

Personality Traits and Internet Consumption of College Students Based on Moderating Test and LASSO Regression

Guangjie Du *

School of Physics and Materials Science, Guangzhou University, Guangdong, China, 510006

* Corresponding author: gd_dukongkit@outlook.com

Abstract. With the burgeoning population of internet users, the prevalence of online shopping in China is also on the rise, encompassing college students as well. However, the progression of online shopping faces numerous challenges, one of which pertains to the impact of individual personality traits. To address this issue, this study aims to establish a comprehensive framework elucidating the connection between college students' personality traits and online consumption behavior, incorporating social support as a crucial regulatory and controlling variable. Moreover, employing the LASSO regression model, a multiple regression equation is derived, utilizing personality traits and social support as independent variables and internet consumption level as the dependent variable (Formula 9). The outcomes reveal that social support exhibits a substantial moderating effect on neuroticism, extraversion, openness, and conscientiousness, and the degree of social support significantly influences the extent of online consumption. However, it is important to note that social support does not manifest a significant moderating effect on agreeableness. This research endeavors to offer valuable insights to college students, enabling them to manage their consumption behavior effectively and rationally in accordance with their unique circumstances.

Keywords: Personality Traits, Network Consumption, Principal Component Analysis, Moderating Effect, LASSO Regression.

1. Introduction

With the proliferation of Internet users, online shopping in China has witnessed a significant rise, and college students are actively participating in this trend. However, the growth of online shopping is accompanied by several challenges. Notably, the diverse personality traits of consumers play a significant role in shaping their online consumption behavior [1]. Among college students, certain irrational consumption behaviors such as impulsive shopping, blind purchases, and compulsive spending are prevalent. Some students even resort to online loans to sustain their lifestyles. The increasing number of college students relying on online loans highlights the presence of irrational consumption behaviors to a considerable extent [2]. While scholars have highlighted the moderating effect and emphasized the importance of social support in strengthening educational guidance towards responsible consumption [3], limited research exists that delves into multiple regression analysis to assess the relationship between social support and personality traits. To address this gap, we employ the widely recognized Big Five personality scale developed by American psychologists McCrae and Costa. This scale encompasses neuroticism, openness, extroversion, agreeableness, and conscientiousness, serving as the prevailing tool for personality evaluation [4]. Consequently, this paper utilizes the Big Five personality traits to explore their influence on Internet consumption while investigating the potential moderating effect of social support. To achieve this, we first collected data through a questionnaire survey and conducted initial reliability and validity analyses. We then employed principal component analysis to reduce the dimensionality of the questionnaire variables, enabling streamlined subsequent analysis. Subsequently, Pearson correlation analysis and moderation tests were conducted to examine the relationship between social support and Internet consumption levels. Finally, employing the LASSO regression model, we derived a multiple regression equation (Formula 9) with personality traits and social support as independent variables and Internet consumption levels as the dependent variable.

2. Collection and preliminary analysis of questionnaire survey data

2.1. Questionnaire design and data collection

Based on the literature suggestions of Yan Biobin et al. (2003) [5] and Huang Wenyan et al. (2012) [6] regarding the investigation of personality traits and Internet consumption, this study employs a questionnaire designed specifically for full-time college students. The survey encompasses four key areas, namely basic information, the Big Five personality traits, social support, and Internet consumption levels. The Big Five personality traits are evaluated through a set of 12 questions, while social support and Internet consumption levels are assessed using 6 and 5 questions, respectively. After thoroughly cleaning the questionnaire data, a total of 307 valid samples were obtained. The interviewees consisted solely of college students, maintaining an equal male-to-female ratio and ensuring a balanced representation of various majors. Notably, more than 60% of the respondents belonged to the third and fourth academic years, with monthly living expenses ranging from 1,000 to 2,000 yuan.

2.2. Reliability and validity analysis [7]

In this study, the degree of consistency of questionnaire research variables in each measurement item was assessed using Cronbach's Alpha reliability coefficient. For variables to demonstrate good reliability, a Cronbach's Alpha coefficient exceeding 0.7 is desired. The scale reliability analysis results were obtained through the utilization of SPSS analysis software, and these findings are presented in Table 1.

Table 1. Analysis results of scale reliability

Dimension	Cronbach's Alpha	Number of questions
Neurosis	0.928	12
Extroversion	0.939	12
Openness	0.945	12
Amenity	0.931	12
Conscientiousness	0.941	12
Social support	0.888	6
Online consumption	0.902	5

Table 1 clearly demonstrates the robustness of the questionnaire, as evidenced by the reliability coefficients exceeding 0.8 for all dimensions of the Big Five personality, social support, Internet consumption behavior, and other indicators. Remarkably, the highest coefficient achieved an impressive value of 0.945, further reinforcing the strong reliability of the questionnaire.

The primary emphasis of this study centers on establishing the structural validity, which pertains to the capacity of each item to effectively measure the intended variable. To substantiate the structural validity of the scale, an Exploratory Factor Analysis (EFA) test was conducted using the gathered data. The analysis software employed for this purpose was SPSS, enabling the acquisition of KMO and Bartlett's test analysis outcomes, as illustrated in Table 2.

Table 2. Analysis results of scale reliability

The Kaiser-Meyer-Olkin measure of sample adequacy		0.952
Bartlett sphericity test	Approximate chi-square	13017.95
	<i>df</i>	1770
	<i>Sig.</i>	0

Table 2 presents the results of the preliminary analysis, demonstrating that the Kaiser-Meyer-Olkin (KMO) measure yielded a value of 0.952, surpassing the recommended threshold of 0.7. Additionally, the significance of Bartlett's sphericity test (*Sig.* < 0.001) confirms that the questionnaire data satisfied the necessary prerequisites for conducting principal component analysis. Hence, subsequent principal component analysis was performed.

2.3. Principal component analysis (PCA)[8]

Principal Component Analysis (PCA) is a multivariate statistical technique employed to achieve "dimensionality reduction." Its purpose is to transform the variables in the Big Five personality questionnaire into distinct dimensions, which describe personality traits. This transformation aims to evaluate the validity of the questionnaire structure. Each principal component is an independent linear combination of the original variables from the questionnaire, preserving a significant portion of the information contained in the original variables.

To begin, we standardized the variable data for each questionnaire item. Then, we calculated the covariance matrix, denoted as R :

$$R = (r_{ij})_{n \times n} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix} \quad (1)$$

Next, we determined the eigenvalues and corresponding eigenvectors of matrix R . These eigenvectors are denoted by A :

$$A = \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nm} \end{pmatrix} \quad (2)$$

Using the feature vectors, we constructed n new index variables as follows:

$$\begin{cases} y_1 = u_{11}\bar{x}_1 + u_{21}\bar{x}_2 + \cdots + u_{n1}\bar{x}_n \\ \quad \quad \quad \cdots \cdots \\ y_n = u_{1n}\bar{x}_1 + u_{2n}\bar{x}_2 + \cdots + u_{nn}\bar{x}_n \end{cases} \quad (3)$$

In Formula 3, y_1 represents the first principal component, y_2 represents the second principal component, and so on, up to y_n , which represents the NTH principal component.

For factor extraction, we employed SPSS analysis software, following the process, with a criterion of retaining factors with characteristic roots greater than 1. During factor rotation, we utilized maximum variance orthogonal rotation for factor analysis.

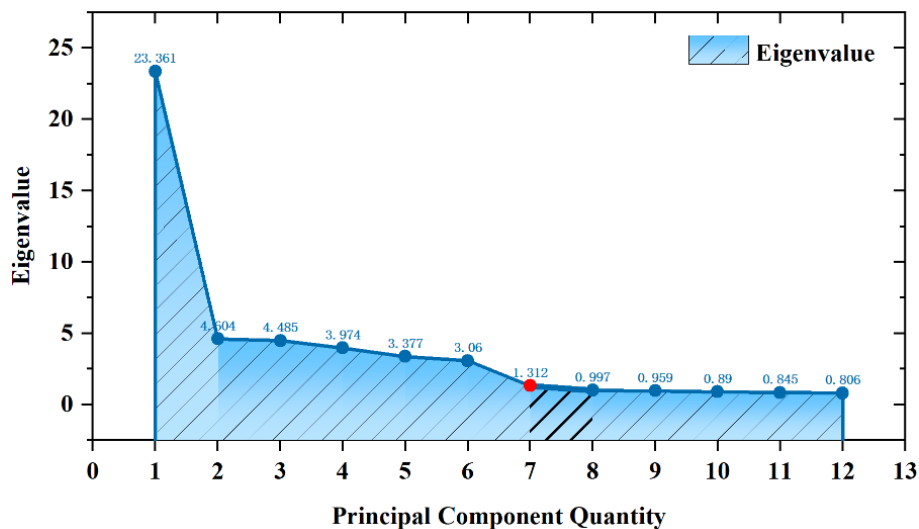


Figure 1. Principal component analysis (PCA) lithotripsy

The Macadam diagram (Figure 1) visually represents the extent to which each principal component interprets the variation in the data. By observing the diagram, we concluded that with eight principal components, the eigenvalue for total variance interpretation was below 1. Consequently, we extracted

a total of seven principal components, with a cumulative variance interpretation rate of 63.62% (greater than 50%). These results indicate that the seven selected factors—neuroticism, extraversion, openness, agreeableness, conscientiousness, social support, and Internet consumption behavior—adequately represent the questionnaire and demonstrate its favorable structural validity.

3. The influence of social support on online consumption based on the moderating test

3.1. Correlation test

By employing correlation analysis, we can initially assess the association between social support and personality traits. However, in order to ascertain the specific functional relationship, further regression analysis is required. One of the primary methods for correlation analysis is the Pearson correlation test.

The measurement indices utilized in the Pearson correlation test consist of the P and the r . The P signifies the significance of the test results, while the r represents the correlation coefficient between variables. In this study, SPSS software was employed to conduct correlation analysis between personality traits and social support. The correlation coefficients ranged from 0.113 to 0.589, demonstrating a significant positive correlation, as depicted in Figure 2. Thus, it can be preliminarily inferred that social support exhibits a strong correlation with online consumption behavior. However, further investigation is necessary to explore the specific personality traits through which social support significantly influences online consumption behavior.

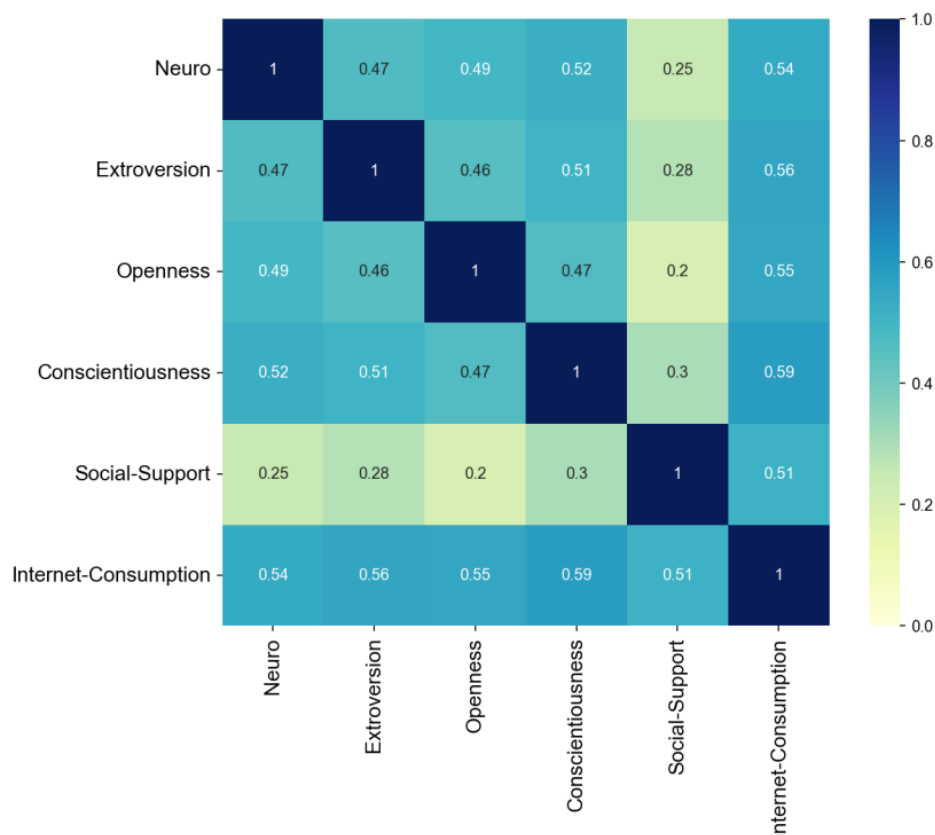


Figure 2. The correlation between social support and personality traits

3.2. Moderating effect test

Moderating Effect refers to a phenomenon where the causal relationship between two variables undergoes changes based on different values of the moderating variable. The measurement and testing

of the moderating effect are closely linked to the measurement level of both the independent variable and the moderating variable [9].

In our study, we focus on exploring the influence of social support as a moderating variable on online consumption behavior across various personality traits. To ensure the accuracy of our experimental results and eliminate the impact of other variables, we consider gender, major, grade, household registration, and monthly living expenses as control variables. Additionally, we treat neuroticism, extroversion, openness, agreeableness, and conscientiousness as independent variables, while online consumption behavior (Y) serves as the dependent variable for adjustment testing.

To enhance the explanatory power of the regression equation coefficients, we adopt the recommendations put forth by Aiken & West (1991) [10], Cohen et al. (2003) [11], and Hayes (2013) [12]. Accordingly, we center the personality traits and social support variables, thereby calculating the moderating function terms.

Centralization of variables can be achieved by subtracting the mean value (μ) from the value (x) of a personality trait index, as shown in formula 4:

$$x' = x - \mu \quad (4)$$

Consequently, we obtain the following centralized variables: Neuroticism (X_1), Extraversion (X_2), Openness (X_3), Agreeableness (X_4), Conscientiousness (X_5), and Social Support (Z). The subsequent step involves conducting hierarchical regression for the adjustment test.

In the hierarchical regression analysis, assuming an index X_m ($m = 1, 2, 3, 4, 5$) corresponds to both the online consumption behavior (Y) and personality traits, we can derive formula 5:

$$Y = aX_m + hZ + cX_mZ + e \quad (5)$$

Formula 5 can be rearranged as formula 6:

$$Y = hZ + (a + cZ)X_m + e \quad (6)$$

It is evident that, for a fixed value of social support (Z), this represents the linear regression of online consumption behavior (Y) on the personality trait (X_m). X_mZ denotes the interaction term, whereas the relationship between online consumption behavior (Y) and personality trait (X_m) is described by the regression coefficient, specifically the moderating effect term $a + cZ$. This term is a linear function of social support (Z), and the magnitude of c indicates the size of the moderating effect exerted by social support. In the subsequent analysis, we mainly aim to estimate and test the significance of c . A significant value of c would indicate a substantial moderating effect of social support on the relationship between personality traits and online consumption behavior.

3.3. Analysis of the results of the adjustment effect model

By employing hierarchical regression analysis and conducting T-tests and significance tests, the analysis table (Table 3) reveals the moderating effect of social support on the relationship between Big Five personality traits and online consumption behavior. Examining Table 3, it becomes evident that the interaction terms "neuroticism * social support", "extroversion * social support", "openness * social support" and "conscientiousness * social support" demonstrate statistically significant P , indicating a significant influence of social support on the online consumption behavior of individuals possessing these specific personality traits. Conversely, the interaction term "agreeableness * social support" does not exhibit statistically significant results, suggesting that social support does not significantly interfere with the online consumption behavior of individuals characterized by agreeableness, considering a certain level of other personality traits such as neuroticism, extroversion, openness, and conscientiousness.

Table 3. Significance test of the moderating effects of personality traits and social support

Interaction Term	Correlation Coefficient	Standard Error	t	P
Neurotic * Social Support	0.135	0.054	2.507	0.013**
Extraversion * Social Support	0.158	0.045	3.513	0.001***
Extraversion * Social Support	0.123	0.047	2.626	0.009***
Agreeableness * Social Support	0.048	0.042	1.148	0.252
Conscientiousness * Social Support	0.102	0.038	2.696	0.007***

We conducted a comprehensive analysis of the interactive factors that demonstrated significant interference, and we determined the varying influence of social support on online consumption behavior at different levels (refer to Figure 3-6). By examining the amplitude difference (slope) of personality traits' impact on online consumption behavior across various levels of social support, we can assert that regardless of whether individuals exhibit high or low levels of neuroticism, extraversion, openness, and conscientiousness, different levels of social support positively contribute to the extent of online consumption behavior. Notably, the linear slope associated with high social support is significantly more pronounced compared to that of low social support, highlighting the substantial positive influence exerted by social support on online consumption behavior.

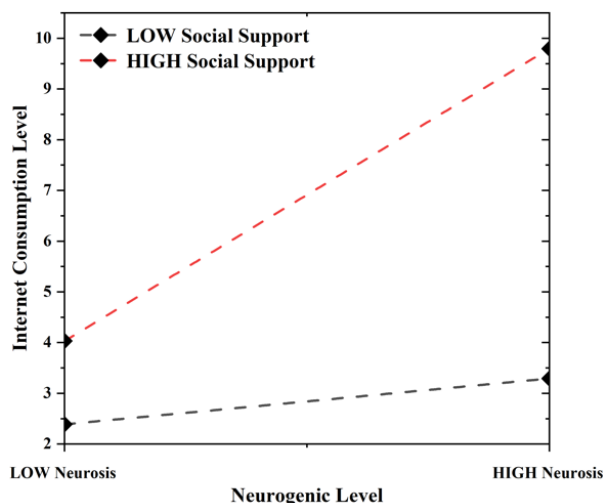


Figure 3. The difference in the amplitude of neurosis to Internet consumption behavior under different levels of social support

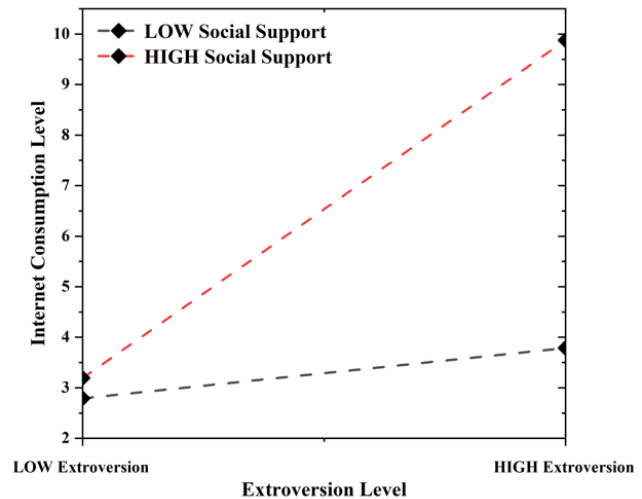


Figure 4. The difference in the amplitude of extroversion to Internet consumption behavior under different levels of social support

4. Online consumption model based on LASSO regression model [13]

4.1. The establishment of LASSO internet consumption model

First introduced by LASSO in 1996, the Lasso regression method is a variant of linear regression that incorporates L1 regularization terms. This approach is particularly useful for feature selection and achieving model sparsity. By applying Lasso regression, the model can automatically identify the most significant independent variables while shrinking the coefficients of less relevant variables to zero.

To begin, let's denote the argument matrix as X_{ij} , where i represents the sample index and j represents the argument index. In this case, our independent variable matrix consists of six indicators: neuroticism, extraversion, openness, agreeableness, conscientiousness, and social support.

To construct the model, we will employ the LASSO regression framework. Here, Y_i represents the dependent variable, which is the level of network consumption. Our objective is twofold: to find the optimal coefficient estimates by fitting the available data and to make predictions based on this model.

The formula for the Lasso regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad (7)$$

Here, Y_i denotes the dependent variable (network consumption level), β_0 represents the intercept term, β_j refers to the coefficient of the independent variable X_{ij} , p denotes the number of independent variables ($p = 6$), and ε_i signifies the error term.

In order to obtain the most suitable coefficient estimates, Lasso regression minimizes the following objective function:

$$\text{minimize} \frac{1}{2N} \sum_{i=1}^N (Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \cdots + \beta_p X_{ip}))^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (8)$$

In equation 8, N represents the number of samples. The first term denotes the sum of squared errors, while the second term represents the L1 regularization term. The regularization parameter λ , allows us to control the level of regularization intensity. By solving the optimization problem defined by the objective function, we can derive the best coefficient estimates and establish the Lasso regression model.

4.2. Solution and result analysis of LASSO model

According to the model established above, the LASSO function in Matlab is used to solve and predict the model, and the cross-verified elastic network fitting MSE diagram (Figure 7) and the actual value and prediction result diagram (Figure 8) are obtained. In Figure 10, the vertical lines highlighted in blue and green represent two distinct regularization parameters, namely LambdaMinSE and Lambda1SE. These specific values are crucial in our analysis. As we compare the two, our preference leans towards LambdaMinSE due to its association with a smaller mean square prediction error.

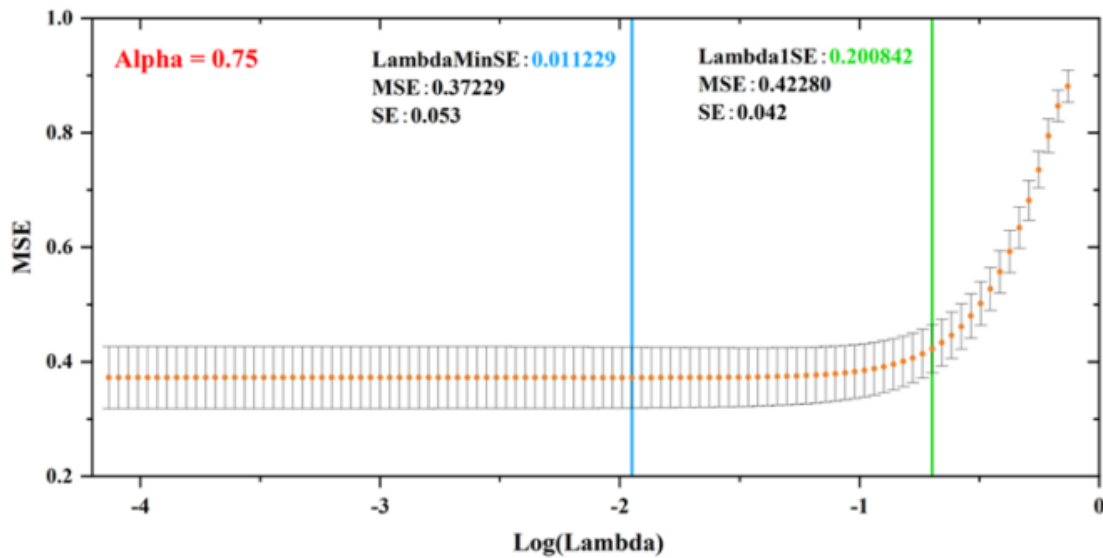


Figure 7. Cross-validated elastic network fitting MSE

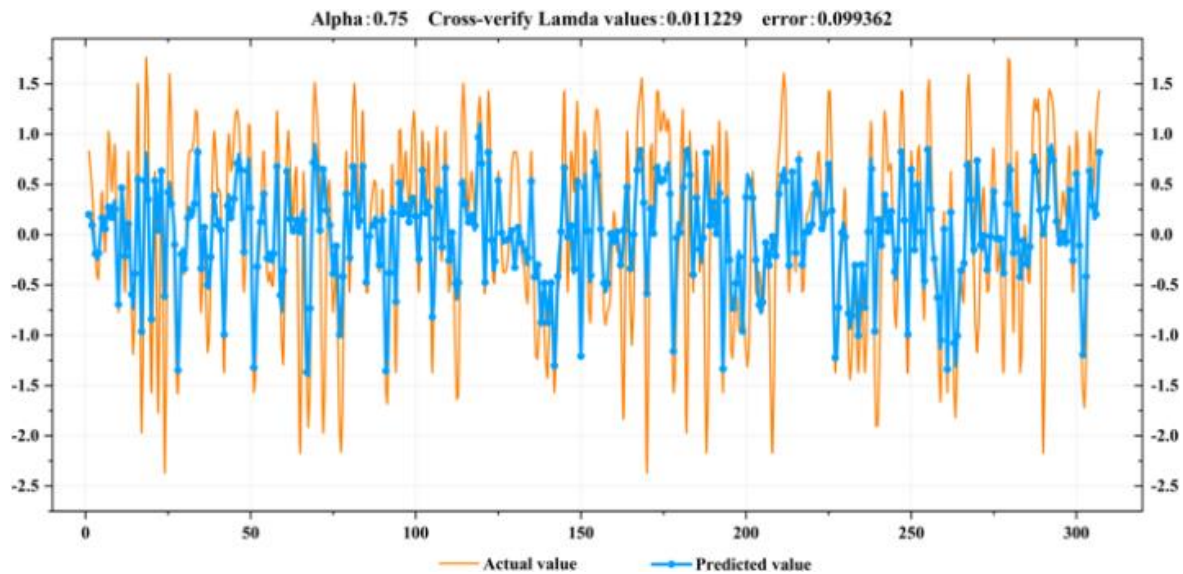


Figure 8. Comparison of actual values and predicted results of LASSO regression model

According to model solving, the non-standardized formula of the prediction model (values 1 to 5 of each variable) can be obtained as follows:

$$\begin{aligned} \text{Internet} - \text{Consumption} = & -0.353 + 0.107 \times \text{Amenity} + \\ & 0.182 \times \text{Conscientiousness} + 0.195 \times \text{Openness} + 0.138 \times \text{Extroversion} + \\ & 0.13 \times \text{Neurosis} + 0.32 \times \text{Social} - \text{Support} \end{aligned} \quad (9)$$

Based on Formula 9, the varying levels of Internet consumption can be partially attributed to different personality traits and levels of social support. The agreeableness coefficient, which stands at 0.107, suggests that individuals with higher agreeableness tend to exhibit a greater inclination towards online consumption. Similarly, the coefficient of conscientiousness, at 0.182, indicates that individuals with a higher level of conscientiousness show a greater propensity for online consumption. Individuals with higher openness, as indicated by the coefficient of 0.195, are more likely to engage in online consumption and may demonstrate a greater willingness to explore new online services and products. Moreover, the extraversion coefficient of 0.138 suggests that individuals with a higher level of extraversion may favor online activities such as social networking and online shopping. On the other hand, the neurological coefficient, measuring 0.13, suggests that individuals with higher neuroticism may exhibit greater susceptibility to emotional fluctuations, leading to increased online consumption during periods of emotional instability. Lastly, the social support coefficient of 0.32 implies that individuals with stronger social support are more inclined to fulfill their social and emotional needs through online consumption.

5. Conclusion

This paper establishes a relationship model between personality traits and Internet consumption, with social support serving as the moderating control variable. By employing the LASSO regression model, a multiple regression equation (Equation 9) is derived, incorporating personality traits and social support as independent variables and Internet consumption level as the dependent variable. Notably, the predictive test conducted in this study indicates the robustness of the model, as evidenced by the low prediction mean square error (MSE) of only 0.011. The findings reveal that social support exerts a significant moderating effect on neuroticism, extraversion, openness, and conscientiousness, while the level of social support significantly influences the extent of online consumption. However, no significant moderating effect of social support on agreeableness is observed. Moreover, it is observed that college students exhibit a greater tendency towards online consumption when they possess higher levels of social support or lower levels of neuroticism. The utilization of Equation 9

can offer valuable guidance to college students, enabling them to engage in their consumption behavior in a more scientific and rational manner, tailored to their individual circumstances.

References

- [1] Liu Ling. Research on Influencing Factors of College Students' Online Consumption Behavior [D]. Anhui University of Science and Technology, 2015.
- [2] Fu Shun, Pei Ping, Gu Tianzhu. Consumption Behavior, online loan intention and online loan platform Preference of college students: An empirical analysis based on Questionnaire survey data of colleges and universities in Jiangsu Province [J]. Lanzhou Academic Journal, 2019 (11): 108 - 120.
- [3] Wang Lingxiao. The influence of personality traits on college students' Internet consumption behavior research [D]. Central south forestry university of science and technology, 2020. The DOI: 10.27662 /, dc nki. GZNL. 2020.000196.
- [4] Li Hongyan. A brief introduction to "Big Five" Personality factor Model [J]. Journal of Shaanxi Normal University, 2002, 31 (6): 89 - 91.
- [5] A surname. The Impact of Social Support on Subjective Well-being of College Students [D]. South China Normal University, 2003.
- [6] Huang Wenyan, Wen Shisong. An empirical study on Factors influencing college students' willingness to consume online [J]. Science and Technology Management Research, 2012, 32 (03): 165 - 169.
- [7] Zeng Wuyi, Huang Bingyi. Analysis of reliability and validity of questionnaire [J]. Statistics & Information Forum, 2005 (06): 13 - 17.
- [8] He Xiaoqun. Multivariate Statistical Analysis. Beijing: China Renmin University Press, 2012.
- [9] Wen Zhonglin, Hou Jietai, Zhang Lei. Comparison and application of moderating Effect and mediating effect [J]. Acta Psychologica Sinica, 2005 (02): 268 - 274. (In Chinese).
- [10] Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA: Sage Publications.
- [11] Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression and correlation for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- [12] Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Press.
- [13] Ke Zhenglin. Application of Lasso and its related methods in multiple linear regression model [D]. Beijing Jiaotong University, 2011.