METHODOLOGIES AND APPLICATION

Influencing factors analysis and development trend prediction of population aging in Wuhan based on TTCCA and MLRA-ARIMA

Congjun Rao[1] · Yun Gao[1]

Abstract

With the rapid development of the economy, the problem of population aging has become increasingly prominent. To analyse the key factors affecting population aging effectively and predict the development trend of population aging timely are of great significance for formulating relevant policies scientifically and reasonably, which can mitigate the effects of population aging on society. This paper analyses the current situation of population aging in Wuhan of China and discusses the main factors affecting the population aging quantitatively, and then establishes a combination prediction model to forecast the population aging trend. Firstly, considering the attribute values of the primary influence factors are multi-source heterogeneous data (the real numbers, interval numbers and fuzzy linguistic variables coexist), a two-tuple correlation coefficient analysis method is proposed to rank the importance of the influencing factors and to select the main influencing factors. Secondly, a combination prediction model named Multiple Linear Regression Analysis-Autoregressive Integrated Moving Average is established to predict the number and the proportion of aging population in Wuhan. By using the statistical data of Wuhan in the past 20 years, this combination prediction model is used for empirical analysis, and a prediction result of the number and the proportion of aging people in Wuhan in the future is obtained. Based on these quantitative analysis results, we propose some countermeasures and suggestions on how to alleviate the population aging of Wuhan from aspects of economic development, pension security system design and policy formulation, which provide theoretical basis and method reference for relevant population management departments to make scientific decisions.

Keywords Population aging · Influencing factors analysis · TTCCA · MLRA-ARIMA

1 Introduction

In recent years, the proportion of aging people in the global population has been increasing year by year, and the population age structure has been aging. In China, there were two baby peaks, from 1949 to 1957, and from 1962 to 1971. At the first peak, the birth rate was between 32 and 38 per thousand. At the second peak, the birth rate was between 30 and 44 per thousand. The two baby peaks have had a long-term impact on the age structure of China's population. With the changing situation at home and abroad, China became one of the countries with a large number of aging populations in 1999. The aging age

structure of the population will change the social economy, consumption and labor force, and have a negative impact on the steady economic development in the new normal state. Population aging has a complex and long-term impact on national development and social operation (Wiener and Tilly 2002; Modigliani 2005; Lyons et al. 2018; Schön and Stähle 2020).

The population aging refers to the aging process of population age structure caused by the increasing proportion of population aged 60 (or 65)-years-old or older in the total population. Generally speaking, when the number of people aged 60 or above in a certain region or a country is more than 10% of its total population (or the proportion of people aged 65 or above is more than 7%), the country or region has entered an aging society. Population aging will bring a series of problems to individuals, families and society. For example, the increasing aging population will bring social and economic aspects of challenges, such as

✉ Congjun Rao
  cjrao@163.com

1 School of Science, Wuhan University of Technology, Wuhan 430070, People's Republic of China

demographic dividend gradually disappear, the working population will shrink, the burden coefficient of young adults will increase, medical costs will rise, elderly dependency ratio will rise, the elderly consumer demand will be increased, the average savings rate will decline and social investment will reduce. The population aging will lead to higher health care expenditures of families, which increases the demand for financial resources of families with older people and increases the economic burden on more families and family members, and the burden on families and society to provide for the aged will increase. With the aging of the population becoming more serious, the health problems of the elderly population will change, such as increased rates of cancer, bone fractures, cardiovascular diseases, depression and dementia, as well as increased disease complications, which will significantly reduce the quality of health and the level of life expectancy of the elderly population. Thus, population aging is not only a serious social problem, but also a highly valued economic problem.

As the capital of Hubei Province, the central city of central China and the new first-tier city, Wuhan has a fast-economic development, a favorable geographical location, a long history and abundant educational resources. Since 1993, the degree and scale of population aging in Wuhan was increasing and expand. Wuhan introduced a talent introduction policy named "Millions of college students stayed in Wuhan" in 2017. The implementation of this policy will have an impact on the net migration rate and density of population, cause the change of population age structure, and further affect the aging degree of population in Wuhan. As an important political, economic and cultural development center, if the problem of population aging is not properly solved, it will inevitably affect the development of Wuhan's economy in the future. According to the Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020), in 1993, the population number of Wuhan aged 60 and older and its proportion were 710,000 and 10.3%, respectively. Since the beginning of the twenty-first century, both the number and the proportion of aging population in Wuhan have shown a trend of continuous increase. In 2010, the number of aging populations in Wuhan exceeded 1 million and rose to 1.2743 million, with the proportion of aging population reaching 16%. In 2016, the number of people aged 60 or older in Wuhan rose to 1.72752 million, with the proportion of aging population rising to 20.7% over 20%. It can be seen from the above data that the number and the proportion of aging population in Wuhan are increasing year by year, and the aging degree is becoming more and more serious.

In Wuhan, the aging population shows a trend of large number and rapid increase, but the prediction of the aging population of Wuhan and the analysis of the influencing factors are not satisfactory, which will affect the government's scientific decision-making and the formulation and implementation of relevant policies to a certain extent. In addition, the level of economic development in Wuhan is not in line with the speed of population aging. At the same time, the continuous aggravation of aging will inevitably lead to the gradual shortage of labor force, which will also test the old-age security system in Wuhan. Thus, we need to make a reasonable prediction and comprehensive analysis on the aging degree and development trend of population in Wuhan, so as to provide theoretical basis and method reference for relevant population management departments to make scientific decisions.

Based on the above research background, the purpose of this paper is to select the main influencing factors of the population aging in Wuhan, predict the aging population and its proportion in the future, and provide decision-making support for the government to formulate appropriate policies in an effective and timely manner, so as to eliminate the adverse impact of the population aging on social and economic development.

2 Literature review

On the problem of population aging, many scholars have made many valuable discussions from different perspectives. They mainly focus on the development status and trend of population aging, the prediction of future aging population, the analysis of influencing factors of population aging and countermeasures, and so on (Wang 2004; Zhang 2009; Lee and Shin 2019; Li et al. 2021).

2.1 Research status on influencing factors of population aging

The research on the problem of population aging should not only be limited to the problem of population prediction, but also should trace back to the source and find the main influencing factors of population aging fundamentally, and then take this as the basis to predict and analyse the population aging, so as to formulate effective and feasible countermeasures to eliminate the impact of population aging on social and economic development. Therefore, it is very important to study the influencing factors of population aging.

In analyzing the main influencing factors of population aging, the most commonly used methods include correlation coefficient method, principal component analysis method, gray relational degree analysis method, MLRA model, and so on. Wei et al. (2018) discussed the relationship between population policy and economic growth by considering the necessity of solving the problem of

population aging. He (2011), respectively, used the gray relational model and the principal component analysis model to select the main influencing factors of population aging. Then, Chen and Hao (2014) selected the main influencing factors of population aging by constructing MLRA model. Li (2017) firstly used the Thiel index method to decompose the differences among different regions, and then, from the perspective of panel data, the estimation method of Feasible Generalized Least Squared (FGLS) was used to test the influencing factors of population aging in combination with panel data (Hu et al. 2012; Xu et al. 2017; Wang and Gan 2017; Xiao et al. 2020b).

The above analysis on the influencing factors of population aging is done under the condition that the attribute values of the influencing factors are all in the form of real numbers. However, in actual decision-making, the values of indexes that have a great impact on population aging cannot be obtained in the form of accurate real numbers. Instead, only a value range can be estimated according to the actual situation, or a rank value can be expressed by fuzzy linguistic variables (Rao and Zhao 2009). Therefore, in the actual analysis of the main factors of population aging in Wuhan, there will be the multi-source and heterogeneous data for the attribute values of primary influencing factors, i.e., the real numbers, the interval numbers and the fuzzy linguistic variables coexist. But most of the traditional analysis methods are no longer applicable to this kind of multi-source heterogeneous data with real numbers, interval numbers and fuzzy linguistic variables.

In this paper, on the basis of the attribute value types of primary influencing factors, when the attribute values are the multi-source heterogeneous data with the real numbers, interval numbers and fuzzy linguistic variables, we apply the two-tuple model (Herrera and Martínez 2000; Herrera et al. 2005; Zhang 2013; Rao et al. 2016a, b), and transform all original multi-source heterogeneous data into two-tuple data, and then propose a new method named TTCCA to select the main factors influencing the population aging in Wuhan.

The two-tuple model was first proposed by Herrera and Martínez (2000). In this model, a linguistic fuzzy variable is regarded as a continuous variable within its definitional domain and a dual combination formed by a linguistic fuzzy variable, and a real number is used to express the linguistic assessment information (Rao et al. 2017a). Subsequently, some extended models were proposed, i.e., 2-tuple hybrid ordered weighted averaging (THOWA) operator (Rao et al. 2015), interval 2-tuple linguistic VlseKriterijumska Optimizcija I Kaompromisno Resenje in Serbian (ITL-VIKOR) method (You et al. 2015), gray linguistic 2-tuple weighted averaging (GLTWA) operator (Rao et al. 2016b), 2-tuple linguistic Data Envelopment

Analysis (DEA)(Geng et al. 2017), linguistic 2-tuple gray correlation degree model (Rao et al. 2017a), 2-tuple linguistic generalized aggregation (I2LGA) operator (Liu and Chen 2018), Hesitant 2-tuple linguistic Bonferroni operators (Wang et al. 2019), 2-tuple fuzzy linguistic approach (Muhuri and Gupta 2020), two-tuple mixed correlation degree (Rao et al. 2020a), multi-source heterogeneous multi-attribute decision-making (MSHMADM) method based on the linguistic 2-tuple (Xiao et al. 2020c, d). These studies show that the data process in the two-tuple model and its extended models can effectively avoid the information loss and information distortion in the process of information gathering comparing with some existing congeneric methods. In this paper, just based on the advantages of two-tuple model in information processing, we will propose a new method named TTCCA by combining the two-tuple model with the traditional correlation coefficient analysis method to solve the selection problem of the main factors influencing the population aging in Wuhan.

2.2 Research status on development trend prediction of population aging

The research on development trend prediction of population aging mainly used some quantitative methods (Hou 2012; Wang and Liu 2012; Zhao et al. 2015; Chen 2016; Li 2016) such as gray prediction models, parameter prediction methods of regression analysis, matrix prediction method, population age shift model, time series model, neural network model, nonparametric regression method, MLRA model, support vector regression, Bayesian Hierarchical space–time model, and so on.

Gillen and Spore (1994) made the first in-depth study of population aging by assuming that the global fertility rate remains at the replacement level, thus predicted the number of global population. In 2003, Lutz et al. (2003) proposed a population age shifting model to predict the population by setting the total fertility rate, and compared the prediction results of different schemes. Jiang (2012) used the population age shifting model and the population development equation model to predict various population age indexes in China. Gray Model (1,1) (GM(1,1)), as an important prediction model, is also widely used in population prediction. Liang (2017) also used an extended gray prediction model based on GM(1,1) to predict the aging population.

The Leslie matrix prediction method is also widely used in the prediction of population aging trend. Smit et al. (2006) made a medium and long-term prediction of the elderly population by Leslie matrix prediction method, and then analyzed the future aging problem from different perspectives such as the total population, structure and aging coefficient. Liu (2016) used the Leslie matrix

prediction method to predict the size and structure of population in China. Based on this, Meng (2012) proposed a combination model of Leslie matrix and time series analysis to predict the aging population. For the study of population aging, time series models are also commonly used. For example, Sun and Wu (2015) predicted the number of future aging population and its proportion through the ARIMA, and proposed corresponding countermeasures and suggestions for solving various social problems caused by population aging. Gong et al. (2007) and Chen et al. (2014) proposed parametric autoregression models for analysis and prediction, and compared the prediction results with first-order autoregressive model (AR(1)) to improve the prediction accuracy of aging population.

The parametric regression analysis models are also often used in the study of population aging, commonly used are exponential model and Logistic regression model. The exponential model is based on linearization of the model, the ordinary least squares (OLSs) are used to estimate the parameters of the linear regression model, and then the historical data of population aging are used for fitting and predictive analysis. For example, Chen and Yu (2006), Zhu and Pang (2009) used the Logistic regression model to predict the population and obtained a good prediction effect. The method of MLRA is used to establish a model and predict the population future through the OLS according to the population data of a certain region or the whole country.

To a certain extent, the traditional prediction methods have limitations in the prediction accuracy and scope of application. Many scholars have applied the artificial neural network method to the problem of population aging. For example, Chen et al. (2012) proposed a radial neural network prediction model to solve the population prediction problem in Hunan Province and obtained a good prediction effect. In addition, Lv and Xuan (2012) predicted the aging coefficient of Beijing's population by constructing a Vector Autoregressive model (VAR).

In the above analysis of the aging population trend prediction, many studies extrapolated the time trend based on simple statistical data and simple target variables, without combining the comprehensive impact of population fertility, death, migration, medical care and other factors. Such prediction results will have an impact on the reasonable prediction of population aging. In view of this problem, this paper will establish a combination prediction model to predict the development trend of population aging in Wuhan.

The rest of this paper is organized as follows. Section 3 proposes a new TTCCA method to select the main influence factors of population aging under the information environment of multi-source heterogeneous data (the real numbers, interval numbers and fuzzy linguistic variables coexist). Section 4 establishes a combination prediction model based on MLRA and ARIMA to predict the population aging trend and the proportion of aging population, and makes an empirical analysis according to the population statistical data of Wuhan in the past 20 years. Based on these quantitative analysis results, Sect. 5 provides some countermeasures and suggestions on how to alleviate the population aging of Wuhan. Section 6 concludes the paper.

The flow chart regarding the scheme of the proposed methods in our paper is shown in the following Fig. 1.

3 Main influencing factors analysis of population aging in Wuhan based on TTCCA method

In this section, we analyse the present situation of population aging in Wuhan first, and then present a new TTCCA method based on traditional correlation coefficient analysis method to select the main influence factors of population aging.

3.1 Situation analysis of population aging in Wuhan

First of all, the population aging situation in Wuhan is statistically analyzed. By consulting the Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020), the data information of population aged 60 and above in Wuhan from 1993 to 2016 can be obtained (see Table 1). When formulating policies, the government should not only consider the number of population aged 60 and above,
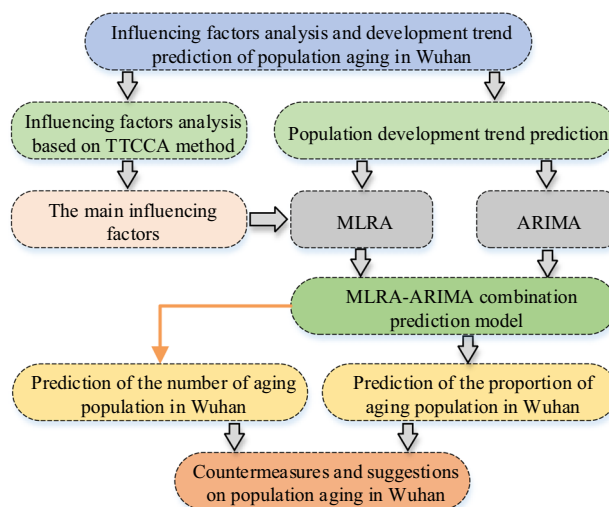

Fig. 1 Flow chart regarding the scheme of the proposed methods

Table 1 The number and the proportion of population aged 60 and above in Wuhan from 1993 to 2016
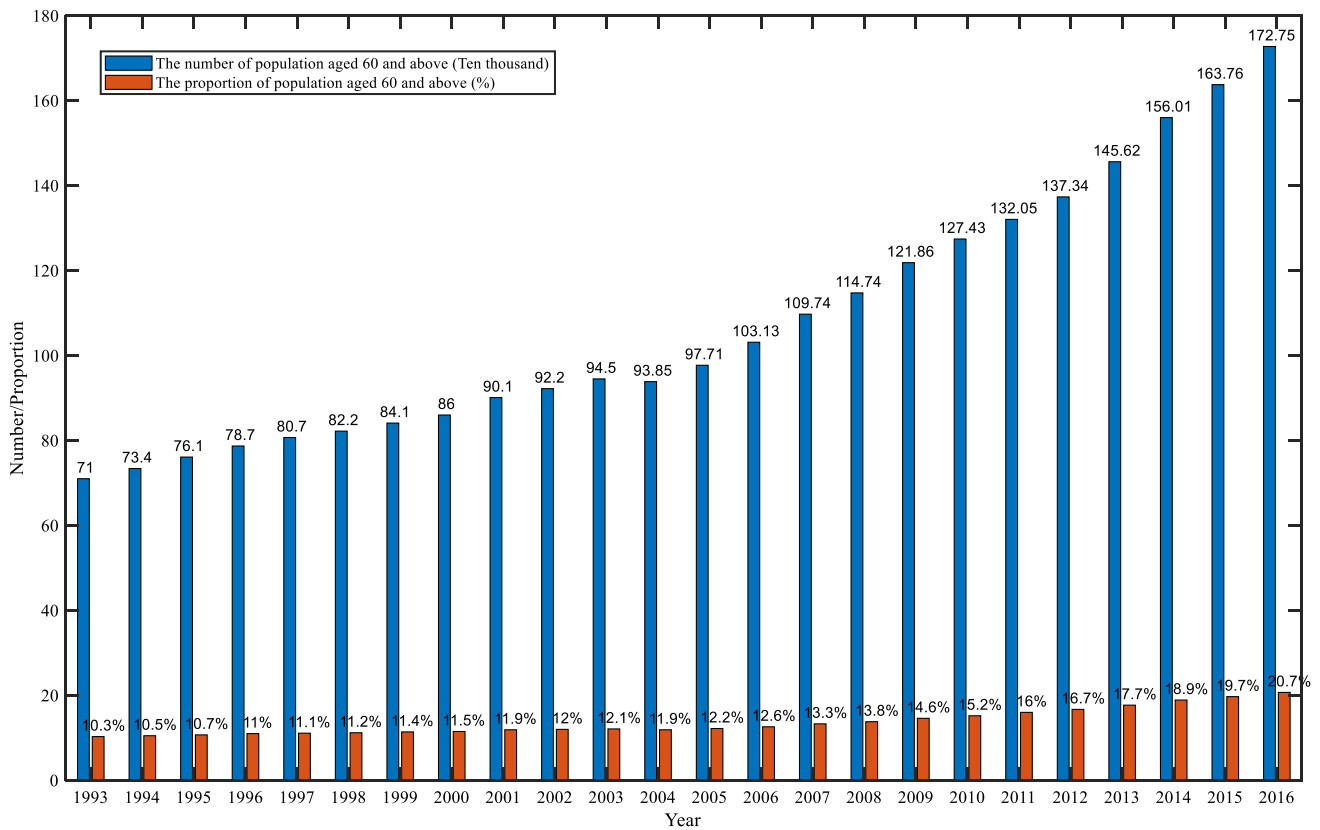
Years	The number of population aged 60 and above (ten thousand)	The proportion of population aged 60 and above (%)	Years	The number of population aged 60 and above (ten thousand)	The proportion of population aged 60 and above (%)
1993	71.000	10.3	2005	97.710	12.2
1994	73.400	10.5	2006	103.130	12.6
1995	76.100	10.7	2007	109.740	13.3
1996	78.700	11.0	2008	114.740	13.8
1997	80.700	11.1	2009	121.860	14.6
1998	82.200	11.2	2010	127.430	15.2
1999	84.100	11.4	2011	132.050	16.0
2000	86.000	11.5	2012	137.340	16.7
2001	90.100	11.9	2013	145.620	17.7
2002	92.200	12.0	2014	156.010	18.9
2003	94.500	12.1	2015	163.760	19.7
2004	93.850	11.9	2016	172.750	20.7

The data is from the Statistical Yearbook of Wuhan from 1993 to 2016 (see Wuhan Bureau of Statistics (2020))

but also consider whether the proportion of aging population is reasonable. Based on the data of population aged 60 and above in Wuhan, the proportion of population aged 60 and above in each year can be further obtained, as shown in Table 1 and Fig. 2.

As can be seen from Table 1 and Fig. 2, since Wuhan entered the aging city in 1993, the aging population in

Fig. 2 The changing trend diagram of the number and the proportion of population aged 60 and above in Wuhan

Wuhan has increased year by year, showing a rising trend. From 710,000 in 1993 to 860,000 in 2000, to 1,274,300 in 2010, and to 1,727,500 in 2016. From 1993 to 2016, the number and the proportion of population aged 60 and above in Wuhan increased continuously. Compared with 1993, the proportion of people aged 60 and above increased by 10.4% points in 2016. It is foreseeable that this proportion will continue to rise in the future, and the problem of population aging will gradually become more prominent. The number and the proportion of population aged 60 and above are rising in Wuhan, and the population development is facing the problem of population structure change. Therefore, it is necessary to further analyse the population age structure, and the results are shown in Fig. 3.

It can be seen from Fig. 3 that the population structure of Wuhan has changed significantly. In 2016, the proportion of population aged 0 to 14 in the total population was 12.07%, and that of population aged 15 to 59 was 67.23%. Compared with the sixth census in 2010, the proportion of population aged 0 to 14 in the total population increased by 2.07% points, while the proportion of population aged 60 and above rose by 5.5% points. Compared with the fifth census in 2000, the proportion of population aged 0 to 14 in the total population decreased by 5.23% points, and the

proportion of population aged 60 and above in the total population increased by 9.2% points. The proportion of population aged 15 to 59 in the total population showed a parabolic shape, which reached the highest in 2005 and is now in the declining stage, that is, the demographic dividend of Wuhan is decreasing, and it can be predicted that the proportion will continue to decline in the future stage. In general, the population age structure of Wuhan has changed significantly, and the population aging is becoming more and more serious.

In addition, the birth rate, death rate and natural growth rate are often the main influencing factors of population aging, so it is necessary to make a statistical analysis of its changing trend. From the Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020), we can get the related data of birth rate, mortality and natural growth rate from 1997 to 2016. Before 2009, the birth rate, mortality and natural growth rate fluctuate around 8‰, 6‰ and 2.5‰, respectively. But since 2010, the birth rate and natural growth rate showed an obvious upward trend. The natural growth rate fell sharply in 2004, when the death rate exceeded the birth rate because of special factors.
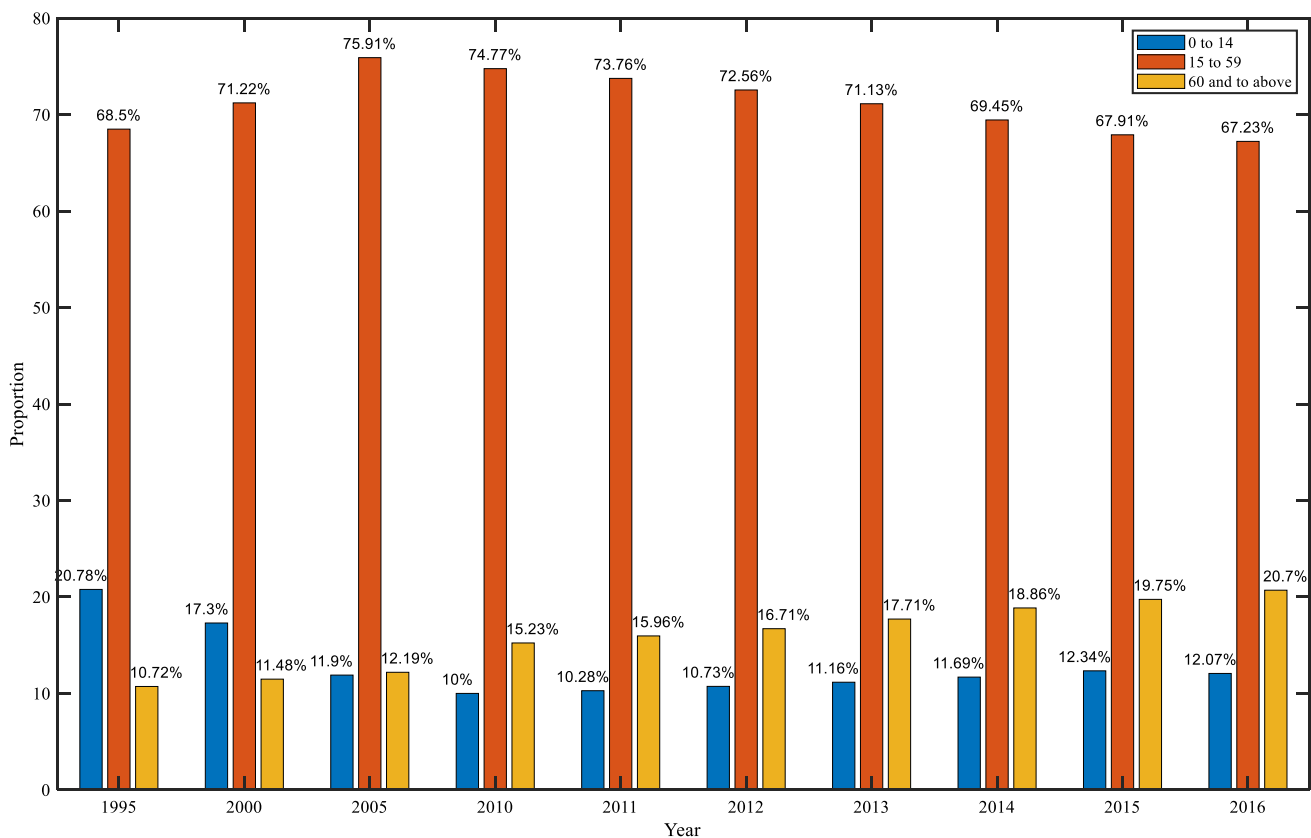

Fig. 3 The age structure of population in the main years of Wuhan

3.2 Analysis of main factors based on TTCCA method

In practical decision-making, due to the complexity of the environment, the fuzziness of human thought and the limit factors of data acquisition conditions, we cannot get accurate values expressed in real numbers for some indexes with large influence on population aging, and only estimate a value range (e.g., the average annual consumption level of residents), or use fuzzy language variables to represent a level value (e.g., physical health level, employment stability, etc.). Therefore, in the actual analysis of the main factors influencing the population aging, there will be a case that the attribute values of the preselected influencing factors are the multi-source heterogeneous data, i.e., the real numbers, interval numbers and fuzzy linguistic variables coexist. For such kind of index data, the classical statistical methods (e.g., correlation coefficient analysis, principal component analysis, etc.) cannot be directly used to extract the main influencing factors.

For the decision-making problem with multi-source heterogeneous data information, it is necessary to transform the multi-source heterogeneous data for the need of decision. Some traditional processing methods transform the fuzzy linguistic variables into real numbers, or transform all real numbers and fuzzy linguistic variables into triangular fuzzy numbers (or interval numbers, or trapezoidal fuzzy numbers), and then use the fuzzy TOPSIS method, the VIKOR method and so on to rank alternatives. The deficiency of these methods is liable to occurs the information loss or information distortion in the data conversion process. However, the two-tuple model proposed by Herrera and Martínez in 2000 can overcome this deficiency. The two-tuple regards a linguistic phrase as a continuous variable in its domain, and it is in the form of a binary form consisting of a predetermined linguistic phrase and a real value to express the linguistic assessment information integration after received all the information, which can effectively avoid information loss and distortion in the process of information gathering and operation in linguistic assessment, and the calculation precision and reliability of information processing are superior to other similar methods (Herrera and Martínez 2000; Herrera et al. 2005; Zhang 2013; Rao et al. 2017a, 2020b, c). Based on the information processing advantages of two-tuple model, this section proposes a new method named TTCCA method based on the traditional correlation coefficient analysis to select the main factors influencing the population aging in Wuhan.

3.2.1 The two-tuple model

(1)   The definition and operations of two-tuple.

Definition 1 (Herrera and Martínez 2000) A two-tuple is expressed by a binary form (s_k, a_k), where s_k and a_k are defined as follows.

(1) s_k is the k-th element in a predefined linguistic evaluation set S, where $S = \{s_0, s_1, \ldots, s_t\}$ consists of $t + 1$ linguistic fuzzy variables. If $k > l$, then $s_k > s_l$.

(2) a_k is a numerical value that represents the deviation between the evaluation result and s_k, such that $a_k \in [-0.5, 0.5)$.

Definition 2 (Herrera and Martínez 2000; Rao et al. 2017a) The operations of two-tuple (s_k, a_k) are defined as follows.

(1)   Comparison operation. For any two two-tuples (s_k, a_k) and (s_l, a_l), if $k \geq l$, then

$$(s_k, a_k) \geq (s_l, a_l).$$

(2)   Max operator and Min operator. When $(s_k, a_k) \geq (s_l, a_l)$, we have

$$\max\{ (s_k, a_k), (s_l, a_l)\} = (s_k, a_k),$$
$$\min\{(s_k, a_k), (s_l, a_l)\} = (s_l, a_l).$$

(3)   Distance operator. For any two two-tuples $A:(s_k, a_k)$ and $B: (s_l, a_l)$, the distance between A and B is defined as

$$D(A, B) = \frac{|a_k + k - l - a_l|}{l - 1}.$$

(2)   Transformation of multi-source heterogeneous data

In the actual decision-making, considering that the multi-source heterogeneous data have different dimensions and orders of magnitude, it is necessary to process the original index value information to eliminate the adverse effect of dimensions and orders of magnitude on the decision result. The basic data processing method is to standardize the original index values and then convert the standardized data into two-tuples. Next, we give the method of standardizing the original index values, and then provide the methods of how the linguistic fuzzy variables,

real numbers and interval numbers are converted into two-tuples.

(1)   Standardization of original decision matrix

Suppose that there are m preselected influencing factors affecting population aging in Wuhan. The attribute values of these m preselected influencing factors are multi-source heterogeneous data (the real numbers, interval numbers and fuzzy linguistic variables coexist). We give the time series statistical index values of each index for consecutive p years. The matrix formed by all attribute values of p years is denoted as $C = (c_{ij})_{m \times p}$. Next, we standardize the data in the matrix C. Let the matrix C be

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \ldots & c_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mp} \end{pmatrix}.$$

For the attribute values in the form of real numbers, we use the following Eq. (1) to standardize

$$z_{ij} = \frac{\max_i c_{ij} - c_{ij}}{\max_i c_{ij} - \min_i c_{ij}}. \tag{1}$$

For the attribute values in the form of interval numbers, the attribute value is denoted as $[c_{ij}^L, c_{ij}^U]$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$, we use the following Eq. (2) to standardize

$$\begin{cases} z_{ij}^L = \dfrac{c_{ij}^L}{\sum_{i=1}^n c_{ij}^U}, \\ z_{ij}^U = \dfrac{c_{ij}^U}{\sum_{i=1}^n c_{ij}^L}. \end{cases} \tag{2}$$

By using the Eq. (1), the attribute values in the form of real numbers can be standardized as $z_{ij} \in [0, 1]$, and by using the Eq. (2), the attribute values in the form of interval numbers can be standardized as $[z_{ij}^L, z_{ij}^U]$, where $z_{ij}^L \in [0, 1]$ and $z_{ij}^U \in [0, 1]$. The standardized matrix is denoted as $Z = (z_{ij})_{m \times p}$.

Next, we provide the methods of how the linguistic fuzzy variables, real numbers and interval numbers are converted into two-tuples.

(i) For a linguistic fuzzy variable $s_k \in S$, we can use the method given by the following Definition 3 to convert it into a two-tuple.

Definition 3 (Herrera and Martínez 2000) Let $s_k \in S$ be a linguistic fuzzy variable, the corresponding two-tuple can be converted by the following function θ:

$$\begin{aligned} &\theta : S \to S \times [-0.5, 0.5), \\ &\theta(s_k) = (s_k, 0). \end{aligned} \tag{3}$$

(ii) For a real number $\delta \in (0, 1)$, we can use the method given by the following Definition 4 to convert it into two-tuple.

Definition 4 (Herrera and Martínez 2000) Let $S = \{s_0, s_1, \ldots, s_t\}$ be a linguistic evaluation set, and $\delta \in (0, 1)$, then $\delta \in (0, 1)$ can be converted into an equivalent two-tuple by the following function Δ:

$$\begin{aligned} &\Delta : [0, 1] \to S \times [-0.5, 0.5], \\ &\Delta(\delta) = (s_k, a_k). \end{aligned} \tag{4}$$

such that

$$\begin{cases} k = round(\delta * t), \\ a_k = \delta * t - k. \end{cases} \tag{5}$$

where "$round$" is a rounding operation.

From Definition 4, we can use an inverse function Δ^{-1} of function Δ defined in the following Definition 5 to return a two-tuple (s_k, a_k) into the corresponding real number $\delta \in (0, 1)$.

Definition 5 (Herrera and Martínez 2000) Let $S = \{s_0, s_1, \ldots, s_t\}$ be a linguistic evaluation set, and (s_k, a_k) be a two-tuple, then there is a function Δ^{-1}, which can convert the two-tuple (s_k, a_k) into the corresponding real number $\delta \in (0, 1)$.

$$\begin{aligned} &\Delta^{-1} : S \times [-0.5, 0.5) \to [0, 1], \\ &\Delta^{-1}(s_k, a_k) = \frac{k + a_k}{t} = \delta. \end{aligned} \tag{6}$$

(iii) For an interval number $c_{ij} = [c_{ij}^L, c_{ij}^U]$, we can use the following Definition 6 to convert it into a two-tuple.

Definition 6 (Herrera and Martínez 2000; Rao et al. 2017a) Let $S = \{s_0, s_1, \ldots, s_t\}$ be a linguistic evaluation set, and $I = [a, b]$ be an interval number, the intersection of $I = [a, b]$ and the linguistic fuzzy variable s_k is denoted as

$$r_k = \max_x \min\{\mu_I(x), \mu_{s_k}(x)\}, \quad k \in [0, 1, \ldots, t], \tag{7}$$

where $\mu_I(x)$ and $\mu_{s_k}(x)$ are the membership functions of I and s_k, respectively, i.e.,

$$\mu_I(x) = \begin{cases} 1, & x \in [a, b], \\ 0, & \text{else}. \end{cases} \tag{8}$$

$$\mu_{s_k} = \begin{cases} \dfrac{x - b_k}{c_k - b_k}, & x \in [b_k, c_k], \\ 0, & \text{else}, \\ \dfrac{x - d_k}{c_k - d_k}, & x \in [c_k, d_k]. \end{cases} \tag{9}$$

Then the corresponding equivalent real number δ of (s_k, a_k) is defined as

$$\delta = \frac{\sum_{k=0}^{t-1} k * r_k}{\sum_{k=0}^{t-1} r_k}. \tag{10}$$

and we can use the method given by Definition 4 to convert δ into the corresponding two-tuple (s_k, a_k).

If we use triangular fuzzy number to express linguistic fuzzy variable s_k, that is, $s_k = (b_k, c_k, d_k)$, then we have

$$\begin{cases} b_0 = 0, \\ b_k = \dfrac{k-1}{t}, \\ c_k = \dfrac{k}{t}, \\ d_k = \dfrac{k+1}{t}, \\ d_t = 1. \end{cases} \tag{11}$$

In this paper, we set $t = 4$, then we have

$S = \{(0, 0, 0.25), (0, 0.25, 0.5), (0.25, 0.5, 0.75), (0.5, 0.75, 1), (0.75, 1, 1)\}.$

3.2.2 The TTCCA method

From the original decision matrix $C = (c_{ij})_{m \times p}$, based on the traditional correlation coefficient analysis method (Asim et al. 2019; Xie et al. 2020) and the two-tuple given by Sect. 3.2.1, we proposed a TTCCA method. The basic steps are as follows.

Step 1: Standardize the data in the original decision matrix $C = (c_{ij})_{m \times p}$. Use Eqs. (1) and (2) to standardize the real numbers and interval numbers, respectively, and the standardized decision matrix is denoted as $Z = (z_{ij})_{m \times p}$.

Step 2: Use the methods given by Definition 3 to Definition 6 to convert all data in $Z = (z_{ij})_{m \times p}$ into two-tuples, and the two-tuple matrix is denoted as $H = (h_{ij})_{m \times p}$, where $h_{ij} = (s_{ij}, a_{ij})$.

Step 3: Determine the reference sequence and the compared sequence. Take a decision element of p consecutive years as a reference sequence, which is denoted as

$h_0 = (h_0(1), h_0(2), \ldots, h_0(p))$
$= ((s_{0,1}, a_{0,1}), (s_{0,2}, a_{0,2}), \ldots, (s_{0,p}, a_{0,p})).$

In the two-tuple matrix $H = (h_{ij})_{m \times p}$, the elements in each column constitute a time series, which is called a compared sequence and is denoted as

$h_i = (h_i(1), h(2), \ldots, h_i(p))$
$= ((s_{i,1}, a_{i,1}), (s_{i,2}, a_{i,2}), \ldots, (s_{i,p}, a_{i,p})),$
$i = 1, 2, \ldots, m.$

Step 4: Calculate the two-tuple correlation coefficient

$$r_{0i} = \frac{\sum_{j=1}^{p} [\Delta^{-1}(s_{0,j}, a_{0,j}) - \bar{h}_0][\Delta^{-1}(s_{i,j}, a_{i,j}) - \bar{h}_i]}{\sqrt{\sum_{j=1}^{p} [\Delta^{-1}(s_{0,j}, a_{0,j}) - \bar{h}_0]^2} \sqrt{\sum_{j=1}^{p} [\Delta^{-1}(s_{i,j}, a_{i,j}) - \bar{h}_i]^2}}, \tag{12}$$

where $\bar{h}_i = \frac{1}{p} \sum_{j=1}^{p} \Delta^{-1}(s_{i,j}, a_{i,j})$, $\bar{h}_0 = \frac{1}{p} \sum_{j=1}^{p} \Delta^{-1}(s_{0,j}, a_{0,j})$, and Δ^{-1} is the converted operation given by Definition 5.

Step 5: Rank the influence degree of all m preselected influencing factors. According to ranking result of two-tuple correlation coefficient r_{0i} calculated in Step 4, the larger the value after taking the absolute value of r_{0i}, the greater the influence of the influencing factor i on population aging.

Similar to the traditional correlation coefficient method, when $0.7 \leq |r_{0i}| < 1$, the factor i is called highly correlated with population aging. When $0.4 \leq |r_{0i}| < 0.7$, the factor i is called moderately correlated with population aging. When $0.2 \leq |r_{0i}| < 0.4$, the factor i is called low correlation with population aging. When $|r_{0i}| < 0.2$, the correlation is called to be very low or close to zero.

In fact, from the TTCCA method given by Eq. (12), we can see that if the attribute values of all m preselected influencing factors are not multi-source heterogeneous data, but all of them are real numbers, then the Eq. (12) just become the traditional calculation formula of correlation coefficient analysis method. In other words, the TTCCA method proposed in this section is an extension and generalization of the traditional correlation coefficient analysis method.

3.3 Application case analysis

3.3.1 Data acquisition and processing

(1) Data in the form of real numbers: the main factors of influencing population aging in Wuhan from 2010 to 2015 include per capita Gross Domestic Product (GDP) (Yuan), birth rate (‰), death rate (‰), natural growth rate (‰), two-child rate (%) and population density (people/per square kilometer), where "Yuan" is the unit of Chinese currency. The index values of these influencing factors are real numbers, which can be obtained from Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020).

(2) Data in the form of interval numbers: due to the gap between the rich and the poor, the annual consumption expenditure of residents in Wuhan cannot reasonably reflect the consumption expenditure of residents in the whole city. If the influence factor of annual consumption expenditure is only used to analyse, the result of modeling will inevitably produce a big deviation. In order to make the analysis more accurate, the household consumption

expenditure is expressed by the interval number within the scope of the annual consumption expenditure.

Here, the average annual consumption level of Wuhan residents from 2010 to 2015 is converted into interval numbers (see Table 2) for analysis.

When making judgment or prediction on the problem to be studied, the judgment method based on the knowledge and experience of experts is called Delphi method (Ma et al. 2011) (also known as expert investigation method). In essence, the Delphi method is an anonymous feedback consultation method. The general process is to sort out, summarize and make statistics after obtaining the opinions of experts on the issues to be decided, and then give anonymous feedback to all experts and solicit opinions again, focus opinions and give feedback again, until the consensus is reached. The advantage of this method is that everyone's views will be collected, which can avoid some possible disadvantages of group decision-making, so that people with special status have no chance to control the will of the group, and can give full play to the role of all experts. Thus, in this paper, the Delphi method is used to determine the index values for the qualitative indexes, which cannot be determined in the form of exact real numbers.

(3) Data in the form of fuzzy linguistic variables: Considering that the physical condition and the employment stability degree are two qualitative indexes, so the Delphi method can be used to determine their index values. The decision-making process of Delphi method is as follows:

Step 1: Specify the requirements and decision objectives, and inform the selected experts in writing. The selected experts must master the knowledge of the specific field. The number of experts is generally 8–20.

Step 2: Each expert makes his own decision based on his own experience and knowledge, explains the reasons and basis of the decision in writing, and gives a written reply to the person that is in charge of the decision-making.

Step 3: The person in charge of decision-making summarized and sorted out the experts' decision-making opinions, and, respectively, explained the basis and reasons

Table 2 The average annual consumption level of residents in Wuhan

Years	The average annual consumption level	The interval number for the average annual consumption level
2010	14,490.070	[12,000, 15,000]
2011	17,140.960	[15,001, 18,000]
2012	18,813.140	[18,001, 20,000]
2013	20,157.320	[20,001, 22,000]
2014	22,002.000	[22,001, 24,000]
2015	24,943.000	[24,001, 26,000]

Table 3 The evaluation decision sheet for physical condition given by experts

Related variables	Medical care expenditure	Life insurance expenditure	Residents' minimum living allowance
Values of related variables	a_{ij}	b_{ij}	c_{ij}
Expert's comprehensive assessment value	s_k		

of different decision-making values, and then returned to the experts, and asked the experts to modify the original decision.

Step 4: When the experts make the second decision-making, they should analyse the reasons and basis for the second decision according to various decision-making opinions.

Repeat above steps, all experts revise their decision result again and again until they are basically in agreement.

Now we use above Delphi method to determine the evaluation values of several qualitative indexes.

Among the influencing factors of population aging in Wuhan, the index values of physical condition and employment stability degree cannot be expressed in real numbers, and generally evaluated in the form of fuzzy linguistic variables such as "good", "average", "poor" and so on (Rao et al. 2015, 2017b; Xiao et al. 2020a; Peng et al. 2020; Tian et al. 2019, 2020). Here the Delphi method can be used to determine the index values of these two factors. Because the physical condition is related to the variables such as medical care expenditure, life insurance expenditure, residents' minimum living allowance, and so on. And these variables can be quantified by the exact real numbers. In practical decision-making, each invited expert will give the comprehensive assessment value s_k according to the values of medical care expenditure (a_{ij}), life insurance expenditure (b_{ij}) and residents' minimum living allowance (c_{ij}). Thus, the expert's evaluation decision sheet (see Table 3) can be obtained. The evaluation value is generally divided into 5 grades, that is, very good, good, average, poor, very poor. Summarize and sort out the evaluation results of all experts, and send them to all the experts again. Let all experts revise their original decision according to the overall evaluation results, and provide their revised decision-making opinions and their basis and reasons. After repeated consultation, induction and modification, the process is over until all the experts are basically in agreement. The final decision result is denoted as $s_k, k = 0, 1, \ldots, 4$, where the corresponding fuzzy linguistic variables are "very poor", "poor", "average", "good" and "very good".

The employment stability degree is related to the number of new jobs, the unemployment rate and the quantity of employment. And these three variables can be quantified by the exact real numbers. Similar to Table 3, we give the expert's evaluation decision sheet (see Table 4) on employment stability degree. Each invited expert will give the comprehensive assessment value s_h based on their own knowledge and experience according to the values of the number of new jobs (x_{ij}), the unemployment rate (y_{ij}) and the quantity of employment (z_{ij}). The decision-making process of applying Delphi method is similar to the decision-making process of physical condition given above. The final decision result is denoted as s_k, $k = 0, 1, \ldots, 4$, where the corresponding fuzzy linguistic variables are "very low", "low", "average", "high" and "very high".

3.3.2 Decision-making process

From Sect. 3.3.1, the index values of per capita GDP (Yuan), birth rate (‰), death rate (‰), natural growth rate (‰), two-child rate (%) and population density (people/per square kilometer) are all real numbers, the index values of average annual consumption level of residents are interval numbers, and the index values of physical condition and employment stability degree are fuzzy linguistic variables given by Delphi method. All index values are listed in Table 5.

The specific decision-making process based on TTCCA method is as follows.

(1) Convert the multi-source heterogeneous data matrix into two-tuple matrix. By using the transformation methods given by Definition 3 to Definition 6, we can covert all multi-source heterogeneous data in Table 5 into two-tuples, and the two-tuple matrix are given in Table 6.

(2) Determine the reference sequence and the compared sequence.

Use the aging population (A_0) in six consecutive years as a reference sequence, the values of per capita GD (A_1), birth rate (A_2), death rate (A_3), natural growth rate (A_4), two-child rate (A_5), population density (A_6), average annual consumption level of residents (A_7), physical condition (A_8) and employment stability degree (A_9) in six consecutive years as 9 compared sequences.

(3) Calculate the two-tuple correlation coefficient between compared sequence $h_i(j)(i = 1, 2, \ldots, 9)$ and reference sequence $h_0(j)$.

By using the Eq. (12) in Sect. 3.2.2, we obtain the two-tuple correlation coefficient as follows.

$r_{01} = 0.980$, $r_{02} = 0.983$, $r_{03} = -0.718$,
$r_{04} = 0.901$, $r_{05} = 0.937$, $r_{06} = 0.002$,
$r_{07} = -0.961$, $r_{08} = -0.180$, $r_{09} = -0.109$.

(4) Ranking of schemes.

By Step (3), the ranking result of the correlation coefficient obtained after taking the absolute value is as follows.

$|r_{02}| > |r_{01}| > |r_{07}| > |r_{05}| > |r_{04}| > |r_{03}| > |r_{08}| > |r_{09}| > |r_{06}|$.

Thus, the ranking result of the influence degree for all m preselected influencing factors is as follows.

birth rate $\succ$ per capita GD $\succ$ average annual consumption level of residents $\succ$ two-child rate $\succ$ natural growth rate $\succ$ death rate $\succ$ physical condition $\succ$ employment stability degree $\succ$ population density

According to above ranking result of the 9 influencing factors, the birth rate has the largest influence on the population aging in Wuhan, followed by the per capita GDP of Wuhan, and then followed by average annual consumption level of residents and two-child rate, and the population density has the least influence on the population aging in Wuhan. Obviously, $0.9 < |r_{0i}| < 1$, $i = 1, 2, 4, 5, 7$, is satisfied, which means the first five factors, i.e., birth rate, per capita GD, average annual consumption level of residents, two-child rate and natural growth rate, are highly relevant to population aging in Wuhan. Thus, these five factors are selected as the main factors of influencing population aging in Wuhan.

4 Combination prediction of population aging trend in Wuhan based on MLRA-ARIMA

In this section, the MLRA model and ARIMA are integrated, and combined with the weighted combination idea, a combination prediction model is established to predict the

Table 4 The evaluation decision sheet for employment stability degree given by experts

Related variables	The number of new jobs	The unemployment rate	The quantity of employment
Values of related variables	x_{ij}	y_{ij}	z_{ij}
Expert's comprehensive assessment value	s_h		

Table 5 The original index values information

Year	Aging population (ten thousand)	Per capita GDP (Yuan)	Birth rate (‰)	Death rate (‰)	Natural growth rate (‰)	Two-child rate (%)	Population density (people/per square kilometer)	Average annual consumption level of residents	Physical condition	Employment stability degree
2010	127.430	58,961	9.390	7.800	1.590	15.790	985	[12,000, 15,000]	s_2	s_1
2011	132.050	68,315	9.490	7.420	2.070	17.230	1180	[15,001, 18,000]	s_3	s_4
2012	137.340	79,482	10.720	5.540	5.180	19.970	1911	[18,001, 20,000]	s_1	s_4
2013	145.620	89,000	11.280	4.980	6.300	19.550	1203	[20,001, 22,000]	s_0	s_4
2014	156.010	98,000	12.220	4.970	7.250	21.600	1206	[22,001, 24,000]	s_1	s_1
2015	163.760	104,132	12.700	5.750	6.950	27.130	1238	[24,001, 26,000]	s_4	s_4

Table 6 Two-tuple matrix

Factors	2010	2011	2012	2013	2014	2015
A_0	$(s_4, 0.000)$	$(s_3, 0.491)$	$(s_3, -0.091)$	$(s_2, -0.002)$	$(s_1, -0.147)$	$(s_0, 0.000)$
A_1	$(s_4, 0.000)$	$(s_3, 0.172)$	$(s_2, 0.183)$	$(s_1, 0.34)$	$(s_1, -0.457)$	$(s_0, 0.000)$
A_2	$(s_4, 0.000)$	$(s_4, -0.121)$	$(s_2, 0.393)$	$(s_2, -0.284)$	$(s_1, -0.420)$	$(s_0, 0.000)$
A_3	$(s_0, 0.000)$	$(s_1, -0.463)$	$(s_3, 0.194)$	$(s_4, -0.014)$	$(s_4, 0.000)$	$(s_3, -0.102)$
A_4	$(s_4, 0.000)$	$(s_4, -0.339)$	$(s_1, 0.463)$	$(s_1, -0.329)$	$(s_0, 0.000)$	$(s_0, 0.212)$
A_5	$(s_4, 0.000)$	$(s_3, 0.492)$	$(s_3, -0.474)$	$(s_3, -0.326)$	$(s_2, -0.049)$	$(s_0, 0.000)$
A_6	$(s_4, 0.000)$	$(s_3, 0.158)$	$(s_0, 0.000)$	$(s_3, 0.058)$	$(s_3, 0.045)$	$(s_3, -0.093)$
A_7	$(s_2, -0.245)$	$(s_2, -0.217)$	$(s_2, -0.181)$	$(s_2, -0.162)$	$(s_2, -0.144)$	$(s_2, -0.126)$
A_8	$(s_2, 0.000)$	$(s_3, 0.000)$	$(s_1, 0.000)$	$(s0, 0.000)$	$(s_1, 0.000)$	$(s_4, 0.000)$
A_9	$(s_1, 0.000)$	$(s_4, 0.000)$	$(s_4, 0.000)$	$(s_4, 0.000)$	$(s_1, 0.000)$	$(s_4, 0.000)$

future number and the proportion of aging population in Wuhan. Specifically, firstly, the five main factors influencing the population aging selected in Sect. 3 are taken as independent variables, and the number of aging population and the proportion of aging population are taken as dependent variables, respectively, to establish the multivariate prediction models for forecasting the number of aging population and the proportion of aging population, respectively. Then, and the number of aging population and the proportion of aging population are taken as dependent variables, respectively, two univariate prediction models based on ARIMA are established to predict the number of aging population and the proportion of aging population, respectively. Finally, based on the predicted results of the above two models for the aging population and the proportion of aging population, combined with the weighted combination idea, a combination prediction model named MLRA-ARIMA is established to predict the number and the proportion of aging population in Wuhan in the next 10 years.

4.1 MLRA-ARIMA combination prediction model

MLRA model is to establish multiple linear regression equations of dependent variables to multiple independent variables based on the actual observed values of dependent variables and multiple independent variables, and verify the effectiveness of the model by testing and analyzing the significance of the comprehensive linear influence of each variable on dependent variables. ARIMA is a model established by transforming non-stationary time series into stationary time series, and then make a regression of dependent variables to its lag value and the present value and lag value of random error terms. The former is a multivariate prediction model, while the latter is a univariate prediction model. This section will combine these two prediction models and establish a combination prediction model to predict the future number and the proportion of aging population in Wuhan.

4.1.1 MLRA model

In Sect. 3, by quantitative analysis, it is concluded that birth rate, per capita GDP, average annual consumption level of residents, two-child rate and natural growth rate

are the main factors of affecting population aging in Wuhan. And the two main variables reflecting the trend of population aging in Wuhan are the number of aging population and the proportion of aging population. In order to effectively analyse the trend of the number of aging population (or the proportion of aging population) with the change of these five main influence factors, we take these five main influence factors as the independent variables, and the number of aging population (or the proportion of aging population) as the dependent variable to establish a prediction model to predict the number of aging population (or the proportion of aging population) based on MLRA (Mori and Suzuki 2018; Rao and Yan 2020; Wang et al. 2020b). The modeling steps are as follows:

Step 1: Model establishment

Take the number of aging population (or the proportion of aging population) as the dependent variable, which is denoted as y (or y'), and the per capita GDP, birth rate, two-child rate, average annual consumption level of residents and natural growth rate as five independent variables, which are denoted as $x_1, x_2, \ldots, x_5$, respectively. Then the MLRA model is express as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_5 x_5 + e, \tag{13}$$

where β_0 is a constant term, $\beta_1, \beta_2, \ldots, \beta_5$ are the regression coefficients of y to $x_1, x_2, \ldots, x_5$, e is an unobservable random variable, which is called an error term with the mean value is zero, and the variance is greater than zero,

Use the OLS (Zhang et al. 2016) to estimate the parameter β in the model (13), then the sum of the squared errors of the given data is

$$Q(\beta) = (y - x\beta)^T (y - x\beta).$$

To find the value of β that minimizes $Q(\beta)$, we can get the least squares estimation of β, which is denoted as $\hat{\beta}$, where $\hat{\beta} = (X^T X)^{-1} X^T Y$.

Substitute $\hat{\beta}$ into Eq. (13), we can obtain the estimated value of y, i.e.,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_5 x_5,$$

and the fitting value of the given data is

$$\hat{Y} = X\hat{\beta}.$$

Step 2: Hypothesis testing of regression model

After the MLRA model is established, the hypothesis of linear relationship between dependent variables and multiple independent variables needs to be tested for significance, that is, the established MLRA model needs to be tested for significance.

(1)   Fit goodness test of regression equation

The determination coefficient R^2 refers to the proportion of the dependent variable y that can be explained by the free variable, and its mathematical model is

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}, \tag{14}$$

where $\text{SST} = \sum_{i=1}^{5} (y_i - \bar{y})^2$ is squares sum of dispersion, and $\text{SSE} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 i$ is squares sum of residuals.

The value range of R^2 is from 0 to 1, and the closer the value of R^2 approaches 1, the better the fitting effect of the established regression equation on the actual observed value is. And the closer the value of R^2 approaches 0, the worse the fitting effect of the established regression equation on the actual observed value is.

(2)   Residual analysis

The Durbin Watson (DW) test (Grose and King 1991) can be used to verify the quality of a model. The expression for the DW statistic is

$$\text{DW} = \frac{\sum_{t=2}^{5} (e_t - e_{t-1})^2}{\sum_{t=1}^{5} e_t^2}, \tag{15}$$

where the value range of DW statistic is from 0 to 4, t is the time, and residuals are collected in chronological order.

4.1.2 ARIMA time series model

Section 4.1.1 establishes a MLRA model to predict the future trend of the number and the proportion of aging population in Wuhan. This section will establish a high-precision prediction models based on ARIMA (Santos et al. 2019; Poornima and Pushpalatha 2019) according to the existing time series data. The modeling process is as follows:

If a linear random process x_t can be expressed as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \tag{16}$$

then the ARMA(p,q) model in the form of a delay operator can be expressed by

$$\varphi(B) = \theta(B)\varepsilon_t,$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, is called a p-order autoregressive coefficient polynomial, and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is called a q-order moving average coefficient polynomial.

The judgment criteria of the ARMA are shown in Table 7.

When the sequence is not a stationary sequence or has an obvious linear trend, the difference method is needed to eliminate the tendency. The first-order difference is $\nabla x_t = x_t - x_{t-1}$, the p-order difference is

$\nabla^p x_t = \nabla^{p-1} x_t - \nabla^{p-1} x_{t-1}$, and the k-step difference is $\nabla^k = x_t - x_{t-k}$.

Thus, the model of ARIMA can be expressed by

$$\nabla^d x_t = \sum_{i=0}^{d} (-1)^d C_d^i x_{t-i}.$$

If it is expressed by delay operator, we have

$$\nabla^d x_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t.$$

The above method is to establish a mathematical model based on curve fitting and parameter estimation for one or several groups of time series data, and to predict the future data by testing the fitting effect of the model.

4.1.3 MLRA-ARIMA combination prediction model

(1) Basic principle of combination prediction

The combined prediction model can integrate different types of individual prediction models, which can reduce the systematic error of model prediction and improve the prediction effect (Zhu et al. 2019; Qu et al. 2019; Mao et al. 2020; Wang et al. 2020a). The modeling principle of linear combination prediction model is as follows:

$$y_{0t} = W_1 y_{1t} + W_2 y_{2t} + \cdots + W_n y_{nt},$$

where y_{0t} is the combination prediction value of the period t, $y_{1t}, y_{2t}, \ldots, y_{nt}$ is the prediction values of different n single prediction models at time t, and $W_1, W_2, \ldots, W_n$ is the corresponding combination weights.

The key of linear combination prediction model is to determine a reasonable weight W_i, which is determined by the minimum variance principle of combination prediction error.

When $n = 2$,

$$W_1 = \sigma^2 / (\sigma_1^2 + \sigma_2^2), \quad W_2 = 1 - W_1,$$

where $\sigma_i^2 (i = 1, 2)$ is the residual variance of the i-th prediction model.

When $n > 2$,

$$W_i = \frac{1}{Q_i} \quad (i = 1, 2, \ldots, n),$$

where Q_i is the sum of the square residuals of the i-th single prediction model.

(2) The combination prediction model

Combine the MLRA model established in Sect. 4.1.1 with the ARIMA established in Sect. 4.1.2, we establish a combination prediction model, and the basic steps are as follows:

Step 1: Take the five main factors influencing the population aging selected in Sect. 3 as the independent variables, and the number of aging population and the proportion of aging population as the dependent variables to establish multivariate linear regression models, respectively. Then, use the ARIMA to predict the index values of five main influence factors for the future 10 years, and substitute the predicted values into the MLRA model (13), then the predictive value of the number and the proportion of aging population in Wuhan in the next 10 years can be obtained.

Step 2: Based on the historical data of the number and the proportion of aging population in Wuhan from 1993 to 2016, establish the ARIMA time series model to predict the number and the proportion of aging population in Wuhan in the next 10 years.

Step 3: Combine with the prediction results of the two models in Step 1 and Step 2, the combination prediction values of the number and the proportion of aging population in Wuhan in the next 10 years are given. The combination prediction formula is:

$$y = W_1 y_1 + W_2 y_2, \tag{17}$$

where y_1 and y_2, respectively, represent the predicted value of the number of aging population (or the proportion of aging population) in Wuhan in the same year calculated by the MLRA model and the ARIMA, and W_1 and W_2, respectively, represent the weight of the predicted value of the two models.

Step 4: Model test

For the combination prediction model, it is necessary to test the feasibility that whether the model can be used to predict future. The common test method is the mean error test. The calculation formula is as follows.

$$\eta = \frac{1}{n} \sum_{i=1}^{n} |\Delta_i|, \tag{18}$$

where $|\Delta_i|$ is the absolute error of the i-th real value and the fitting value. The calculation method of $|\Delta_i|$

Table 7 The judgment criteria of the ARMA

Model	Autocorrelation coefficient	Partial autocorrelation coefficient
AR(p)	Trailing	q-order truncation
MA(q)	q-order truncation	Trailing
ARMA(p,q)	Trailing	Trailing

is $|\Delta_i| = \frac{|\hat{y}_i - y_i|}{y_i}$, where y_i is the actual real value and $\hat{y}_i$ is the fitting value of the combination model.

4.2 Empirical analysis

4.2.1 Prediction of the number and the proportion of aging population in Wuhan based on MLRA model

From model (13), take the proportion of aging population as the dependent variable, and the five main factors (per capita GDP, birth rate, two-child rate, average annual consumption level of residents and natural growth rate) influencing the population aging selected in Sect. 3 as the independent variables, and then we establish the MLRA model.

The Statistical Product and Service Solutions (SPSS) software is used for parameter estimation of the MLRA model. According to the preliminary regression results, it can be found that although the determination coefficient is 0.997, the P value of several variables is greater than 0.05, which indicates that the regression model is not significant. Therefore, the multicollinearity test of independent variables is carried out by using the method of variance inflation factor, and the relative results are shown in Table 8.

From Table 8, we can see that the VIF values of variables x_1 and x_4 are much greater than 10, so we can conclude that there is multicollinearity among the independent variables. In addition, we use the conditional number method of eigenvalues to verify the results by using the R software, and the conditional number is 638.8108, which is much greater than 100, indicating that there is a strong multicollinearity. Therefore, the stepwise regression method and the optimal subset regression method are used to select relevant indexes, respectively. The stepwise regression method shows that when the model has only the independent variables x_1, x_2, x_4 and x_5, the AIC of the model is the smallest and the added value of residual sum of squares is the least. Moreover, the optimal subset regression method shows that the maximum adjusted determination coefficient and the minimum Cp value are 0.996 and 3.848, respectively, after removing the variable

x_3. Therefore, after deleting the variable x_3, the MLRA model is obtained as follows.

$$y = 9.508 + 0.788x_1 - 0.139x_2 + 0.003x_4 + 0.334x_5.$$

(19)

Firstly, the significance of the model (19) is judged. The model fitting results are shown in Table 9.

As can be seen from Table 9, the VIF values of the independent variables of the model are all less than 10, which means there is no multicollinearity among the independent variables. Moreover, the P values are all less than 0.05, that is, all independent variables are significant, and the multicollinearity of the original model is eliminated.

Secondly, the fit goodness test is carried out for the MLRA model (19). We can conclude from the fit goodness test by Eq. (14) and DW test by Eq. (15) that the standard error of the model is small, and the adjusted value of R^2 is 0.936, which means the established model has passed the F test with a significance level of 0.01. Therefore, it can be considered that the fitting degree of the model is good, and the established MLRA model (19) is reasonable. The test value of DW test is 1.160, which indicates that the residual and independent variables are independent of each other.

Similarly, take the number of aging populations as the dependent variable, and the five main factors influencing the population aging selected in Sect. 3 as the independent variables to establish another MLRA model given by Eq. (13). The stepwise regression method is used for the selection of relevant indexes, and the model is as follows.

$$y = 88.980 + 9.846x_1 - 3.644x_2 + 0.023x_4 + 2.219x_5.$$

(20)

For the above MLRA model (20), the multicollinearity test, the fit goodness test and the DW test are all passed.

In addition, another statistical software, i.e., R Software, is used for regression modeling, and the obtained regression results were exactly the same as models (19) and (20).

By predicting the indexes of the main factors influencing the population aging in Wuhan in the next 10 years, together with the MLRA model (19), the prediction results are shown in Table 10.

Table 8 The result of multicollinearity test

Variables	Coefficient	t	Statistical significance	VIF
(Constant)	0.090	13.904	0.000	
x_1	0.008	9.487	0.000	42.896
x_2	-0.001	-1.354	0.197	10.879
x_4	-2.602×10^{-5}	-0.378	0.711	90.763
x_5	0.003	1.958	0.070	19.614
x_3	0.000	1.274	0.223	16.722

Table 9 The final model fitting results

Variables	Coefficient	t	Statistical significance	VIF
(Constant)	9.508	13.904	0.000	
x_1	0.788	9.487	0.000	9.896
x_2	0.139	-1.354	1.19×10^{-11}	8.879
x_4	0.003	-0.378	0.0378	8.763
x_5	0.334	1.958	0.007	1.614

Similar to the analysis method above, we can obtain the prediction results of the number of aging populations in Wuhan, which is listed in Table 11.

4.2.2 Prediction of the number and proportion of aging population in Wuhan based on ARIMA

The data of population aged 60 and above in Wuhan from 1993 to 2016 are taken as the original sequence, and the ARIMA is established to analyse and predict the number of aging population in Wuhan. The calculated results are shown in Table 12.

The comparison graph drawn by the original real values and the predicted values are given in Fig. 4. From Fig. 4, we can observe that the curve corresponding to the predicted values matches the curve corresponding to the original values very well.

(1)   Stability test

The original sequence is taken the logarithm operation, and the autocorrelation test is conducted according to the sequence diagram of the logarithmic sequence. It is found that the autocorrelation coefficient decayed to 0 very slowly with the increase of the number of delayed phases, and the reverse increase trend occurred after the number of delayed phases was 10, so the sequence can be determined as a non-stationary sequence. Further, the unit root test is performed on the original sequence, and it is found that at the significance level of 0.05, the unit root test quantity was

Table 10 Prediction results of the proportion of aging population in Wuhan

Years	Prediction value (%)	Years	Prediction value (%)
2017	21.669	2022	29.078
2018	22.923	2023	31.067
2019	24.149	2024	33.340
2020	25.608	2025	35.818
2021	27.212	2026	38.610

3.2163, which is greater than the critical value of the unit root test at the significance level of 1% to 10%. The null hypothesis is accepted, indicating that the sequence is non-stationary. The above test results show that the original sequence is non-stationary, and the order of the non-stationary sequence can be determined.

(2)   Determine the order of difference

The first-order difference of the original sequence is found to be non-stationary. The results show that the value of the second difference fluctuates around 0. Intuitively speaking, the second difference has no obvious trend and no periodicity. Therefore, the sequence of second-order difference is considered to be a stationary sequence. Further, the autocorrelation test was performed for the second-order difference sequence, and the results are shown in Fig. 5.

It can be seen from Fig. 5 that the autocorrelation coefficient fluctuates around 0 and is basically close to 0, so the sequence after the second difference can be judged as a stationary sequence. Then, the unit root test is conducted to further confirm its stationarity. The test results are shown in Table 13.

It can be concluded from Table 13 that the unit root test value of the second-order difference is less than the critical value at the significance level of 1%, and the null hypothesis is rejected at the significance level of 0.05. Therefore, the second-order difference sequence of the logarithmic sequence is a stationary sequence, i.e., $d = 2$. To sum up, the sequence can be considered as a second-order differential stationary non-white noise sequence, so the ARIMA can be established.

(3)   Parameter estimation of the model

By calculating, it is found that the models of ARIMA(1,2,3) and ARIMA(1,2,1) both meet the requirements. Therefore, a more suitable model should be selected from these two models. The fitting analysis of ARIMA(1,2,3) and ARIMA(1,2,1) is made. The fitting analysis results of these two models are shown in Table 14 and Table 15, respectively.

By comparing the results in Table 14 and Table 15, it can be seen that the indexes of Akaike's information criteria and Schwarz criterion for ARIMA(1,2,3) are all smaller than that of ARIMA(1,2,1), so it can be considered that the ARIMA(1,2,3) is better, and the fitting result of ARIMA(1,2,3) is

$$x_t = 0.0016 - 0.3371 x_{t-1} - 0.1 x_{t-2} + \varepsilon_t. \tag{21}$$

Table 11 Prediction results of the number of aging population in Wuhan

Years	Prediction value (ten thousand)	Years	Prediction value (ten thousand)
2017	182.204	2022	245.182
2018	194.175	2023	260.214
2019	204.387	2024	277.307
2020	217.583	2025	294.973
2021	229.516	2026	314.255

Table 12 Test the predicted values of population aged 60 and above

Years	Original sequence	Predicted value	Error value	Error rate	Years	Original sequence	Predicted value	Error value	Error rate
1993	71.000	–	–	–	2005	97.700	96.172	1.538	1.574
1994	73.400	–	–	–	2006	103.100	101.08	2.050	1.988
1995	76.100	75.788	0.312	0.410	2007	109.700	109.685	0.055	0.050
1996	78.700	78.708	− 0.008	− 0.010	2008	114.700	114.415	0.325	0.283
1997	80.700	81.400	− 0.700	− 0.867	2009	121.900	120.31	1.55	1.272
1998	82.200	82.887	− 0.687	− 0.836	2010	127.400	128.551	− 1.121	− 0.880
1999	84.100	84.186	− 0.086	− 0.102	2011	132.100	133.714	− 1.664	− 1.260
2000	86.000	86.474	− 0.474	− 0.551	2012	137.300	137.308	0.032	0.023
2001	90.100	88.463	1.637	1.817	2013	145.600	144.599	1.021	0.701
2002	92.200	93.751	− 1.551	− 1.682	2014	156.000	154.200	1.810	1.160
2003	94.500	95.223	− 0.723	− 0.765	2015	163.800	165.988	− 2.228	− 1.361
2004	93.900	96.553	− 2.703	− 2.880	2016	172.800	172.392	0.358	0.207
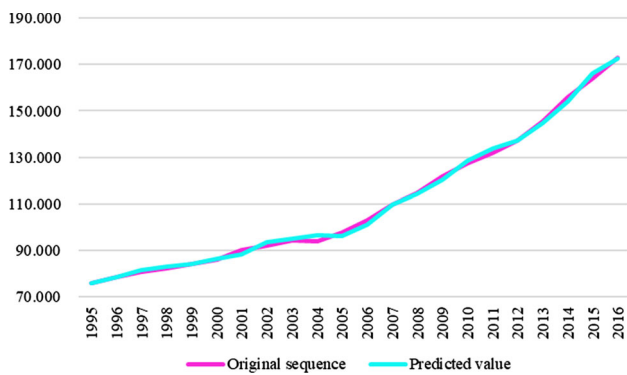
Fig. 4 A comparison graph drawn by the real values and the predicted values

(4)   Applicability test of the model

Before the above model (21) is used to predict the future, the adaptability of the model needs to be tested. The residual test results are shown in Fig. 6.

The autocorrelation coefficient, partial autocorrelation coefficient, Q-Stat value and P value all show that the residual sequence does not exist autocorrelation, the residual sequence is stable, and the model fitting is successful, indicating that the model (21) has passed the test. So we can use the ARIMA(1,2,3) given by Eq. (21) to
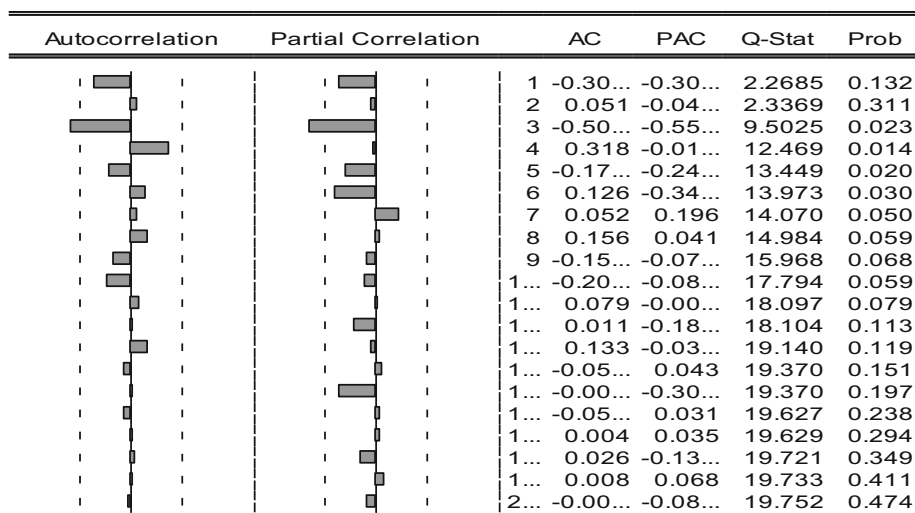
predict the number of aging population in Wuhan from 2017 to 2026. The predicted result is listed in Table 16.

In addition, the population in Wuhan from 1993 to 2016 is collected from the Statistical Yearbook of Wuhan, and the proportion of aging population in Wuhan from 1993 to 2016 is calculated. The ARIMA is used to predict the proportion of aging population in Wuhan in the future. After taking the logarithm of the original sequence, the second-order difference is made, and the second-order difference sequence is used for modeling. The established model is ARIMA(3,2,3), and the fitting effect of the model is good, and the accuracy is high. The results of the proportion of aging population in Wuhan in the next 10 years are shown in Table 16.

4.2.3 Combination prediction of population aging trend in Wuhan

In Sects. 4.2.1 and 4.2.2, the number and the proportion of aging population in Wuhan are modeled and predicted by using the models of MLRA and ARIMA, and the average relative error are relatively small (that is, the prediction accuracy is relatively high). For example, when the MLRA model is used to predict the proportion of aging population in Wuhan, the fitting precision is 93.6. % When the ARIMA is used to predict the number and the proportion of

Fig. 5 Autocorrelation test diagram of the second-order difference

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.30...	-0.30...	2.2685	0.132
		2	0.051	-0.04...	2.3369	0.311
		3	-0.50...	-0.55...	9.5025	0.023
		4	0.318	-0.01...	12.469	0.014
		5	-0.17...	-0.24...	13.449	0.020
		6	0.126	-0.34...	13.973	0.030
		7	0.052	0.196	14.070	0.050
		8	0.156	0.041	14.984	0.059
		9	-0.15...	-0.07...	15.968	0.068
		1...	-0.20...	-0.08...	17.794	0.059
		1...	0.079	-0.00...	18.097	0.079
		1...	0.011	-0.18...	18.104	0.113
		1...	0.133	-0.03...	19.140	0.119
		1...	-0.05...	0.043	19.370	0.151
		1...	-0.00...	-0.30...	19.370	0.197
		1...	-0.05...	0.031	19.627	0.238
		1...	0.004	0.035	19.629	0.294
		1...	0.026	-0.13...	19.721	0.349
		1...	0.008	0.068	19.733	0.411
		2...	-0.00...	-0.08...	19.752	0.474

Table 13 The unit root test result of the second-order difference

	t test statistic	Probability
Unit root test	− 4.847	0.001
Critical value of the test		
Significance level of 1%	− 3.832	
Significance level of 5%	− 3.030	
Significance level of 10%	− 2.655	

aging population in Wuhan, the relative error is within 10%. Therefore, it is concluded that both MLRA and ARIMA are applicable to the short-term prediction of aging population in Wuhan.

In order to obtain a better prediction result, reduce the systematic error caused by the model, and improve the prediction effect, a combined prediction model can be established to predict the aging trend of population in Wuhan. Due to the high accuracy of the MLRA and ARIMA in this paper, the combination prediction model
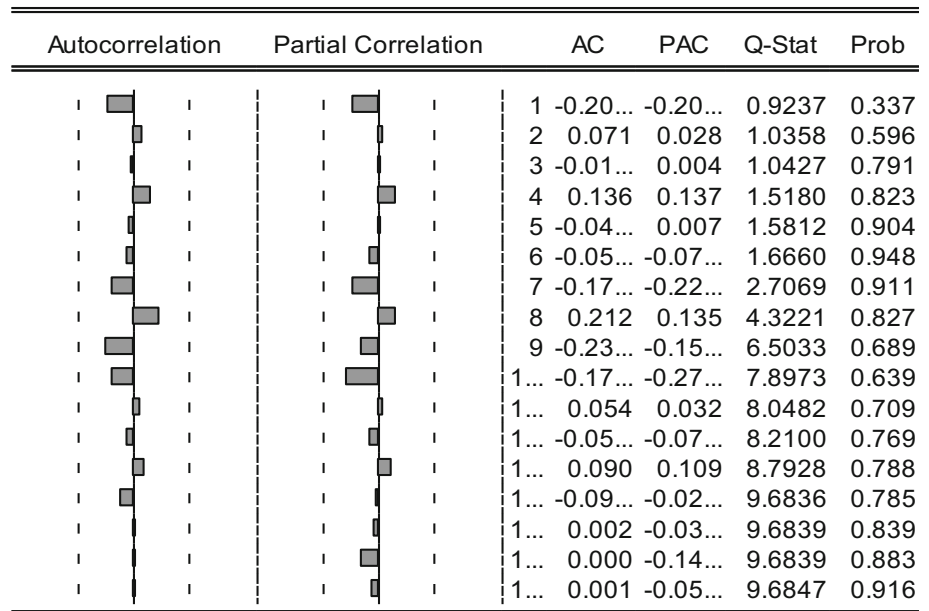
Table 14 The fitting results of ARIMA(1,2,3)

Variable factors	Coefficient	Standard error	t statistic	Probability
C	0.002	0.001	2.210	0.040
AR(3)	− 0.337	0.222	− 1.516	0.038
MA(3)	− 0.100	0.128	− 7.793	0.000
R square	0.762	Mean of dependent variables		0.000
Adjusted R square	0.736	Variance of dependent variable		0.018
Regression of SE	0.016	Akaike's information criteria		− 5.779
Sum of squares for residuals	0.005	Schwarz criterion		− 5.630
Logarithmic likelihood value	58.487	H-Q information criteria		− 5.747
DW statistic	2.227			

Table 15 The fitting results of ARIMA(1,2,1)

Variable factors	Coefficient	Standard error	t statistic	Probability
C	0.112	0.019	5.820	0.000
AR(1)	0.597	0.196	3.044	0.008
MA(3)	− 0.981	0.060	− 16.290	0.000
R square	0.557	Mean of dependent variables		0.079
Adjusted R square	0.502	Variance of dependent variable		0.033
Regression of SE	0.023	Akaike's information criteria		− 4.555
Sum of squares for residuals	0.009	Schwarz criterion		− 4.406
Logarithmic likelihood value	46.274	H-Q information criteria		− 4.530
DW statistic	0.002			

Fig. 6 The residual test diagram of the ARIMA(1,2,3)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.20...	-0.20...	0.9237	0.337
		2	0.071	0.028	1.0358	0.596
		3	-0.01...	0.004	1.0427	0.791
		4	0.136	0.137	1.5180	0.823
		5	-0.04...	0.007	1.5812	0.904
		6	-0.05...	-0.07...	1.6660	0.948
		7	-0.17...	-0.22...	2.7069	0.911
		8	0.212	0.135	4.3221	0.827
		9	-0.23...	-0.15...	6.5033	0.689
		1...	-0.17...	-0.27...	7.8973	0.639
		1...	0.054	0.032	8.0482	0.709
		1...	-0.05...	-0.07...	8.2100	0.769
		1...	0.090	0.109	8.7928	0.788
		1...	-0.09...	-0.02...	9.6836	0.785
		1...	0.002	-0.03...	9.6839	0.839
		1...	0.000	-0.14...	9.6839	0.883
		1...	0.001	-0.05...	9.6847	0.916

Table 16 The predicted results

Years	The proportion of aging population (%)	The number of aging population (ten thousand)	Years	The proportion of aging population (%)	The number of aging population (ten thousand)
2017	21.732	182.320	2022	29.863	245.181
2018	23.074	192.847	2023	31.982	261.405
2019	24.560	204.271	2024	34.357	279.157
2020	26.205	216.737	2025	36.993	298.583
2021	27.944	230.335	2026	39.922	319.8730

(17) proposed in Sect. 4.1.3, i.e., $y = W_1 y_1 + W_2 y_2$, is used to predict the number and the proportion of aging population in Wuhan in the next 10 years. Here we think the predicted values obtained by these two models are of the same importance, namely, we set $W_1 = W_2 = 0.5$.

Before using the combination prediction model for prediction, the error analysis on the three established prediction models (MLRA, ARIMA, combination prediction model). The model error analysis on the number and the proportion of aging population in Wuhan is shown in Table 17.

It can be seen from Table 17 that in the error analysis, the average relative error of the combination prediction model is the smallest, so the combination prediction model can be used to predict the number and the proportion of aging population in Wuhan in the next 10 years.

Based on the prediction results of the MLRA in Sect. 4.1.1 and the prediction results of the ARIMA in Sect. 4.1.2, the combination prediction model given by Eq. (17) is used to predict the number and the proportion of aging population in Wuhan in the next 10 years. The prediction results are shown in Table 18.

We can draw the graph of the prediction results of the combination prediction model listed in Table 18, which is shown in Fig. 7.

As can be seen from Fig. 7, the proportion of aging population in Wuhan grew slowly from 2017 to 2020, but the proportion grew faster and faster after 2020, and the proportion of aging population in Wuhan will still increase year by year in the next 10 years. The aging population has an obvious rising trend, and the rising speed is relatively fast. By 2026, the proportion of aging population in Wuhan will reach 39%.

5 Countermeasures and suggestions on population aging in Wuhan

Section 3 finds out the main factors influencing the population aging in Wuhan via empirical analysis. Section 4 analyses and forecasts the number and the proportion of aging population in Wuhan by establishing a combination prediction model, and concludes that the number and the proportion of aging population in Wuhan will increase year

Table 17 Comparative analysis of fitting effect of three models

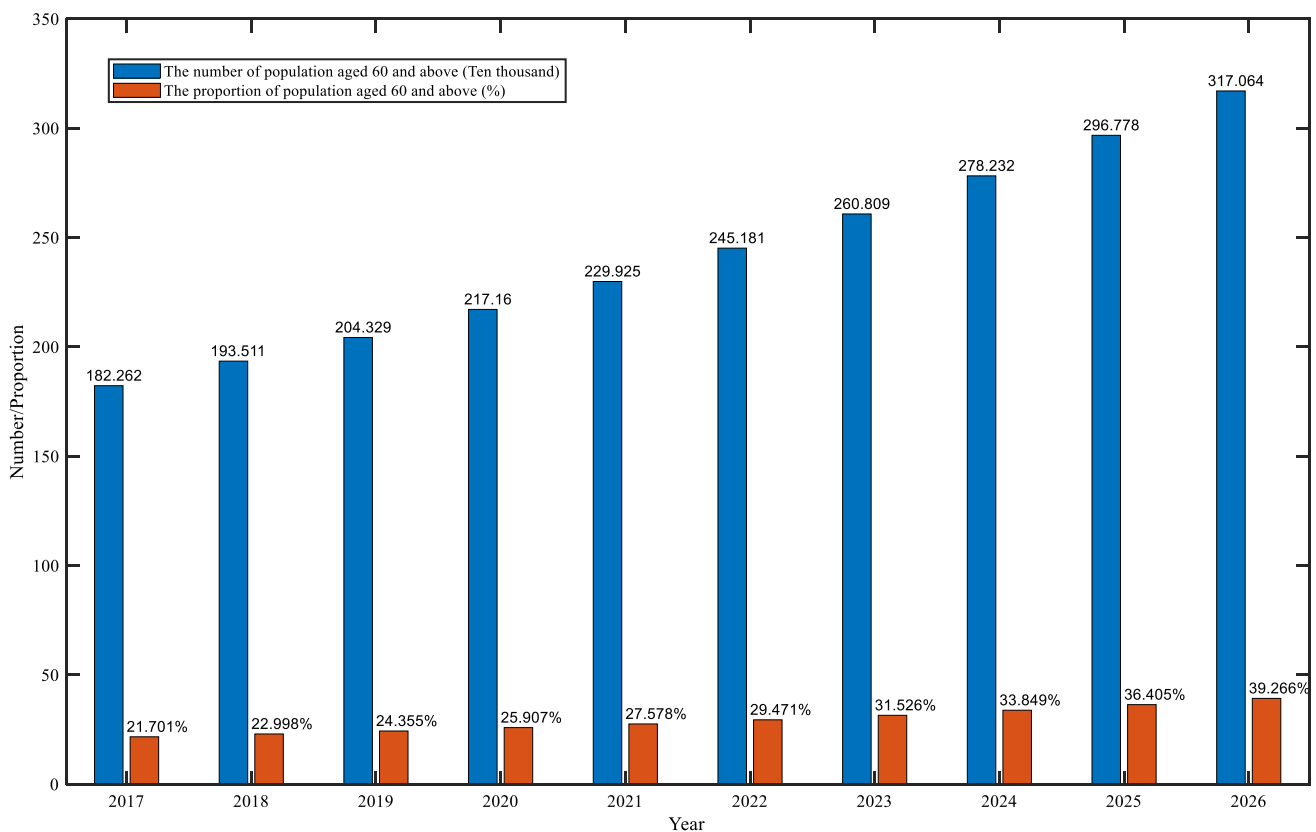| Models	Models for predicting the proportion of aging population			Models for predicting the number of aging population		
	MLRA	ARIMA	Combination prediction model	MLRA	ARIMA	Combination prediction model
Average relative error (%)	1.267	0.844	0.710	2.070	1.155	1.085

Table 18 The prediction results of the combination prediction model

Years	The number of aging population (ten thousand)	The proportion of aging population (%)	Years	The number of aging population (ten thousand)	The proportion of aging population (%)
2017	182.262	21.701	2022	245.181	29.471
2018	193.511	22.998	2023	260.809	31.526
2019	204.329	24.355	2024	278.232	33.849
2020	217.160	25.907	2025	296.778	36.405
2021	229.925	27.578	2026	317.064	39.266

by year in the next 10 years, and the trend of increase is obvious. Based on these quantitative analysis results, this section will provide some countermeasures and suggestions on how to alleviate the problem of population aging in Wuhan from the aspects of economic development, pension policy, pension security system, and so on.

(1) Enhance the recognition degree of population aging in Wuhan

Fig. 7 The comparison diagram of the combination prediction model

From 1993 to now, Wuhan is one of the many aging cities, and the aging degree of the population is continued unabated. Thus, population aging has become an important issue that cannot be ignored in the social and economic development of Wuhan. As a new first-tier city, Wuhan's future development strategy and planning should take the population aging into consideration.

As can be seen from Table 1, the proportion of aging population in Wuhan exceeded 10% in 1993 and increased year by year in the next 23 years. It can also be seen from the historical data that the number of aging people aged 60 and above increased from 710,000 in 1993 to 1,727,500 in 2016. The increase of the number of aging people more than doubled with an increasing trend every year. As can be seen from the prediction results of the proportion and the number of aging populations in Table 18, the proportion of aging population in Wuhan is getting higher and higher in the next 10 years. In 2026, the proportion of aging population in Wuhan will reach 39%. These research results fully illustrate that the population aging problem poses a severe challenge to the economic and social development and management of Wuhan. The government needs to pay more attention to the population aging and take the aging problem as an important problem to maintain the development and stability of Wuhan. In a wider range of areas, comprehensive consideration, coordinated operations, take population aging as a strategic problem to plan and deploy early, formulate effective and feasible countermeasures to eliminate the adverse effects of population aging on social and economic development.

(2) Maintain the steady and sustained economic growth of Wuhan

In Sect. 3, when analyzing the main influencing factors of population aging in Wuhan, it can be found that the per capita GDP of Wuhan is one of the most important influencing factors of population aging by calculating the two-tuple correlation coefficient. Therefore, economic development has an important impact on the population aging of Wuhan. From 1997 to 2016, Wuhan's per capita GDP increased year by year, and the increase rate became larger after 2007. Therefore, maintaining the steady and sustained growth of Wuhan's economy is crucial to solving the problem of population aging.

First, economic development can ensure the financial support of implementing the aging policy. Second, from the data in Table 5 in Sect. 3, the residents' consumption level of Wuhan is growing every year, from 14,490 Yuan in 2010 to 24,943 Yuan in 2015, the residents' consumption level increases more than 10,000 Yuan in 5 years, which means the economic growth also contributed to the Wuhan residents' average annual consumption level. Sustained economic growth is a necessary condition to ensure

the continuous increase of Wuhan residents' consumption level. From Sect. 3.2, among the main factors influencing population aging in Wuhan, the residents' medical care expenditure is one of the important indexes of measuring the Wuhan residents' physical condition. From the Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020), we can obtain the data of Wuhan residents' medical care expenditure (see Table 19).

As can be seen from Table 19, the medical care expenditure of Wuhan residents is increasing year by year, which also indicates that sufficient capital investment is essential to establish a medical security system in line with the actual situation of Wuhan. As a new first-tier city, Wuhan's political, economic and cultural status is self-evident. Only economic development can promote the progress of the city, and only economic development can provide better economic guarantee for solving the problem of population aging in Wuhan.

(3) Improve Wuhan's family planning policy

From Sect. 3, we concluded that the birth rate and the two-child rate have a great impact on the population aging in Wuhan according to the quantitative analysis of the MLRA model given by Eqs. (19) and (20). From the Data in Statistical Yearbook of Wuhan, before 2010, the birth rate in Wuhan was controlled at around 8‰. After 2010, the birth rate began to increase continuously. By 2015, the birth rate in Wuhan reached 12.7‰. At the same time, before 2010, the two-child rate in Wuhan was 13%. After 2010, the two-child rate increased rapidly, reaching over 20% in 2014 and 27.13% in 2015. It can be seen that Wuhan is actively responding to the national call and fully implementing the two-child policy. But the birth rate of Wuhan city in 2016 fell to 11.48‰, which shows that although the two-child rate is rising, the other fertility rate, such as the one-child rate and three-child rate, have fallen.

Therefore, our government should continue to attach great importance to population and family planning, strengthen research on population development strategies, scientifically grasp the trend of population development and determine the concept of giving priority to investment in the all-round development of human beings. Our government must improve the family planning policy further, actively support the national two-child policy and keep the birth rate rising reasonably, perfect the family planning management and service system, make give priority to the family planning to share the fruits of reform and development, and strive to make the population development in harmony with the economic and social development.

(4) Vigorously develop the aging industry in Wuhan

As can be seen from Table 18, the population aging degree of Wuhan is becoming more and more serious. The

Table 19 The data of Wuhan residents' medical care expenditure

Year	2010	2011	2012	2013	2014	2015
Residents' medical care expenditure (Yuan)	903.600	1272.360	1379.370	1437.220	1445.120	1491.840

proportion of aging population has increased from 10.3% in 1993 to 20.7% in 2016, doubling the proportion of aging population. From our prediction result, in the next 10 years, the proportion of aging population in Wuhan is still increasing year by year. Compared with 1996, the proportion of aging population in Wuhan will double to 25.907% in 2020, and reach 39.266% in 2026. The growing proportion of aging people will place additional burdens on individuals, families and society. Therefore, to develop the aging industry and to encourage the elderly people to re-employment can ensure that the elderly have economic income, and also can ease the pressure of the family and society.

The variables mentioned in Sect. 3, such as medical care expenditure, life insurance expenditure and the quantity of employment, reflect the incomplete development of the aging market. Therefore, the government should attach importance to the development opportunities and formulate relevant plans to expand the aging market, so as to increase economic income and improve the aging market. At the same time, the priority areas of the aging industry in Wuhan should be determined. From Sect. 3, we know that the medical care expenditure, life insurance expenditure and the quantity of employment of Wuhan are the main factors influencing the population aging. Therefore, the development of the aging industry in Wuhan should give priority to the medical care market, insurance market and job creation.

(5) Appropriately increase the capital investment for the elderly in Wuhan

In Sect. 3, when analyzing the main influencing factors, it is found that the growth rate of residential GDP does not correspond to the growth rate of residents' medical care expenditure and residents' minimum living allowance. According to the data in Statistical Yearbook of Wuhan (Wuhan Bureau of Statistics 2020), the growth rates of the three from 2011 to 2015 are calculated, and the growth rates are shown in Table 20.

From Table 20, we can see that the growth rate of per capita GDP is significantly different from that of medical care expenditure and residents' minimum living allowance, and both the growth rates of residents' minimum living allowance and medical care expenditure are lower than the growth rate of per capita GDP (except 2011). It can be seen

that while the economy is growing, it is also necessary to appropriately increase the capital investment for the population aging, and appropriately increase the growth rate of the residents' minimum living allowance, so that the residents' minimum living allowance can maintain basic life and medical care.

(6) Strengthen the implementation of Wuhan's talent introduction policy

Wuhan implemented the policy of "one million college students studying in China" in 2017. This implementation of the talent introduction policy is of great significance to alleviate the problem of population aging. The processes and trends of demographic age structure change (as can be seen from Fig. 3) are influenced by population movements and migration in addition to births and deaths. The population density of Wuhan has been increasing year by year since 1997 (with a sharp increase in 2012), which indicates that the population in per unit land area of Wuhan has been increasing year by year. The net migration of population in Wuhan before 2010 is relatively large, the net migration rate of population before 2010 is positive, and the net migration rate before 2005 presents an increasing trend, indicating that the net migration of population is of great significance to the change of population age structure in Wuhan. The net migration rate of Wuhan is negative after 2010, which indicates that the population migration rate of Wuhan is greater than the population outflow rate. The implementation of talent introduction policy can increase the proportion of population aged 15 to 59 in the total population, leading to the change of population age and the change of the proportion of aging population. Therefore, the government should strengthen the implementation of talent introduction policy.

6 Conclusions

This paper analyses the main factors of influencing population aging in Wuhan and discusses the development trend prediction of population aging in Wuhan. First, under the data information environment that the attribute values of the preselected influence factors are multi-source heterogeneous data (the real numbers, interval numbers and fuzzy linguistic variables coexist), a TTCCA method is proposed

Table 20 The data of growth rates

Years	Growth rate of residents' minimum living allowance	Growth rate of residents' medical care expenditure	Growth rate of per capita GDP
2011	0.250	0.408	0.159
2012	0.151	0.084	0.164
2013	0.081	0.042	0.120
2014	0.036	0.006	0.101
2015	0.035	0.032	0.063

to rank the importance of the influencing factors and to select the main influencing factors. Then, a combination prediction model is established to predict the population aging trend in Wuhan, and the predicted values of the aging population in Wuhan in the next 10 years are given. Finally, based on the results of quantitative analysis, some countermeasures and suggestions on how to alleviate the problem of population aging in Wuhan are provided.

The contribution of this paper is as follows: (i) In the analysis of the main factors that influence the population aging problem in Wuhan, when the attribute values of the preselected influence factors are the multi-source heterogeneous data, we propose a new TTCCA method based on the traditional correlation coefficient analysis, which provides a new effective approach to solve the optimization selection problem of main influencing factors under the information environment of multi-source heterogeneous data. (ii) Combining MLRA with ARIMA, together with the weighted combination idea, a combination prediction model named MLRA-ARIMA is established to predict the number and the proportion of the aging population in Wuhan. The combination prediction model overcomes the shortcomings of the single model in prediction accuracy and prediction scope, thus effectively enhancing the reliability of the prediction results.

In the future research work, we will fully consider the impact of COVID-19, and establish some new models based on our proposed models in this paper for the prediction of the aging population and the analysis of the main factors influencing the population aging in Wuhan.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Asim A, Nasar R, Rashid T (2019) Correlation coefficient of intuitionistic hesitant fuzzy sets based on informational energy and their applications to clustering analysis. Soft Comput 23:10393–10406

Chen W (2016) China's demographic estimation using the generalized stable population model. Popul J 38(1):5–13

Chen MH, Hao GC (2014) Research on regional difference decomposition and influence factors aging in China. China Popul Resour Environ 24(4):136–141

Chen YG, Yu B (2006) Three models for predicting population growth—theoretical foundation, application methods, and revised expressions. J Central China Normal Univ (Nat Sci) 40(3):452–456

Chen YH, Li YS, Su CG (2012) Radial basis function neural network model applied in the forecast of population aging taking Hunan Province as an example. Econ Geogr 32(4):34–39

Chen GH, Cai YF, Li F (2014) The prediction of China's population aging trend and analysis of the structure based on nonparametric auto-regression model. Northwest Popul 4:81–87

Geng XL, Qiu HQ, Gong XM (2017) An extended 2-tuple linguistic DEA for solving MAGDM problems considering the influence relationships among attributes. Comput Ind Eng 112:135–146

Gillen P, Spore D (1994) Functional and residential status transition among nursing home residents. J Gerontol: Med Sci 1:33–42

Gong YL, Zhang DS, Wu XQ (2007) Nonparametric autoregression prediction model on population growth rate. Appl Stat Manag 26(1):38–42

Grose SD, King ML (1991) The locally unbiased two-sided Durbin–Watson test. Econ Lett 35:401–407

He JN (2011) Principal component analysis on the indirected influencing factors of population aging in Shanxi Province. Future Dev 34(6):108–111

Herrera F, Martínez L (2000) A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans Fuzzy Syst 8(6):746–752

Herrera F, Martinez L, Sanchez PJ (2005) Managing non-homogeneous information in group decision making. Eur J Oper Res 166(1):115–132

Hou DQ (2012) Prediction and analysis of population in Hubei Province based on Leslie model. Wuhan University of Technology, Wuhan

Hu AG, Liu SL, Ma ZG (2012) Population Aging, population growth and economic growth: evidence from China's provincial panel data. Popul Res 36(3):14–26

Jiang YY (2012) Population prediction based on age shift algorithm. Stat Decis 13:82–84

Lee HH, Shin K (2019) Nonlinear effects of population aging on economic growth. Jpn World Econ 51:100963

Li JM (2016) Analysis and prediction of Chinese aging based on Bayesian hierarchy spatio-temporal model. Stat Res 33(8):89–94

Li LL (2017) Analysis on regional differences and influencing factors of population aging in China. J Huazhong Agric Univ (Soc Sci Ed) 6:94–102

Li P, Rao CJ, Goh M, Yang ZQ (2021) Pricing strategies and profit coordination under a double echelon green supply chain. J Clean Prod 278:123694

Liang X (2017) Prediction of aging population based on grey prediction model. Harbin Institute of Technology, Harbin

Liu YZ (2016) The total population and structure of China based on Leslie model. Renmin University of China, Beijing

Liu PD, Chen SM (2018) Multiattribute group decision making based on intuitionistic 2-tuple linguistic information. Inf Sci 430–431:599–619

Lutz W, O'neill BC, Scherbov S (2003) Europe's population at a turning point. Science 299(5615):1991–1992

Lv SG, Xuan DP (2012) The prediction of aging coefficient in Beijing. Stat Res 29(3):27–31

Lyons AC, Grable JE, Joo SH (2018) A cross-country analysis of population aging and financial security. J Econ Ageing 12:96–117

Ma ZL, Shao CF, Ma SQ et al (2011) Constructing road safety performance indicators using Fuzzy Delphi Method and Grey Delphi Method. Expert Syst Appl 38(3):1509–1514

Mao SH, Zhu M, Wang XP, Xiao XP (2020) Grey–Lotka–Volterra model for the competition and cooperation between third-party online payment systems and online banking in China. Appl Soft Comput 95:106501

Meng C (2012) Population prediction based on leslie matrix and time series analysis. Jilin University, Changchun

Modigliani F (2005) Rethinking pension reform. Cambridge University Press, Oxford

Mori YC, Suzuki TJ (2018) Generalized ridge estimator and model selection criteria in multivariate linear regression. J Multivariate Anal 165:243–261

Muhuri PK, Gupta PK (2020) A novel solution approach for multiobjective linguistic optimization problems based on the 2-tuple fuzzy linguistic representation model. Appl Soft Comput 95:106395

Peng JJ, Tian C, Zhang WY, Zhang S, Wang JQ (2020) An integrated multi-criteria decision-making framework for sustainable supplier selection under picture fuzzy environment. Technol Econ Dev Econ 26(3):573–598

Poornima S, Pushpalatha M (2019) Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft Comput 23:8399–8412

Qu SJ, Zhou YY, Zhang YL, Wahab MIM, Zhang G, Ye YY (2019) Optimal strategy for a green supply chain considering shipping policy and default risk. Comput Ind Eng 131:172–186

Rao CJ, Yan BJ (2020) Study on the interactive influence between economic growth and environmental pollution. Environ Sci Pollut Res 27(31):39442–39465

Rao CJ, Zhao Y (2009) Multi-attribute decision making model based on optimal membership and relative entropy. J Syst Eng Electron 20(3):537–542

Rao CJ, Goh M, Zhao Y et al (2015) Location selection of city logistics centers under sustainability. Transport Res Part D: Transport Environ 36:29–44

Rao CJ, Zhao Y, Zheng JJ (2016a) An extended uniform-price auction mechanism of homogeneous divisible goods: supply optimisation and non-strategic bidding. Int J Prod Res 54(13):4028–4042

Rao CJ, Zheng JJ, Wang C et al (2016b) A hybrid multi-attribute group decision making method based on grey linguistic 2-tuple. Iranian Journal of Fuzzy Systems 13(2):37–59

Rao CJ, Goh M, Zheng JJ (2017a) Decision mechanism for supplier selection under sustainability. Int J Inf Technol Decis Mak 16(1):87–115

Rao CJ, Xiao XP, Goh M, Zheng JJ, Wen JH (2017b) Compound mechanism design of supplier selection based on multi-attribute auction and risk management of supply chain. Comput Ind Eng 105:63–75

Rao CJ, He YW, Wang XL (2020a) Comprehensive evaluation of non-waste cities based on two-tuple mixed correlation degree. Int J Fuzzy Syst. https://doi.org/10.1007/40815-020-00975-x **(in press)**

Rao CJ, Lin H, Liu M (2020b) Design of comprehensive evaluation index system for P2P credit risk of "three rural" borrowers. Soft Comput 24(15):11493–11509

Rao CJ, Liu M, Goh M, Wen JH (2020c) 2-stage modified random forest model for credit risk assessment of P2P network lending to "Three Rurals" borrowers. Appl Soft Comput 95:106570

Santos DSD, de Oliveira JFL, Neto PSGD (2019) An intelligent hybridization of ARIMA with machine learning models for time series forecasting. Knowl-Based Syst 175:72–86

Schön M, Stähle N (2020) When old meets young? Germany's population ageing and the current account. Econ Model 89:315–336

Smit MGD, Kater BJ, Jak RG et al (2006) Translating bioassay results to field population responses using a Leslie-matrix model for the marine amphipod Corophium volutator. Ecol Model 196(3–4):515–526

Sun L, Wu SP (2015) An empirical study on the impact of population aging on household consumption in China. Stat Decis 9:98–101

Tian C, Peng JJ, Zhang S, Zhang WY, Wang JQ (2019) Weighted picture fuzzy aggregation operators and their applications to multi-criteria decision-making problems. Comput Ind Eng 137:106037

Tian C, Peng JJ, Zhang WY, Zhang SJ, Wang Q (2020) Tourism environmental impact assessment based on improved AHP and picture fuzzy PROMETHEE II methods. Technol Econ Dev Econ 26(2):355–378

Wang L (2004) International comparison study on the aged trend and reasons in China. Popul Econ 1:6–11

Wang GX, Gan YH (2017) China's aging population and regional economic growth. Chin J Popul Sci 3:30–42

Wang YM, Liu CY (2012) Preliminary research on aging population and flexible retirement policy of Shanghai. IERI Procedia 2:455–459

Wang LD, Wang YJ, Pedrycz W (2019) Hesitant 2-tuple linguistic Bonferroni operators and their utilization in group decision making. Appl Soft Comput 77:653–664

Wang RH, Nan GF, Chen L, Li MQ (2020a) Channel integration choices and pricing strategies for competing dual-channel retailers. IEEE Trans Eng Manage. https://doi.org/10.1109/TEM.2020.3007347 **(in press)**

Wang TF, Shi P, Wang GX (2020b) Solving fuzzy regression equation and its approximation for random fuzzy variable and their application. Soft Comput 24:919–933

Wei J, Ni XM, He AS (2018) Study on the relationship between population policy and economic growth in the context of aging. Syst Eng-Theory Pract 38(2):337–350

Wiener JM, Tilly J (2002) Population ageing in the United States of America: implications for public programmes. Int J Epidemiol 31(4):776–781

Wuhan Bureau of Statistics (2020) Wuhan Statistical Yearbook of Wuhan. http://www.whtj.gov.cn/. Accessed 29 March 2020

Xiao QZ, Gao MY, Xiao XP, Goh M (2020a) A novel grey Riccati–Bernoulli model and its application for the clean energy consumption prediction. Eng Appl Artif Intell 95:103863

Xiao QZ, Shan MY, Gao MY, Xiao XP, Goh M (2020b) Parameter optimization for nonlinear grey Bernoulli model on biomass energy consumption prediction. Appl Soft Comput 95:106538

Xiao QZ, Shan MY, Xiao XP, Rao CJ (2020c) Evaluation model of industrial operation quality under multi-source heterogeneous data information. Int J Fuzzy Syst 22(2):522–547

Xiao XP, Duan HM, Wen JH (2020d) A novel car-following inertia gray model and its application in forecasting short-term traffic flow. Appl Math Model 87:546–570

Xie WY, Xu ZS, Ren ZL, Herrera-Viedma E (2020) The probe for the weighted dual probabilistic linguistic correlation coefficient to invest an artificial intelligence project. Soft Comput 24:15389–15408

Xu X, Zhao Y, Zhang X (2017) Spatial variation of population aging and associated factors in Jiangsu Province. Sci Geogr Sin 3737(12):1859–1866

You XY, You JX, Liu HC, Zhen L (2015) Group multi-criteria supplier selection using an extended VIKOR method with interval 2-tuple linguistic information. Expert Syst Appl 42:1906–1916

Zhang Y (2009) Research on status, trend and strategy of population aging in Gansu Province. Lanzhou University, Lanzhou

Zhang H (2013) Some interval-valued 2-tuple linguistic aggregation operators and application in multiattribute group decision making. Appl Math Model 37(6):4269–4282

Zhang XY, Ullah A, Zhao SW (2016) On the dominance of Mallows model averaging estimator over ordinary least squares estimator. Econ Lett 142:69–73

Zhao W, Tao T, Zio E (2015) System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. Appl Soft Comput 30:792–802

Zhu XZ, Pang FY (2009) Application of autoregressive and logistic discrete model in Chinese population forecast. Stat Decis 13(13):157–159

Zhu JM, Wu P, Chen HY et al (2019) Carbon price forecasting with variational mode decomposition and optimal combined model. Physica A 519:140–158