Asian Journal of Probability and Statistics

Volume 22, Issue 4, Page 31-40, 2023; Article no.AJPAS.100429
ISSN: 2582-0230

Combination Prediction of Income Gap between Urban and Rural Residents in China Based on IOWA Operator

Yuwan Huang [a*]

[a] *School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, China.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJPAS/2023/v22i4490

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/100429

Received: 14/03/2023
Accepted: 18/05/2023
Published: 26/05/2023

Original Research Article

Abstract

Aims: Based on the income data of urban and rural residents in China from 1998 to 2021, the income gap variables of urban and rural residents were constructed, and the combination prediction method was used to predict the income gap between urban and rural residents in China.
Methodology: Grey prediction model GM (1,1), Holt-winter seasonless exponential smoothing model and autoregressive moving average ARIMA model were used to construct an order weighted arithmetic mean combination model induced by IOWA with the minimum sum of error squares. Then, by building new weights, three individual forecasting models and combination forecasting models are used to forecast the income gap between urban and rural residents in the next five years.
Results: The results show that the accuracy of the combined prediction model is significantly better than that of the single prediction model.
Conclusion: It can be known that the income gap between urban and rural residents will widen further in the future, with an average growth rate of 4.55%.

**Corresponding author: Email: huangyw0530@163.com;*

Keywords: IOWA operator; combination prediction; Grey prediction model; ARIMA; HM model.

1 Introduction

With the continuous development of economy and society, the gap between urban and rural areas cannot be ignored. The gap between urban and rural areas can often be visually demonstrated by the gap in disposable incomes of urban and rural residents. In the past 20 years, China's urbanization rate has made a great breakthrough, reaching 64.72% in 2021. According to the data of the 7th population census, China's urban population is 902 million, and the rural population is 509 million [1]. The permanent urban population is much larger than that of rural residents. Moreover, there is still a lack of high-paying jobs in rural areas and low income of farmers in China. Development imbalance between regions and between urban and rural areas is a problem that cannot be ignored in China. Therefore, the prediction of the income gap between urban and rural residents can know its future development trend in advance, and provide reference for the future direction of government policy implementation.

2 Literature Review

At present, there are grey forecasting methods GM (1,1), ARMA, ARMIA and LSTM to predict the income gap between urban and rural residents. However, a single prediction method often has problems in the accuracy of prediction. Therefore, in order to improve the accuracy of prediction, some scholars consider combining a variety of single prediction methods to improve the accuracy and accuracy of prediction.

In the review of relevant literature, Zhang et al. [2] identified the long-term factors of urban-rural income gap through theoretical models. The time series econometrics model is used to analyze the income gap between urban and rural areas. Combined with the macroeconomic forecasting model, the trend of urban-rural income gap is predicted. Zheng et al. [3] used ARIMA model to fit the change trend of urban-rural income gap in Yunnan Province in the past, and predicted the trend of urban-rural income gap in Yunnan Province in the next three years. Gu et al. [4] took Jiangsu Province as the research object and built ARMA-GM linear combination model to forecast and analyze the income gap between urban and rural residents in the next five years. Zhang et al. [5] defined a metric to characterize the extreme deviation of the combined model. Simulation results on gold and crude oil price data show that the proposed combined prediction method can effectively combat the overfitting problem of traditional methods. Based on the SARIMA model, Li et al. [6] established the combined prediction model combined with Kalman filter (K-SARIMA) and the combined prediction model combined with BP neural network (BP-Sarima) respectively, and compared and analyzed the built models according to the performance evaluation indexes of MAPE, RMSE, RRSE and other prediction models.

At present, the theory and research of portfolio forecasting are relatively complete. For example, Tang [7] applied the method of portfolio forecasting to investment decision-making. Chen [8] established a multi-objective interval combination prediction model by using IOWA operator and correlation coefficient, and verified the effectiveness of the model through case analysis. Chen [9] proposed the induced ordered weighted geometric average (IOWGA) operator on the basis of the concept of OWGA operator, established a new combinatorial prediction model, and presented a mathematical programming method to determine the weight coefficient of IOWGA. In order to maintain the integrity of interval number and improve the prediction accuracy of interval number, Yang et al. [10] proposed an interval type combined prediction method which combined the improved correlation coefficient and the induced generalized ordered weighted multiple mean (IGOWMA) operator. Zhou [11] used three single models (stepwise regression model, GM (1,1) and exponential smoothing model) to predict the urban-rural income gap in Zhejiang Province, and then used the induced ordered weighted arithmetic mean (IOWA) combined prediction model based on different optimality criteria to predict the income gap, which has a good effect. You et al. [12] proposed a combined prediction model based on IOWA operator to forecast the grain output in our country in the next five years. Guo et al. [13] built a combined grey neural network prediction model based on IOWA operator by establishing grey prediction model and BP neural network model and taking the minimum sum of error squares as the optimization model.

This paper constructs the difference between the per capita disposable income of urban residents and the per capita disposable income of rural residents during 1998-2021 to reflect the income gap between urban and rural residents. The grey prediction GM (1,1), Holt-winter seasonless index smoothing and ARIMA were used to predict the income gap between urban and rural residents respectively. Combining with IOWA operator, the combined prediction model was constructed and compared with the single prediction. Finally, the income gap between urban and rural residents in the next five years was predicted.

3 Data Selection

The data in this paper come from China Statistical Yearbook. The urban per capita disposable income/yuan and rural per capita disposable income/yuan from 1998 to 2021 are selected, and the absolute income gap between urban and rural residents is obtained by calculating the difference between them. The trend chart is shown below.

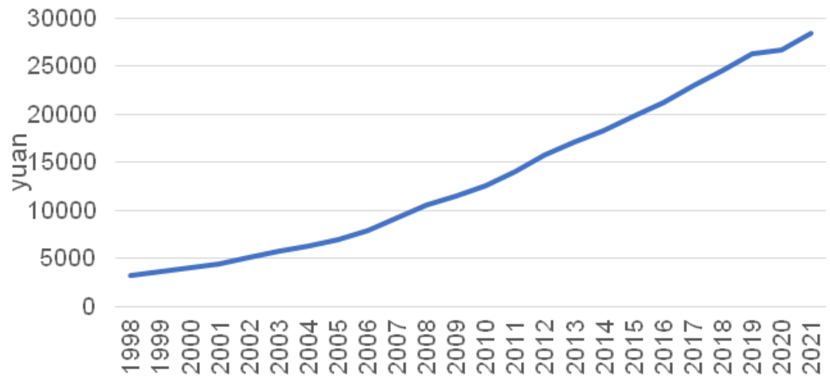

Fig. 1. Trend chart of income gap between urban and rural residents

4 Single Prediction Model Analysis

4.1 Grey prediction model

In this paper, the original sequence is the income gap between urban and rural residents. MATLAB software is used to accumulate the data to generate a new sequence, which is to whiten the gray process.

The original time series is $x = (\mathrm{x}(1), x(2), \cdots x(n))$, which is changed into the new series $y = (\mathrm{y}(1), y(2), \cdots y(n)), y(k) = \sum_{i=1}^{k} x(i)\,, k = 1,2, \cdots, n$ after accumulation processing. $Z(k), Z(k) = (y(k) + y(k+1))/2$ is generated from the adjacent mean of the new sequence y, and matrix B and vector C are constructed from $Z(k)$.

$$B = \begin{bmatrix} -\frac{1}{2}(z(1) + \mathrm{z}(2)) & 1 \\ -\frac{1}{2}(\mathrm{z}(2) + \mathrm{z}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(\mathrm{z}(\mathrm{n}-1) + z(\mathrm{n})) & 1 \end{bmatrix}, C = \begin{bmatrix} x(2) \\ \mathrm{x}(3) \\ \vdots \\ x(\mathrm{n}) \end{bmatrix}$$

The development coefficient a and gray action b can be obtained from matrix B and vector C, and the differential equation in whitening form can be established for the new sequence: $\frac{dy}{dt} + ay(\mathrm{t}) = b$, and the time response function can be obtained from a and b:

$$y(k+1) = \left(x(1) - \frac{\mathrm{b}}{\mathrm{a}}\right) e^{-\mathrm{a}k} + \frac{\mathrm{b}}{\mathrm{a}}, (k = 1,2, \cdots, n)$$

In this paper, the development coefficient a=-0.0285 and grey action b=29197.083 are calculated using MATLAB software. Therefore, the time response function is:

$$y(k+1) = 1027706.05e^{0.0285k} + 1024459.05, (k = 1,2,\cdots,n)$$

A posteriori difference ratio of 0.0056<0.3 indicates good prediction accuracy, and the prediction results are shown in Table 2.

4.2 Holt-winter has no seasonal exponential smoothing

Exponential smoothing method is a special weighted average method, which can obtain the predicted value of the next period by assigning different weights to the observed value and the predicted value. The Holt-winter non-seasonal exponential smoothing method was used by Eviwes software to predict the time series without four seasons but with time trends such as the income gap between urban and rural residents in this paper.

Using Eviwes software, the smoothing coefficient α value was 0.92, β value was 0.41, the trend item was 1356.757, and the intercept item was 28439.81. Therefore, the prediction formula is:

$$F = 1356.757k + 28439.81$$

The predicted results are shown in Table 1.

Table 1. Predicts the results

Year	Actual	GM (1,1)		Holt-winter		ARIMA	
		Predicted	**Accuracy**	**Predicted**	**Accuracy**	**Predicted**	**Accuracy**
1998	3247	3247	1.0000	3247.00	1.0000	—	—
1999	3609.8	2051.89	0.5684	4018.64	0.8867	—	—
2000	3973.54	2933.612	0.7383	4259.94	0.9279	3972.60	0.9998
2001	4417.05	3840.796	0.8695	4505.86	0.9799	4336.86	0.9818
2002	5123.48	4774.177	0.9318	4900.07	0.9564	4816.24	0.9400
2003	5715.16	5734.513	0.9966	5665.79	0.9914	5649.21	0.9885
2004	6308.27	6722.58	0.9343	6290.01	0.9971	6267.23	0.9935
2005	7012.13	7739.181	0.8963	6892.50	0.9829	6876.54	0.9807
2006	7888.72	8785.139	0.8864	7633.37	0.9676	7633.71	0.9677
2007	9275.56	9861.303	0.9369	8595.42	0.9267	8610.40	0.9283
2008	10550.5	10968.54	0.9604	10204.81	0.9672	10258.20	0.9723
2009	11465.39	12107.75	0.9440	11637.01	0.9850	11647.91	0.9841
2010	12506.6	13279.87	0.9382	12528.47	0.9983	12491.13	0.9988
2011	14033	14485.83	0.9677	13549.49	0.9655	13538.45	0.9648
2012	15737.4	15726.61	0.9993	15217.81	0.9670	15258.78	0.9696
2013	17037.4	17003.23	0.9980	17115.34	0.9954	17150.93	0.9933
2014	18354.9	18316.71	0.9979	18433.77	0.9957	18406.43	0.9972
2015	19773.1	19668.13	0.9947	19721.65	0.9974	19703.78	0.9965
2016	21252.8	21058.57	0.9909	21148.79	0.9951	21149.12	0.9951
2017	22963.7	22489.16	0.9793	22663.55	0.9869	22669.52	0.9872
2018	24633.8	23961.06	0.9727	24472.01	0.9934	24495.84	0.9944
2019	26338.1	25475.47	0.9673	26214.16	0.9953	26220.00	0.9955
2020	26702.2	27033.62	0.9876	27968.27	0.9526	27970.65	0.9525
2021	28481	28636.76	0.9945	27966.11	0.9819	27837.37	0.9774
Average accuracy			0.9492		0.9776		0.9799

4.3 ARIMA

The ARIMA model is used to predict the income gap between urban and rural residents in the next five years. Since the data is not stable, the sequential data is stable after two differences. P value =0.00<0.05 after unit root test, so d value in the ARIMA model is 2. The difference sequence and AR (1) and MA (1) were simultaneously performed least-squares regression. The significance level of AR (1) did not pass, and the T-test of MA (1) passed. The least-squares regression of the difference sequence and MA (1) passed, and $R^2 = 0.3$. Through Q test of the model, P values of adjoint probability are all greater than 0.05, so it is a white noise process. Unit root test was performed on model residuals with P value =0.0013<0.05, so model ARIMA(0,2,1) passed the fitting.

4.4 Single prediction result

It can be basically seen from Table 1 that ARIMA has a good prediction accuracy for data in this paper, with an average accuracy of 0.9799. The average accuracy of Holt-winter seasonless exponential smoothing method is 0.9776, and the grey prediction method is the lowest. However, the accuracy of the three single prediction methods is basically above 0.8, except for grey prediction, the accuracy of the actual value in 1999 is 0.59. Overall, the forecast is good.

5 IOWA Model Introduction and Analysis

5.1 IOWA operator

Let's say there are n two-dimensional arrays $((v_1, \alpha_1), (v_2, \alpha_2), \cdots, (v_n, \alpha_n))$, and $W = (w_1, w_2, \cdots, w_n)^T$ is the weighted vector, and W satisfies $\sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1,2, \cdots, n$. Let

$$IOWA_W\big((v_1, \alpha_1), (v_2, \alpha_2), \cdots, (v_n, \alpha_n)\big) = \sum_{i=1}^{n} w_i \, a_{\alpha-index(i)}$$

Then $IOWA_W$ is an n-dimensional induced ordered weighted arithmetic mean operator. $\alpha_1, \alpha_2, \cdots, \alpha_n$ is the induction variable, $\alpha - index(i)$ is induced variable from big to small order of the ith a subscript.

5.2 IOWA combinatorial prediction model

Assume that there are n prediction methods, x_t is the actual observed value in the t period, x_{it} is the predicted value in the t period of the i forecasting method, e_{it} is the absolute prediction error in the t period of the i forecasting method, α_{it} is the prediction accuracy in the t period of the i forecasting method. Among them,

$e_{it} = x_t - x_{it},\ i = 1,2, \cdots, n,\ t = 1,2, \cdots, N$. Then:

$$\alpha_{it} = \begin{cases} 1 - \left|\dfrac{x_t - x_{it}}{x_t}\right|, \left|\dfrac{x_t - x_{it}}{x_t}\right| < 1 \\ 0, \left|\dfrac{x_t - x_{it}}{x_t}\right| \geq 1 \end{cases}$$

The predicted precision α_{it} of x_{it} is taken as the induction variable, and it is arranged from large to small, so as to generate n two-dimensional arrays $((v_1, \alpha_1), (v_2, \alpha_2), \cdots, (v_n, \alpha_n))$. Given the weight vector $W = (w_1, w_2, \cdots, w_n)^T, w_i \geq 0 (i = 1,2, \cdots, n)$ satisfies $\sum_{i=1}^{n} w_i = 1$, the predicted value of orderly weighted arithmetic average combination induced by IOWA is:

$$\widehat{x_t} = \sum_{i=1}^{n} w_i x_{\alpha-index(it)}$$

Let $e_{\alpha-index(it)} = x_t - x_{\alpha-index(it)}$, $i = 1,2,\cdots,n$, $t = 1,2,\cdots,N$ be the induced prediction error, then the prediction error of IOWA combination prediction in the t period is:

$$x_t - \widehat{x_t} = x_t - \sum_{i=1}^{n} w_i x_{\alpha-index(it)}$$

$$= \sum_{i=1}^{n} w_i x_t - \sum_{i=1}^{n} w_i x_{\alpha-index(it)}$$

$$= \sum_{i=1}^{n} w_i e_{\alpha-index(it)}$$

The total prediction error sum of squares of IOWA induced ordered weighted arithmetic mean combination prediction in the sample period is:

$$U = \sum_{t=1}^{N} (x_t - \sum_{i=1}^{n} w_i x_{\alpha-index(it)})^2 = \sum_{t=1}^{N} (\sum_{i=1}^{n} w_i e_{\alpha-index(it)})^2$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j (\sum_{t=1}^{N} e_{\alpha-index(it)} e_{\alpha-index(jt)}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \bar{E}_{ij}$$

Among them, $\bar{E}_{ij} = \bar{E}_{ji} = \sum_{t=1}^{N} e_{\alpha-index(it)} e_{\alpha-index(jt)}$; $i,j = 1,2,\cdots,n$. $\bar{E} = (\bar{E}_{ij})_{n\times n}$ is denoted as IOWA's prediction error information matrix of order n.

Therefore, IOWA combined prediction model based on the squared error and minimum prediction error can be expressed as:

$$minU = W^T \bar{E} W$$

$$s.t \begin{cases} R_n^T W = 1 \\ W \geq 0 \end{cases}$$

Among them, $W = (w_1, w_2, \cdots, w_n)^T$ is the N-dimensional combination weight vector, and $R_n = (1,1,\cdots,1)^T$ is the n-dimensional vector.

5.3 Construction of IOWA composite model

Through the prediction of grey prediction model, Holt-winter seasonless exponential smoothing model and ARIMA model for the sample period, the prediction accuracy of the three models was ranked from highest to lowest with the prediction accuracy as the induction value. The results are shown in Table 2.

By calculating the sum of the squared errors of the predicted values of the three individual prediction models, the minimum prediction error information matrix of the sum of the squared errors is obtained. By bringing the error information matrix into IOWA's combined prediction model, the combined prediction model of the income gap between urban and rural residents can be obtained. The prediction error information matrix is shown as follows:

$$E = \begin{bmatrix} 1321567553 & 1342571661 & 1318556636 \\ 1342571661 & 1373919498 & 1347025868 \\ 1318556636 & 1347025868 & 1323494863 \end{bmatrix}$$

Table 2. Ranking of prediction accuracy

Year	Actual	The highest		The second		The lowest	
		Predicted	Accuracy	Predicted	Accuracy	Predicted	Accuracy
2000	3973.54	3972.60	0.9998	4259.94	0.9279	2399.61	0.7383
2001	4417.05	4336.86	0.9818	4505.86	0.9799	3840.80	0.8695
2002	5123.48	4900.07	0.9564	4816.24	0.9400	4774.18	0.9318
2003	5715.16	5734.51	0.9966	5665.79	0.9914	5649.21	0.9885
2004	6308.27	6290.01	0.9971	6267.23	0.9935	6722.58	0.9343
2005	7012.13	6892.50	0.9829	6876.54	0.9807	7739.18	0.8963
2006	7888.72	8785.14	0.8864	7633.37	0.9676	7633.71	0.9677
2007	9275.56	9861.30	0.9369	8610.40	0.9283	8595.42	0.9267
2008	10550.5	10258.20	0.9723	10204.81	0.9672	10968.54	0.9604
2009	11465.3	11637.01	0.9850	11647.91	0.9841	12107.76	0.9440
2010	12506.6	12491.13	0.9988	12528.47	0.9983	13279.87	0.9382
2011	14033	14485.83	0.9677	13549.49	0.9655	13538.45	0.9648
2012	15737.4	15726.62	0.9993	15258.78	0.9696	15217.81	0.9670
2013	17037.4	17003.24	0.9980	17115.34	0.9954	17150.93	0.9933
2014	18354.9	18316.72	0.9979	18406.43	0.9972	18433.77	0.9957
2015	19773.1	19721.65	0.9974	19703.78	0.9965	19668.13	0.9947
2016	21252.8	21149.12	0.9951	21148.79	0.9951	21058.57	0.9909
2017	22963.7	22669.52	0.9872	22663.55	0.9869	22489.16	0.9793
2018	24633.8	24495.84	0.9944	24472.01	0.9934	23961.07	0.9727
2019	26338.1	26220.00	0.9955	26214.16	0.9953	25475.48	0.9673
2020	26702.2	27033.62	0.9876	27968.27	0.9526	27970.65	0.9525
2021	28481	28636.76	0.9945	27966.11	0.9819	27837.37	0.9774

Table 3. Prediction results of IOWA combination

Year	Predicted	Accuracy	Year	Predicted	Accuracy
2000	4081.48	0.9728	2011	14131.01	0.9930
2001	4400.90	0.9963	2012	15549.33	0.9880
2002	4868.30	0.9502	2013	17045.72	0.9995
2003	5708.47	0.9988	2014	18350.71	0.9998
2004	6281.38	0.9957	2015	19714.88	0.9971
2005	6886.45	0.9821	2016	21149.00	0.9951
2006	8348.69	0.9417	2017	22667.26	0.9871
2007	9387.28	0.9880	2018	24486.81	0.9940
2008	10237.97	0.9704	2019	26217.79	0.9954
2009	11641.14	0.9847	2020	27387.80	0.9743
2010	12505.28	0.9999	2021	28382.62	0.9965

Therefore, IOWA's combined prediction model with the minimum sum of error squares is:

$$minU = W^T EW$$

$$s.t.\begin{cases} w_1 + w_2 + w_3 = 1 \\ w_1 \geq 0, w_2 \geq 0, w_3 \geq 0 \end{cases}$$

Then,

$$minU = 1321567553{w_1}^2 + 1373919498{w_2}^2 + 1323494863{w_3}^2 + 2 \times 1342571661 w_1 w_2 + 2 \times 1318556636 w_1 w_3 + 2 \times 1347025868 w_2 w_3$$

$$s.t.\begin{cases} w_1 + w_2 + w_3 = 1 \\ w_1 \geq 0, w_2 \geq 0, w_3 \geq 0 \end{cases}$$

Using LINGO software to solve linear programming, $w_1 = 0.6210576, w_2 = 0.1920936E - 06, w_3 = 0.3789422, U = 1320427000$. IOWA combined prediction model was constructed according to the weight obtained:

$$\hat{x}_t = 0.621x_{\alpha-index(1t)} + 0.379x_{\alpha-index(2t)} + 0.192E - 06x_{\alpha-index(3t)}$$

$$t = 1,2,\cdots,22$$

5.4 IOWA model evaluation

In this paper, six kinds of errors and average accuracy are used to construct an error evaluation system to evaluate the prediction results of three single prediction and combination prediction models. The error evaluation system includes: sum of squares error, average absolute error, average relative error, root mean square error, root mean square relative error, root mean square percentage error and average accuracy.

Sum of squares error: $F_{SSE} = \sum_{t=1}^{N}(x_t - \hat{x}_t)^2 = \sum_{t=1}^{N} e_t^2$

Average absolute error: $F_{MAE} = \frac{1}{N}\sum_{t=1}^{N}|x_t - \hat{x}_t| = \frac{1}{N}\sum_{t=1}^{N}|e_t|$

Average relative error: $F_{MRE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t-\hat{x}_t}{x_t}\right| = \frac{1}{N}\sum_{t=1}^{N}|\varepsilon_t|$

Root mean square error: $F_{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_t^2}$

Root mean square relative error: $F_{RMSRE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\frac{x_t-\hat{x}_t}{x_t})^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \varepsilon_t^2}$

Root mean square percentage error: $F_{MSPE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(\frac{x_t-\hat{x}_t}{x_t})^2} = \frac{1}{N}\sqrt{\sum_{t=1}^{N} \varepsilon_t^2}$

Average accuracy: $F_{MA} = \frac{1}{N}\sum_{t=1}^{N}(1 - \left|\frac{x_t-\hat{x}_t}{x_t}\right|) = \frac{1}{N}\sum_{t=1}^{N}(1 - |\varepsilon_t|)$

Table 4. Prediction effect of the model

Evaluation index	GM	H-W	ARIMA	IOWA
SSE	6380644.50	3363218.65	3399093.99	1109120.92
MAE	444.289	270.127	294.493	154.928
MRE	0.051	0.0224	0.0219	0.014
RMSE	538.544	390.991	370.755	224.532
RMSRE	0.078	0.031	0.027	0.021
MSPE	0.017	0.007	0.006	0.004
MA	0.949	0.978	0.980	0.986

It can be seen from the calculation results in the Table 4 that, under each evaluation index, the prediction error of IOWA induced ordered weighted arithmetic average combination prediction model is significantly lower than that of the other three single prediction models. For example, under the absolute square error, the error result of IOWA combined prediction model is 154.928, which is smaller than that of the other three prediction models, and the average accuracy is 0.986. Therefore, it can be considered that the prediction accuracy of the combined prediction model in this paper is better than that of other single prediction models. Therefore, the author will focus on using IOWA combination prediction model to forecast the income gap between urban and rural residents in the next five years, and the prediction result is of certain reliability.

5.5 Prediction of income gap of Chinese urban and rural residents

It can be learned from the above analysis that the prediction accuracy of the IOWA combination prediction model is higher than that of the other three individual prediction models, so we use this model to forecast the income gap of urban and rural residents of our country in the next five years

Considering that IOWA's combined prediction model relies on the actual value in the sample period and takes the prediction accuracy as the induction variable, the future prediction cannot rely on the actual value. Therefore, it is necessary to calculate a new weight to predict the income gap between urban and rural residents in the future. The idea is as follows: Firstly, the prediction accuracy of each single prediction model in each period is considered. Those with high prediction accuracy in each period are assigned large weight above, those with the second highest prediction accuracy are assigned the second largest weight, and those with the lowest prediction accuracy are assigned the least weight. Secondly, the weight given by each single forecasting method in each period is summed up to find a simple average; Finally, the final weights of three single forecasting methods are obtained. In this paper, the weight of grey prediction model is 0.2258, that of Holt-winter non-seasonal exponential smoothing model is 0.3823, and that of ARIMA (0,2,1) model is 0.3919. The prediction formula is as follows:

$$\hat{x}_t = 0.2258x_{t-GM(1,1)} + 0.3823x_{t-HW} + 0.3919x_{t-ARIMA}$$

$$t = 22,24,\cdots,26$$

Table 5. Future predicted value of income gap between urban and rural residents

Year	GM	HW	ARIMA	IOWA
2022	30286.19	29796.57	29868.51	29935.33
2023	31983.26	31153.32	31256.01	31381.00
2024	33729.33	32510.08	32643.52	32837.72
2025	35525.83	33866.84	34031.02	34305.84
2026	37374.20	35223.59	35418.53	35785.67

As can be seen from the Table 5, the income gap between urban and rural residents in China will rise rapidly in the next five years. According to the results of the combined prediction model, the average growth rate of the income gap between urban and rural residents is 4.55%, and the gap will gradually widen, reaching about 36,000 yuan in 2026.

6 Conclusion

In this paper, three single forecasting methods, namely grey forecasting model, Holt-winter non-seasonal exponential smoothing model and autoregrestic moving average ARIMA model, are selected to forecast the time series samples of the income gap between urban and rural residents, and IOWA induced orderly weighted average operator combination prediction model is constructed by three single forecasting models. The combined prediction model significantly improves the prediction accuracy of this paper, and predicts the urban-rural income gap in the next five years. The results show that the income gap between urban and rural areas in China will be further widened in the future. Therefore, to coordinate urban and rural development, it is necessary to pay attention to the problem of inter-regional development imbalance.

Acknowledgements

The work was supported by the Graduate Research Innovation Foundation of School of Statistics and Applied Mathematics, Anhui University of Finance and Economics (ACYC2021376).

Competing Interests

Author has declared that no competing interests exist.

References

[1] China Population Census Yearbook; 2020
National Bureau of Statistics. China Population Census Yearbook; 2020.

[2] Yanqun Zhang, Haiyuan Wan. Determinants and trend forecast of income gap between urban and rural residents [J]. Journal of Quantitative & Technical Economics. 2019;36(03): 59-75. (In Chinese)

[3] Xuhai Zheng, Tao Zhang. Analysis and Prediction of urban-rural Income Gap based on ARIMA Model: A Case study of Yunnan Province [J]. Rural Economics and Science and Technology, 2018, 29(15): 139-141. (In Chinese)

[4] Gu ZY. Forecast of urban-rural income gap in Jiangsu Province based on ARMA-GM combination model.] Journal of Changshu University of Technology, 2012,26(08): 18-21. (In Chinese)

[5] Yifei Zhang, Sheng Cheng, Jue Wang et al. Tradeoff between global and extreme Deviations: An improved combinatorial predictive optimization method [J/OL]. System Engineering Theory and Practice: 1-15 [2023-05-18]. (In Chinese)

[6] Panfeng Li, Zujun Ma, Hao Sun. Prediction of blood supply and demand based on SARIMA Combination prediction model [J/OL]. Industrial engineering and management: 1-18 [2023-05-18]. (In Chinese)

[7] Xiaowo Tang, Yongkai Ma, Yong Zeng, et al. Research on Modern Portfolio Forecasting and Portfolio Investment Decision Method and Its Application [M]. Beijing: Science Press; 2003. (In Chinese)

[8] Huayou Chen, Xiang Li, Lei Jin, Mengjie Yao. Prediction method of interval combination based on Correlation Coefficient and IOWA Operator [J]. Statistics and Decision. 2012;(06):83-86. (In Chinese)

[9] Huayou Chen, Zhaohan Sheng. A new combinational prediction method based on IOWGA Operator [J]. Journal of Management Engineering. 2005(04):39-42. (In Chinese)

[10] Ye Yang, Hongjun Yuan, Lingyun Hu. Based on IGOWMA operator interval type variable weight coefficient of combination forecast model [J]. Journal of Yanbian University (Natural Science Edition). 2023,49 (01):8-15. (In Chinese)

[11] Zhou TN. Combination prediction of urban-rural income gap in Zhejiang Province based on IOWA operator. Journal of Changchun Institute of Technology (Natural Science Edition). 2019;22(01):111-117-123. (In Chinese)

[12] Wenqian You, Kejun Zhuang. Prediction of grain yield combination based on IOWA operator [J]. Journal of Chongqing Technology and Business University (Natural Science Edition). 2020;37(05):80-87. (In Chinese)

[13] Lijin Guo, Jiahao Zhang. Prediction of wheat yield based on IOWA operator combined grey neural network [J]. Food and Oils. 202;35(10):26-30. (In Chinese)

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/100429