UDC 519.233.5

Qi D.

Research on factors affecting differences in investment attractiveness based on HLM model

Recommended for publication by professor Bure V. M.

1. Abstract. This article analyzes the influencing factors and formation mechanism of investment attractiveness differences from the national and provincial levels, which is of great significance to promoting economic development and improving people's living standards. Based on the Hierarchical Linear Model's principles and methods and based on the survey data of mainland China's provinces and provincial administrative regions, empirical analysis is carried out on the influencing factors and formation mechanisms of investment attractiveness differences. We discovered that the investment gravitational difference has a complicated mechanism of action, and it is jointly solved from multiple levels and perspectives such as total import and export trade, taxation, and per capita income.

2. The model.

2.1. Overview of Hierarchical linear Model. The design of the sections using commands Hierarchical Linear Model (HLM) was proposed by Bryk and Raudenbush in 1992. It is based on the ordinary linear regression analysis model and can process individual and group-level data simultaneously. When processing the seal data, first establish a regression equation with the first level (individual level) explanatory variables, then use the intercept and slope in the equation as dependent variables, and then use the second level (group level) explanatory variables as independent variables, perform a second regression. The basic form is

$$L_1 : Y_{ij} = \beta_{0j} + \beta_{1j} X_m + r_{ij}, \tag{1}$$

$$L_2 : \beta_{0j} = \gamma_{00} + \gamma_{01} W_n + \mu_{0j}, \tag{2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_n + \mu_{1j}. \tag{3}$$

Qi Dongfang – graduate student, Saint Petersburg State University; e-mail: st073409@student.spbu.ru, phone: +7(981)810-21-63

In formulas (1) to (3): L_1 and L_2 represent two levels respectively; Y_{ij} is the dependent variable, i represents each unit of the first level, j represents each unit of the second level; X_m, W_n are the first and second levels, respectively level explanatory variables, m and n ask the number of independent variables at the first and second levels respectively; r_{ij}, μ_{0j}, μ_{1j} are random effects, indicating the unexplained part of the independent variable; intercept β_{0j} is the average value of Y_{ij}, that is, the group average—the result of the centering processing of the number or the total average; β_{1j} represents the regression slope.

In this article, the maximum likelihood estimation method is used to estimate the regression equations' parameters at each level through the HLM software.

2.2. Model construction. Taking "investment attractiveness" as the explained variable, to eliminate the variance of the data and smooth the data, it is necessary to perform logarithmic processing on the income data and construct a multi-level linear model shown in table 1. To use a multi-level linear model to study the income differences of rural households, we first need to test the model's usability.

1. As described in Model 1, this model is ANOVA model and does not include any independent variables. It only has random effects at the individual level. It is to determine how much of the overall variation of the investment attractiveness logarithm lny is caused by the group level to test the multi-level model's feasibility for research.

ANOVA model:

$$L_1 : \ln Y_{ij} = \beta_{0j} + r_{ij}, \quad L_2 : \beta_{0j} = \gamma_{00} + \mu_{0j}, \tag{4}$$

where $\ln Y_{ij}$ is the logarithm of investment attractiveness; β_{0j} is the average value of the j-th two-story unit $\ln Y_{ij}$; r_{ij} is the variation of the j-th two-story unit $\ln Y_{ij}$; r_{00} is the total average of all two-level units $\ln Y_{ij}$; μ_{0j} is the second layer Random component of the equation.

2. Model 2 is a random coefficient regression model, which only contains individual-level variables, but allows the intercept and regression coefficients of the individual-level equations to vary randomly between each group level to measure the individual level's effect on investment attractiveness.

Random coefficient regression model:

$$\begin{aligned} L_1 : \ln Y_{ij} &= \beta_{0j} + \beta_{nj} X_n + r_{ij}, \\ L_2 : \beta_{0j} &= \gamma_{00} + \mu_{0j}, \\ \beta_{nj} &= \gamma_{n0} + \mu_{nj}, \end{aligned} \tag{5}$$

where X_n is the n-th independent variable of the first level; β_{nj} is the regression slope of the n-th independent variable of the first level; r_{nj} is the average value of β_{nj}; μ_{nj} is the random component of β_{nj}.

3. Model 3 adds group-level variables to the intercept term based on controlling individual-level variables, measures the group's direct effect on individuals, and allows individual-level variables to vary randomly between groups. We regard it as a particular type of complete model form:

$$\begin{aligned} L_1 : \ln Y_{ij} &= \beta_{0j} + \beta_{nj} X_n + r_{ij}, \\ L_2 : \beta_{0j} &= \gamma_{00} + \gamma_{0m} W_m + \mu_{0j}, \\ \beta_{nj} &= \gamma_{n0} + \mu_{nj}, \end{aligned} \tag{6}$$

where W_m is the m independent variables of the second level, and r_{0m} is the regression slope of the m-th independent variable of the second level.

4. Model 4 is a complete model, through the result of the interaction of the combination of individuals to explore the difference in investment attractiveness process, that is, individuals to the investment attractiveness of indirect adjustment, the second level of variable coefficient symbol and the first level of corresponding variable coefficient symbol the same play a strengthening role, on the contrary, play a weakening role:

$$\begin{aligned} L_1 : \ln Y_{ij} &= \beta_{0j} + \beta_{nj} X_n + r_{ij}, \\ L_2 : \beta_{0j} &= \gamma_{00} + \gamma_{0m} W_m + \mu_{0j}, \\ \beta_{nj} &= \gamma_{n0} + \gamma_{nm} W_m + \mu_{nj}, \end{aligned} \tag{7}$$

where r_{nm} is the slope used by the m-th independent variable in the second layer to explain the slopes of the respective variables in the first layer, reflecting the cross-level interaction.

3. Result analysis.

3.1. ANOVA model analysis. The analysis results are shown in table 1. In the variance analysis, the between-group variance is 1,364 and is statistically significant at the 0.1% level, indicating that there are significant differences between the groups, with obvious stratification characteristics. The intra-group correlation coefficient $ICC = \mathrm{Var}\,(U_0)\,/\,[\mathrm{Var}\,(\mathrm{U}_0) + \mathrm{Var}(R)] = 1.364/(1.364 + 0.370) = 78\%$ indicates that 78% of the differences between groups are caused by the group level, which is not negligible. Therefore, the multi-layer linear model can be used for analysis.

Table 1. Parameter Estimation

Random Effect	**Standard Deviation**	**Variance Component**	**df**	**Chi-square**	**P-value**
Intrcpt1, U_0	1.168	1.364	21	2032.411	0.000
level-1, R	0.608	0.370			

Table 2. HLM analysis of investment attractiveness

	Model 2		**Model 3**		**Model 4**	
Fixed effect	coefficient	standard error	coefficient	standard error	coefficient	standard error
x_1	−0.103	0.076	−0.155**	0.05	−0.089	0.050
x_2	1.559***	0.363	1.700***	0.436	0.975***	0.159
x_3	0.002***	0.0003	0.0002***	0.0003	0.00052***	0.0003
x_4			0.498***	0.117	1.818**	0.428
x_5			0.517***	0.107	−1.418**	0.466
x_1x_4			0.242*	0.116	0.374**	0.141
x_1x_5			−0.478***	0.136	−0.673***	0.154
x_2x_4			−0.136	0.147	1.392**	0.522
x_2x_5			0.292	0.193	−2.334***	0.572
x_3x_4			−0.008	0.008	−0.0015*	0.007
x_3x_5			−0.029**	0.011	−0.0015	0.009
Var(R)	0.712***		0.098***		0.468***	
level1 Var(R)	0.336		0.365		0.310	

In table 2 "$*$" means $P < 0.05$, "$**$" means $P < 0.01$, "$***$" means $P < 0.001$; x_1 is cash income per capita; x_2 – volume of work performed for the type of activity "Construction"; x_3 – GRP(Gross Regional Product); x_4 – taxes levied; x_5 – total import and export trade.

3.2. Random coefficient regression model analysis. In the analysis of Model 2, Model 3, and Model 4, the paper expresses the explanatory variables' variance components at the individual level.

It only shows the changes in the variance between groups and the first-level residuals. The fixed effect represents the influence effect of the independent variables of level 1 and level 2 on the investment attractiveness difference. The random effect means the investment attractiveness difference that the independent variable does not explain. The positive influence on investment attractiveness is x_1, but it is not significant. The ones that have a significant negative impact on investment attractiveness are x_2 and x_3.

3.3. Complete mode analysis. The direct effect of x_4 and x_5 is the significant positive effect of x_4 and x_5. Under the cross-level interaction: x_4 has a moderately significant strengthening effect on the investment attractiveness of x_2, indicating that the enhancement of x_4 can promote the influence of x_2 on investment attractiveness. In the same way, x_4 also played a highly significant strengthening effect on x_2. And x_5 has a high significant weakening effect on x_1 and x_2, and x_4 has a low significant weakening effect on x_3.

4. Conclusion. The difference in investment attractiveness is the result of a combination of multiple factors. It is neither an individual regional issue nor a pure national policy issue, but a regional economic development, policy system, historical background and many other interrelated issues. All aspects are resolved together.

References

1. Bure V. M., Parilina E. M. Probability Theory and Mathematical Statistics. First Edition, SPb: "Lan", 2013, 416 p.

2. National Bureau of Statistics [Internet resource]: `URL:http://www.stats.gov.cn/tjsj/ndsj` (date: 15.02.21).

3. Qi D. Study of the investment attractiveness of China's regions // Control processes and stability. 2020. Vol. 7. No 1. P. 423–428.