7 CAN THE FUTURES PRICE OF AGRICULTURAL PRODUCTS PREDICT THE SCALE OF CHINA'S AGRICULTURAL PRODUCTION?

Xianning WANG[1,2,*]
Xikai HUANG[1, 2]
Longkun TIAN[1,2]
Huiyan ZHOU[1,2]

Abstract

The change in the production scale of agricultural products not only affects the income of agricultural producers and the management decisions of agriculture-related enterprises, but also affects national food security; therefore, the accurate prediction of the production scale of agricultural products cannot be ignored. Agricultural futures as a financial derivative have precedence; their price fluctuation is the result of the role of multiple parties, which, to a certain extent, can respond to and affect the production of agricultural products. Based on the high-frequency characteristics of agricultural futures prices and the growth cycle of agricultural products, this paper selects the high-frequency monthly futures price data of soybean and corn as the research object and compiles the growth cycle futures price data of agricultural products, selects the mixed-frequency data regression model to predict the scale of agricultural product production, and takes the benchmark prediction model as a reference to comprehensively compare the prediction effect.The conclusions of this paper are as follows: 1. the mixed-frequency data regression model for agricultural futures prices can predict the scale of agricultural production in China, and the direct prediction using mixed-frequency data can tap the potential information contained in the high-frequency data, thus improving the prediction accuracy; 2. there is a negative effect between monthly agricultural futures prices and the related agricultural production in the period of March to May near the harvest, especially in the recent month, which is the most obvious.

Keywords: agricultural production scale; agricultural futures prices; mixed-frequency data; mixed-frequency data regression models.

JEL Classification: C53, E37, G17, Q1, C23, E23

Introduction

The agricultural futures market plays an important role in regulating and safeguarding the market to iron out risky fluctuations in supply and marketing, promote the quality and efficiency of agricultural production, and steadily realize national food security. As the future price of

[1]. *School of Economics and Management, Chongqing Normal University, No. 37 of Middle Road of University Town, Shapingba District, 401331, Chongqing, China*

[2] *Big Data Marketing Research and Applications Center, No. 37 of Middle Road of University Town, Chongqing Normal University, Shapingba District, 401331, Chongqing, China*

* *Corresponding author: Email: xianningwang@sina.com*

agricultural products, agricultural futures prices have direct and indirect guiding effects on the current scale of agricultural production. Firstly, the price transmission mechanism of agricultural production has the ability to better respond to market demand； secondly, the hedging mechanism of agricultural futures trading can reduce the loss and risk of agricultural practitioners in the case of a mismatch between supply and demand. However, although academics and the industry have long recognized the virtual and investment attributes of the agricultural futures market and the potential relationship between the physical and industrial development of agricultural production, and in particular the former's role in foreseeing and pointing to the latter in advance, there is a lack of quantitative research that takes into account the mixing of data, time span, forecasting technology, accuracy requirements, etc.Therefore, whether agricultural futures prices can predict the scale of agricultural production has become an important area worth exploring in China.

Domestic and foreign scholars are mature and rich in research on forecasting the scale of China's agricultural production, but most of them first homogenize the data of the factors to be considered, identify the relationship between the influencing factors, and establish a forecasting model and method with better accuracy. Objectively speaking, the statistical frequency of agricultural futures price data is relatively high, the statistical frequency of agricultural production scale data is relatively low, and the homogenization data processing method will lead to human intervention of data information. First, the low-frequency processing of high-frequency data will lead to the short-term fluctuation of ignoring the influencing factors;secondly, the high-frequency processing of low-frequency data will increase the human intervention in the data, and third, the excessive cleaning of data has potential impacts on the selection of the model and the calculation of accuracy. The potential impact of excessive data cleaning on model selection and accuracy calculations. The mechanism of action, data quality, and model selection are the three elements that determine the feasibility of forecasting, while the quality of data in empirical studies often gives way to the other two. Numerous studies at home and abroad can support the mechanism of agricultural futures prices on the production scale of Chinese agricultural products, and there are also many attempts at model selection. However, the research on constructing forecasting models by directly applying the mixed frequency data of agricultural futures prices and China's agricultural production scale needs to be further deepened, and the specific analysis of the differential forecasting changes of different agricultural futures needs to be further researched, which has great theoretical value and practical significance.

Different from the traditional prediction model, the mixed-frequency prediction model can not only directly incorporate the sample data of different frequencies into the same regression model and use the data overrun of high-frequency explanatory variables to predict low-frequency explanatory variables, but also flexibly select the specific form of lag order and weight function according to the sample data of low-frequency and high-frequency variables, so as to increase the prediction accuracy and improve the prediction effect. Domestic and foreign research has been applied in the economic field and has achieved better progress. For example, Clements *et al*. (2009) use monthly and quarterly data to construct a MIDAS model to predict the U.S. quarterly GDP; Zheng and Wang. (2013) predict China's economic cycle by constructing a mixed-frequency data district system transfer dynamic factor model that can comprehensively utilize China's monthly and quarterly data; Gong and Chen. (2014) use a multivariate mixed-frequency data regression model to predict China's monthly CPI; Liu *et al*. (2011) used a mixed data sampling model to forecast and predict China's quarterly GDP and empirically analyzed export is the main factor causing China's economic growth deceleration during the financial crisis. Therefore, this paper, drawing on previous studies, focuses on mixed-frequency data to improve the timeliness and accuracy of agricultural futures prices in forecasting the scale of agricultural production in China.

In order to answer the question of whether agricultural commodity futures prices can predict the production scale of Chinese agricultural commodities, this paper compiles agricultural commodity cycle futures data according to the growth cycle of agricultural commodities and tries to construct a mixed-frequency forecasting model and method for the production scale of Chinese agricultural commodities in view of the dimensions of mixed-frequency data, time span, forecasting technology, and accuracy requirements. The monthly futures market prices of two major agricultural products, soybean and corn, are selected as the research object, and the agricultural product cycle futures price data are compiled. The production scale of agricultural products is expressed by the agricultural product output, so as to empirically analyze the timeliness and accuracy of the prediction of the agricultural product production scale by agricultural product futures prices and to comprehensively compare the prediction effect by taking the benchmark prediction model as a reference.

Literature review

At present, many scholars have predicted the scale of agricultural production through multi-methods such as Bastiaanssen et al. (2003). By the high-resolution radiometer AVHRR measurements of 20 satellites in the Geographic Information System to cover the annual crop rotation cycle, predicting the entire Indus Basin along the Pakistani crop yield, the study found that the model of wheat, rice, sugarcane agricultural production prediction effect relative to the cotton, while providing better spatial details for agricultural production decision makers than traditional area-level data; De Wit *et al*. (2007) used ensemble Kalman filtering to assimilate microwave remote sensing inversions of soil moisture content into the WOFOST (world food studies) model to predict winter wheat yields in southwestern Europe, and found that the use of this assimilation method can improve regional winter wheat. Yu *et al*. (2007) combined a logistic model with a stepwise regression model to predict soybean yields and found that the combined prediction model could significantly improve the prediction accuracy. Ji and Zhu (2008) applied the climatic factors in northern China to the prediction of agricultural yields in the northern region of China, and estimated the yields of agricultural products such as maize, rice and wheat through the meteorological yield prediction method, and the prediction analysis found that the climatic factors had a significant effect on the prediction of yields.

And Simelton *et al*. (2009) used a mixed-effects model to estimate agricultural yields of maize, rice, and wheat by using meteorological yield prediction methods and found that climatic factors had a greater impact on agricultural yields. Simelton et al. used a mixed-effects model to confirm that the use of irrigation water had a serious impact on the agricultural yields of grains in China by considering the area of arable land, the number of people working in the agricultural industry, the capital used in the agricultural industry, agricultural technology, and the infrastructure of the agricultural industry. Becker *et al*. (2010) combined a daily surface reflectance dataset corrected by a bi-directional reflectance distribution function with crop statistics to develop a regression prediction model for winter wheat agricultural production in Kansas and applied the regression model directly to the prediction of winter wheat agricultural production in the Ukraine. The study found that the prediction method is simple, has limited data requirements, and is able to provide indications of pre-harvest yield shortages and surpluses of winter wheat in areas with fewer daily surface reflectance data; Xiao *et al*. (2010) developed a grey support model for peanut agricultural production. grey support vector machine combination model on peanut agricultural output, and found that the prediction accuracy of the grey support vector machine combination model was significantly higher than that of the two single models, grey prediction and support vector machine; Zhang and Zou (2011) predicted Chinese grain agricultural output through five model averaging methods, namely, S-AIC, S-BIC, MMA, JMA, and OPT, and confirmed that China's grain agricultural output was predicted based on the model averaging methods with a The prediction accuracy is high, and reasonable screening of models before combining models can improve the prediction accuracy; Chen (2013) combined the grey GM(1,1) model and Markov model to predict the grain output of Qingdao city, and the prediction results show that the grey Markov model has a high prediction accuracy for medium- and long-term prediction of the grain output which is more fluctuating; Prentovic *et al*. (2015) predicted Chinese grain agricultural output by a combination of model-averaging and model-averaging methods in an important walnut producing area（Novi Sad region in northern Serbia) to construct a walnut yield regression prediction model and used Pearson correlation analysis to study the relationship between annual walnut yield and biotic factors (e.g., airborne pollen data), and abiotic factors (e.g., meteorology), and found that walnut agricultural yield depends largely on the weather conditions, and the amount of airborne pollen also plays an important role.

In order to avoid the technical trends in agricultural yield data, Ghosh *et al*.(2015) shown that rice yield simulated using observed weather data was used as a baseline for prediction analysis through the Integrated Resource and Environment-Rice Crop Simulation Model, and it was found that the level of prediction of rice yield increased significantly with the inclusion of weather observations, and the simulation model reduced the uncertainty in the prediction of rice yield; In terms of agricultural yield and considering the influencing factors of yield, Khan *et al*.(2017) predicted the yield of agricultural products from the perspective of the relationship between agricultural research and development, and the prediction results show that there was a significant correlation between agricultural research and development and agricultural products' yield; Li (2017) used the grey GM(1,1) model as the basic method to construct a grey linear combination prediction model for the demand of Beijing's agricultural products logistics, and the study confirms that the grey linear combination model has a better fitting degree; Chen and Yu (2017) built a new dimensional unbiased grey Markov forecasting model for agricultural production, and the analysis results show that the model has a better forecasting accuracy and is suitable for medium- and long-term forecasting of agricultural production.

Most of the above scholars predict agricultural production from the aspects of scientific and technological equipment, crop growth, agricultural production, etc., and few of them use agricultural futures to predict the production of related agricultural products from the financial point of view. On the relationship between agricultural futures and spot prices, many scholars have studied how, He *et al*. (2013) used the VEC model, Granger causality test, impulse response analysis and BEKK model to empirically analyze the price discovery function and volatility spillover effect of China's cotton futures and spot market. The results of the study show that there exists a long-term equilibrium relationship and a two-way Granger-led relationship between futures prices and spot prices. However, the futures market has a stronger guiding effect on the spot market and a stronger information effect than the spot market. In addition, both markets have strong volatility lag effects and significant mutual volatility spillover effects, but the volatility spillover effect of the futures market on the spot market is significantly larger than the volatility spillover effect of the latter on the former; Li *et al*. (2018) adopted typical futures varieties in China as the samples, and use econometric modelling to analyze the relationship between futures prices and current spot prices. The study shows that for mature futures products, there is a short-term leading relationship between futures and spot prices on the basis of basically maintaining synchronous movements. Liu *et al*. (2006) conducted an empirical study on soybean varieties of the Commodity Exchange by using the cointegration test and other methods. It is concluded that the long-run equilibrium relationship between futures prices and spot prices affects the short-run price fluctuations and returns them to the long-run equilibrium state; futures prices and spot prices show strong interaction, and there exists a bidirectional Granger causality, i.e., there exists a reciprocal price-led relationship between the two.

Based on the relationship between agricultural futures prices and spot prices and the price discovery function of agricultural futures markets, this paper tries to use high-frequency monthly agricultural futures price data to predict the scale of agricultural production. Because the National Bureau of Statistics on the production and supply of agricultural products accounting for the highest frequency only annual value, and most of the agricultural products are cooked once a year, this prediction of the object of its own characteristics decided that the highest frequency can only be the annual value, but the authors of the traditional prediction model used above can not deal with the different frequencies of the variables in the face of the mixed-frequency data modelling needs to be the same frequency of the data processing, which will produce the results will be This results in a serious loss of potential information in the high-frequency data, which leads to an increase in the prediction error. In order to overcome the limitations of the traditional model and make maximum use of the inherent volatility of high-frequency data and the trend of low-frequency data, this paper chooses the mixed-frequency data regression prediction model to directly incorporate the sample data of different frequencies into the same prediction model for prediction.

Based on the excellent attributes of the mixed-frequency data regression prediction model, in recent years scholars at home and abroad have applied the mixed-frequency data regression prediction model to the prediction of various fields, Clements *et al*. (2009) used monthly and quarterly data to construct the MIDAS model to predict the quarterly GDP of the U.S.; Zheng and Wang (2013) predicted China's economic cycle by constructing the mixed-frequency data area system transfer dynamic factor model that can comprehensively use China's monthly data and quarterly data; Gong and Chen (2014) used multivariate mixed-frequency data regression model to forecast China's monthly CPI; Liu *et al*.(2011) took monthly investment, consumption, and export data to forecast quarterly GDP, and concluded that export is the main factor causing China's deceleration of economic growth in the period of financial crisis, and that the MIDAS model has a comparative advantage of accuracy in short-term forecasting of China's macroeconomic aggregates, and has significant feasibility and feasibility in real-time forecasting. The MIDAS model has a comparative advantage of accuracy in short-term forecasting of China's macroeconomic aggregates, and has significant feasibility and timeliness in real-time forecasting; Liu *et al*. (2010) used Monte Carlo simulation and China's macroeconomic fluctuation model to prove the validity of the MIDAS model in the forecasting of China's macroeconomy; Qin *et al*. (2022) utilized the MIDAS model to forecast the price of land. The use of MIDAS model by the above scholars has verified the timeliness and effectiveness of the model, but so far the use of the model is mostly focused on the economic field, and there is little prediction in the agricultural field. Based on the impact of agricultural futures prices on the supply side of agricultural products and the effectiveness and forward-looking nature of the MIDAS model, this paper aimed to use agricultural futures prices to predict the production scale of relevant agricultural products through the mixed-frequency data regression prediction model.

Regression prediction model for one-dimensional mixed-frequency data

The one-dimensional mixed-frequency data regression prediction model considers the effect of only one explanatory variable on the explanatory variables and investigates the dynamic relationship between the two. The annual production of agricultural products isY_t^Q , the monthly agricultural product futures price is$X_{d_m,t}^m$, and the monthly agricultural commodity futures price can be observed for m values in the year [t-1, t], i.e., m is the frequency multiplicative difference between the high-frequency data and the low-frequency data.

The mixed-frequency data regression forecasting model for annual production forecasts of agricultural products can be expressed as follows:

$$Y_t^Q = \alpha + \beta W(L^{\frac{1}{m}};\theta)\, X_{d_m,t}^m + \mu_t \qquad (1)$$

where α is a constant term, and β are the coefficients, and μ_t is the model error term;Y_t^Q refers to the annual production of agricultural products in period t;$X_{d_m,t}^m$ refers to the futures price of agricultural products in period t;W$\left(L^{\frac{1}{m}};\theta\right)$ is a polynomial function on the weight function $\omega_i(\theta)$ a polynomial function with the expression $W\left(L^{\frac{1}{m}};\theta\right)=\sum_{k=0}^{k}\omega\,(k;\theta)\,L^{\frac{k}{m}}$ and $L^{\frac{k}{m}}X_{d_m,t}^{m}=X_{d_m,t-\frac{k}{m}}^{m}$; θ are the parameters of the polynomial weights, k is the maximum lag order of the high-frequency explanatory variables, and k=0 denotes the monthly agricultural futures price in October and k=1 denotes the monthly agricultural futures price in the 11th month.

The mixed-frequency data sampling model with h steps forward (MIDAS(m, K, h)) has the advantage of forecasting and correcting annual data in real time compared with the basic mixed-frequency data sampling model and the traditional homoscedastic forecasting model. Generally speaking, the traditional homoscedastic forecasting model uses annual data when forecasting annual data, and the acquisition of real-time annual data has a certain time lag, MIDAS(m, K, h) can make full use of the updated monthly data to make real-time forecasts of annual data, and constantly update and correct the forecast results. When h = 1, the data from the first 3 quarters of year t and before are used to forecast the annual data in year t. When h = 4, annual data for year t can be forecast using quarterly data for year t-1 and earlier.

The more general form of the forward h-step MIDAS model is:

$$Y_t^Q=\alpha+\beta W\left(L^{\frac{1}{m}};\theta\right)X_{d_m,t-\frac{h}{m}}^{m}+\mu_t \qquad (2)$$

Since agricultural production is inertial and annual agricultural production in period t is influenced by its autoregressive period, the model should incorporate an autoregressive term. Therefore, the AR-MIDAS model is expressed as:

$$Y_t^Q=\alpha+\sum_{j=1}^{p}\gamma_j\,Y_{t-j}+\beta W\left(L^{\frac{1}{m}},\ \theta\right)X_{d_m,t}^{m}+\mu_t \qquad (3)$$

where j denotes the autoregressive order of the annual production of agricultural products and p is its maximum autoregressive order; γ_j is the effect of the respective regression period of the annual production of agricultural products on the current period.

Polynomial weights can effectively reduce the parameters to be estimated in the model, in this paper, we discuss the effects of six polynomial weights on the prediction accuracy of the model for sampling mixed-frequency data and select the optimal form of polynomial weights.

The general weight function constructed using the Beta density function takes the form:

$$\omega_i(\theta)=\omega_i(\theta_1,\ \theta_2,\ \theta_3)=\frac{f(x_i,\ \theta_1,\ \theta_2)}{\sum_{i=1}^{i^{maX}}f(x_i,\ \theta_1,\ \theta_2)}+\theta_3 \qquad \text{bb} \qquad (4)$$

where: i is the lag order of the weight function, and i^{maX} is the maximum lag order of the weight function, the i The range of variation of $i=0,\ 1,\ \ldots$ and i^{maX}, $x_i=i/i^{maX}$. $f(x_i,\ \theta_1,\ \theta_2)=x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}\Gamma(\theta_1+\theta_2)/\Gamma(\theta_1)\Gamma(\theta_2)$.$\Gamma(\theta)=\int_0^{\infty}e^{-x}x^{\theta-1}dx$.

Based on this function, respectively, take θ_3 to 0 and θ_1 as 1 to get two forms of weight function $\omega_i(\theta)$ which are the Beta weight function. One is the Beta weight function $\omega_i(\theta)$ that is, when $\theta_3=0$ when $\omega_i(\theta)=\omega_i(\theta_1,\ \theta_2)=f(x_i,\ \theta_1,\ \theta_2)/\sum_{i=1}^{i^{maX}}f(x_i,\ \theta_1,\ \theta_2)$; two is the Beta-Non-Zero weight function $\omega_i(\theta)$, i.e., when $\theta_1=1$ when $\omega_i(\theta)=\omega_i(1,\ \theta_2,\ \theta_3)=f(x_i,\ 1,\ \theta_2)/\sum_{i=1}^{i^{maX}}f(x_i,\ 1,\ \theta_2)+\theta_3$.

The Exponential Almon weight function (Exp Almon) is of the form:

$$\omega_i(\theta)=\omega_i(\theta_1,\ \theta_2,\ \ldots,\ \theta_Q)=\exp\left(\theta_1 i+\theta_2 i^2+\cdots+\theta_Q i^Q\right)/\sum_{i=1}^{i^{maX}}\exp\left(\theta_1 i+\theta_2 i^2+\cdots+\theta_Q i^Q\right) \quad (5)$$

In this paper, a three-parameter almonte index weighting function is used to make predictions about the scale of production of agricultural products.

The Almon weight function (Almon) is expressed as:

$$\beta\omega(K;\theta_0, \theta_1, \theta_2, \theta_3) = \sum_{p=0}^{3} \theta_p k^p \quad (6)$$

The segmentation function (Step) weights are of the form:

$$\beta\omega(k;\theta) = \theta_1 I_{i\in[a_0, a_1]} + \sum_{p=2}^{p} \theta_p I_{i\in[a_{p-1}, a_p]} \quad (7)$$

where $I_{i\in[a_{p-1}, a_p]}$ is a schematic function which takes the value of 1 when $a_{p-1} \ll i \ll a_p$ takes the value of 1 when, and 0 otherwise.

The Unrestricted Mixed Data Sampling Model (U-MIDAS) does not have the restriction of polynomial weights in the base model and its model can be expressed as:

$$Y_t^Q = \alpha + B(\beta, L^{\frac{1}{m}}) X_{d_m,t}^m + \mu_t \quad (8)$$

where $B(\beta, L^{\frac{1}{m}}) = \sum_{k=0}^{k} \beta_k L^{\frac{k}{m}}$, the β_k denotes the effect of each lag of monthly agricultural futures prices on current agricultural production.

Comparison of prediction effect and accuracy

4.1 Selection of data and forecasting indicators

Based on the price discovery function of the agricultural futures market and the guiding role of agricultural production, this paper first selects the monthly futures prices of the soybean index and corn index from January 2005 to December 2022 as the high-frequency independent variables of soybeans and corn, and the annual production of soybeans and corn agricultural products from 2005 to 2022 as the low-frequency dependent variables, and establishes a model for forecasting; the monthly agricultural product futures price and annual production data of corn and soybeans from 2005 to 2019 are used as estimation samples for the model, on the basis of which out-of-sample forecasts of the annual production of corn and soybeans from 2020 to 2022 are made. In the analysis process, the root mean square error (RMSE) index is used as the basis for measuring the model's strengths and weaknesses because the RMSE index can better reflect the model's prediction accuracy, and the smaller the RMSE is, the higher the model's prediction accuracy is.Then, according to the characteristics of the growth cycle of soybeans and corn, we compile the monthly data of the futures price index of the corresponding agricultural products' growth cycle, take it as the independent variable, select the corresponding annual production as the dependent variable, and establish the model for prediction.

Specifically, the monthly data compilation of the agricultural growth cycle futures price index is based on the specific time of harvesting of the corresponding agricultural products in China as the basis for the division of the compilation. Take soybeans as an example. China's soybean sowing is mainly divided into spring sowing soybeans and autumn sowing soybeans. Generally speaking, spring-sown soybeans are planted before and after Qingming and harvested in September; autumn-sown soybeans are generally planted after the wheat harvest and harvested from late September to mid-October.As the end of October approaches, the country's soybeans have been harvested. At this time, the use of a mixed-frequency data regression prediction model in accordance with the calendar year and month system, with monthly data on the current year's production forecasts to join the November-December futures price data, will lag behind the real-time forecasting and prediction accuracy, so this paper considers the current year's monthly futures price of soybeans in November as the cycle of the starting month and the next year's monthly futures price of soybeans in October as the cycle of the termination of the month. Compiling agricultural growth cycle futures price data.

The high-frequency explanatory variables of this paper, soybean index, corn index futures price monthly data, annual data source: Tongdaqin financial software, and low-frequency explanatory variables soybean and corn annual production data source: China Statistical Yearbook.

4.2 Empirical analysis

This paper attempts to use high-frequency monthly agricultural futures price data to predict the scale of agricultural production, while the national statistical calibre published on the production and supply of agricultural products in the account of the highest frequency is only the annual value. In the face of mixed-frequency data modeling, if you use the traditional forecasting model, the frequency of the data processing needs to be the same, which will result in the results of the high-frequency data of the potential information of the serious loss of the prediction error increasing. Using the MIDAS model, Liu Jinquan, Liu Han, et al. [15] demonstrated that because the mixed-frequency data model can directly use the mixed-frequency data to construct the model, it avoids the loss of information caused by data summing or interpolation and the inflated human information, maximizes the use of the existing high-frequency data information, and improves the validity of the estimation of the macro-econometric model and the accuracy of the prediction. In order to overcome the limitations of the traditional model and maximize the use of high-frequency data inherent volatility and low-frequency data trend, this paper selects the mixed-frequency data regression forecasting model to directly incorporate different frequency sample data into the same forecasting model for prediction.

(1) Analysis of mixed-frequency real-time prediction results for corn and soybeans

Different mixed-frequency data sampling models are constructed using the above six parameter weighting forms and monthly agricultural futures prices and annual agricultural production of corn and soybean, respectively. In parameter estimation, this paper uses the principle of minimizing the root mean square error (RMSE) of out-of-sample prediction to determine the optimal weighting function, the optimal lag order for high-frequency monthly agricultural futures prices, and the optimal autoregressive order for low-frequency annual agricultural production. Due to the limitation of space, only representative high-frequency lag orders and low-frequency autoregressive orders are selected for display after empirical verification. The out-of-sample prediction accuracy of each mixed-frequency data model is shown in Table 1 and Table 3.

Table 1 Out-of-sample RMSE values for different mixed-frequency models for real-time forecasting of annual production at monthly futures prices for maize

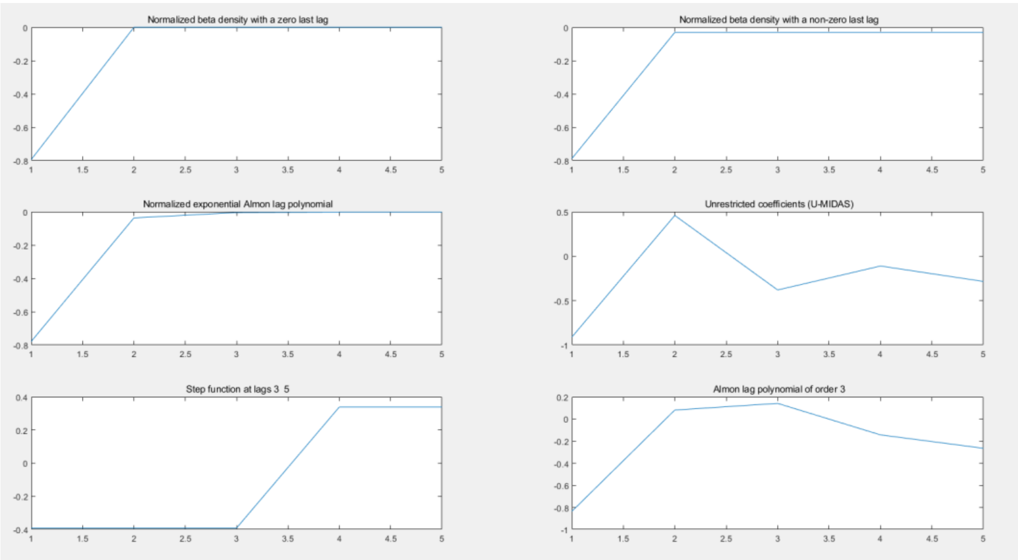
weights	Corn Monthly Futures Price Lag Order							
	3	4	5	6	7	8	9	10
	Annual maize production autoregressive order of 1							
Beta	1.0504	1.0504	1.0504	1.0504	1.0504	1.0504	1.0504	1.0504
Beta NN	1.3688	1.0991	1.0729	3.7492	0.9507	1.0438	0.9224	0.9576
Exp Almon	1.1464	1.0504	1.1474	1.7156	1.0504	1.8708	1.0504	1.7639
Almon	12.5736	1.7584	1.3556	3.0333	5.7706	0.8615	2.9468	2.1124
Stepfun	2.2184	3.5228	4.5048	3.6281	3.7793	3.3304	4.1562	4.6905
U-MIDAS	1.8117	1.7584	0.2603	0.5528	3.1838	4.6369	14.447	17.586
	Annual maize production autoregressive order of 2							
Beta	2.0321	2.0321	2.0321	2.0321	2.0321	2.0321	2.0321	2.0321

Beta NN	2.757	4.2662	2.4244	2.1845	2.3802	2.9707	3.739	1.964
Exp Almon	2.5507	2.0673	2.0474	1.8189	2.0465	2.9706	2.0321	2.1411
Almon	1.3542	4.3325	3.3515	1.7924	4.3199	1.9267	2.492	1.8004
Stepfun	2.6588	4.2662	3.6034	4.157	4.5108	4.7833	3.7139	5.413
U-MIDAS	2.2217	4.3325	2.5322	3.7301	6.1848	5.7255		

From the above table, it can be seen that the values of RMSE of the six models change with the changes in the autoregressive order of the low-frequency variable annual corn yield and the lagged order of the high-frequency variable monthly corn futures price, and it can be found that the autoregressive unrestricted mixed-frequency data sampling model (AR(1)-UMIDAS(12,5)) has a higher accuracy of predicting annual corn yield during the period of 2020–2022, when the autoregressive order P is changed to the first order and the lagged order K is changed to the fifth order. For the 2022 period, the RMSE value is 0.2603. The empirical results show that the maximum influence of annual corn production on subsequent annual corn production is one year, and the influence of monthly corn futures prices on annual corn production reaches its maximum in the fifth month and then declines.

Figure 1 shows the fluctuations of the model Beta-MIDAS, Beta Non-Zero-MIDAS, Exp Almon-MIDAS, Stepfun-MIDAS, Almon-MIDAS weights, and U-MIDAS model coefficients under the optimal autoregressive order and optimal lag order.

Figure 1: Fluctuation of five model weight functions and U-MIDAS model coefficients for maize with lag order 5



As can be seen in Figure 1, the five coefficients estimated by the AR(1)-UMIDAS(12,5) model fluctuate in the range of (-1, 0.5), i.e., there are both positive and negative multiplier effects of monthly corn futures prices on the annual corn production, and this effect lasts for five months.By comparing the various mixed-frequency data sampling models described above, it was determined that the AR(1)-UMIDAS(12,5) model has a comparative advantage in terms of out-of-sample prediction accuracy for annual maize yields, and the parameter estimation results of the model are shown in Table 2.The estimation results indicate that annual maize yields are affected by their own variations in one year and that the total effect of annual maize yields on their own is 36.2836%. The monthly futures price of corn has both positive and negative effects on annual corn production, which lasts for 5 months and is dominated by negative effects.The direction and extent of the impact of monthly corn futures prices on annual corn production are consistent with the reality in China. In the short term, because of the role of the supply and demand theorem, when the demand is relatively stable, the supply and price are in an inverse relationship; when the supply increases, the price will fall, and when the supply decreases, the price will rise. Because the futures market has precedence, the futures price of corn shows a corresponding response in the first few months of the corn harvest, which is reflected in the parameter estimation by the fact that the total effect in the last five months is negative, especially in the last one month, where the negative effect is the largest.

Table 2. Results of parameter estimation of Almon-AR(1)-MIDAS model for maize

model parameter	Estimated results
α	3.421675
β_1	-0.91016
β_2	0.458811
β_3	-0.38068
β_4	-0.11085
β_5	-0.28418
γ_1	0.362836

Notes:α Refers to the constant term of the model;β_i Refers to the effect of the monthly corn futures price lag period i on annual corn production in the current period;γ_j Refers to the effect of autoregressive period j of annual corn production on the current period.

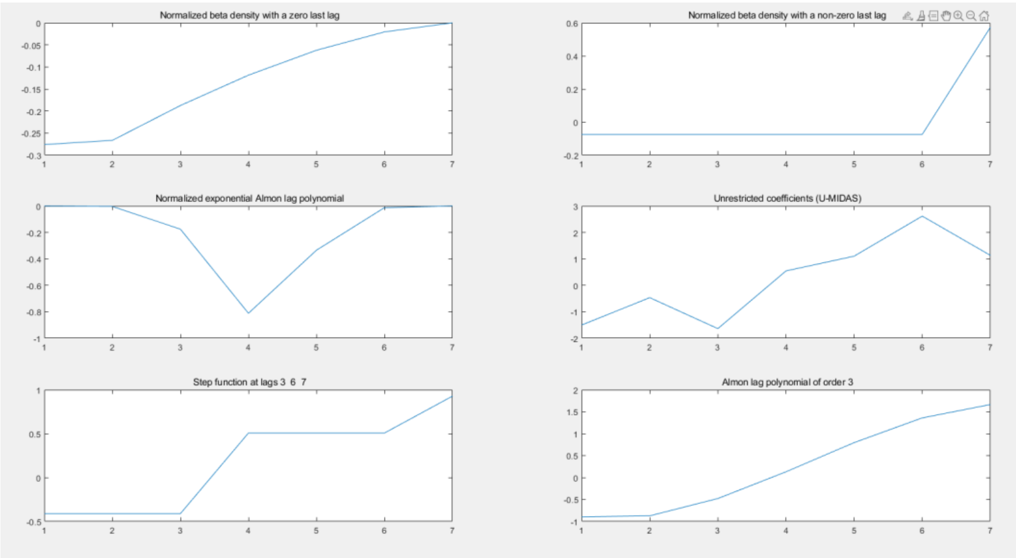
Table 3 Out-of-sample RMSE values for different mixed-frequency models for monthly soybean futures prices for real-time forecasting of annual production

weights	Soybean Monthly Futures Price Lag Order							
	3	4	5	6	7	8	9	10
	Annual soybean production autoregressive order of 1							
Beta	19.3595	19.3595	18.5131	18.0583	19.1802	19.2577	19.282	19.2911
Beta NN	19.3308	18.1055	18.9865	16.8182	21.9726	17.9658	19.3555	18.9986
Exp Almon	19.3985	18.3953	18.3039	19.3925	19.3844	18.9097	19.3784	20.9541
Almon	18.8557	22.4298	17.5793	18.435	22.3482	21.2697	28.8087	26.5025
Stepfun	17.5185	18.1055	14.8063	18.9492	19.0352	18.356	20.0845	20.9228
U-MIDAS	19.1232	22.4298	23.5919	20.9306	21.7922	21.3325	22.3765	
	Annual soybean production autoregressive order of 2							
Beta	19.846	19.846	19.846	17.6888	19.2154	19.3628	19.4094	19.4235

Beta NN	24.9919	19.5483	18.4475	18.085	19.2015	19.7569	30.5684	21.5013
Exp Almon	19.8459	19.327	18.4061	18.5066	18.5067	20.3521	23.7552	23.2828
Almon	25.8314	29.6763	17.4063	17.9145	20.7022	17.4936	27.6774	23.651
Stepfun	19.2942	19.5483	17.2902	19.5638	19.544	19.9781	22.0213	21.9643
U-MIDAS	24.9919	29.6763	29.3709	15.2208	13.4479			

The results presented in Table 3 indicate that the autoregressive unrestricted mixed-frequency data sampling model (AR(2)-UMIDAS(12,7)) has a high accuracy in predicting the annual yield of maize for the period of 2020–2022, when the autoregressive order P is varied to the 2nd order and the lagged order K is varied to the 7th order, at which time the RMSE value is 13.4479. This empirical result indicates that the maximum period of influence of annual soybean production on subsequent annual soybean production is 2 years, and the influence of monthly soybean futures prices on annual soybean production reaches its maximum in the 7th month.

Figure 2 Fluctuation of five model weight functions and U-MIDAS model coefficients for soybean lag order 7



As can be seen in Figure 2, the AR(2)-UMIDAS(12,7) model estimates seven coefficients with a range of fluctuations of (-2, 3), i.e., there are both positive and negative multiplier effects of monthly soybean futures prices on annual soybean production, and this effect lasts for seven months.

By comparing the various mixed-frequency data sampling models described above, it was determined that the AR(2)-UMIDAS(12,7) model has a comparative advantage in terms of out-of-sample prediction accuracy for annual soybean yield, and the parameter estimates of the model are shown in Table 4.The estimation results indicate that the annual yield of soybeans is affected by its own changes over a period of 2 years and that the annual yield of soybeans itself has a total impact effect of 56.3607%. The monthly futures price of soybeans has both positive and negative effects on annual soybean production, which lasts for 7 months and is predominantly positive. As in the analysis of corn, in the short term, because of the supply and demand theorem, when demand is relatively stable, supply and price are inversely related; when supply increases, prices will fall, and when supply decreases, prices will rise. Because the futures market is a precursor, the futures price of corn will show a corresponding reaction in the first few months of the corn harvest, but in the long run, or if we consider a longer period of time, the production of agricultural products is in a positive relationship with the price because the price of the relevant agricultural products will lead to an increase in the sowing area of the farmers to increase their production, and the decline in the price of the relevant agricultural products will lead to a decrease in the sowing area of the farmers to reduce their production. Thus, when monthly futures prices are lagged to the 7th order, the total effect is seen to be positive.

Table 4 Parameter estimation results of AR(2)-UMIDAS model for soybean

model parameter	Estimated results
α	2.444767
β_1	-1.50011
β_2	-0.46429
β_3	-1.63508
β_4	0.54619
β_5	1.104169
β_6	2.617954
β_7	1.134593
γ_1	0.224407
γ_2	0.3392

Notes:α Refers to the constant term of the model;β_i Refers to the effect of the monthly corn futures price lag period i on annual corn production in the current period;γ_j Refers to the effect of autoregressive period j of annual corn production on the current period.

(2) Mixed-frequency prediction of maize and soybean and analysis of baseline model results

Benchmark model is used to compare and analyse the advantages and disadvantages of the prediction of the corn AR (1)-UMIDAS (12,5) and soybean AR (2)-UMIDAS (12,7) models constructed in this paper, some simple macroeconomic forecasting models, and this paper mainly adopts the 2 models of OLS and AR as the benchmark model. Benchmark model in the process of forecasting are used in the same frequency of low-frequency data, this paper in accordance with the mixed-frequency data regression forecasting model of the time interval from 2005 to 2019 corn, soybean annual agricultural commodity futures prices and annual production data as a research sample of the benchmark model to establish the model for estimation, and out-of-sample forecasts of the annual production of corn and soybeans in 2020-2022.This paper compares the strengths and weaknesses of the corn AR(1)-UMIDAS (12,5) and soybean AR(2)-UMIDAS (12,7) models by the ratio of their forecasting accuracy (RMSE) to that of the benchmark model. As denotes the ratio of the RMSE of the maize AR(1)-UMIDAS(12,5) model to that of the corresponding benchmark model, in this paper b represents the two benchmark models, OLS and AR, respectively, and if the value is less than 1, it indicates that the prediction accuracy of the corn AR(1)-UMIDAS(12,5) model is better than the corresponding benchmark model; denotes the ratio of the RMSE of the soybean AR(2)-UMIDAS(12,7) model to the corresponding benchmark model. If the value is less than 1, then it means that the prediction accuracy of the soybean AR(2)-UMIDAS(12,7) model is better than the corresponding benchmark model. The out-of-sample forecasting accuracy (RMSE) of the benchmark model for the annual corn and soybean yields in 2020-2022 and the ratio of the RMSE of the corn AR(1)-UMIDAS (12,5) and soybean AR(2)-UMIDAS (12,7) models to the benchmark model are shown in Tables 5 and 6 below. In the comparison table of the two varieties of corn and soybean, it can be clearly seen that the prediction accuracy of the mixed-frequency data regression prediction model is significantly better than the benchmark model.

Table 5 Comparative analysis of the maize AR(1)-UMIDAS(12,5) model with the baseline model

baseline model	RMSE	$cRMSE_b$
OLS	3.228498	0.080625728
AR	3.829701	0.067968753

Table 6 Comparative analysis of the soybean AR(2)-UMIDAS(12,7) model with the benchmark model

baseline model	RMSE	$sRMSE_{\mathrm{b}}$
OLS	37.34077	0.360139869
AR	19.87546	0.67660824

(3) Analysis of the results of the forward h-step mixing short-term forecasts for corn and soybeans

On the basis of the above analysis, the corn AR(1)-UMIDAS (12,5) and soybean AR(2)-UMIDAS (12,7) models constructed in this paper have a comparative advantage in terms of out-of-sample prediction accuracy for the annual production of corn and soybean in 2020–2022. Therefore, this paper constructs forward h-step mixed-frequency data regression prediction models for corn AR(1)-UMIDAS(12,5,h) and soybean AR(2)-UMIDAS(12,7,h) on the basis of the corn AR(1)-UMIDAS(12,5) and soybean AR(2)-UMIDAS(12,7,h) models, which not only make out-of-sample predictions but also apply the latest corn and soybean production data in a timely manner. It can also apply the latest monthly agricultural futures price data for corn and soybeans in time to provide real-time reports on annual corn and soybean production and constantly update and revise the forecasts of annual corn and soybean production.This paper compares the out-of-sample forecasting accuracy of corn AR(1)-UMIDAS (12,5,h) and soybean AR(2)-UMIDAS (12,7,h) with the two benchmark models according to the above data division. From Table 7, it can be seen that when h=1 versus h=2, i.e., the prediction accuracy of the corn AR(1)-UMIDAS(12,5,1) and corn AR(1)-UMIDAS(12,5,2) models is better; from Table 8, it can be seen that when h=1, i.e., the soybean AR(2)-UMIDAS(12,7,1) model has a better prediction accuracy, which indicates that the mixed-frequency data regression forecasting model has the comparative advantage of accuracy in short-term forecasting,and Liu Han and Liu Jinquan [4] have also concluded that the MIDAS model has the comparative advantage of accuracy in short-term forecasting of China's macroeconomic aggregates when using monthly investment, consumption, and export data to forecast quarterly GDP. This implies that when applying the mixed-frequency data regression forecasting model in and out of the annual production forecasts of agricultural products, the use of the latest monthly agricultural product futures price data will improve the forecasting accuracy of the model.

Table 7 Comparative analysis of the maize AR(1)-UMIDAS (12,5,h) model with the baseline model

h	RMSE	$cRMSE_{OLS}$	$cRMSE_{AR}$
1	2.0617	0.638594	0.538345
2	2.7007	0.836519	0.705199
3	6.6516	2.060277	1.736846
4	12.991	4.023853	3.392171
5	10.2876	3.186497	2.686267
6	12.2451	3.792816	3.197404
7	8.4794	2.626423	2.214115
8	8.7195	2.700792	2.27681
9	5.6555	1.751743	1.476747
10	5.1292	1.588726	1.339321

Table 8 Comparative analysis of the soybean AR(2)-UMIDAS (12,7,h) model with the benchmark model

h	RMSE	$sRMSE_{OLS}$	$sRMSE_{AR}$
1	15.8018	0.423178	0.795041
2	28.1909	0.754963	1.418377
3	36.8071	0.985708	1.851887
4	31.8568	0.853137	1.602821
5	35.3278	0.946092	1.777458
6	31.2545	0.837007	1.572517
7	27.1578	0.727296	1.366399
8	20.5735	0.550966	1.035121
9	26.3858	0.706622	1.327557
10	29.2271	0.782713	1.470512

Main findings

This paper takes soybean and corn as examples of two agricultural products, according to the growth cycle of the agricultural products themselves, to prepare the agricultural products growth cycle futures data using a mixed-frequency data regression prediction model to empirically study whether agricultural products futures prices can predict the scale of agricultural production, compare the prediction accuracy with the baseline model, and come up with the following conclusions.

1. Agricultural commodity futures price using a mixed-frequency data regression model can foresee the scale of China's agricultural commodity production, and using mixed-frequency data for direct forecasting can explore the potential information contained in high-frequency data, thus improving forecasting accuracy.
1. 2.Monthly agricultural futures prices have both positive and negative effects on the annual production of the relevant agricultural products; specifically, monthly agricultural

futures prices have a negative effect on the annual production of the corresponding agricultural products in the period of 3-5 months before the harvest, and monthly agricultural futures prices have a positive effect on the annual production of the relevant agricultural products after May.

2. Mixed-frequency forecasting using agricultural cycle futures price data can predict the scale of agricultural production months in advance before the release of authoritative data, thus providing a data reference and decision-making basis for agricultural producers, agriculture-related enterprises, and even government agencies.

Acknowledgements

It was supported by the Chongqing Social Science Planning Doctor and Cultivation Project, Grant/Award Number: 2020PY41; Chongqing Basic Research and Frontier Exploration Project, Grant/Award Number: cstc2019jcyj msxmX0592; Humanities and Social Sciences Research Project of Chongqing Education Commission, Grant/Award Number: 23SKGH066; Project of Reform Research of Chongqing Higher Education and Teaching, Grant/Award Number: 233174; Chongqing Normal University Ph.D. Start-up Fund Project, Grant/Award Number: 19XLB001.

References

Bastiaanssen, W. and Ali, S., 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture Ecosystems and Environment*, 94, pp.321-340. https://doi.org/10.1016/S0167-8809(02)00034-8

Becker, R.I., Vermote, E., Lindeman, M. and Justice C., 2010. A generalised regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sensing of Environment*, 114, pp.1312-1323. https://doi.org/10.1016/j.rse.2010.01.010

Chen, B. and Yu, H., 2017. Agricultural product yield prediction model based on new dimensional unbiased grey Markov[J]. Electronic Science and Technology, 30, pp.30-33. https://doi.org/10.16180/j.cnki.issn1007-7820.2017.06.009

Chen, H., 2013. Grain yield prediction of Qingdao city based on grey Markov model. *Computer Simulation*, 5, pp.429-433. https://doi.org/10.3969/j.issn.1006-9348.2013.05.099

Clements, M. and Galvao, B., 2009, Forecasting US Out put Growth Using Leading Indicators: an Appraisal using MIDAS Models. *Journal of Applied Econometrics*, 24(7), pp.1187-1206. https://doi.org/10.1002/jae.1075

De Wit, A.J.W. and van Diepen, C.A., 2007. Crop model data assimilation with the Ensemble Kalman filter for imoproving regional crop yield forecast. *Agricultural and Forest Meteorology*, 146, pp.38-56. https://doi.org/10.1016/j.agrformet.2007.05.004

Ghosh, K., Singh, A., Mohanty, U.C., Acharya, N., Pal, R.K., Singhe K.K. and Pasupalak, S., 2015. Development of a rice yield prediction system over Bhubaneswar, India: combination of extended range forecast and CERES-rice model. *Royal Meteorological Society*, 22, pp.525-533.

Gong Y., Chen Q. and Zheng, X., 2014. A study on short-term forecasting of CPI based on mixed-frequency model. *Statistical Research*, 31(12), pp.25-31. https://doi.org/10.19343/j.cnki.11-1302/c.2014.12.004

He X., Zhang S., 2013. Price discovery and volatility spillover effects of cotton futures and spot market in China. *Systems Engineering Theory and Practice*, 33(07), pp.1723-1728. https://doi.org/10.3969/j.issn.1000-6788.2013.07.012

Jing, Q. and Zhu, D., 2008. Impacts of Changes in Climatic and Its Variability on Food yield in Northeast China. *Acta Agronomica Sinica*, 34, pp.1588-1597. https://doi.org/10.3321/j.issn:0496-3490.2008.09.015

Khan, F., Salim, R., Bloch, H. and Islam, N., 2017. The public R&D and productivity growth in Australia's broadacre agriculture: is there a link. *The Australian Journal of Agricultural and Resource Economics Society Inc*, 61, pp.285-303. https://doi.org/10.1111/1467-8489.12202.

Liu, F. and Liu, Y., 2006. An empirical study on the relationship between futures price and spot price volatility--Taking agricultural product soybean as an example[J]. *Finance and Trade Economics*, 08, pp.77-81. https://doi.org/10.19795/j.cnki.cn11-1166/f.2006.08.016

Liu, H. and Liu, J., 2011. Real-time forecasting and short-term forecasting of China's macroeconomic aggregates--an empirical study based on a mixed-frequency data forecasting model. *Economic Research*, 46(03), pp.4-17. https://doi.org/CNKI:SUN:JJYJ.0.2011-03-002

Liu, J., Liu, H. and Yin, C., 2010. Application of mixed-frequency data model to China's macroeconomy--an empirical study based on the MIDAS model. *Economic Science*, 179(05), pp.23-34. https://doi.org/10.19523/j.jjkx.2010.05.004

Li, X., 2017. Demand forecasting of agricultural products logistics based on grey linear combination model. *Journal of Beijing Jiaotong University*, 16, pp.120-126. https://doi.org/ 10.3969/j.issn.1672-8106.2017.01.014

Li, Y. and Mu, B., 2018. Empirical study on the relationship between commodity futures and spot price leading in China[J]. *Business Research*, 10, pp.28-33. https://doi.org/10.13902/j.cnki.syyj.2018.10.004

Prentovic, M., Radisi, P., Smith, M. and Sikoparija, B., 2015. predicting walnut (Juglans spp.) crop yield using meteorological and airborne pollen data. *Annals of Applied Biology*, 165, pp.249-259. https://doi.org/10.1111/aab.12132

Qin, M., Qi, X., Song, Y. et al., 2022. Land price prediction based on a mixed-frequency error correction model. *Statistics and Decision Making*, 38(03), pp.178-183. https://doi.org/10.13546/j.cnki.tjyjc.2022.03.033

Simelton, E., Fraser, E.D.G., Termansen, M., Forster, P.M. and Dougill, A.J., 2009. Typologies of crop-drought vulnerability ability: an empirical analysis of the socio-economic factors that influence the sensitivity and resilience to drought of three major food crops in China (1961-2001). *Environmental Science & Policy*, 12(4), pp.438-452. https://doi.org/10.1016/j.envsci.2008.11.005

Xiao, P., Ma, J., Geng, L., 2010. Yield prediction of agricultural products based on grey support vector machine combination model. *China Agricultural Mechanisation*, 1, pp.44-47. https://doi.org/10.3969/j.issn.1006-7205.2010.01.013

Yu, X., Ge, J., Liu, C., 2007. Application of combined prediction method in soybean yield prediction in Heilongjiang reclamation area. *Practice and Understanding of Mathematics*, 37, pp.27-32. https://doi.org/10.3969/j.issn.1000-0984.2007.24.005

Zhang, X.Y. and Zou, G.H., 2011. Model averaging method and its application in forecasting[J]. *Statistical Research*, 28, pp.97-102. https://doi.org/10.3969/j.issn.1002-4565.2011.06.018

Zheng T. and Wang X., 2013. Mixed-frequency data measurement and real-time analysis of China's economic cycle. *Economic Research*, 48(06), pp.58-70. doi:http://dspace.xmu.edu.cn:8080/dspace/handle/2288/56945