td>
</tr>
<tr>
<td>$\sum_2$</td>
<td>0.4616</td>
<td>0.1362</td>
<td>0.2508</td>
<td>0.7706</td>
<td>0.041</td>
<td>67.00</td>
</tr>
</tbody>
</table>

4.3.2. Time-Varying Influence of Public Attention and Public Attitude on Garlic Price Fluctuation Based on the TVP-VAR Model

(1) The analysis of stochastic volatility
Figure 7 shows the posterior mean with corresponding one-standard-deviation error bands. The plots of variables of public attention, public attitude and price fluctuation have different responses. The stochastic volatility of public attitude is maintained at a stable level, where the stochastic volatility of public attention reached its peak at the end of 2015 and then gradually tended to zero, the stochastic volatility of garlic price fluctuation peaked in June 2016, October 2019 and December 2020. The stochastic volatilities of public attention, public attitude and price fluctuation show different patterns for the same time period. Therefore, it is necessary to analyze the dynamic impulse responses of public attention and public attitude to garlic price fluctuation at the different periods and different lag intervals.

(2) Equal interval impulse responses of garlic price fluctuations to changes in network public opinions

The dynamic impulse response of public attention and public attitude to garlic price fluctuation are shown in Figure 8. It shows the impulse responses of garlic prices after being impacted by one-unit standard deviation shock of network public opinions at different lag orders, from lag order 1, 3 and 5.

Figure 8. The dynamic impulse response of public attention and public attitude to garlic price fluctuation.

From Figure 8, it can be seen that the impulse responses of garlic prices to changes in public attitudes and public attention have obvious time-varying characteristics with the change of network public opinions, indicating that they have different impacts on garlic price fluctuations according to different periods. In terms of direction, when BI receives a positive shock of one-unit standard deviation, it brings a positive response to R, while when Read receives a positive shock of one-unit standard deviation, R responds negatively. This is mainly because when public attitudes are biased toward positive sentiment, garlic mediators will raise prices to gain more profits, while excessive public attention will prompt people to plant garlic or sell garlic at low prices. Therefore, positive sentiment has contributed to the rise of garlic price, and excessive attention will inhibit the rise of garlic.
prices to some extent. In terms of intensity, with the lag order is 1, the impulse response intensity of R affected by BI is around 0.02, while the maximum impulse response intensity of R affected by Read is around $-0.006$. The impact of BI on R is stronger than that of Read, indicating that public attitude has a greater impact on garlic price fluctuation than public attention. In terms of duration, the responses of R affected by BI and Read are different in different lag periods. When the lag order is 1, the intensity of the impulse responses of R affected by both BI and Read are the largest. When the lag order is 3, the intensity of the impulse responses of R affected by both BI and Read are significantly weaker. At the lag order of 5, the impulse responses of R affected by BI and Read are essentially zero. This indicates that with the increase in lag order, the influence of BI and Read on R rapidly decreases, and both have short-term effects.

(3) Variable-interval impulse responses of garlic price fluctuations to changes in garlic network public opinion

In this paper, three time points, February 2016, August 2017 and July 2019, were selected to analyze the influence of public attitudes and public attention on garlic price fluctuation, which correspond to the three peaks of R. As shown in Figure 8, the impulse response trends of R are affected by BI and Read at different time points, which are basically consistent with the impulse response trends of R at different lag orders, but the response intensity varies over time. The results of time point impulse response of public attitude and public attention to garlic price fluctuation can be seen from Figure 9.

![Figure 9. Time point impulse response of public attitude and public attention to garlic price fluctuation.](image)

In terms of the impulse response trend, the impulse response of R affected by BI first rapidly rises and then gradually decreases to a zero value, the impulse response affected by Read first rapidly falls and then the impulse response gradually rises and eventually converges to a zero value. In term of the impulse response intensity, the impulse response of BI affected by R reaches its maximum intensity at a lag order of 1, around 0.02, and then gradually decreases to zero at a lag order of 9. The impulse response intensity of garlic price fluctuations is positive throughout the sample interval, indicating that positive sentiment is conducive to higher garlic prices. The impulse response for the change in Read reaches its maximum intensity at lag order 1, ranging from $-0.004$ to $-0.007$, and the impulse response then gradually rises to essentially zero at lag order 9. The impulse response intensity is essentially negative throughout the sample interval, indicating that public attention has a dampening influence on garlic price increases. In terms of different time points, the impulse response of R affected by BI is largest in July 2019, while in February 2016, the impulse response intensity of R is the smallest. The impulse response intensity of R affected by Read in August 2017 is largest, and the intensity is least in July 2019. This is mainly due to the rapid development of the Internet in recent years and the increasing influence of social media and news on people; hence, public attitudes have increasingly more influence on garlic price fluctuations. In addition, due to the increase in channels and resources to
access information, farmers no longer blindly follow the trend of planting garlic, so the influence of public attention on garlic price fluctuation is weaker than in previous years.

4.4. Robustness Test

To verify the robustness of the above results, the different periods of time series data were used in analysis. The fluctuation of garlic price is divided into two periods according to the principle of valley-to-valley, where the first period runs from January 2015 to July 2018 and the second runs from August 2018 to December 2020. A VAR model was used to re-estimate the influence of public attention and public attitude on garlic price fluctuation in the two fluctuation periods, and the impulse responses are shown in Figure 10. The results show that public attitude has a short-term promoting influence on garlic price fluctuation in both the first and second periods, and public attention has an inhibiting influence on garlic price fluctuation after the order lag of 3. Among them, in the second period, the influence of public attention on garlic price fluctuation is promotion prior to inhibition, although there is a discrepancy with the result of the TVP-VAR model, which is consistent with the analysis in Section 1; that is, there is a promotion influence on garlic price fluctuation when attention is low, and an increasing inhibitory influence on garlic price fluctuation as attention increases. This demonstrates that the model estimation results in this paper are highly robust, and the conclusions are credible.

![Figure 10. The impulse response of different periods.](image)

5. Conclusions

In order to study the influence of network public opinions on the price fluctuations of small agricultural products, this paper takes garlic as an example, crawling through garlic price-related information on garlic forums from January 2015 to December 2020 using web crawler technology, which included the text content and readings of individual posts, and sentiment analysis methods were used to construct numerical indicators of online public opinion, public attitude indicator (BI) and public attention indicator (Read), and the dynamic influence of public attitude and public attention on garlic price fluctuations was analyzed using a TVP-VAR model. The NLP-TVP-VAR model results support the original hypothesis that both public attitude and public attention have a short-term influence on garlic price fluctuation, and the influences differ in direction and intensity, which is also...
consistent with the conclusions in [47]. In terms of direction, the public attitude has a significant positive influence on garlic price fluctuation, whereas the public attention has an inhibiting influence. In terms of intensity, the intensity of garlic price fluctuations is more strongly affected by public attitude than public attention. In addition, the influences of public attitude and public attention on garlic price fluctuations vary at different time points. In recent years, the influence of public attitude on garlic price fluctuation has become stronger, while the influence of public attention on garlic price fluctuation has been weaker than in previous years.

Thus, to perfectly understand the factors influencing price fluctuation of small agricultural products and further increase the timeliness and accuracy of price fluctuation prediction of small agricultural product in addition to stabilize small agricultural product prices, we present the following countermeasures and suggestions:

1. The comprehensive information platform for the whole industrial chain of small agricultural products can be improved by collecting and releasing relevant information in a timely manner. A comprehensive information platform for the whole industrial chain should be constructed to realize the comprehensive collection, management and utilization of production and market circulation-related information, such as industrial dynamics, market conditions and natural disaster forecasts and scientifically respond to farmers’ rational entry or exit from the market;

2. An early warning system should be constructed for price fluctuations of small agricultural products, and corresponding measures taken in a timely manner. Since network public opinions can be crawled in real time, an early warning system for fluctuations in small agricultural product prices can be improved based on network public opinions. In addition to the conventional specific data including planted area, production, import and export prices and market prices, the value of network public opinions can be added to construct early warning indicators, which can allow more accurate predictions and provide basic support for the early warning model.

6. Limitations and Future Research

Network public opinions on small agricultural products includes not only forum information, but also news information. Different types of network public opinion information have different rates of dissemination and influence, which, in turn, have different impacts on the price fluctuations of small agricultural products. Therefore, it is necessary to further collect online public opinion information from different platforms and explore its impact on the price fluctuation of small agricultural products. In addition, we will further study the influence of public attentions and public attitudes on the price fluctuation of small agricultural products, and use network public opinion information to predict and provide early warnings on small agricultural prices, so as to provide the basis for farmers and the government to avoid market risks and, thus, more effectively guarantee the stability of small agricultural prices.

Author Contributions: Conceptualization, J.M.; methodology, Q.W.; software, X.L.; validation, X.L., J.M. and Q.W.; formal analysis, Q.W.; investigation, J.M.; resources, X.L.; data curation, J.M.; writing—original draft preparation, X.L.; writing—review and editing, Q.W.; visualization, X.L.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. 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 data that support the findings of this study are available from the correspond author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.
Appendix A

Figure A1. Recursive residuals of the linear VAR model.

Figure A2. The results of MCMC algorithm.
Appendix B

Table A1. ADF stability test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>t-Value</th>
<th>p-Value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>-5.498864</td>
<td>0.0000 ***</td>
<td>stable</td>
</tr>
<tr>
<td>Read</td>
<td>-10.75708</td>
<td>0.0001 ***</td>
<td>stable</td>
</tr>
<tr>
<td>BI</td>
<td>-7.753960</td>
<td>0.0000 ***</td>
<td>stable</td>
</tr>
</tbody>
</table>

Note: (1) The selection criteria of test form is based on minimum AIC and SC values; (2) *** is statistically significant at the level of 1%.

Table A2. Determination of the VAR model lag.

<table>
<thead>
<tr>
<th>Lag</th>
<th>LogL</th>
<th>LR</th>
<th>FPE</th>
<th>AIC</th>
<th>SC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-32.26060</td>
<td>NA</td>
<td>0.00045</td>
<td>0.808018</td>
<td>0.908375 *</td>
<td>0.847615 *</td>
</tr>
<tr>
<td>1</td>
<td>-13.5338</td>
<td>18.25724 *</td>
<td>0.000441 *</td>
<td>0.785642 *</td>
<td>1.187068</td>
<td>0.944031</td>
</tr>
<tr>
<td>2</td>
<td>-8.530128</td>
<td>8.928882</td>
<td>0.000499</td>
<td>0.908619</td>
<td>1.611114</td>
<td>1.595041</td>
</tr>
<tr>
<td>3</td>
<td>-2.499846</td>
<td>10.20509</td>
<td>0.000550</td>
<td>0.999995</td>
<td>2.003559</td>
<td>1.395665</td>
</tr>
<tr>
<td>4</td>
<td>3.890919</td>
<td>10.22522</td>
<td>0.000601</td>
<td>1.080279</td>
<td>2.384912</td>
<td>1.595041</td>
</tr>
<tr>
<td>5</td>
<td>11.49202</td>
<td>11.46012</td>
<td>0.000637</td>
<td>1.233322</td>
<td>2.729024</td>
<td>1.756875</td>
</tr>
<tr>
<td>6</td>
<td>16.26460</td>
<td>6.755039</td>
<td>0.000741</td>
<td>1.253397</td>
<td>3.151676</td>
<td>2.005740</td>
</tr>
</tbody>
</table>

Note: * represents the optimal lag period selected by the corresponding method.

References


44. Hassounieh, I.; Radwan, A.; Serra, T.; Gil, J.M. Food scare crises and developing countries: The impact of avian influenza on vertical price transmission in the Egyptian poultry sector. Food Policy 2012, 37, 264–274. [CrossRef]


