The Journal of Engineering

The 2nd 2018 Asian Conference on Artificial Intelligence Technology (ACAIT 2018)

Effect of demographic structure on resource utilisation using term frequency–inverse document frequency algorithm – evidence from China

eISSN 2051-3305
Received on 19th July 2018
Accepted on 26th July 2018
E-First on 16th October 2018
doi: 10.1049/joe.2018.8331
www.ietdl.org

Yulin Liu[1] ✉, Zhihui Li[2], Xingmin Yin[2], Leifei Lyu[3]
[1]*Centre for Public Economy & Public Policy Research, Chongqing University, Chongqing, People's Republic of China*
[2]*Economics and Business Administration, Chongqing University, Chongqing, People's Republic of China*
[3]*School of Public Affairs, Chongqing University, Chongqing, People's Republic of China*
✉ *E-mail: lylmx@cqu.edu.cn*

Abstract: China has stepped into middle-income status, which is characterised by demographic imbalance and huge discrepancy of resource utilisation, resulting in an imbalance of regional development. Using term frequency–inverse document frequency algorithm, the authors extract the resource keywords and construct the resource utilisation index. The authors then explore the impact of demographic structure on resource utilisation. The main conclusions are as follows: (i) according to the Moran Index, there are strong spatial autocorrelations in resource utilisation at the provincial level in China; (ii) the empirical results demonstrate that the regression coefficients of spatial lags are significant. This provides compelling evidences that the neighbouring regions have obvious spatial spillover effects on local resource utilisation; (iii) the authors' findings also reveal that the total dependency ratio which reflects the demographic structure has a negative and significant effect on resource utilisation. With other explanatory variables held constant, the child-age dependency ratio and old-age dependency ratio perform differently in the robustness testing. Ultimately, targeted and systematic policy suggestions are proposed to improve resource utilisation and adjust the demographic structure.

1 Introduction

The concept of 'middle-income status' is firstly proposed by the World Bank publications entitled 'An East Asian Renaissance: Ideas for Economic Growth' in 2007. Since then, the apprehension that China may fall into the middle-income trap has not eased. Especially in recent years, China's economic growth rate has also slowed down gradually. Simultaneously, the cumulative problems in rapid development are also exposed, such as aging, resource shortage, and environmental pollution. From the experiences of developed countries such as Europe and America, developing the economy at the expense of massive resource consumption can neither be sustainable nor bring real development. Actually, China still plays the role of the world's largest energy consumer according to the BP Statistical Review of World Energy 2016. Approximately 4.3 billion tons of standard coal was consumed in 2015, accounting for 23% of global consumption and 34% of global net growth. Although China's energy consumption per unit GDP has decreased year by year, it is also less than that of most developed countries, such as USA and Japan. Therefore, we conclude that there is indeed excessive resource utilisation in China. Just when China's economy has entered a new normal, natural resources are close to being exhausted. The environmental carrying capacity is also on the verge of the upper limit, causing a severe environmental damage. As everyone knows, the purpose of economic development is to bring benefit to the people rather than harm. These resource problems will become self-inflicted wounds in the long term. As the final bearers, people have the responsibility and duty to address these problems for their own interest. It is worth noting that resource problems are not only serious in China but also throughout the world. In this context, resource problems have been hot issues for scholars since the advent of resource crisis. Economic development needs resources. In addition, economic development benefits people. Hence, along this line, it is worthwhile to study resource problems on the demographic side to provide a new visual angle.

Without the advantages of low wages or high skills and with limited natural resources, how China can make the next jump to high-income status? This article will explain the reasons why natural resource consumption is so serious from the perspective of demographic structure in middle-income status. China's demographic structure has changed dramatically since the implementation of the family planning policy in the 20th century. The total dependency ratio reached 62.6% in 1982, although it has dropped to 36.9% in 2015. Superficially speaking, this means that the responsibility of caring for the old and raising children has decreased greatly. In the sixth national population census, people aged 60 and over reached 13.26% of the population in China. This means that China has undoubtedly entered the aging society. Simultaneously, China's total fertility rate has dropped under the replacement standard and decreased progressively. Admittedly, China is facing the twin problems of low fertility and an ageing population. If nothing is done, the phenomenon of the low birth rate and ageing will become more common in the near future. The potential drive of China's economy is also greatly hindered. Finally, it is not beneficial to the solution of the resource problems. Therefore, the factors of the demographic structure should be researched if China wishes to pursue connotative development versus resource-based economic development.

The aim of this paper is to reveal the main factors, such as demographic structure, that influence resource utilisation on a macro level and the trend of their spatial variance in China, quantitatively measure the intensity of the effects of these factors, and deepen the understanding of the spatial association between the resource utilisation and demographic structure. From the perspective of the demographic structure, we obtain a special interpretation of excessive resource consumption.

2 Literature review

Studies on population and resources have attracted the attention of many scholars for a long time. These thoughts can be traced back to William Petty. As he said, 'Labour is the father of wealth. Land is the mother of wealth'. Through continuation and evolution, research achievements have been enriched. Looking at this research topic, we decided to take 'economy growth' as a bridge and conduct a literature review from two aspects: demographic structure and economic growth and resources and economic growth.

Terms from the existing research involving demographic structure and growth are very abundant. The demographic structure mainly influences economic growth through the demographic dividend, savings rate, and human capital. This can be summarised as follows: (i) the increase of the elderly population indicates that adults need to undertake a heavier responsibility of support. The demographic dividend also gradually subsided, resulting in a weak economic growth, and by postponing retirement and utilising educational resources, the second demographic dividend can be tapped to serve the economic growth [1]. (ii) People at different ages have different savings capacities. The young and the elderly are restricted in their ability to obtain decent incomes by external factors, such as laws and energy. Therefore, their savings ratio is extremely low. Due to the uncertainty, adults are inclined to increase savings so that they can hedge risks. Generally, the change of demographic structure brings out the fluctuation of the savings ratio, which impacts investment and ultimately economic growth [2]. (iii) Parents are willing to improve the quality of children within the constraints of family planning, and education is an important way to promote the accumulation of children's human capital. Human capital has the characteristics of knowledge and technology, which plays a key role in technological innovation. Technological innovation will affect total factor productivity, thereby stimulating the economy [3, 4].

Most scholars have focused their research on the relationship between resources and economic growth. The effect of natural resources on growth has been a topic widely discussed in the economic literature at present [5, 6]. So far, there is no obvious evidence that abundant resources are a 'Gospel' or 'Curse' for the economic growth. In reality, the relationship between resources and economic growth has different performances in different countries and different periods [7–9]. In this regard, most related studies have concentrated along three aspects: (i) natural resources have guaranteed material production. Capital and labour are also needed in the process of resource utilisation. The service and the support of related industries are indispensable too, and economic growth will be promoted by relying on industry-linkage effects. (ii) The benefits of resources can provide a source of capital accumulation for a country's industrialisation. The utilisation of resources can also set the conditions for the sustained increase in economy through a series of system designs. However, some scholars have put forward different ideas against it, such as the crowding-out effect, terms of trade deterioration, and the rent-seeking effect [10–12]. They think the countries that have abundant resources may benefit from the prices of resource products to achieve rapid growth in the short term. However, an excessive reliance on natural resources will lead to economic stagnation in the long term. (iii) Recently, scholars have come to realise that the relationship between the natural resources and economic growth is not linear. It may also be 'V' or 'N' type in some research results [13, 14]. Of course, a 'Gospel' or 'Curse' depends on other factors, such as the quality of the institution and the efficiency of the allocation of production factors [15, 16].

Although studies on demographic structure and economic growth and resources and economic growth are very abundant, there are some key issues that are neglected in these research findings. Few studies have examined the mechanism between demographic structure and resources. Currently, most researchers have studied the relationship among resources and demographic size, distribution, and so on. Based on the case of Shanghai, one scholar attributes the effect of resources to demographic size and urbanisation [17]. Additionally, some scholars have systematically evaluated the spatial pattern and change rule of the land carrying capacity under the condition of demographic distribution at multiple levels. On the basis of a literature review, we found that most results take carrying capacity as the research angle, such as mineral resources carrying capacity and water resources carrying capacity [18, 19]. In addition, massive research objects are single and are not comprehensive or integrated, such as water resources, land resources, and mineral resources. As a result, the conclusions obtained are scattered, fragmented, and lack rigour [20, 21]. In terms of research methods, a large majority of articles are seldom analysed by the mainstream economic methods but are mainly concentrated on simulation models [22].

Generally, previous studies ignore the diversification of resource types as well as the regional propinquity characteristic of the spatial distribution. The resource index approach has theoretical and statistical problems, because it employs arbitrary methods for selecting variables. Few papers have considered the spatial aspects of resource utilisation. Compared with the previous studies, this paper mainly makes progress in the following aspects. First, the resource utilisation index that we constructed gives a comprehensive reflection of a variety of resources, such as energy, electricity, and so on. Some scholars present the C-LOG algorithm and MapReduce algorithm to investigate it [23, 24]. However, we find that term frequency–inverse document frequency (TF–IDF) algorithm has greater advantage in text mining. Hence, TF–IDF algorithm will be applied to construct the resource index. Second, we investigate whether there is a spatial effect of resource utilisation based on the Moran Index at the provincial level in China. Therefore, the theories and approaches of spatial econometrics will be applied to the topic of demographic structure and resource utilisation. Ultimately, more scientific and rigorous conclusions will be obtained.

3 Materials and methods

3.1 Construction of resource index

This article focuses on the construction of index with the following steps: indicator screening, construction of indicator system, and weight endowing.

3.1.1 Indicator screening: TF–IDF is commonly used to endow weights in information retrieving and data mining. A large pool of tests proves that TF–IDF performs well in automatic keyword extraction. Therefore, this article adopts TF–IDF algorithm as a computer technology keyword automatic extraction method, supplemented by other technical methods. The implementation process is completed by the following steps:

- Collecting corpus. Corpus pool mainly consists of 50–100 articles related to 'resources' and 50–100 articles related to other topics.
- Using the term frequency tool to analyse the corpus pool, two files can be generated: freq_count.txt and stru_count.txt, which are converted into Excel files for convenience.
- Calculating the word frequency and computing TF value using the following formula:

$$\mathrm{TF}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

where $n_{i,j}$ indicates the frequency of that word in the file. The denominator is the sum of the frequency of all the words in the file.

Based on the calculation results of TF values in Table 1, the denominator is large when calculating TF-a1, TF-a2, and TF-b1, which obtains minimum TF values: 2.13×10^{-6}, 1.96×10^{-5}, and 2.00×10^{-6}, respectively. Data loss may occur due to sparse events, and such TF values are not in the minority. In this way, TF values calculated by method b are the most important, though all possible method combinations are calculated.

- The stop-word filtering of term frequency constructs a document vector, which is a candidate word set of indicators.

Table 1 Part examples of the results of TF

Word	uniCount	TF-a1	TF-a2	TF-b1	TF-b2
	3872	—	—	—	—
resource endowment	2	2.19×10^{-6}	1.96×10^{-5}	4.52×10^{-5}	0.0023
resource structure	7	7.34×10^{-4}	6.87×10^{-3}	6.22×10^{-4}	0.0062
resource consumption	1	2.13×10^{-6}	2.96×10^{-3}	2.00×10^{-6}	0.0044
resource scarcity	1	3.42×10^{-4}	4.60×10^{-3}	1.40×10^{-3}	0.0019
resource tax	41	4.16×10^{-6}	3.40×10^{-3}	7.52×10^{-3}	0.0058
resource contribution	3	5.29×10^{-6}	7.36×10^{-3}	9.41×10^{-5}	0.0021
resource library	5	4.21×10^{-4}	2.26×10^{-3}	2.52×10^{-3}	0.0068

Table 2 Part examples of the results of IDF

Word	uniCount	The number of documents-a-50	The number of documents-b-100	The number of documents-c-50	IDF-a	IDF-b	IDF-c
	91,336						
water resource	2	1	3	3	1.39794	1.39794	1.0991
cultivated land	7	2	3	2	1.221849	1.39794	1.221849
energy consumption	1	1	2	2	1.39794	1.522879	1.221849
forest coverage	1	5	9	5	0.930819	1.39103	0.853872
mineral resource	41	1	4	4	1.221849	1.39794	1.221849
rich resource	3	1	5	5	1.09691	1.045757	0.79588
grain production	5	27	31	5	1.39794	1.221849	0.920819

Table 3 Candidates of evaluating indicator

Number	Evaluating indicator
1	per capita grain
2	per capita water
3	per capita cultivated area
4	per capita energy production
5	per capita forest area
6	energy consumption per unit of GDP
7	electricity consumption per unit of GDP
8	per capita energy use

- Normalise the candidate words of the indicator and retain those with higher word frequency or full specification in the synonyms.
- Calculate the inverse document IDF value using the following formula:

$$\mathrm{IDF} = \mathrm{Log}(N/n_k + L) \tag{2}$$

where the value of L is obtained by experiments, N is the total number of documents in the document set, and n_K is the number of documents with feature items.

The IDF calculation requires the following two parts:

1. The total number of documents in the corpus: According to the structure and needs of the corpus, three different types of pools can be constructed:

   a. The corpus consists of 50 main documents for calculating word frequency. Each document is independent and the total number of documents is 50.
   b. The corpus pool is composed of 50 documents for calculating word frequency, plus other 50 documents. Each document is independent and the total number of documents is 100.
   c. The corpus pool is composed of 50 documents for calculating word frequency and combines them as 1 document, plus other 49 documents. The total number of documents is 50.

2. The number of documents containing certain words: We use C language to calculate the number of documents containing certain words. We get 3983 entries from stored and reorganised array of structures, which consist of two members, namely the string member word[50] (for storing each term) and the integer member count (for storing the number of documents containing each word, initialised to 0).

During runtime, three different corpus pools a, b, and c, the value of the string array fileadd[ ] with complete file name for storing the storage path of all documents need to be adjusted to point to three different corpus pools. The IDF statistics are shown in Table 2.

- Calculate the TF–IDF value based on the existing results. The formula is:

$$\mathrm{Weight}_{\text{TF-IDF}} = \mathrm{TF} * \mathrm{IDF} \tag{3}$$

- According to the descending order of the TF–IDF value, select the index reference word at the priority of high value.
- Based on index terms, artificially construct the index terms.

Finally, based on the experimental data, the core words were extracted from the corpus and artificially integrated and expressed in a unified manner. The evaluation index candidates obtained are listed in Table 3.

3.1.2 Construction of resource utilisation index system: From the perspective of participants, this study establishes a class model. Starting with lexical analysis, this study selects resource indicators, and establishes a class model. We then abstract indicator items after screening through ‘candidate key classes’. With continuous abstractions and generalisations, a complete class diagram model with a hierarchical structure is obtained. Combining TF–IDF keyword automatic extraction index screening results, we finally get the resource utilisation index system in Table 4.

3.1.3 Indicator weights assigning: The weights of multi-characteristic resource evaluation indicators are expected to reflect the subjective opinion of scholars and reveal the nature of the data. Based on optimisation, this study will use the weights of the evaluation indicators obtained from the subjective and objective weighting methods to form a final combined weighting methods,

Table 4 Resource utilisation index system

First-grade index	Second-grade index	Third-grade index
resource utilisation	resource endowment	per capita grain
		per capita water
		per capita cultivated area
		per capita energy production
		per capita forest area
	resource consumption	energy consumption per unit of GDP
		electricity consumption per unit of GDP
		per capita energy use

Table 5 Weights of the resource utilisation index system

First-grade index	Second-grade index	Weight	Third-grade index	Weight
resource utilisation	resource endowment	0.5	per capita grain	0.22
			per capita water	0.26
			per capita cultivated area	0.19
			per capita energy production	0.19
			per capita forest area	0.14
	resource consumption	0.5	energy consumption per unit of GDP	0.37
			electricity consumption per unit of GDP	0.28
			per capita energy use	0.35

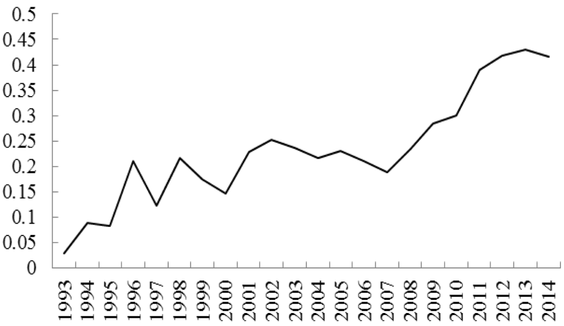

Fig. 1 *Global Moran's I of resource utilisation in 1993–2014*

so that it cannot only objectively reflect the importance of each index, but also reflect the subjectivity of decision makers. The method of combination weighting is as follows.

Let the final combined weight be $w_j = \alpha \times a_j + (1 - \alpha) \times b_j$, where a_j is the objective weight of the *j*th attribute and b_j is the subjective weight of the *j*th attribute. The objective weight in this article is the weighted average obtained from the independent coefficient method and the weights obtained by the coefficient of variation method. α is the undetermined coefficient in this equation. The calculation process is as follows:

$$\alpha = \frac{n}{n-1} G_{\text{AHP}} \tag{4}$$

In the formula, G_{AHP} is the difference coefficient of each component of the analytic hierarchy process (AHP) method:

$$G_{\text{AHP}} = \frac{2}{n}(1p_1 + 2p_2 + \cdots + np_n) - \frac{n+1}{n} \tag{5}$$

where n is the number of indicators and $p_1, p_2, \ldots, p_n$ are the rearranged $W_1, W_2, \ldots, W_n$ components in the AHP from small to large.

The combination weighting method takes advantages of both the subjective and the objective weighting methods, making the determination of weights in multi-index comprehensive evaluation more reasonable. The process of solving the objective weights, G_{AHP} and α, is done once in MATLAB programming. Based on this, through the combination of weighting methods, the distribution of the weights of the resource utilisation index system in this project is shown in Table 5.

3.2 Moran's I test for resource utilisation

According to the theory and the method of spatial econometrics, the empirical analysis process described in this section can be summarised as follows. Spatial statistical approaches are used to judge whether there is spatial autocorrelation in samples. If it exists, the spatial econometric model will be selected using the approach of the Lagrange multiplier diagnosis and the robust Lagrange multiplier diagnosis during the next work. Then, the model parameters are estimated by means of the maximum likelihood estimation. Otherwise, a classical regression model may be an appropriate analysis method.

Geographically, natural resources are unevenly distributed. Restricted by the technology, region, and historical factors, mankind has regionally exploited and utilised natural resources. It is also reasonable to suspect that spatial autocorrelation may be an issue [25]. To test whether there is spatial autocorrelation in resource utilisation, the most common method is the Moran Index, and the formula can be defined as follows:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij}} \tag{6}$$

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{7}$$

where S^2 is the sample variance, $\boldsymbol{W}_{ij}$ is the spatial weight matrix, and x_i is an attribute value for resource utilisation in region i. Commonly, the Moran Index value ranges from −1 to 1. $I > 0$ indicates a positive correlation. Otherwise, it implies a negative correlation.

Generally, resource utilisation is closely related to geographical factors. Therefore, it is necessary to incorporate geographical factors into the spatial weight matrix. Therefore, a spatial weighting matrix will be established, as follows:

$$\boldsymbol{W}_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases} \tag{8}$$

$\boldsymbol{W}_{ij} = 1$ if regions i and j are contiguous and $\boldsymbol{W}_{ij} = 0$ if regions i and j are not contiguous.

The global Moran Index of resource utilisation in China is calculated using formula (6), as shown in Fig. 1.

There is also a trend of enhancement, indicating that the resource utilisation in spatial agglomeration has strengthened. Therefore, it can be concluded that spatial autocorrelation of resource utilisation between Chinese provinces is an objective reality. This study also provides overwhelming evidence by applying spatial econometric methods in the following sections.

3.3 Selection and estimation of spatial lag model (SLM) and spatial error model (SEM)

Compared with the basic assumptions of classical econometrics, spatial econometrics holds that the sample observations are not completely independent, but there is some kind of spatial correlation that is embodied in spatial dependence and spatial heterogeneity. Spatial dependence refers to the fact that one observation associated with a location, which we might label, depends on other neighbouring observations in locations $i \neq j$. The term spatial heterogeneity is a special case of observed or unobserved heterogeneity, a familiar problem in standard econometrics [26]. Therefore, spatial econometric models can be divided into two basic types: SLM and SEM. The SLM is mainly applicable to situations where the actions of a region make an impact on other regions, and this impact is represented by the spatial lag operator in the SLM. The SEM is mainly adapted to the case that spatial heterogeneity is represented by a random error term. Namely, spatial heterogeneity can be identified as an unobserved random shock to the dependent variable. Combined with the theme of this research, the spatial econometric model can be set as follows:

$$\begin{cases} y_{it} = \tau + \rho \boldsymbol{W}_{ij} y_{it} + \beta x_{it} + \mu_{it} \\ \mu_{it} = \lambda \boldsymbol{W}_{ij} \mu_{it} + \varepsilon_{it} \end{cases} \tag{9}$$

where ρ represents a regression parameter to be estimated, μ_{it} denotes the stochastic disturbance in the relationship, y_{it} represents the resource utilisation of the ith province in tth year, x_{it} describes a series of core explanatory variables, the parameter λ is a coefficient of the spatially correlated errors, $\boldsymbol{W}_{ij}$ denotes the spatial weighting matrix, and ε_{it} is subject to the standard normal distribution.

From the general model in (7), we can derive special models by imposing restrictions. For example, setting $\lambda = 0$ produces a regression model with spatial dependence. Letting $\rho = 0$ results in a regression model with spatial heterogeneity.

Though the Moran Index is a useful approach to test for spatial correlation in the regression model, it cannot discriminate between two types of spatial correlations. Therefore, it is necessary to establish a set of criteria to determine which model more accurately represents reality. In addition, the rigour and scientific nature of the model selection are closely related to the accuracy of the conclusions. To determine if there are spatial lag and spatial error components in the models, two types of Lagrange multiplier (LM) diagnoses (LM diagnoses and robust LM diagnoses) are developed for identifying spatial heterogeneity in the error term or spatial dependence in the dependent variable. In brief, LM diagnoses cover LME for 'spatial error' and LML for 'spatial lag'. Likewise, robust LM diagnoses contain robust LME and robust LML, and the robust LM diagnoses are obtained by eliminating the spatial lag dependence or error dependence from the LM diagnoses.

Additionally, the spatial lag term in the SLM violates the hypothesis that explanatory variables in the classical econometric model are completely exogenous, and the error term of stochastic disturbance in the SEM goes against the hypothesis that stochastic disturbance should be independent and identically distributed. In the event of spatial lag dependence or error dependence, the ordinary least squares (OLS) do not yield unbiased and consistent estimates [27]. In terms of estimation methods, instrumental variable (IV) method, the generalised method of moments (GMM), and the maximum likelihood method (MLE) can address such issues. Through the calculation process, applying the IV method is very simple, and the estimated value of the parameter is often beyond the scope of its definition. It is not required to assume the distribution of stochastic disturbance by using GMM in advance. However, how to determine an appropriate IV is very subjective when dealing with endogenous variables, and it brings the accuracy down. Applying MLE could address this well. Therefore, it is a good choice to use MLE for estimating SLM and SEM.

3.4 Index selection and data description

The dependent variable in our model is resource utilisation. We selected the resource utilisation index as a proxy variable for resource utilisation. There is no data of the 'resource utilisation' as the statistical calibre at present in China. Therefore, we have constructed index of resource utilisation using the TF–IDF algorithm. The resource utilisation index is mainly composed of two aspects: resource endowment and resource consumption. The resource endowment consists of per capita grain, per capita water, per capita cultivated area, per capita energy production, and per capita forest area. Resource consumption includes energy consumption per unit of GDP, electricity consumption per unit of GDP, and per capita energy use. Combined with a subjective and objective weighting method, we get the appropriate weights of indices. After the weighting calculation, the resource utilisation index is finally obtained.

The core independent variable in this paper is the demographic structure. The demographic structure mentioned in this paper refers to the age structure of demographics. We measured the demographic structure by using population variables in three different approaches (total dependency ratio, child-age dependency ratio, and old-age dependency ratio). The total dependency ratio is an important manifestation of demographic structure. The higher the total dependency ratio, the more serious the burden working-aged people bear. The total dependency ratio consists of the child-age dependency ratio and old-age dependency ratio. In the mathematic formula, child-age dependency measures the proportion of young residents between the ages of 0 and 14 per 100 working-age people between the ages of 15 and 64. Similarly, the old-age dependency ratio shows the proportion of older residents over the age of 65 per 100 working-age people between the ages of 15 and 64.

On the basis of relative literature review, population scale, human capital per capita, economic scale, and physical capital per capita are appropriately chosen as control variables. Population scale is measured by the permanent population at the end of a year. In general, resource utilisation is often constrained by the population scale, and as the population scale increases, resource utilisation increases.

Human capital can reflect the population quality well. Commonly, we take the index of average educational attainment as the proxy variable of human capital. The average educational attainment can be calculated from the following formula: $H = \sum_{i=1}^{5} P_i H_i / P$.

Here P_i represents the number of persons aged 6 and over for whom i is the highest level of school attained; $i = 1$ for no school, 2 for primary, 3 for junior, 4 for senior, and 5 for higher; P refers to the total population aged 6 and over; H_i denotes the number of years for education in ith group.

We take the gross domestic product per capita at a constant price as the replacement variable for economic scale. The stock is estimated by applying the perpetual inventory method (PIM).The mathematical formula of PIM is $K_t = K_{t-1}(1 - \delta_t) + I_t$. The annual depreciation rate δ_t is deemed as 10.96% at the provincial level. The gross fixed capital formation is an alternative variable of investment I_t. The data of capital stock K_0 in the base period originates from Shan's estimation [28].

Considering the availability and consistency of data, we selected 30 provinces in China (excluding Chongqing, Hong Kong, Macao, and Taiwan) covering the period from 1993 to 2014. All original data are from the China Statistical Yearbook, China Energy Statistical Yearbook, China Statistical Yearbook on Environment, and China National Bureau of Statistics. Through a series of calculation processes, the corresponding research data will be obtained. To eliminate the impact of price factors, all the data related to price are calculated using the constant prices of 1993. At the same time, we compute simple data statistics, as shown in Table 6.

Given that the resource utilisation is spatially dependent, we separately estimated it using SAR and SEM models with panel data. The SAR and SEM models are developed as follows:

$$\begin{cases} \mathrm{RU}_{it} = \tau_0 + \rho \boldsymbol{W}_{ij}\mathrm{RU}_{it} + \tau_1 L_{it} + \tau_2 H_{it} + \tau_3 \mathrm{GDP}_{it} + \tau_4 K_{it} + \varepsilon_{it} \\ \varepsilon_{it} \sim N\left(0, \sigma_e^2 I_n\right) \end{cases} \quad (10)$$

$$\begin{cases} \mathrm{RU}_{it} = \gamma_0 + \gamma_1 L_{it} + \gamma_2 H_{it} + \gamma_3 \mathrm{GDP}_{it} + \gamma_4 K_{it} + \mu_{it} \\ \mu_{it} = \lambda \boldsymbol{W}_{ij}\mu_{it} + \varepsilon_{it} \\ \varepsilon_{it} \sim N\left(0, \sigma_e^2 I_n\right) \end{cases} \quad (11)$$

where τ and γ refer to the regression parameters to be calculated in order to describe the cause–effect correlation between the dependent variable and the explanatory variable In (10) and (11); $\boldsymbol{W}_{ij}$ represents the spatial weight matrix, which is 30×30 and consists of 0 and 1; ρ and λ refer to the spatial autoregressive coefficients used to measure the spatial dependence of the variable.

4 Results

As illustrated in the previous section, the global and local Moran Index show that there is a kind of spatial agglomeration in the spatial distribution of resource utilisation between provinces in China. Although a spatial correlation analysis has a certain application, there is no further explanation of the factors and impact mechanisms that affect the utilisation of resources. It is certainly worth adopting the MLE. For comparative purposes, the results obtained by applying the OLS are also reported in Table 7. Subsequently, the LM diagnosis could be applied by using MATLAB software. Based on the results of the test, a further analysis will be shown.

Table 7 reports coefficient estimates and standard errors from the OLS. Before the regression estimates, we should differentiate between the fixed effects (FE) model and the random effects (RE) model in the panel data. In such a situation, the RE model is preferred under the null hypothesis due to higher efficiency. While under the alternative, the FE model is at least consistent and preferred. The Hausman test is an appropriate method to handle this problem. For this reason, we carry out the Hausman test. It is indicated that the null hypothesis should be accepted, which means the RE model is the preferred approach. As we expected, the estimated coefficient of the total dependency ratio is −0.053, as seen in Table 7. It shows that the total dependency ratio has a negative relationship with the resource utilisation. The reason is that people without a job, such as children and the old, do not engage in material production. The increase of the total dependency ratio means a relative decline in the labour force, and labour is the source of power for the economic development. Resources are also closely related to the economic growth. When the total dependency ratio increases, economic development will be impeded. The existing resource utilisation pattern will be changed as well.

We test whether the main results are robust by adding control variables and alternative indicators of demographic structure: the child-age dependency ratio and the old-age dependency ratio. By including alternative indicators, we note that our main results are also robust. The effect of the old-age dependency ratio is still significant. Then, we are aware of a potential problem about bias from omitted variables with the use of a parsimonious model specification. Therefore, we check the robustness of our results with the inclusion of further controls: population scale, human capital per capita, economic scale, and physical capital per capita. We find that our main results pertaining to demographic structure are still robust. Population scale is negatively associated with resource utilisation, while physical capital per capita is significantly positive in columns (3) and (4) of Table 7.

Table 8 shows the results of LM and robust LM diagnoses. The probability values in LM-lag and LM-error are both very significant. In addition, the probability value in R-LM-error is larger than that in R-LM-lag. This result suggests that R-LM-lag is more significant than R-LM-error. Accordingly, we can recognise that SLM is more appropriate in the paper.

Therefore, we will further analyse the resource utilisation in China from the perspective of demographic structure by applying SLM. We add the individual effects and time effects in all models to control the effects of these unobservable factors. As a comparison, the results are reported by using SLM and SEM in Table 9.

Compared with the results of the panel RE model, the estimation coefficients are in accordance with the SLM. The space lag coefficient in columns (1), (2), (3), and (4) are statistically significant, which implies a positive impact. In general, the resource utilisation at the provincial level shows a strong spillover effect geographically. The spatial interaction of resource utilisation at the provincial level can be transmitted between adjacent areas. It also verifies that if we ignore the spatial spillover effect, the results may be biased or inconsistent through the OLS. Corresponding proposals also lack essential accuracy.

As reported in Table 9, the estimated coefficient of the total dependency ratio is significant in column (1). This confirms that there is a strong relationship between demographic structure and resource utilisation. Similarly, to check the robustness of our results, we put some control variables and alternative indicators of demographic structure in the SLM. Overall, our main results pertaining to demographic structure are robust. The estimated coefficients of the old-age dependency ratio are very significant at the 1% level. In addition, the marginal effect of the old-age dependency ratio is greater than that of the child-age dependency ratio in the process of resource utilisation. This result implies that China has entered into a society with a low birth rate and serious aging at present. By the end of 2014, women's total birth rate had dropped to under the replacement rate. The proportion of the people aged 60 and over to the total population in China reaches 15.5%. The decrease of fertility, with the increasing aging population, causes a direct negative growth of the labour force. Furthermore, accompanying the rising cost of child-rearing and pension issues of the aged, the decline of the labour force brings a severe burden to working people and increases labour costs. Ultimately, the vitality of economic development is seriously impeded, which affects resource utilisation.

In terms of control variables, population scale has a significant negative effect on resource utilisation. This confirms a fundamental fact that relying on a large number of demographic dividends creates the prosperity of China's economy. Similarly, natural resources are also a scarce production factor. As a result, when the demographic dividend can achieve economic growth, people will have no desire to improve resource utilisation efficiency. This ultimately hampers resource utilisation.

Physical capital per capita has a significant and positive effect on resource utilisation, which implies that physical capital is favourable for resource utilisation. As to the reasons, resources and physical capital are both indispensable factors of production in the process of economic growth. In addition, physical capital is also an essential base of other factors. The other factors can work well only combined with physical capital. For instance, if exceptional resources lack the support of physical capital, resource advantages will not be shown. Next, the present price transmission mechanism has been obstructed between the primary resource commodity and finished goods. The current resource prices did not really reflect the real scarcity of resources and environmental losses, resulting in inefficient and excessive resource utilisation. A report by the United Nations Environment Programme concludes that China has become the world's largest resource consuming country. Many fossil fuels, such as coal, oil, and gas, will have been depleted in 30 years. The current growth pattern is at the expense of high resource utilisation since China's reform and opening-up. Thus, compared with the resource factor, the demand for physical capital is very strong. This is also in accordance with the previous social development in China. Currently, China has been in a crucial period of economic transformation and upgrading. The serious restraints of natural resources make the extensive style of economic

Table 6 Descriptive statistics of variables in 30 provinces during 1993–2014

Variable name	Abbreviation	Obs	Mean	Std. dev	Min	Max
resource utilisation index	RU	660	53.28	6.53	28.01	78.64
total dependency ratio	DR	660	41.27	9.09	19.3	68.03
child-age dependency ratio	CDR	660	30.13	10.18	9.6	59.26
old-age dependency ratio	ODR	660	11.14	2.56	4.97	21.9
population scale	*L*	660	0.42	0.27	0.02	1.14
human capital per capita	*H*	660	7.72	1.43	2.21	12.03
economic scale	GDP	660	1.08	0.96	0.12	5.61
physical capital per capita	*K*	660	4.58	6.1	0.05	36.6

Table 7 Estimation results (OLS)

	(1)	(2)	(3)	(4)
DR	−0.053***		0.019	
	(−3.44)		(0.59)	
CDR		−0.011		−0.044
		(−0.7)		(−1.37)
ODR		0.454***		0.445***
		(6.53)		(5.98)
L			−7.988***	−5.08**
			(−3.33)	(−2.10)
H			0.467*	−0.582*
			(1.81)	(−1.93)
GDP			−0.302	−0.39
			(−0.79)	(3.35)
K			0.163***	0.199***
			(3.08)	(3.84)
Cons	55.369***	48.457***	51.72***	55.68***
	(42.61)	(31.94)	(14.77)	(15.97)
N	660	660	660	660
Province	30	30	30	30

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively. The standard error is in parentheses. For simplicity, the results of FE are not reported, since the RE model is chosen through the Hausman test in each case.

Table 8 LM and robust LM diagnoses

Spatial dependence test	Probability value
LM spatial lag	0.000
robust LM spatial lag	0.000
LM spatial error	0.000
robust LM spatial error	0.329

growth hard to carry on. The pursuit of effective, quality sustainable development is a good way to lead China's economy to prosperity.

5 Conclusions

This paper aims to investigate the impact of demographic structure on resource utilisation in middle-income status. To carry out this research, using TF–IDF algorithm, we extract the resource keywords and construct the resource utilisation index using TF–IDF algorithm. Then, we establish the global and local Moran Index test, and the results show that there is a strong spatial autocorrelation in resource utilisation at the provincial level in China. Additionally, after LM and robust LM diagnoses, we construct an SLM based on panel data in the period from 1993 to 2014 for 30 provinces in China. We can easily find that the space lag coefficient ρ is significant at the 1% level, which shows a strong spillover effect geographically. In the robustness test, we take the child-age dependency ratio and the old-age dependency ratio as substitution variables for demographic structure. Controlling a series of potential factors that affect resource utilisation, the results demonstrate that the old-age dependency ratio is significant and a positive sign for resource utilisation. In addition, in terms of control variables, population scale and physical capital per capita both have extremely significant impacts on resource utilisation but show the opposite direction.

At present, China has been in the key period of making the next jump to high-income status. In order to improve and balance the efficiency of resource utilisation, as well as promoting demographic transition, the critical question of what further measures the government might be prepared to take will feature so far. Based on the above conclusions, some targeted and systematic policy suggestions have been proposed on two sides. On the one hand, the government should eradicate the restrictions of the family planning policy and change fertility behaviour from compulsory to market orientated. Although China has carried out the universal two-child policy to ease pressure from the aging population in 2016, the trend of population reduction will not be changed for a long period. A cost–benefit analysis would be used to achieve the goal of improving demographic structure. Additionally, much importance should be given to the construction of human capital, not only to its investment and accumulation but also to the building of a complete system. On the other hand, we should explore and compile a natural resource balance sheet with the aim of obtaining a clear view of the stock of natural resources. The reasons why the phenomenon of severe resource waste has occurred in China is that a system of legal accountability and punishment with regard to natural resources is not complete. Consequently, a clear and unified

Table 9 Estimation results (MLE)

	SLM				SEM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DR	−0.025*		0.001		−0.026		−0.011	
	(−1.71)		(0.05)		(−1.21)		(−0.33)	
CDR		0.002		−0.049		−0.02		−0.079**
		(0.12)		(−1.62)		(−1.09)		(−2.28)
ODR		0.347***		0.360***		0.350***		0.395***
		(5.16)		(5.04)		(4.74)		(5.02)
ρ/*λ*	0.407***	0.347***	0.370***	0.339***	0.408***	0.330****	−0.373***	0.344***
	(9.42)	(7.64)	(8.20)	(7.44)	(9.08)	(6.82)	(7.35)	(6.59)
L			−6.589***	−4.122*			−5.722**	−3.039
			(−2.84)	(−1.77)			(−2.53)	(−1.3)
H			0.083	−0.763***			−0.021	−0.917***
			(0.34)	(−2.67)			(−0.07)	(−2.82)
GDP			−0.24	−0.32			−0.123	−0.382
			(−0.68)	(−0.92)			(−0.32)	(−1.03)
K			0.146***	0.177***			0.148***	0.199***
			(−2.95)	(3.63)			(2.87)	(3.88)
Cons	32.365***	30.677***	35.003***	39.638***	54.240***	49.905***	55.632***	59.019***
	(11.96)	(11.13)	(9.00)	(10.04)	(38.57)	(30.30)	(15.12)	(16.09)
N	660	660	660	660	660	660	660	660
province	30	30	30	30	30	30	30	30
individual effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
time effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively. The standard error is in parentheses. The parameters represent the space lag coefficient and the space error coefficient.

natural resource property rights system should be established as soon as possible.

6 Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 71773011), the Fundamental Research Funds for the Central Universities of China (Grant No. 2015CDJSK), and the basic Research and Innovation Development Fund for Central Universities of China (Grant No. 2017CDJSK01PT04).

7 References

[1] Cai, F.: ‘Demographic transition, demographic dividend, and Lewis turning point in China’, *Econ. Res. J.*, 2010, **4**, pp. 4–13

[2] Modigliani, F., Shi, L.C.: ‘The Chinese saving puzzle and the life-cycle hypothesis’, *J. Econ. Lit.*, 2004, **22**, pp. 145–170

[3] Edmund, S.P.: ‘Investment in humans, technological diffusion and economic growth’, *Stud. Macroecon. Theory*, 1980, **56**, pp. 133–139

[4] Paul, M.R.: ‘Increasing return and the long-run growth’, *J. Political Econ.*, 1986, **94**, pp. 1002–1037

[5] Chen, F.N., Shen, L.: ‘Review of resources economics based on CNKI literatures’, *Resour. Sci.*, 2013, **7**, pp. 1339–1346

[6] Alexander, V.B., Galina, I.B., Stanislav, S.B.: ‘Resource demand growth and sustainability due to increased world consumption’, *Sustainbility*, 2015, **7**, pp. 3430–3440

[7] Gavin, W.: ‘The origins of American industrial success, 1879–1940’, *Am. Econ. Rev.*, 1990, **80**, pp. 651–668

[8] David, P., Wright, G.: ‘Increasing returns and the genesis of American resource abundance’, *Ind. Corp. Change*, 1997, **6**, pp. 203–245

[9] Gylfason, T., Zoega, G.: ‘Natural resources and economic growth: the role of investment’, *World Econ.*, 2006, **29**, pp. 1091–1115

[10] Corden, W.M., Neary, J.P.: ‘Booming sector and de-industrialization in a small open economy’, *Econ. J.*, 1982, **92**, pp. 825–848

[11] Gylfason, T.: ‘Natural resources, education and economic development’, *Eur. Econ. Rev.*, 2001, **45**, pp. 847–859

[12] Jonathan, I.: ‘The varieties of resource experience: natural resource export structures and the political economy of economic growth’, *World. Bank. Econ. Rev.*, 2005, **19**, pp. 141–174

[13] Marcus, J.K., Sarah, M.B.: ‘Conditioning the ‘resource curse’: globalization, human capital, and growth in oil-rich nations’, *Comp. Polit. Stud.*, 2011, **44**, (6), pp. 747–770

[14] Tamat, S., Law, S.H., Yaghoob, J.: ‘Resource curse: new evidence on the role of institutions’, *Int. Econ. J.*, 2014, **28**, pp. 191–206

[15] Halvor, M., Karl, M., Ragnar, T.: ‘Institutions and the resource curse’, *Econ. J.*, 2006, **116**, pp. 1–20

[16] Shao, S., Yang, L.L.: ‘Endogenous technological progress and regional economic growth’, *Econ. Res. J.*, 2011, **2**, pp. 112–123

[17] Wu, J., Hu, D.W., Wang, M., *et al.*: ‘Analysis of population evolvement and its effects on resource and environment in Shanghai, China’, *Popul. Resour. Environ.*, 2011, **4**, pp. 164–168

[18] Ren, C.F., Guo, P., Li, M., *et al.*: ‘An innovative method for water resources carrying capacity research – metabolic theory of regional water resources’, *J. Environ. Manag.*, 2016, **167**, pp. 139–146

[19] Wang, R., Cheng, J.H., Zhu, Y.L., *et al.*: ‘Research on diversity of mineral resources carrying capacity in Chinese mining cities’, *Resour. Policy*, 2016, **47**, pp. 108–114

[20] Iratxe, C., Olivier, P., Franck, V.: ‘Common patrimony: a concept to analyze collective natural resource management- the case of water management in France’, *Ecol. Modell*, 2017, **137**, pp. 126–132

[21] Ren, C.F., Li, R.H., Guo, P.: ‘Two-stage DEA analysis of water resource use’, *Sustainability*, 2017, **9**, p. 52

[22] Hasbagen Li, B.S., Bao, Y.: ‘Theoretical model and empirical researches of regional land carrying capacity’, *Sci. Geogr. Sin.*, 2008, **2**, pp. 189–194

[23] Lakshitha, D., Gamini, D., Ravindra, R.: ‘C-LOG: A chamfer distance based algorithm for localisation in occupancy grid-maps’, *CAAI Trans. Intell. Technol.*, 2016, **1**, pp. 272–284

[24] Li, Z., Feng, Q., Chen, W., *et al.*: ‘RPK-table based efficient algorithm for join-aggregate query on MapReduce’, *CAAI Trans. Intell. Technol.*, 2016, **1**, pp. 79–89

[25] Donald, J.L., Nyakundi, M.M., Tesfa, G.: ‘A Bayesian spatial econometric analysis of SNAP participation rates in Appalachia’, *J. Reg. Anal. Policy*, 2012, **42**, pp. 198–209

[26] Anselin, L.: ‘*Spatial econometrics: methods and models*’ (Kluwer Aeademie Publishers, Dordereht, Netherlands, 1988)

[27] Anselin, L., Bera, A.: ‘*Spatial dependence in linear regression models with an introduction to spatial econometrics*’ (Handbook of Applied Economic Statistics, Marcel Dekker, New York, United States of America, 1988)

[28] Shan, H.J.: ‘Estimating the capital stock of China: 1952–2006’, *J. Quant. Tech. Econ.*, 2008, **10**, pp. 17–31