Using Partial Least Square Approach to Study the Factors affecting the Quality of Life in Data Mining

Liping Lin[1,2,3,4],Weixia Liu[1,4],Dachuan Zheng[1,4], Genping Yu[2,3,4,5], Libin Song[6,7], Zhongyan Lin[2,3,4,5*]

[1]Straits Institute of Minjiang University, Fuzhou, Fujian, China

[2]E-Commerce Branch of Fujian Collaborative Innovation Institute, Fuzhou, Fujian, China

[3]Internet Innovation Research Center of Minjiang University, Fuzhou, Fujian, China

[4]Internet Innovation Research Center-The College's Research Base of Humanities and Social Science in Fujian Province, Minjiang University, Fuzhou, Fujian, China

[5]International Digital Economy College of Minjiang University, Fuzhou, Fujian, China

[6]Fuzhou Digital Healthcare Industry Technology Innovation Center, Fuzhou, China

[7]Fuzhou Comvee Network &Technology Co., Ltd. Fuzhou, China

**Corresponding Author.*

Abstract

The objective of this research was to build a model to study the influencing factors of quality of lifes. Data were collected using the World Health Organization Questionnaire on Quality of Life (WHOQOL-BREF) from the Internet sent to six provinces and two cities. The data were analyzed using Partial Least Squares (PLS) approach. The test of the measurement model in terms of the validity and reliability of the measures was followed by testing of the hypothesized relationship. Results indicate that the hypothesized relationship was as follows: physiological factor, psychological factor, environmental factor, and social relationship factor have a significant positive impact on the of the Internet active people's quality of life, and psychological factor has the greatest influence on the quality of life of the internet active population.

Keywords: *Quality of Life (QOL), Partial Least Squares (PLS), model*

I. Introduction

The research on quality of life began in 1958. American economist J. K. Calbrith mentioned the conception of quality of life in his book Rich Society. Since then, quality of life has gradually become a special research field. Since the 1960s, the study on the quality of life has developed rapidly. It was introduced into China in the early 1980s and has attracted increasingly more attention in China's medical field.

Quality of Life (QOL) is defined by the World Health Organization Quality of Life (WHOQOL) Group as individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns [1]. This definition essentially embodies Maslow's hierarchy of needs theory, which combines the objective external environment with subjective internal feelings.

According to the "44th Statistical Report on the Development of Internet in China" issued by the China Internet Information Center (CNNIC), by December 2020, China's internet population had accounted for about one-fifth of the world's Internet users and the number of mobile internet users was 986 million in China. Therefore, the study of active Internet users is becoming increasingly important. Nowadays, due to the short history of the development of the Internet itself, there are still fewer studies on the portraits of the internet active people. Therefore, the purpose of this study is to model the relationship of QOL and its influencing factors using the data collected from six provinces

and two cities in China using the WHOQOL-BREF questionnaire. The data were analyzed using Partial Least Squares (PLS) software. In this study, a series of hypotheses on the QOL and physiological factor, psychological factor, environmental factor, and social relationship factor are proposed.

The rest of this study is organized as follows: In Section 3, the data sources and the method which is used to empirically analyze the hypotheses are presented. In Section 4, the measurement model in terms of the validity and reliability of the measures is presented to a significant level that supports the hypothesized relationship. Finally, through structural modeling, the influencing factors and their impact degree are obtained and targeted measures to improve the QOL of the Internet active population are put forward.

II. Basic theory

Structural equation modeling (SEM) was proposed by Swedish statistician Karl G. Joreskog in the 1970s as a multivariate statistical method to deal with and test the relationship between measurement models (the relationship between variables and indicators) and structural models (the relationship between variables) effectively. One of its developments is that it uses observable variables as indicators of latent variables to measure latent variables indirectly, so that variables that cannot be directly measured in social science research can also be included in the scope of empirical research. At the same time, because different researchers interpret the data differently, different interpretation models can be obtained. Structural equation models can verify different empirical models and obtain more optimized models through comparison [2]. Therefore, structural equation models are widely used in economic, financial, psychological, and behavioral fields.

Partial Least Squares (PLS) is proposed by Wood which was a new method based on PLS regression in the 1980s. It belongs to the "soft model" of structural equation models. PLS does not need variables to conform to normal distribution and requires less sample size. Generation estimation is a powerful processing capability for high-dimensional data. PLS allows simultaneous modeling of multiple independent latent variables and indicators to describe the relationship between the measurement model and the structural model. It no longer distinguishes between dependent and independent variables, but between exogenous and endogenous variables. PLS can solve the problem of measurement error, explain the variation structure of endogenous potential variables to the maximum, effectively improve the analysis of the model, and has been widely used in psychology, social sciences and other research fields.

III. Method

3.1 Data Sources

The questionnaire used in this study was based on WHOQOL-BREF, which was sent to six provinces: Fujian, Guangdong, Jiangsu, Zhejiang, Anhui and Jiangxi and two cities: Beijing, Shanghai in China through the internet. Using the method of random sampling, 300 distributed questionnaires of the internet active population each region and every questionnaire were independent IP. It finally received a number of 2,400 questionnaires.

3.2 Data Analysis

As described in the work of the study, the PLS approach was used to analyze the model. Smart PLS 3.0 software was used to analyze the data, and through its output, the validity and reliability of the measurement model was tested, followed by an examination of the hypothesized relationship. Smart PLS 3.0 software was used to achieve the operations of model path, model recognition, model estimation and model verification in a unified data testing environment.

3.3 Optimal design of WHOQOL-BREF scale

As early as 1991, the World Health Organization (WHO) formulated the principles and methods to be followed in the QOL assessment of international cooperation projects, and adopted a critical iterative process jointly undertaken

by health professionals and QOL researchers and international experts from all over the World, as well as the development and application of multiple languages and cultures. The WHO Questionnaire on QOL contains 276 core items at minimum. It covers 30 aspects of QOL and has been gradually developed into WHOQOL-100, which covers 24 aspects of QOL. There are four questions in each aspect, plus four questions concerning overall health and QOL. A total of 100 questions cover six areas: physiology, psychology, independence, social relationship, environment, and the other like spiritual pillars/religious/personal beliefs. By 1996, the WHO developed the WHO QOL Measurement Scale (WHOQOL-BREF) to facilitate use of the scale in research. WHOQOL-BREF retains the comprehensiveness of WHOQOL-100. The scores of each field of the summary table are highly correlated with those of the related fields of the WHOQOL-100 scale. The summary table can replace the WHOQOL-100 in measuring the scores of each field related to QOL. WHOQOL-BREF contains 24 core items, including four areas: physiology, psychology, social relationship, and environment and two questions reflecting overall health status and overall QOL, totaling 26 questions [3].

The questionnaire used in this study was designed on the basis of WHOQOL-BREF and consists of two parts: The first part gathers basic information from the respondents, and the second part contains 27 core items, including physiological field, psychological field, environmental field, and social relationship field. The Likert five-subscale is used to express them in terms of very poor, poor, not bad, good, and very good. Three new topics have been added to the design of the core items in WHOQOL-BREF, namely family friction, appetite, and scoring individual QOL, and the traffic conditions in the field of environment have been deleted. Among the additions, since family belongs to the field of social relations, family friction is classified as the field of social relations. Appetite refers to the requirement of people to eat which apparently is a physiological and psychological response, but in this study appetite is attributed to the field of physiology because the questionnaire was put on the Internet, the length of time and the first impression of the population in reading the questionnaire was taken into account. Scoring individual QOL is a kind of evaluation and measurement of the overall QOL, which is classified as the overall QOL. Finally, in view of the empirical analysis of the impact of various fields on the QOL, the overall health status is divided into physiological areas for PLS analysis in this study.

Table 1 Constructs of QOL Based on WHOQOL-BREF

Construct	***Facets incorporated within construct***	***Measurement item***
Physiological factor	Health status (Q10)	PF1
	Pain and discomfort (Q11)	
	Dependence on medical means (Q12)	
	Energy (Q18)	PF2
	Mobility (Q23)	PF3
	Sleep (Q24)	PF4
	Activities of daily living (Q25)	PF5
	Ability to work (Q26)	PF6
	Appetite (Q35) (newly added)	PF7
Psychological factor	Positive feeling (Q13)	PsF1
	Spiritual pillar (Q14)	PsF2
	Attention (Q15)	PsF3
	Body and appearance (Q19)	PsF4
	Self-esteem (Q27)	PsF5
	Negative Feeling (Q33)	PsF6
Environmental factor	Social security (Q16)	EF1
	Environmental conditions (Q17)	EF2
	Sources of economy (Q20)	EF3
	Access to information (Q21)	EF4
	Participation opportunities for leisure (Q22)	EF5
	Housing environment (Q31)	EF6
	Medical services and social security (Q32)	EF7
Social relationship factor	Personal relations (Q28)	SR1
	Sexual life (Q29)	SR2
	Satisfaction of social support needs (Q30)	SR3
	Family friction (Q34) (newly added)	
QOL	Total QOL (Q9)	QOL1
	QOL score (Q36) (Newly added)	QOL2

IV. Model Hypothesis

According to Table 1, we present the following hypotheses:

H1: Physiological factor has a positive impact on QOL.

H2: Psychological factor has a positive impact on QOL.

H3: Environmental factor has a positive impact on QOL.

H4: Social relationship factor has a positive impact on QOL.

The original hypothesis H_0 refers to the above four domains not being related to QOL.

V. Results

5.1 Convergent validity

Whether the measurement items measure the same construct and how to test the degree to the multiple items of measurement model usually use the discriminant validity. Convergent validity of measurement model, also called aggregation validity, Hair et al. suggested that it can be evaluated by loading of the factor, composite reliability (CR), and average variance extracted (AVE).(Hair et al.,2010) [4].

Table 2 Measuring aggregation validity of model

Construct	***Measurement item***	***Loading***	***CR***	***AVE***
Physiological factor	PF1	0.713	0.846	0.522
	PF2	0.757		
	PF3	0.752		
	PF4	0.660		
	PF5	0.791		
	PF6	0.694		
	PF7	0.682		
Psychological factor	PsF1	0.809	0.831	0.545
	PsF2	0.817		
	PsF3	0.719		
	PsF4	0.665		
	PsF5	0.773		
	PsF6	0.627		
Environmental factor	EF1	0.713	0.841	0.512
	EF2	0.721		
	EF3	0.740		
	EF4	0.729		
	EF5	0.704		
	EF6	0.713		
	EF7	0.686		
Social relationship factor	SR1	0.817	0.679	0.612
	SR2	0.695		
	SR3	0.828		

Construct	*Measurement item*	*Loading*	*CR*	*AVE*
QOL	QOL1	0.860	0.687	0.762
	QOL2	0.885		

The load of observation variables on the common factors of a construct can be present by factor loading. As shown in Table 2, the loading of all items exceeded 0.6 which was recommended value by Chin and other researchers (Chin et al., 1997) [5]. How the degree of the construct indicators indicate the latent construct is measured by Composite reliability (CR) .The rang of CR value of the model was from 0.679 to 0.846, and almost CR value was more than 0.7 which was recommended by Hair et al. (Hair et al., 2010) [4]; others, like social relationship factor and QOL, were 0.679 and 0.687, respectively, which were close to the recommended values of 0.7, indicating that the construct had good reliability. Average variance extracted (AVE) value reflects the overall amount of variance in the indicators occupied by the latent construct. The range of AVE values of the model was from 0.522 to 0.762, higher than 0.5 which was recommended bu Hair as well (Hair et al., 2010) [4]. In summary, the convergent validity of the measurement model of the construct was good.

5.2 Discriminant validity

How to detect the degree of the measurement items which differentiate among constructs usually depends on the discriminant validity. Comparing the value between the square root of Average Variance Extracted (AVE) and the correlations among constructs is often used to test the discriminant validity. (Fornell et al., 1981) [6]. As shown in Table 3, all the correlation values in the row were lower than the square roots of AVE, which illustrated the measurement model had adequate discriminatory.

Table 3 Distinguishing Validity of constructs

Constructs	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
(1) Psychological factor	0.739				
(2) Environmental factor	0.742	0.745			
(3) QOL	0.760	0.707	0.873		
(4) Physiological factor	0.778	0.699	0.732	0.780	
(5) Social relationship factor	0.674	0.657	0.640	0.645	0.783

5.3 Reliability analysis

Cronbach's Alpha value (Cronbach coefficient method) which was proposed by Lee Cronbach in 1951 is to determine whether the model is reliability or not. Generally, if the value of Cranbach's Alpha is above 0.6, it means the reliability of the measurement is high, which is based on the concept of Nunnally (1978). Table 4 showed that the alpha values of the measurement items were higher than 0.6, which revealed that the measurement model was reliable.

Table 4 Reliability Test of Model

Construct	*Measurement items*	*Cronbach's alpha*	*Loading range*	*Number of items*[a]
Physiological factor	PF1, PF2, PF3, PF4, PF5, PF6, PF7	0.846	0.660–0.791	7(7)
Psychological factor	PsF1, PsF2, PsF3, PsF4, PsF5, PsF6	0.831	0.627–0.817	6(6)
Environmental factor	EF1, EF2, EF3, EF4, EF5, EF6, EF7	0.841	0.686–0.740	7(7)
Social relationship factor	SR1, SR2, SR3	0.679	0.695–0.828	3(3)
QOL	QOL1, QOL2	0.687	0.860–0.885	2(2)

5.4 Structural model

The structural model shows that the relationships of causality among the constructs of the model [7]. To test the significance of the path coefficients among constructs, the value of beta and significance was obtained with a bootstrapping method using 5,000 samples (Hair et al. 2014) [8]. Both path coefficients and the R^2 value indicate how well the data support the hypothesized model [7].

Table 5 showed the output of Smart PLS 3.0 software, in which the P value of each path is 0, indicating that the result rejected the original hypothesis H_0 and supported the path hypothesis of the model; that is, the constructs were positively correlated with the QOL. R^2 is a resolvable coefficient used to measure the fitting degree of the regression line to the observed value. The closer the value is to 1, the more significant the regression effect of the model. Figure 1 showed that R^2 of QOL is 0.658, which illustrated that the effect of the model was more remarkable.

Table 5 Output of Structural Model

Hypothesis	***Route***	***Standard beta***	***Standard Error***	***T value***	***P value***
H1	Physiological Factor →QOL	0.257	0.022	11.820	0
H2	Psychological Factor→QOL	0.326	0.024	13.540	0
H3	Environmental Factor→QOL	0.209	0.022	9.494	0
H4	Social Relationship factor→QOL	0.117	0.018	6.389	0

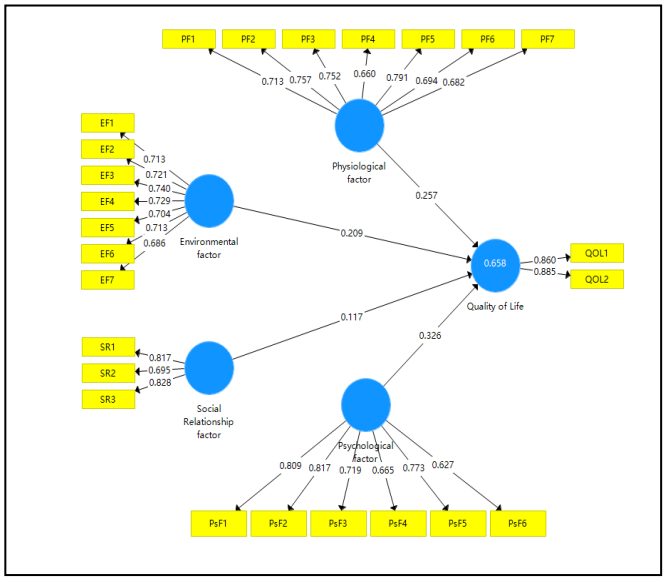

Fig1. QOL structure model

VI Conclusions

Using PLS-SEM and Smart PLS software tools, an empirical analysis of the influencing factors of QOL was undertaken in this study. Results showed that all the influencing factors were positively correlated with the QOL of the population.

Previous literature showed that when studying the specific disease groups' QOL, the influencing factor of the QOL of the group was physiological health but not psychological health. Different from these studies, herein, it was concluded that physiological health had a much smaller impact on the QOL than psychological health through

empirical analysis, which showed that with changing times, people paid more attention to their inner experiences and feelings.

Compared with psychological factor, physiological factor had less influence on the QOL, which focused on recessive or sub-health. However, it cannot be ignored that physical health directly affected the QOL through external forces, so we should also pay attention to the healthy habits of life, eating habits, and work and rest habits, and vigorously promote physical exercise to improve the QOL through physical fitness.

At the same time, it was shown that environmental factors effect on the QOL much greater, so attention should be paid to the construction of recreational facilities, the convenience of health protection, and the smooth operation of information-related software and hardware. The influence of social relationship factor was slightly weak. Under the conditions of the Internet environment, groups tend to relieve their emotions through online chat tools, and their satisfaction with interpersonal relationships and social support is relatively low.

Acknowledgements

This work was supported by the National Key R&D program, the Intergovernmental International Science, Technology and Innovation Cooperation Project (EU Horizon Project No. 2017YFE0118800); Central Financial support for Local Science and Technology Development Special (No. 2018L3007); Foreign cooperation project of Fujian Provincial Science and Technology Plan(NO.2020I0040); Financial Special Project of Fujian Province“ Internet cross-domain data analysis and crowd portrait”; the project of Internet Innovation Research Center-The College's Research Base of Humanities and Social Science in Fujian Province(Minjiang University) (IIRC20200108).

References

[1] Group T W. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychological medicine, 1998, 28:551-558.
[2] Tao He. Research on PLS algorithm of structural equation model. Tianjin University, 2006.
[3] Yuantao Hao, Jiqian Fang. Introduction to the Chinese version of the World Health Organization Quality of Life Scale and instructions for its use. China Tissue Engineering Research, 2000, 4 (8): 1127-1129.
[4] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis. Upper Saddle River, NJ: Prentice-Hall.
[5] Chin, W. W., Gopal, A., & Salisbury, W. D. (1997). Advancing the theory of daptive structuration: The development of a scale to measure faithfulness of appropriation. Information Systems Research, 8, 342–367.
[6] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18, 39–50.
[7] Sang, S., Lee, J. D., & Lee, J. (2010). E-government adoption in Cambodia: A partial least squares approach. Transforming Government: People, Process, and Policy, 4, 138–157.
[8] Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., 2014. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Sage Publication, Los Angeles.