***I. J. Mathematical Sciences and Computing*, 2023, 4, 29-43**
Published Online on December 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmsc.2023.04.04

Study on Influencing Factors of Regional Economy Based on Multilevel Model

Mengjin Yu
School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, 454000, China
E-mail: ymj99086@163.com

Huiyun Bai*
School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, 454000, China
E-mail: 867862063@qq.com
*Corresponding Author

Received: 12 July, 2023; Revised: 21 August, 2023; Accepted: 15 September, 2023; Published: 08 December, 2023

Abstract*:* Based on the relevant data of economic development of 31 provinces in China from 2016 to 2021, this paper selects the level of economic development as the response variable. The explanatory variables are selected from the aspects of location, human capital, industrial structure, foreign trade, system, science and technology and employment level. Aiming at the data with hierarchical structure, a multilevel model is proposed for analysis. It is concluded that human capital, industrial structure, foreign trade and system have a significant impact on the level of economic development. The impact of technology and employment level on the level of economic development is not significant. At the same time, this paper also makes a comparative analysis of the linear regression model, random intercept model and random intercept-slope growth model, which shows the advantages of random intercept-slope growth model.

Index Terms: Correlation analysis, Economic development level, Linear regression, Random intercept-slope growth model, Random intercept model

1. Introduction

Since the reform and opening up, China's economy has been growing rapidly. With the process of building a modern socialist power, economic development has entered a new situation and opened up a new pattern.

With the rapid development of economy, there are also many problems. As the resources, policies, environment, population, technology, capital and other aspects of each region are different [1], rapid development will lead to a certain gap in each region, resulting in the current situation of unbalanced development. Such unbalanced development runs counter to the sustainable development pursued by our country. To a certain extent, the economic gap will promote the cooperation between various regions, which can first develop the areas with advantages, and then promote the development of the surrounding areas. But in the long run, the gap will eventually lead to the continuous influx of technology, capital and population to the developed areas, thus widening the gap [2]. In order to alleviate the various problems caused by economic differences, the state has adopted different policies one after another, such as the large-scale development of the western region to promote regional development, the revitalization of the northeast old industrial base, and the subsequent rise of the east, and other measures have been promulgated, all of which prove the efforts made by the state. Under the implementation of these measures, the overall development differences of the country have also been significantly improved, which also shows to a certain extent. The importance and urgency of alleviating the imbalance of regional development.

The problem of imbalance in regional development is prominent, not only in different regions at different stages of development, but also in economic disparities within the same region. The research on the causes and influencing factors of regional economic differences is of great significance to the coordinated development of China's economy.

Regional economic differences have an important impact on the coordinated development of China's economy, which has also been studied by many scholars.

Yingmin Cai, Xu Wang and Liran Xiong [3] uses clustering analysis and self-organizing feature map (SOM) neural network of hybrid genetic algorithm to cluster the differences of regional economic development, and evaluates the clustering results. The empirical results are good.

Yanyan Li [4] on the basis of summarizing and analyzing the previous research work, the basic model of regional economic development difference based on intelligent hybrid algorithm is constructed, and the principal component analysis and spatiotemporal change analysis of regional economic development are carried out. Finally, simulation experiments are carried out and the results are analyzed.

Hao Xu [5] based on the coefficient of variation and Theil index, this paper calculates and studies the differences of regional economic development in Shaanxi Province from 2000 to 2020. The results show that in the year of study, the differences in economic development of various cities in Shaanxi Province show a trend of expanding at first and then shrinking. It is also found that among the three major regions of Shaanxi Province, Guanzhong region has the greatest difference in economic development, followed by northern Shaanxi, and southern Shaanxi is the relatively stable. From the perspective of the economic development of Shaanxi Province as a whole, Guanzhong accounts for 60% and 80% of GDP in Shaanxi Province, while northern Shaanxi is similar to southern Shaanxi, each occupying about 15%.

Xin Chen, Yuanhua Liu, Cuicui Shi [6] based on the provincial panel data of Shanghai and Anhui Province from 2005 to 2018, the random forest method was used to select other important influencing factors except environmental variables, and then the regression model was used to find the demarcation point of environmental regulation level. According to the demarcation point, the sample was divided into two stages. Finally, based on the Oaxaca-Blinder decomposition method, the reasons for the differences of economic development under different stages of environmental regulation were discussed.

Ju Pan [7] uses Theil index to study the overall differences in the level of inter-provincial and regional economic development in China, as well as intra-regional and inter-regional differences, and uses quantile regression econometric model to analyze the impact of capital, manpower, economic mobility and economic system factors on regional economic differences. It is found that the differences in inter-provincial economic development first increase and then decrease, and inter-regional differences are the main source of overall differences; differences in the level of industrialization also play a relatively important role in intra-regional economic development differences.

2. Index Selection, Variable Interpretation and Data Sources

2.1. Index Selection

There are many factors causing inter-regional economic differences. This paper consults a large number of related literature and selects the following aspects to study on the basis and experience of predecessors.

The level of economic development is the most direct proof of a region's economic development and an important embodiment of regional economic development. the higher the level of economic development, the better the development prospects for the region, and the faster the speed of development will be. and the lower the level of economic development, then the development prospect of the region is relatively not very good, will not provide a lot of convenient conditions to promote development. This will reflect the regional differences between areas with a high level of economic development and areas with a low level of economic development. Therefore, this paper selects the level of per capita economic development as an index to measure inter-regional economic differences.

(1) location

Location refers to a limited space range selected by people when they are engaged in relevant economic activities and has some specific advantages in geographical location [8]. Generally, according to the different subjects or things, the location can be divided into: economic location, geographical location, natural resources location, industrial location and trade location [9]. It is usually based on economic location, while economic location puts more emphasis on the difference of economic interests due to its geographical location. Relatively well-located cities, there will be more people gathered here, more conducive to economic development, and with the passage of time, the scale will continue to expand. The relatively poor location of the city, not only difficult to grow, but also likely to gradually decline, or even decline. Therefore, the location conditions play an important role in the economic development of the region.

(2) Capital

Capital is generally divided into human capital and material capital. Human capital is an index that reflects the knowledge and skills of the labor force itself [10]. Generally speaking, the higher the human capital, the more knowledge and skills the labor force itself has, the higher the value it can produce, the higher the degree of innovation, and the greater the impact on the level of economic development. China has always emphasized the strategy of strengthening the country with talents and attached great importance to the cultivation of talents. As an important component of capital, material capital also plays a certain role in economic development.

(3) Industrial structure

Industrial structure is an indispensable part of economic development and an important force to promote development. Industrial structure is not only a reflection of the allocation of various factors of production among different industries, but also reflects the possession of different industries to factors of production. There should be different division of labor between different industries, and only when the division of labor is clear and reasonable, can we strengthen the relationship between different industries and make them promote each other, resulting in the effect that one plus one is greater than two, so as to better promote the economic development of the whole region. On the basis of Petty theory, British economist Clark further pointed out: "with the increase of national income and the all-round development of economy and society, the center of gravity of a social employment structure will gradually shift from the primary industry to the secondary industry. And transfer to the tertiary industry through the secondary industry [11].

(4) Foreign trade

Foreign trade has an important impact on the economic development of the region. A country's economic level is closely related to its foreign trade. Healthy foreign trade relations can expand a country's trading market, and export trade mainly occupies foreign markets by exporting domestic products. Import trade mainly introduces foreign advanced technology, culture and management experience, so as to learn from its essence and learn from its strong points to make up for its weaknesses, so as to constantly improve its innovative technological capability and improve itself. Launch more advantageous products, thus occupy the foreign market more quickly, and become a virtuous circle [12]. For different regions of a country, the situation of foreign trade is also very different.

(5) System

System, that is, rules, is a criterion used to restrain people's behavior. It can be a well-defined clause, or it can be a concept of behavior that people have in their long life. A good system will promote the market-oriented economy and play a positive guiding role in economic development [13]. This paper mainly uses the level of marketization to reflect it.

(6) Technology

Science and technology occupies an important position in the regional economic development. Science and technology can reflect the economic vitality of the region to a certain extent, and can well show the size of the innovation ability. generally speaking, the more advanced the science and technology is, the stronger the innovation ability is, the more it can promote the economic development of the whole region. and in areas where science and technology is relatively less developed, the level of economic development will lag slightly.

(7) Employment level

The level of employment also plays a vital role in economic development. Generally speaking, the higher the level of employment, the higher the proportion of the labor force entering the market, the more value will be created, the faster the speed of economic development, and the faster the progress of the region, on the contrary, the lower the level of employment. it shows that a large number of labor force can’t be employed, which is not conducive to the development of the region.

2.2. Variable Interpretation

Table 1. Variable interpretation

Variable category	Variable name	Symbol	Variable description	Unit
Response variable	Economic development level	GDP	gross domestic product	Hundred million yuan
Explanatory variable	Human capital	MC	Number of students in ordinary colleges and universities	Ten thousand people
	Industrial structure	IS	Added value of the tertiary industry/GDP	%
	Foreign trade	FT	Total export volume/GDP	%
	System	DOM	(GDP-expenditure)/GDP	%
	Science and Technology	Technology	Number of patent approvals/ Total population	%
	Employment level	EL	Number of employed people/ Total population	%

Location mainly affects regional economic growth directly or indirectly over a long period of time, so it is not reflected in the selection of indicators. Due to the lack of data, only human capital is selected in the index of capital.

2.3. Data Source

Based on the data from 2016 to 2021, this paper studies the factors affecting the economic development level of 31 provinces in China. The index data used in this paper come from "China Statistical Yearbook", "Statistical Yearbook" of provinces and cities and "National data".

3. Model

3.1. Multilevel Model

There are widespread hierarchical structure phenomena in nature and human society, such as the biological system has a natural hierarchical or ethnic structure [14]; the curriculum system has a hierarchical distinction from shallow to deep, from simple to complex and so on. Since stratification is common in nature, the hierarchical structure of society naturally produces multilevel structural data [15]. For example, when studying the influencing factors of medical patients' hospitalization expenses in various hospitals in a certain area, a sufficient number of medical patients are randomly selected from each hospital in this area, including hospitalization expenses, days of hospitalization, duration of illness, age of illness, mode of treatment, conditions of hospitalization, etc., and analyze the influencing factors of hospitalization expenses.

Multilevel model (multilevel models) is a very effective method to deal with hierarchical data. The so-called hierarchical structure means that low-level units are nested in higher-level units, and it is precisely because of this nesting relationship that individuals in the same high-level units have some similar backgrounds but different from other high-level units. The observed data show that the observed data of individuals in the same high-level unit have a certain aggregation, which is called intra-group correlation in statistics, and the individual observed data between different high-level units have a certain heterogeneity [16]. This characteristic of this kind of data does not meet the individual data independence conditions required by traditional analysis methods such as ordinary linear model, so most of the traditional analysis methods will no longer be applicable, and multilevel model is an effective method to analyze this kind of data.

3.1.1. Empty Model

Taking two-level data as an example, we first establish an empty model, which can be called an unconditional two-level model, which is the first step in establishing a multilevel model. The purpose is to test whether there is a difference between the data, so as to judge whether there is a hierarchical structure between the data. If it does not exist, it is not necessary to use a multilevel model. The general form of the empty model is [17]:

$$\text{level 1: } y_{ij} = \beta_{0j} + e_{ij}$$
$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$
$$\text{General model: } y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \tag{1}$$

For level 1, y_{ij} as response variables, β_{0j} to intercept, represents the mean value of group j, $e_{ij} \sim \mathrm{N}(0,\sigma_{e_0}^2)$ is the residual term of level 1; For level 2, β_{0j} as response variables, γ_{00} to intercept, said that the overall average estimate, $u_{0j} \sim N(0,\sigma_{u_0}^2)$ for level 2 residual items, reflects the random effect of group j on y. The residuals between level 1 and level 2 are independent of each other, namely $Cov(u_{0j}, e_{ij}) = 0$.

From the point of the general model, response variable y_{ij} can be expressed as the sum of the fixed part γ_{00} and random part $(u_{0j} + e_{ij})$, Where γ_{00} description model of the fixed effects, and the variance of random part $\sigma_{u_0}^2$ and $\sigma_{e_0}^2$ description model of random effects. Variance of response variable:

$$Var(y_{ij}) = Var(\gamma_{00} + u_{0j} + e_{ij}) = Var(u_{0j} + e_{ij}) = \sigma_{u_0}^2 + \sigma_{e_0}^2 \tag{2}$$

Thus, the total variance of y_{ij} is decomposed into inter-group variance $\sigma_{u_0}^2$ and intra-group variance $\sigma_{e_0}^2$.

Intra-group correlation coefficient ICC can measure the degree of similarity of variables in the same group and is defined as:

$$ICC = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{e_0}^2} \quad (3)$$

Represents the degree of variation at the inter-group level. The value ranges from 0 to 1. The closer it is to 1, the greater the proportion of inter-group variance in the total variance, indicating that the use of multilevel model is reasonable. The closer it is to 0, the more it indicates that the data does not have hierarchical structure. At this time, the model can be simplified to a fixed effect model.

3.1.2. Random Intercept Model

Random intercept is the introduction of explanatory variables on the basis of the empty model to illustrate the effect of explanatory variables on response variables. This explanatory variable can be horizontal 1 unit or horizontal 2 unit. When only horizontal 1 unit exists, the model is [18].

$$\text{level 1: } y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\text{level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\text{General model: } y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + e_{ij} + u_{0j} \quad (4)$$

In the general model, $(\gamma_{00} + \gamma_{10}x_{ij})$ is a fixed part, indicating the fixed effect of the model, $(e_{ij} + u_{0j})$ is a random part, $e_{ij} \sim N(0, \sigma_{e_0}^2)$ is the residual term of level 1, $u_{0j} \sim N(0, \sigma_{u_0}^2)$ is intercept term horizontal 2 residual, represents the random effect of the model. γ_{00} is the total average estimated value, x_{ij} is the observed value of explanatory variable, and γ_{10} is the regression coefficient of explanatory variable x_{ij}, which does not change with the change of level 2 unit. The residual terms between level 1 and level 2 are independent of each other, that is $Cov(e_{ij}, u_{0j}) = 0$.

3.2. Linear Growth Model

Longitudinal data refers to the data obtained by repeated measurements of the same group of individual objects. The growth model is a very effective model for longitudinal data analysis, which is in principle the same as the multilevel model introduced above for other multi-tier data analysis.

In the two-level growth model for analyzing longitudinal data, level 1 is the repeated observation of each research object, and level 2 unit is the individual research object. The general form of random intercept growth model is as follows [19]:

$$\text{level 1: } y_{ij} = \beta_{0j} + \beta_{1j}time_{ij} + e_{ij}$$
$$\text{level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\text{General model: } y_{ij} = \gamma_{00} + \gamma_{10}time_{ij} + u_{0j} + e_{ij} \quad (5)$$

In the general model, $(\gamma_{00} + \gamma_{10}time_{ij})$ is a fixed effect component of the model, $(u_{0j} + e_{ij})$ is the random effect component of the model. y_{ij} is the ith outcome measurement of the research object j, that is, the response variable, γ_{00} is the initial level of the total average outcome measurement, $time_{ij}$ is the time variable of horizontal 1 unit, and γ_{10} is the regression slope of $time_{ij}$. $e_{ij} \sim N(0, \sigma_{e_0}^2)$ is the residual term of level 1, $u_{0j} \sim N(0, \sigma_{u_0}^2)$ is intercept term horizontal 2 residual. The residual terms between level 1 and level 2 are independent of each other, that is $Cov(u_{0j}, e_{ij}) = 0$.

In the above random intercept growth model, the slope of the time variable is set as a fixed effect, that is, the rate of change of the outcome measurement with the passage of time is a fixed coefficient, which will not change with different research objects. In other words, the starting point of individual research objects can be different, but with the passage of time, their rate of change is the same.

However, in practice, the development of outcome measurement will not only have different starting points, but also vary with time, so the random intercept-slope growth model should be more in line with the actual situation. the general form of the model is as follows [20]:

$$\text{level 1: } y_{ij} = \beta_{0j} + \beta_{1j} time_{ij} + e_{ij}$$

$$\text{level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\text{General model: } y_{ij} = \gamma_{00} + \gamma_{10} time_{ij} + u_{0j} + u_{1j} time_{ij} + e_{ij} \quad (6)$$

In the general model, $(\gamma_{00} + \gamma_{10} time_{ij})$ is a fixed effect component of the model, $(u_{0j} + u_{1j} time_{ij} + e_{ij})$ is the random effect component of the model. y_{ij} is the ith outcome measurement of the research object j, γ_{00} is the initial level of the total average outcome measurement, $time_{ij}$ is the time variable of horizontal 1 unit, and γ_{10} is the regression slope of $time_{ij}$. $e_{ij} \sim N(0, \sigma_{e_0}^2)$ is the residual term of level 1, $u_{0j} \sim N(0, \sigma_{u_0}^2)$ is intercept term horizontal 2 residual, represents the degree to which the initial level of the measurement of the outcome of the j individual deviates from the overall average initial outcome estimated by the model, $u_{1j} \sim N(0, \sigma_{u_1}^2)$ is the slope term horizontal 2 residual, indicates the extent to which the rate of change of the outcome of the j individual deviates from the estimated rate of change of the overall average outcome, $Cov(u_{0j}, e_{ij}) = 0$.

Next, add an explanatory variable of level 2, and the model is as follows [21]:

$$\text{level 1: } y_{ij} = \beta_{0j} + \beta_{1j} time_{ij} + e_{ij}$$

$$\text{level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} x_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} x_j + u_{1j}$$

$$\text{General model: } y_{ij} = \gamma_{00} + \gamma_{01} x_j + \gamma_{10} time_{ij} + \gamma_{11} x_j time_{ij} + u_{0j} + u_{1j} time_{ij} + e_{ij} \quad (7)$$

In the general model, $(\gamma_{00} + \gamma_{01} x_j + \gamma_{10} time_{ij} + \gamma_{11} x_j time_{ij})$ is a fixed effect component of the model, $(u_{0j} + u_{1j} time_{ij} + e_{ij})$ is the random effect component of the model. y_{ij} is the ith outcome measurement of the research object j, x_j is the explanatory variable. $time_{ij}$ is a time variable of level 1 unit, γ_{00} and γ_{10} represent the average initial level and average rate of change of the outcome measurement after the explanatory variable x_j of controlling individual level (level 2), respectively. The coefficients γ_{01} and γ_{11} are the regression slope of the explanatory variable x_j, indicating the variation of the initial level and the rate of change of the outcome measurement among individuals, respectively. $e_{ij} \sim N(0, \sigma_{e_0}^2)$ is the residual term of level 1, $u_{0j} \sim N(0, \sigma_{u_0}^2)$ is intercept term horizontal 2 residual, represents the degree to which the initial level of the measurement of the outcome of the j individual deviates from the overall average initial outcome estimated by the model, $u_{1j} \sim N(0, \sigma_{u_1}^2)$ is the slope term horizontal 2 residual, indicates the extent to which the rate of change of the outcome of the j individual deviates from the estimated rate of change of the overall average outcome, $Cov(u_{0j}, e_{ij}) = 0$.

4. Results

4.1. Data Visualization Analysis

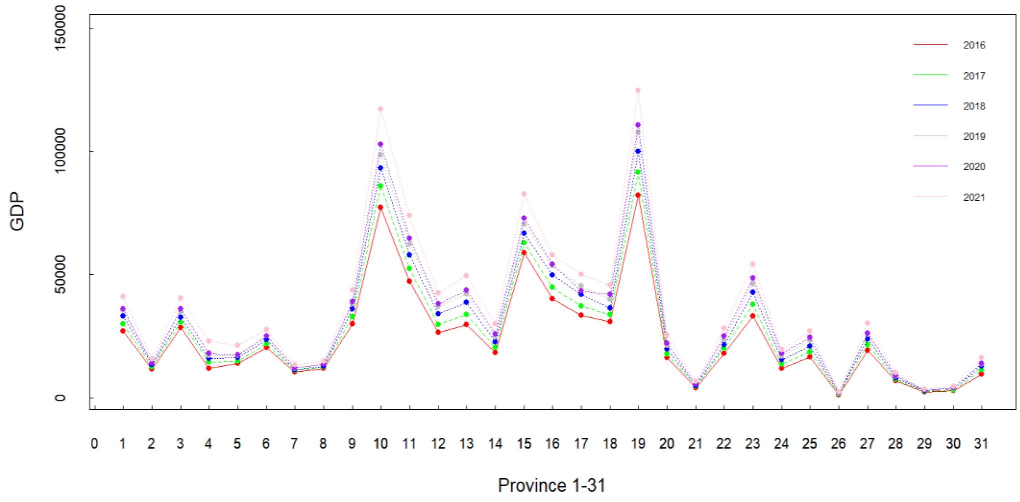

Fig.1. Comparison of Economic Development Levels of Provinces and Cities from 2016 to 2021

Provinces 1 to 31 indicate, in order, Beijing municipal, Tianjin municipal, Hebei province, Shanxi province, Inner Mongolia Autonomous Region, Liaoning province, Jilin province, Shanghai municipal, Jiangsu province, Zhejiang province, Anhui province, Fujian province, Jiangxi province, Shandong province, Henan province, Hubei province, Hunan province, Guangdong province, Guangxi Zhuang Autonomous Region, Hainan province, Chongqing municipal, Sichuan province, Guizhou province, Yunnan province, Tibet Autonomous Region, Shaanxi Province, Gansu Province, Qinghai Province, Ningxia Hui Autonomous Region and Xinjiang Uygur Autonomous Region.

As can be seen from the Fig 1, from an overall point of view, the GDP of each province is increasing year by year, with the lowest in 2016 and the highest in 2021, and the range of increase is not very different from year to year. From the perspective of a single year, taking 2021 as an example, province 19, that is, Guangdong Province, has the highest GDP, the strongest level of economic development, Province 10, that is, Jiangsu Province takes the second place, and Province 26, that is, Tibet has the lowest GDP, and the level of economic development is relatively backward.

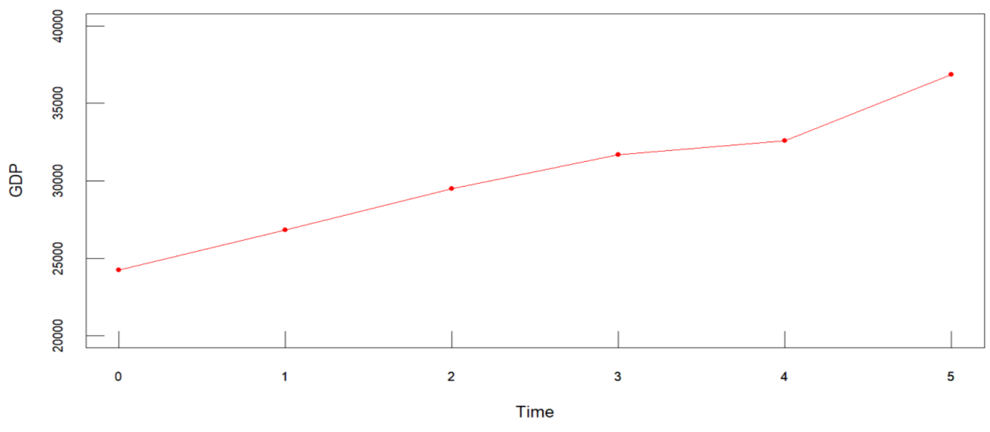

Fig.2. Average trend of GDP from 2016 to 2021

Fig 2 shows the average trend of GDP over the entire study period, with Time0-5 showing 2016, 2017, 2018, 2019, 2020 and 2021, respectively. From the trend in the chart, we can see that the average GDP of the sample in this study showed an upward trend in the study year, and increased linearly in the first four research years. Therefore, this paper uses the linear time model as the basic model in the growth model.

In order to further verify the correlation between explanatory variables and response variables, the author first tests the correlation between explanatory variables and response variables, and the results are as follows:

Table 2. Correlation Test

	GDP	t value	95% confidence interval	p-value
MC	0.8302877	20.208	(0.7795322,0.8702106)	< 2.2e-16
IS	0.03771639	0.51197	(-0.1067422,0.1806156)	0.6093
FT	0.6905506	12.951	(0.6069782,0.7590171)	< 2.2e-16
DOM	0.5379866	8.6572	(0.4271726,0.6328779)	2.401e-15
Technology	0.5781967	9.6128	(0.4737269,0.6666196)	< 2.2e-16
EL	0.2250482	3.1331	(0.08388503,0.35735645)	0.002013

It can be seen from the results that the six explanatory variables selected in this paper have a positive correlation with GDP. Among them, the correlation between MC and GDP is the strongest, which is 0.83, followed by FT, 0.69, while the correlation between IS and GDP is the weakest, only 0.0377.

4.2. Linear Regression Model

First of all, the author first constructs the linear regression model of the economic development level of 31 provinces in China, the model is as follows:

$$GDP = \beta_0 + \beta_1 MC + \beta_2 IS + \beta_3 FT + \beta_4 DOM + \beta_5 Techno\log y + \beta_6 EL \tag{8}$$

After running, the result is as follows:

Table 3. Linear regression results

		Estimate	Std. Error	t value	Pr(>|t|)
Intercept	β_0	-5.586e-17	2.289e-02	0.000	1.00000
MC	β_1	7.153e-01	3.196e-02	22.380	< 2e-16 ***
IS	β_2	-1.262e-01	4.364e-02	-2.891	0.00431 **
FT	β_3	3.128e-01	3.566e-02	8.770	1.37e-15 ***
DOM	β_4	-1.235e-01	3.066e-02	-4.029	8.28e-05 ***
Technology	β_5	2.617e-01	4.466e-02	5.860	2.18e-08 ***
EL	β_6	6.974e-02	5.093e-02	1.369	0.17264

Note: Here *** represents P<0.001, ** represents P<0.01.

It can be seen from the results that MC, IS, FT, DOM, Technology has a significant impact on GDP, in which MC, FT and Technology have a positive effect on GDP. Under the premise that other explanatory variables remain unchanged, each additional unit of MC, GDP will increase 7.153e-01; each additional unit of FT, GDP will increase 3.128e-01; each additional unit of Technology, GDP will increase 2.617e-01. The influence of IS and DOM on GDP is negative. Under the premise that other explanatory variables remain unchanged, each additional unit of IS, GDP will decrease 1.262e-01; each additional unit of DOM, GDP will reduce 1.235e-01. However, the effect of EL on GDP is not significant.

4.3. Application of Multilevel Model

This paper studies the economic development level of 31 provinces in China from 2016 to 2021, taking the sample data of each province as level 1 and different years in each province as level 2 [22]. First of all, an empty model is established to test whether there are differences in the economic development level of 31 provinces in the study year. The model is as follows:

$$\text{level 1: } GDP_{ij} = \beta_{0j} + e_{ij}$$
$$\text{level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$
$$\text{General model: } GDP_{ij} = \gamma_{00} + u_{0j} + e_{ij} \tag{9}$$

Among them, $i = 1, 2, ..., 31$, express in turn Beijing municipal, Tianjin municipal, Hebei province, Shanxi province, Inner Mongolia Autonomous Region, Liaoning province, Jilin province, Shanghai municipal, Jiangsu province, Zhejiang province, Anhui province, Fujian province, Jiangxi province, Shandong province, Henan province, Hubei province, Hunan province, Guangdong province, Guangxi Zhuang Autonomous Region, Hainan province, Chongqing municipal, Sichuan province, Guizhou province, Yunnan province, Tibet Autonomous Region, Shaanxi Province, Gansu Province, Qinghai Province, Ningxia Hui Autonomous Region and Xinjiang Uygur Autonomous Region. $j = 1, 2, ..., 6$, express in turn 2016, 2017, 2018, 2019, 2020 and 2021. γ_{00} is the total intercept and is the fixed effect of the model. $(u_{0j} + e_{ij})$ is the random effect of the model; $\sigma^2_{u_0}$ is the inter-provincial variance; $\sigma^2_{e_0}$ is the intra-provincial variance or the individual level variance.

Run the empty model with R software, and the results are as follows:

Table 4. Random effect of empty model

Random effect		variance	Standard deviation
Inter-provincial variance	$\sigma^2_{u_0}$	609413328	24686
intra-provincial variance	$\sigma^2_{e_0}$	32776888	5725

Table 5. Fixed effect of empty model

Fixed effect		Estimated value	Standard error	t- the statistic	df	P-value
GDP total mean	γ_{00}	30281	4454	6.799	30	1.53e-07 ***

Note: Here *** represents P<0.001.

The test results of the empty model show that the estimated value of the only intercept term γ_{00} is 30281, that is, the estimated value of GDP among provinces is 30281, and passed the significance test of the parameters, which is significant at the level of significance.

From the calculation formula of intra-group correlation coefficient, it can be obtained $ICC = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_{e_0}} = \frac{609413328}{609413328 + 32776888} \approx 0.94896$. It shows that 94.896% of the variation is caused by the difference between different provinces, indicating that the difference of GDP in different provinces is very obvious, and there is significant intra-provincial variation, which is statistically significant. Therefore, the use of multilevel model is reasonable.

Let's add explanatory variables to the empty model to establish a two-level intercept model as follows:

level 1: $GDP_{ij} = \beta_{0j} + \beta_{1j}MC_{ij} + \beta_{2j}IS_{ij} + \beta_{3j}FT_{ij} + \beta_{4j}DOM_{ij} + \beta_{5j}Techno\log y_{ij} + \beta_{6j}EL_{ij} + e_{ij}$

$$\text{level 2: } \beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10},\ \beta_{2j} = \gamma_{20},$$
$$\beta_{3j} = \gamma_{30},\ \beta_{4j} = \gamma_{40},\ \beta_{5j} = \gamma_{50},\ \beta_{6j} = \gamma_{60}$$

General model:

$$GDP_{ij} = \gamma_{00} + \gamma_{10}MC_{ij} + \gamma_{20}IS_{ij} + \gamma_{30}FT_{ij} + \gamma_{40}DOM_{ij} + \gamma_{50}Techno\log y_{ij} + \gamma_{60}EL_{ij} + u_{0j} + e_{ij} \tag{10}$$

Because of the different units among the variables, the author first standardizes the data, and then runs the model, and the results are as follows:

Table 6. Random intercept model results

Fixed effect		Estimated value	Standard error	t value	df	Pr(>|t|)
Intercept	γ_{00}	7.427e-15	7.545e-02	0.000	2.034e+01	1.0000
MC	γ_{10}	5.727e-01	4.709e-02	12.161	1.641e+02	<2e-16 ***
IS	γ_{20}	5.147e-02	3.680e-02	1.399	1.790e+02	0.1637
FT	γ_{30}	1.047e-01	4.911e-02	2.131	1.536e+02	0.0346 *
DOM	γ_{40}	-1.286e-02	4.772e-02	-0.269	1.627e+02	0.7879
Technology	γ_{50}	2.274e-01	2.125e-02	10.701	1.719e+02	<2e-16 ***
EI	γ_{60}	6.691e-03	5.403e-02	0.124	1.484e+02	0.9016
Random effect		Variance	Standard deviation			
Inter-provincial variance	$\sigma_{u_0}^2$	0.17472	0.4180			
intra-provincial variance	$\sigma_{e_0}^2$	0.01046	0.1023			

Note: Here *** represents P<0.001, ** represents P<0.01, * represents P<0.05, . represents P<0.1.

From the results in the table, we can see that the estimated value of the intercept term is 7.427e-15. The impact of GDP on GDP is significant, and their effects are all positive, and the impact of MC is the greatest. Each additional unit of MC, GDP will increase 5.727e-01, followed by Technology, and the impact of FT is relatively minimum. The other three variables IS, DOM, EI in the results show that there is no significant impact on GDP, which is different from the previous correlation test, the author considers that it may be due to the influence of the interaction between explanatory variables.

Next, take the data as vertical data, the different annual data of each province as repeated measurement data, as level 1, the sample data of each province as individuals, for level 2, establish a random intercept growth model, the model is as follows:

$$\text{Level 1: } GDP_{ij} = \beta_{0j} + \beta_{1j} time_{ij} + e_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\text{General model: } GDP_{ij} = \gamma_{00} + \gamma_{10} time_{ij} + e_{ij} + u_{0j} \quad (11)$$

The results are as follows:

Table 7. Results of random intercept growth model

Fixed effect		Estimated value	Standard error	t value	df	Pr(>|t|)
Intercept	γ_{00}	24381.14	4470.87	5.453	30.47	6.17e-06 ***
time	γ_{10}	2360.12	156.97	15.035	154.00	< 2e-16 ***
Random effect		Variance	Standard deviation			
Inter-provincial variance	$\sigma_{u_0}^2$	612648177	24752			
intra-provincial variance	$\sigma_{e_0}^2$	13367631	3656			

Note: Here *** represents P<0.001.

According to the results in the table, the intra-group correlation coefficient $ICC=\frac{\sigma_{u_0}^2}{\sigma_{u_0}^2+\sigma_{e_0}^2}=\frac{612648177}{612648177+13367631}\approx 0.9786$, that is, 97.86% of the variation is caused by inter-provincial heterogeneity.

Next, a more realistic random intercept-slope growth model is constructed. The general form of the model is as follows:

$$\text{Level 1: } GDP_{ij}=\beta_{0j}+\beta_{1j}time_{ij}+e_{ij}$$

$$\text{Level 2: } \beta_{0j}=\gamma_{00}+u_{0j}$$

$$\beta_{1j}=\gamma_{10}+u_{1j}$$

$$\text{General model: } GDP_{ij}=\gamma_{00}+\gamma_{10}time_{ij}+u_{0j}+u_{1j}time_{ij}+e_{ij} \tag{12}$$

The results are as follows:

Table 8. Results of random intercept-slope growth model

Fixed effect		Estimated value	Standard error	t value	df	Pr(>|t|)
Intercept	γ_{00}	24381.1	3617.4	6.740	30.0	1.80e-07 ***
time	γ_{10}	2360.1	342.5	6.891	30.0	1.19e-07 ***
Random effect		Variance	Standard deviation			
Inter-provincial variance	$\sigma_{u_0}^2$	405018050	20125			
time	$\sigma_{u_1}^2$	3567398	1889			
intra-provincial variance	$\sigma_{e_0}^2$	1206021	1098			

Note: Here *** represents P<0.001.

The random intercept growth model is compared with the random intercept-slope growth model. From the above results, with the increase of random slope in the model, the horizontal 1 residual variance decreases from 13367631 to 1206021, a decrease of 90.98%. In order to make it more convincing, the author makes an analysis of variance between the two models, and the results are as follows:

Table 9. Variance analysis

	npar	AIC	BIC	logLik	deviance	Chisq	df	Pr(>Chisq)
random intercept growth model	4	3760.0	3772.9	-1876.0	3752.0			
random intercept-slope growth model	6	3412.3	3431.6	-1700.1	3400.3	351.73	2	< 2.2e-16 ***

Note: Here *** represents P<0.001.

It can be seen that the random intercept-slope growth model considering slope term is better, and there is significant difference.

Furthermore, the random intercept-slope growth model is used to add explanatory variables to evaluate the effect of explanatory variables on response variables. the model is as follows:

$$\text{Level 1: } GDP_{ij}=\beta_{0j}+\beta_{1j}time_{ij}+e_{ij}$$

$$\text{Level 2: } \beta_{0j}=\gamma_{00}+\gamma_{01}MC_j+\gamma_{02}IS_j+\gamma_{03}FT_j+\gamma_{04}DOM_j+\gamma_{05}Technology_j+\gamma_{06}EL_j+u_{0j}$$

$$\beta_{1j}=\gamma_{10}+u_{1j}$$

General model:

$$GDP_{ij} = \gamma_{00} + \gamma_{01}MC_j + \gamma_{02}IS_j + \gamma_{03}FT_j + \gamma_{04}DOM_j + \gamma_{05}Techno\log y_j + \gamma_{06}EL_j + \gamma_{10}time_{ij} + u_{0j} + u_{1j}time_{ij} + e_{ij} \quad (13)$$

In this model, the author assumes that the influence of all explanatory variables on outcome measurement does not change with time, that is, there is no interaction between explanatory variables and time variables. The results are as follows:

Table 10. Random intercept-slope growth model with explanatory variables added

Fixed effect		Estimate	Standard error	t value	df	Pr(>|t|)
Intercept	γ_{00}	-2.617e-01	1.419e-01	-1.844	2.438e+01	0.077310 .
time	γ_{10}	1.047e-01	1.628e-02	6.431	3.494e+01	2.11e-07 ***
MC	γ_{01}	-1.070e-01	5.346e-02	-2.002	1.345e+02	0.047326 *
IS	γ_{02}	-4.450e-02	2.097e-02	-2.122	1.380e+02	0.035635 *
FT	γ_{03}	1.002e-01	2.498e-02	4.011	6.393e+01	0.000161 ***
DOM	γ_{04}	4.422e-02	2.315e-02	1.910	7.363e+01	0.060025 .
Technology	γ_{05}	9.963e-04	2.046e-02	0.049	1.412e+02	0.961240
EL	γ_{06}	2.915e-02	2.981e-02	0.978	8.914e+01	0.330802
Random effect		Variance	Standard deviation			
Inter-provincial variance	$\sigma_{u_0}^2$	0.615870	0.78477			
time	$\sigma_{u_1}^2$	0.006956	0.08341			
intra-provincial variance	$\sigma_{e_0}^2$	0.001442	0.03797			

Note: Here *** represents P<0.001, ** represents P<0.01, * represents P<0.05, . represents P<0.1.

After considering the explanatory variables, the influence of time, MC, IS, FI, DOM on the response variables is significant, while the effects of Technology and EL on the response variables are not significant. Among them, the influence of MC and IS on the response variable is negative, while the influence of other variables on the response variable is positive.

Under the premise of keeping other explanatory variables unchanged, each additional unit of MC, GDP will reduce 1.070e-01, and each additional unit of IS, GDP will decrease-4.450e-02. The effects of these two explanatory variables on GDP are negative and significant. The influence of time variable on GDP is also very great. It can be seen from the results that the P value of time is very significant. In terms of the form of the model, each additional unit of time, GDP will increase 1.047e-01. It can be seen that GDP increases with time, which is a condition that can’t be ignored.

The effects of FT, DOM, Technology and EL on GDP are positive, but the P values of Technology and EL are not significant, and the effect of FT on GDP is very significant. When other explanatory variables remain unchanged, each additional unit of FT, GDP increases 1.002e-01, and each additional unit of DOM increases 4.422e-02.

Table 11. Comparison of results of linear regression model, random intercept model and random intercept-slope growth model

		linear regression model	random intercept model	random intercept-slope growth model
Random effect				
Inter-provincial variance	$\sigma_{u_0}^2$		0.17472	0.615870
intra-provincial variance	$\sigma_{e_0}^2$		0.01046	0.001442
time	$\sigma_{u_1}^2$			0.006956

Fixed effect				
Intercept	γ_{00} (β_0)	1.00000	1.0000	0.077310 .
time	γ_{10}			2.11e-07 ***
MC	γ_{01} (β_1)	< 2e-16 ***	<2e-16 ***	0.047326 *
IS	γ_{02} (β_2)	0.00431 **	0.1637	0.035635 *
FT	γ_{02} (β_3)	1.37e-15 ***	0.0346 *	0.000161 ***
DOM	γ_{03} (β_4)	8.28e-05 ***	0.7879	0.060025 .
Technology	γ_{04} (β_5)	2.18e-08 ***	<2e-16 ***	0.961240
EL	γ_{05} (β_6)	0.17264	0.9016	0.330802

Note: Here *** represents P<0.001, ** represents P<0.01, * represents P<0.05, . represents P<0.1.

Table 12. AIC comparison of results of linear regression model, random intercept model and random intercept-slope growth model

	df	AIC
linear regression model	8	103.6227
random intercept model	9	-134.6078
random intercept-slope growth model	12	-319.9854

Linear regression does not distinguish between intra-provincial variance and inter-provincial variance, the result is that only EL has no significant effect on GDP, others have strong significance, random intercept model gives inter-provincial variance is 0.17472, intra-provincial variance is 0.01046, only MC and Technology have strong significance on GDP, FT has significant effect on GDP, the other three variables have no significant effect on GDP. The provincial variance of the random intercept-slope growth model is 0.001442, which is significantly lower than that of the random intercept model.

The information standard statistics of the three models are compared with AIC, and the results show that the random intercept-slope growth model is better.

5. Summary, Conclusion, and Future Works

Based on the economic development level data of 31 provinces in China from 2016 to 2021, this paper selects GDP as the response variable to study the factors that affect the economic development of provinces from six aspects: human capital, industrial structure, foreign trade, system, science and technology and employment level. By consulting a variety of literature, we finally choose to use the number of students in ordinary colleges and universities to represent human capital, the added value of the tertiary industry / GDP to reflect the industrial structure, the total export / GDP to represent foreign trade, (GDP-financial expenditure) / GDP to represent the system, patent approval / total population to represent the scientific and technological situation, employment population / total population to represent the employment level for specific data analysis.

The advantages and innovations of this paper are as follows:

(1) First, a general linear regression model is established, and then the general multi-level model is adopted considering the multi-level structure of the data.

(2) The data are regarded as longitudinal data and the random intercept-slope development model is fitted.

(3) The model is compared and studied.

The comparison of the first two models can reflect that when dealing with data with hierarchical structure, the multi-level model has significant advantages, can better reflect the reasons for the differences, and distinguish the differences between groups and within groups, so as to better deal with the differences. The comparison of the latter two models can show that even if the hierarchical data have different data characteristics, the applicable model forms are also different, and the conclusions drawn will be different, so as to remind the importance of model selection.

For longitudinal data, the time variable is a factor that can't be ignored, and its influence on the response variable is very significant.

For the selected influencing factors, this paper obtains the impact on GDP, and uses it to guide the government to make corresponding adjustments, which has important practical significance. The details are as follows:

(1) The influence of human capital and industrial structure on GDP is significant and negative. Local governments should continue to invest resources in local talent education to train high-quality talents, but at the same time pay attention to the proportion of investment; they should also pay attention to the situation of the local tertiary industry, pay attention to the proportion of the increase of the tertiary industry, and adjust it in time to make it in the most reasonable state.

(2) Foreign trade and institutions are indispensable to the development of a country, and they have a very significant impact on GDP and play a positive guiding role in economic development. According to the local situation, the local government should formulate trade policies suitable for development and a good economic system that can promote development, so as to maximize the role of promoting economic development.

The economic differences between different regions exist objectively [23]. Local governments should make targeted policies in the light of local actual conditions, strengthen exchanges and development with various regions, and guide reasonable competition between regions. Complementary advantages, clear division of labor, avoid the emergence of huge economic differences, is the coordination and sustainability of China's economic development.

It is worth noting that in the results of this paper, the influence of science and technology and employment level on GDP is not significant, whether this is the result of the interaction between explanatory variables, which is worthy of our further discussion. In addition, when constructing the random intercept-slope growth model, the author makes the assumption that there is no interaction between the explanatory variable and the time variable Time. If there is no such assumption, whether the results will be different or not is a question that can be further studied.

References

[1] He, Pengfei, "Summary on the Difference Study of Chinese Regional Economic Development," Social science trends, no. 02, pp. 54-62, 2021.

[2] Sheng Lei and Zhang Ziyu, "Research on the factors of economic divergence between the north and south of China: from the perspective of regional innovation capability differences," Journal of Huazhong Normal University (Natural Science Edition), vol. 55, no. 05, pp. 727-740, 2021.

[3] Cai Yingmin, Wang Xu and Xiong Liran, "Difference Analysis of Regional Economic Development Based on the SOM Neural Network with the Hybrid Genetic Algorithm," *Computational Intelligence and Neuroscience*, 2021.

[4] Li Yanyan, "Analysis of Regional Economic Development Differences Based on Intelligent Hybrid Algorithm," Complex, 2021.

[5] Xu Hao, "Analysis on the differences of regional economic development in shaanxi Province: based on the coefficient of variation and Theil index," National circulation economy, no. 10, pp. 24-27, 2022.

[6] Chen Xin, Liu Yuanhua and Shi Cuicui, "Research on regional economic development differences by integrating Oaxaca-Blinder decomposition method from the perspective of environment regulation," Resources and industry, vol. 24, no. 03, pp. 63-70, 2022.

[7] Pan Ju, "Analysis of regional Economic Differences Based on the Quantile Regression Measurement Model," Journal of Shenyang University (Social Science Edition), vol. 23, no. 04, pp. 381-387, 2021.

[8] Sun Cuilan, Course of regional economics, Peking University Press, 2008.

[9] Yang Hongru, "Regional economic growth difference and influence factors analysis of Henan province," Jilin University, Economics and management science, 2016.

[10] Wang Chunfang, "Study on regional differences and influencing factors of high-quality economic development in China," Hebei University, Economics and management science, 2022.

[11] Qi Liangshu, economics for development, Beijing: China Development Publishing House, pp. 103-105, 2002.

[12] Mu Lan, "On the Level Measurement, Factors of High-Quality Economic Development in Yangtze River Delta Region," Journal of Dali University, vol. 8, no. 07, pp. 44-53, 2023.

[13] Gao Zhigang, Zhao Zhenjing and Han Yanling, "Measurement of High-Quality Economic Development Level in Western Urban Agglomerations and an Analysis of Its Influencing," Xinjiang finance and economics, no. 03, pp. 5-16, 2023.

[14] Ma Chenchen and Ouyang Jing and Xu Gongjun, "Learning Latent and Hierarchical Structures in Cognitive Diagnosis Models", Psychometrika, 2022.

[15] J.Ch.Wang, Multilayer statistical analysis model, Higher Education Press, 2008.

[16] J.Q.Fang, Statistical Methods for Biomedical Research, 2nd ed., Higher Education Press, 2019.

[17] Li Ying, Cao Jichen, Shi Mei, Meng Yao and Cui Li, "Study on Multilevel Statistical Analysis Model Building," 2011 International Conference on Future Computer Science and Application (FCSA 2011), 2011.

[18] Ren Yan, Huang Yunxiang, Zhang Yuanjin, Yao Minghong, Jia Yulong, Yang Min and Sun Xin, "Multilevel model and its application in evaluation of medicine policy intervention," China Journal of Evidence-based Medicine, vol. 21, no. 12, pp. 1474-1479, 2021.

[19] Li Xiaosong, Zhang Wentong and Ni Zongzan, "Multilevel Models in Analysis of Crossover Design," Chinese Journal of Health Statistics, no. 05, pp. 18-20, 1999.

[20] Shi Lei, Multilevel model and its statistical diagnosis, Beijing: Science Press, 2008.

[21] Wang Jichuan, Xie Haiyi and Jiang Baofa, Multilayer Statistical Analysis Model —— Method and Application, Beijing: Higher Education Press, 2008.
[22] Yu Ze, "Analysis on the quality difference and influencing factors of basic education between urban and rural areas in China based on multilevel model," Yunnan university of finance and economics, Social science Ⅱ, 2020.
[23] Cao Wanru, "Research on unbalanced growth of regional economy in China," Hebei GEO University, Economics and management science, 2022.

Authors' Profiles

Mengjin Yu is a student for Master degree for applied Mathematics in the School of Mathematics and Information Science at Henan Polytechnic University in China and her research interests are in applied statistics.

Huiyun Bai is from Zhengzhou, Henan Province.She is an M.Ec. graduate in Statistics from Beijing Wuzi University.Her main research topics are Economic Statistics and Spatial Statistics.She is currently working as a teacher and teaches "Mathematical Statistics" for postgraduates and "Advanced Mathematics" for undergraduates in School of Mathematics and Information Science, Henan Polytechnic University.

How to cite this paper: Mengjin Yu, Huiyun Bai, "Study on Influencing Factors of Regional Economy Based on Multilevel Model", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.9, No.4, pp. 29-43, 2023. DOI: 10.5815/ijmsc.2023.04.04