Table 7. The hidden variables of the yearly investment, output and employment of Beijing

Year	Investment status (inv.,ξ1)	Output status (GDP, ξ2)	Employment status (emp. ,ξ3)
2003	1.415	1.906	1.797
2004	1.495	0.969	1.629
2005	1.575	1.177	1.159
2006	0.462	1.063	0.642
2007	0.475	0.486	0.172
2008	0.207	0.106	-0.021
2009	0.005	-0.174	-0.189
2010	-1.344	-0.491	-0.259
2011	-1.014	-0.79	-0.621
2012	-0.872	-0.838	-0.753
2013	-0.371	-0.833	-1.09
2014	-0.989	-1.149	-1.09
2015	-1.043	-1.432	-1.378

As there's stronger pertinence between hidden variables and manifest variables, the PLS regression model of hidden variables for manifest variables can be built. With the hidden variables as the dependent variables and the manifest variables as the independent variables, PLS regression can be conducted, and the relationship between hidden variables and manifest variables can be reached:

ξ_1 =0.3636 inv1 - 0.3661 inv2 - 0.3071 inv3

ξ_2 =0.3764 GDP1 + 0.3111 GDP2 - 0.3826 GDP3

ξ_3 =0.2666 emp1 + 0.4320 emp2 - 0.4786 emp3

In addition, there's a certain dependency among investments, output and employment, the regression model among components can be sought or built. For the case, the polynomial of hidden variables can be built. After conducting polynomial regression, the relationship of employment (ξ3), investments (ξ1) and output (ξ2) can be got:

$\xi3 = -0.2407 + 0.2406(\xi1)^2 + 0.8893\xi2$

t-statistical magnitude: (-2.61) (3.32) (14.47)

p-value:0.02 0.008 0.000

F-statistical magnitude=148.89, adjusted determination coefficient $\bar{R}^2$=0.96.

The model fits the data well and the coefficients and the overall model all passed the test, with the statistical significance. From another perspective, it indicated that Beijing employment status is influenced by the investment status, output status and the quantitative relation among the three factors.

5. SUMMARY

The paper puts forward a method of building the multiple regression models under the condition that dependent variables and independent variables are all compositional data. As the compositional data have the fixed-sum constraint, the application condition of the classic least square regression method is completely destroyed, so the classic regression method cannot be used for modeling. The paper spreads the compositional data to the whole real number field through logratio transformation, reduces the dimension and extracts new aggregate variables – hidden variable for the multi-dimensional compositional data after transformation through building the PLS path model, and researches the multiple regression relationship among hidden variables. In the modeling process, the method can meet the fixed-sum constraint of compositional data, overcome the adverse effect of complete multicollinearity on modeling in the compositional data, and highlight the thematic meaning of compositional data and its effect and significance in modeling. To further specify the working process of the multiple compositional data regression modeling method, the paper applied the suggested method, used the structural data of the investments, GDP and employment of Beijing three industries, and built the regression model among the employment status, investment status and GDP of Beijing three industries. The case study indicated that the modeling method raised in the paper can provide an effective technological approach to solve such problems, with important application value.

ACKNOWLEDGEMENTS

This paper was funded by three projects: BIPT-POPME; Development Research Centre of Beijing New Modern Industrial Area (2016); BIPT-ER (2014); URT2017J00013.

REFERENCES

[1] Aitchison J. (1986). The statistical analysis of compositional data. London: Chapman and Hall.

[2] Chin WW. (1998). The partial least squares approach for structural equation modeling. in: G.A. Marcoulides (Ed.) Modern Methods for Business Research, Lawrence Erlbaum Associates, 295-336.

[3] Guinot C, Latreille J, Tenenhaus M. (2001). PLS path modellind and multiple table analysis. Application to the cosmetic habits of women in Ile-de-France. Chemometrics and Intelligent Laboratory Systems 58: 247-259.

[4] Lohmöller JB. (1989). Latent variables path modeling with partial least squares. Physica-Verlag, Heildelberg 34(1): 110-111.

[5] Bayol MP, Foye ADL, Tellier C, Tenenhaus M. (2000). Use of PLS path modeling to estimate the European consumer satisfaction index (ECSI) model. Statistica Applicata – Italian Journal of Applied Statistics 12(3): 361-375.

[6] Wang HW, Huang W. (2013). Linear regression model of compositional data. System Engineering (2).