Volume 2, Issue 4, 2024

Print ISSN: 2959-992X
Online ISSN: 2959-9938

TRENDS IN SOCIAL SCIENCES AND HUMANITIES RESEARCH

Trends in Social Sciences and Humanities Research

Volume 2, Issue 4, 2024

Published by Upubscience Publisher

Trends in Social Sciences and Humanities Research
Print ISSN: 2959-992X Online ISSN: 2959-9938
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

THE IMPACT OF COMPREHENSIVE SOCIAL SUPPORT ON THE TREATMENT OUTCOMES OF PATIENTS WITH DEPRESSION 1-8
YuTing Zhan

RESEARCH ON THE STRATEGY OF ORGANIC INTEGRATION OF CHINESE EXCELLENT TRADITIONAL CULTURE AND IDEOLOGICAL AND POLITICAL COURSES IN COLLEGES AND UNIVERSITIES IN THE CONTEXT OF NEW LIBERAL ARTS EDUCATION 9-13
Wei Zhang*, Jing Jing

ANALYSIS AND PREDICTION OF AIR QUALITY INFLUENCE FACTORS IN CHANGSHA CITY 14-17
WenHui Zeng

AN ANALYSIS OF EARLY WARNING FOR CREDIT CARD CUSTOMER CHURN 18-21
ChangFu Yang

ANALYSIS OF GRASS-ROOTS BEHAVIOR AMONG YOUNG GROUPS 22-27
Chen Chen

STUDY ON THE INFLUENCING FACTORS OF GUANGXI'S TOTAL EXPORTS BASED ON RIDGE REGRESSION AND LASSO REGRESSION 28-32
YuHe Cheng

ANALYSIS AND FORECAST OF THE AVERAGE SALES PRICE OF RESIDENTIAL COMMERCIAL HOUSING 33-37
XinYue Dang

FINANCIAL CREDIT RISK ASSESSMENT BASED ON MACHINE LEARNING 38-43
MingYue Gao

AN EMPIRICAL ANALYSIS OF THE INDUSTRIAL STRUCTURE AND EMPLOYMENT STRUCTURE IN JIANGSU PROVINCE 44-51
Ling Li

ANALYZING REGIONAL ECONOMIC INFLUENCING FACTORS BASED ON DIFFERENT CONTRACTION METHODS 52-61
CaiYun Peng

ANALYSIS OF FACTORS INFLUENCING THE ENGEL INDEX BASED ON REGRESSION MODELS 62-67
GaoBo Peng

APPLICATION AND EFFECTIVENESS ASSESSMENT OF QUALITY CERTIFICATION AND STANDARDISATION IN AGRICULTURAL PRODUCT E-COMMERCE 68-72
YuanBo Jia*, Rui Yan, Khorloo Yundendorj

ANALYSIS AND PREDICTION OF INFLUENCING FACTORS ON THE PROBABILITY OF STROKE 73-77
XinChun Wang

RESEARCH ON THE MARKETING STRATEGY OF DOMESTIC BEAUTY INDUSTRY FROM THE PERSPECTIVE OF DIGITAL ECONOMY 78-83
ZhaoShuo Wu

LSTM MODEL ENHANCED BY KOLMOGOROV-ARNOLD NETWORK: IMPROVING STOCK PRICE PREDICTION ACCURACY 84-89
XiaoXuan Yao

ANALYSIS AND DECISION-MAKING OF REGIONAL ECONOMIC VITALITY AND ITS INFLUENCING FACTORS 90-101
ShaSha Zhang

TREND ANALYSIS AND FORECAST OF HOUSEHOLD APPLIANCE OWNERSHIP AND ELECTRICITY CONSUMPTION IN XIANGYANG 102-106
MinShi Zheng

ANALYSIS OF FACTORS INFLUENCING HUNAN PROVINCE'S GDP TOTAL 107-112
QiBin Zhu

A REFLECTIVE INQUIRY INTO LANGUAGE LARGE-UNIT TEACHING BASED ON CORE LITERACY 113-117
XiaoYu Liu*, JunJun Wu

INDIVIDUAL ENJOYED TEACHING: THE DEMANDS OF TEACHING UNDER THE BACKGROUND OF THE GLOBAL COMMON GOOD 118-122
JunJun Wu*, XiaoYu Liu

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3002

THE IMPACT OF COMPREHENSIVE SOCIAL SUPPORT ON THE TREATMENT OUTCOMES OF PATIENTS WITH DEPRESSION

YuTing Zhan
Department of Psychology, Ningxia University, Yinchuan, 750000, Ningxia, China.
Corresponding Email: Tyndalll163@email.com or 12023130343@stu.nxu.edu.cn

Abstract: This study examines the application and effectiveness of comprehensive social support in the treatment of depression, emphasizing the combined roles of emotional, practical, and informational support. The findings indicate that social support not only enhances treatment adherence and efficacy but also contributes to symptom alleviation and improved quality of life. By analyzing cross-cultural differences, the research further explores the application of social support in diverse cultural contexts, highlighting its significance in depression treatment strategies. Additionally, the study discusses the implementation and optimization of comprehensive social support interventions in clinical practice, aiming to provide a basis for future research directions and policy formulation. The conclusion underscores the necessity of integrating social support into comprehensive depression treatment, offering new perspectives and strategies for improving treatment outcomes.
Keywords: Comprehensive social support; Depression; Treatment

1 INTRODUCTION

Depression, also known as depressive disorder, is one of the most prevalent mental illnesses worldwide [1,2]. According to the World Health Organization (WHO), it is estimated that approximately 340 million people globally suffer from varying degrees of depression[3]. Depression can lead to severe psychological disturbances and adverse emotional states, such as sadness, fatigue, and hopelessness. Individuals with major depressive disorder (MDD) may experience suicidal ideation and even attempt suicide, resulting in significant physical and emotional distress for the patients and imposing a substantial economic burden on society[4].

1.1 Background

Early foundational research established the crucial role of social support in alleviating depressive symptoms, suggesting that supportive relationships significantly reduce the risk of developing depression[5]. These relationships provide emotional solace, promote psychological resilience, and buffer against stress. Currently, it is widely recognized that comprehensive social support plays a significant role in enhancing treatment outcomes for patients with depression. Recent empirical studies build on this foundation, exploring how various types of social support, such as emotional and informational support, contribute to improving treatment outcomes and ameliorating depressive conditions.
Emotional support, which includes empathy, care, and love, has been proven to be highly effective in reducing the severity of depressive symptoms[6]. Informational support, encompassing advice and guidance, assists patients in navigating complex treatment options. Additionally, advances in digital health technologies have expanded the scope of social support interventions. Online communities and telepsychiatry services provide continuous and convenient support, which is crucial for individuals facing challenges within traditional healthcare settings[7]. Digital platforms not only increase the availability of support but also create new avenues for delivering personalized interventions.

1.2 Concept of Social Support

Since the inception of human society, mutual support among individuals has been a fundamental aspect of social interaction. Various disciplines, including medicine, sociology, communication studies, and psychology, have begun to interpret the concept of social support from their respective theoretical perspectives. Early researchers often studied social support qualitatively, viewing it as a broad, unified relationship system. They believed that any existing relationship inherently helps individuals cope with life's challenges[8].
For instance, Cobb[9] defined social support as information that makes individuals feel cared for, loved, and respected, suggesting that they are part of a mutually responsible social network. Cohen and Wills[5] posited that social support involves psychological assistance and material resources provided by social networks to help individuals effectively manage stress.

1.2.1 Types of social support

Research on the concept of social support can be categorized from two perspectives: functional and operational. From the functional perspective, social support refers to the material and emotional assistance an individual receives from their social relationships. From the operational perspective, social support represents a quantifiable measure of an individual's social connections[10]. Subsequently, some scholars began using quantitative methods to differentiate types of social support. For example, Barrera[11] identified six forms of social support: tangible assistance, behavioral support, intimate interaction, guidance, feedback, and positive social interaction. Van der Poel[12] categorized social support into tangible support, instrumental support, emotional support, network support, self-esteem support, and nurturing support. Cutrona and Russell[13] differentiated social support into social integration, tangible support, informational support, emotional support, and self-esteem support, emphasizing that a social support network is a social structure from which individuals can draw various resources (e.g, material, emotional).

Based on the analysis in this study, despite scholars interpreting social support from different angles, the types can be broadly classified into two main categories: objective support and subjective support. Objective support includes material and network support, which exist independently of an individual's subjective experiences and are objectively present. Subjective support encompasses interpersonal emotional support, such as respect, empathy, and understanding in social interactions. While social support has been extensively studied from various perspectives, the specific forms and types of social support-based interventions for depression in the field of clinical psychology remain inadequately defined, warranting further investigation.

1.2.2 Subjects and objects of social support

Scholars have also studied the subjects and objects of social support, viewing it as an exchange of resources between two conscious individuals: the provider and the recipient[14]. Regarding the subjects of social support, Thoits[15] identified family members, friends, and colleagues as key providers of social support. Van der Poel[12] expanded the definition of support providers to include three levels: the state, community, and individuals, thereby defining the subjects of social support within a broader "social network" encompassing both formal and informal relationships.

In studying the objects of social support, two perspectives emerge. One perspective posits that the objects of social support are selective, primarily targeting socially disadvantaged or vulnerable groups[12]. The other perspective argues that social support is a universal social behavior, suggesting that any individual in everyday life can be the recipient of social support[12].

1.3 Research Objectives and Significance

This study aims to delve into the impact of comprehensive social support on the treatment outcomes of depression. As a widespread psychological disorder, depression's severity and prevalence have garnered global attention. Despite the significant roles that pharmacotherapy and psychotherapy play in treating depression, high relapse rates and individual variability in treatment responses remain major challenges in clinical practice[16]. In this context, social support has increasingly attracted researchers' attention as a potential protective factor and intervention tool.

One of the primary factors contributing to depression is a lack of social support, which refers to insufficient social, emotional, or practical assistance within an individual's social network[17]. A deficiency in social support can lead to decreased prefrontal cortex limbic activity in depression patients[18] and abnormal neurobehavioral responses[19,20]. Therefore, systematic research on social support is crucial for identifying individuals at risk of or suffering from depression and provides a scientific basis for developing social support interventions and treatment strategies. Although existing studies have indicated a correlation between social support and the alleviation of depressive symptoms, the specific mechanisms remain unclear[5]. This study aims to systematically review the existing literature to uncover the pathways through which different types of social support influence depression treatment, thereby enriching and expanding the theoretical understanding of social support.

Furthermore, this research offers significant implications for clinical interventions. The integration of comprehensive social support may enhance the effectiveness of depression treatment and improve patients' quality of life. For instance, emotional support from family members, practical assistance from friends, and informational support from professionals can all facilitate patient recovery on various levels[21]. By elucidating the specific roles of social support, this study provides scientific evidence for clinical psychologists and practitioners to design and implement effective social support interventions. Understanding the critical role of social support in depression treatment also aids in shaping public health policies, encouraging the establishment of stronger community support networks, and raising societal awareness and investment in mental health issues[22].

The following sections will first introduce the relationship between social support, mental health, and depression. Subsequently, they will review and integrate relevant research from both cross-sectional and longitudinal studies, incorporating findings from cross-cultural studies to propose a comprehensive intervention model. Finally, the study will outline three future research prospects.

2 LITERATURE REVIEW

2.1 The Relationship Between Social Support and Mental Health

2.1.1 Theoretical foundations

The relationship between social support and mental health has been extensively studied, with several theories providing frameworks for understanding this connection. Two primary theoretical foundations are the Social Support Buffering Hypothesis and the Direct Effect Model of social support.

2.1.2 Social support buffering hypothesis

Cohen and Wills[5] proposed the Social Support Buffering Hypothesis, which posits that social support can mitigate the negative impact of stressful events on an individual's mental health. This hypothesis includes two key elements: the stress-buffering mechanism and individual perception. The stress-buffering mechanism suggests that social support can alleviate the adverse effects of major stressors (such as bereavement, unemployment, or severe illness) through various means. Emotional support provides comfort and empathy, reducing feelings of loneliness and helplessness. Tangible support offers specific assistance (such as financial aid or daily care), easing the individual's burden. Informational support enhances coping capacity by providing problem-solving advice and information[5].

Individual perception refers to the importance of perceived support alongside actual support received. Studies have shown that even in the absence of actual support, the belief that support is available when needed can improve mental health [23].

2.1.3 Direct effect model of social support

The Direct Effect Model suggests that social support can directly enhance mental health, regardless of whether individuals are facing stressful events. This model emphasizes the continuous role of social support in daily life, manifesting in three specific aspects: stable social relationships, social integration, and social influence norms. Stable social relationships, such as close family ties and enduring friendships, contribute to a sense of security and belonging, thereby enhancing overall mental health[10]. Social integration refers to the extent to which individuals are embedded in their social networks, which is closely related to their mental health. Highly integrated individuals often have higher self-esteem and lower feelings of loneliness, which help prevent depression and anxiety[24]. Social influence and norms within support networks can positively impact an individual's behaviors and attitudes. Support from friends and family can encourage healthy lifestyles and promote psychological well-being[22].

2.1.4 The overall impact of social support on mental health

A substantial body of empirical research demonstrates that social support has a significant positive impact on mental health. First, social support effectively reduces psychological stress and negative emotions. Studies have found that individuals with robust social support networks exhibit lower levels of depression and anxiety when facing stressful events[25]. Support from family and friends provides emotional comfort, reducing feelings of loneliness and helplessness, thereby alleviating psychological stress.

Second, social support enhances self-efficacy and coping abilities. By receiving positive feedback and assistance from others, individuals can build confidence in their capabilities, enabling them to more effectively tackle life's challenges and difficulties[26]. This is particularly crucial in clinical practice, as boosting patients' self-efficacy is a key factor in promoting their recovery.

Moreover, social support is closely linked to physical health. Research indicates that social support can indirectly improve physiological health by reducing psychological stress and fostering positive emotions[27]. For example, strong social support can lower the risk of cardiovascular diseases and enhance immune system function, thereby improving overall health.

In summary, social support plays a vital role in promoting mental health. Whether by mitigating the adverse effects of stressful events or providing continuous emotional and practical support in daily life, social support demonstrates significant protective effects. By comprehensively understanding and leveraging the various functions of social support, we can more effectively address mental health issues and improve individuals' overall quality of life.

2.2 Social Support and Depression

2.2.1 The preventive role of social support in depression onset

Social support effectively mitigates individuals' stress responses to life events, thereby preventing the onset of depression. Research has shown that individuals with robust social support networks exhibit lower levels of stress and psychological distress when facing life stressors[5]. Emotional support offers comfort and empathy, helping individuals manage negative emotions and reducing the risk of depression. Warmth from family and understanding from friends can significantly alleviate feelings of loneliness and helplessness, thus reducing psychological stress.

Tangible support provides specific assistance, such as financial aid and help with household chores, easing individuals' daily burdens and preventing depressive feelings resulting from life stress. Tangible support not only offers material assistance but also enhances feelings of security and trust. Informational support, through advice and problem-solving information, helps individuals better cope with stressful events. Advice and experience sharing from colleagues and friends can help individuals manage work and life pressures more effectively[28].

Social support enhances psychological resilience, laying a foundation for preventing depression. Psychological resilience refers to an individual's ability to adapt and recover when facing adversity. Social support can strengthen psychological resilience in various ways, thereby preventing depression[29]. Positive feedback and assistance from a social support network enable individuals to develop more effective coping strategies, reducing the accumulation of negative emotions. Participating in support groups or community activities can help individuals learn and adopt others' coping strategies, enhancing their own coping abilities. Social support provides individuals with stronger confidence and capacity when facing challenges. Research shows that individuals with high self-efficacy exhibit stronger coping abilities and lower depression rates when encountering stressful events[26]. Stable social relationships and support networks provide a strong sense of belonging and security, which help prevent depression. For instance, family care and support from friends can significantly enhance individuals' psychological security, reducing depressive feelings stemming from loneliness and helplessness.

Social support improves social integration, reducing triggers for depression. Social integration refers to the degree of participation and sense of belonging in a social network. Highly socially integrated individuals often have stronger social support networks and higher self-esteem, effectively preventing depression[24]. Active participation in community activities and social interactions can enhance social connectedness, reducing loneliness and lowering the risk of depression. Community activities provide opportunities for social interaction and enhance social responsibility and belonging. Establishing and maintaining a broad social support network enables individuals to access more support and resources, improving their mental health[30]. For example, joining interest groups or engaging in volunteer services can help individuals build new social relationships and receive both emotional and practical support.

2.2.2 The role of social support in depression treatment

Social support significantly promotes treatment adherence among depression patients, meaning patients are more likely to follow their doctor's recommendations. Studies have found that patients receiving support from family and friends are more likely to take their medication on time, attend regular therapy sessions, and adhere to medical advice, thereby improving treatment outcomes[31]. Encouragement and supervision from family members can help patients overcome resistance to treatment, increasing adherence and effectiveness. Family members can assist in managing medication, reminding patients to take their medication on time, and accompanying them to therapy sessions. Friends' care and companionship can boost patients' motivation and confidence in treatment, enhancing the continuity and effectiveness of treatment. Emotional support provides comfort and understanding, helping depression patients alleviate negative emotions and enhance psychological security. Tangible support offers specific help, such as financial assistance and help with household chores, reducing patients' daily burdens and promoting recovery[22].

Social support helps depression patients improve social functioning and reintegrate into society. Through positive social interactions and support, patients can gradually regain self-confidence and social skills, rebuilding effective social relationships and promoting recovery[5]. Joining support groups or community activities allows patients to find others who have experienced depression, gaining emotional resonance and support, thereby improving social functioning and quality of life. Members of support groups can share experiences and coping strategies, offering emotional comfort and encouragement, helping patients rebuild social relationships. Community support can provide various forms of assistance, such as psychological counseling, vocational training, and social activities, helping patients rebuild social functioning and life skills[28]. For example, community psychological counseling services can offer professional psychological support and guidance, helping patients cope with life's challenges and pressures.

2.3 The Impact of Comprehensive Social Support on Depression Treatment Outcomes

2.3.1 Research design and methods

Randomized controlled trials (RCTs) are considered the "gold standard" for evaluating the efficacy of medical interventions, providing reliable evidence of causal relationships. In these studies, participants are randomly assigned to either an experimental group that receives social support interventions or a control group that receives standard treatment. This method minimizes selection bias and the impact of confounding variables. Experimental studies using RCTs have provided direct evidence of the effects of social support interventions. For example, Pfeiffer[32] conducted a study where patients with depression were randomly assigned to receive either standard treatment or enhanced social support. The results showed that those in the social support group experienced significant reductions in depressive symptoms and improved treatment adherence. This finding supports the positive therapeutic effects of social support and suggests that it can be a valuable supplement to traditional pharmacotherapy and psychotherapy.

Prospective cohort studies involve individuals who do not exhibit depressive symptoms at the start of the study, tracking the types and levels of social support they receive over time and its impact on subsequent depression development[32]. These studies help understand how social support acts as a preventive measure, reducing the incidence of depression. Long-term observational studies have shown that the continuity of social support is significantly associated with the long-term reduction of depressive symptoms, emphasizing the importance of maintaining good social relationships in managing depression.

Observational studies, through longitudinal tracking or cross-sectional research, have further explored the correlation between social support and depression treatment outcomes. George[33] followed thousands of patients and found that those

reporting higher levels of social support showed greater reductions in depressive symptoms and better recovery of social functioning after treatment. These studies highlight the role of social support in promoting mental health recovery and maintaining long-term well-being. Cross-sectional studies provide snapshots of the association between social support and depression treatment outcomes at specific points in time, while longitudinal studies reveal the dynamic relationship and long-term effects of social support on depression treatment outcomes over multiple time points.

2.3.2 Specific impacts of different types of social support

Emotional support, including providing comfort and boosting patients' self-esteem and self-worth, can alleviate negative emotions and reduce feelings of isolation, which are crucial for the recovery process in depression[34]. Tangible support, such as assistance with daily living and financial aid, is especially important for patients with limited functional capacity. This type of support helps alleviate daily stress, allowing patients to focus more on their treatment[35]. Informational support, which includes providing information about the illness, treatment options, and healthy lifestyles, can help patients better manage their condition, enhancing treatment autonomy and self-efficacy. Informational support aids patients in making informed treatment decisions, improving treatment adherence[32].

2.4 Cross-Cultural Comparisons

2.4.1 Differences in the role of social support across cultural contexts

Cross-cultural comparisons reveal differences in the effectiveness of social support on depression treatment across various cultural contexts. Cultural background influences individuals' perceptions and expectations of social support, the manner of its provision, and its impact on mental health.

Individualistic vs. Collectivistic Cultures: In individualistic cultures (e.g, the United States and Western European countries), social support primarily manifests as emotional and informational support, emphasizing individual autonomy and independence[36]. Individuals in these cultures are more inclined to seek professional help and formal support networks. In contrast, in collectivistic cultures (e.g, East Asian countries), social support relies more heavily on family and community, emphasizing group harmony and mutual assistance[37]. In these cultures, tangible assistance and emotional comfort often come from close family members and relatives, with support being more long-term and comprehensive.

Impact of Cultural Norms on Emotional Expression and Support Needs: Different cultures have varying levels of acceptance for emotional expression, affecting the communication and effectiveness of emotional support[38]. In some cultures, direct emotional expression may be deemed inappropriate or a sign of weakness, leading to more indirect and subtle forms of emotional support. For example, in Japanese culture, non-verbal support and tacit understanding are often considered key components of emotional support[39].

2.4.2 Case studies in cross-cultural research

Cross-cultural research through specific case studies delves into the concrete effects of social support on depression treatment in different cultural backgrounds. Taylor[40] conducted a cross-cultural comparison study between American and Japanese depression patients' social support systems. The study found that although American patients relied more on emotional and informational support, Japanese patients benefited significantly more from tangible and emotional support from family and community, reflecting the profound influence of cultural background on the forms and effectiveness of support.

Intervention studies in multicultural environments, such as Chu's[41] research on Asian Americans, showed that culturally adaptive interventions significantly improved treatment acceptance and outcomes. The study adjusted intervention content based on cultural background, emphasizing family involvement and culturally sensitive support measures, resulting in significant reductions in depressive symptoms in the treatment group.

2.5 Comprehensive Social Support Intervention Models

2.5.1 Introduction to comprehensive intervention models

Comprehensive social support intervention models aim to provide holistic and personalized intervention plans by integrating emotional, tangible, and informational support. A multi-level support system combines community engagement programs, online support networks, and face-to-face group meetings to create a layered support system. This model not only focuses on patients' mental health but also includes life support and health education[42]. Personalized support involves assessing patients' specific needs and developing individualized support plans. For instance, some patients may require more emotional support, while others need tangible help or informational support[43].

2.5.2 Analysis of successful and unsuccessful cases

Successful Case: The Comprehensive Community Support Program in the United States successfully reduced depressive symptoms among participants by providing customized mental health services, crisis intervention, and continuous social support. This program emphasizes multidisciplinary teamwork, offering comprehensive support ranging from psychological counseling to life skills training, significantly reducing depressive symptoms and improving participants' quality of life[44].

Unsuccessful Case: Some intervention programs fail due to a lack of consideration for the cultural characteristics and actual needs of the target population, resulting in low participation and poor outcomes. For example, a standardized support

program designed in the West was ineffective in immigrant communities because it failed to integrate culturally sensitive support measures, leading to poor treatment adherence and minimal symptom improvement[45].

2.5.3 Methods for evaluating effectiveness

Standardized Assessment Tools: Using standardized mental health scales such as the Beck Depression Inventory (BDI) and the Patient Health Questionnaire (PHQ-9) to regularly assess changes in patients' depressive symptoms. These tools are widely used in clinical research and have good reliability and validity[46].

Qualitative Evaluation: Collecting patients' subjective experiences and feedback on support interventions through in-depth interviews and focus groups. This method helps understand patients' acceptance, satisfaction, and subjective effectiveness of the interventions[47].

Long-Term Follow-Up Studies: Conducting long-term follow-up studies to assess the sustainability and long-term impact of interventions. These studies can reveal the long-term effects of comprehensive social support interventions in reducing depressive symptoms and preventing relapse[48].

3 DISCUSSION

Comprehensive social support has shown significant positive effects in the treatment of depression. Studies have demonstrated that patients who receive increased emotional, tangible, and informational support exhibit notable improvements in treatment adherence, symptom reduction, and overall quality of life. For instance, Dennis Dowswell[43] found in their systematic review that comprehensive social support effectively reduces depressive symptoms and enhances patients' psychological health and life quality.

This study supplements and expands existing social support theories, particularly in their application to depression treatment. Firstly, it further validates the Social Support Buffering Hypothesis[5], emphasizing the crucial role of social support in alleviating stress and improving mental health. Secondly, through cross-cultural comparisons, it reveals the influence of cultural context on the forms and effectiveness of social support, broadening the applicability and depth of social support theories[40]. Based on the findings, the study proposes new theoretical hypotheses, including a multidimensional integration model of comprehensive social support intervention. This model hypothesizes that combining emotional, tangible, and informational support in a multidimensional intervention maximizes the effectiveness of depression treatment. Additionally, the cultural adaptability intervention hypothesis suggests that adjusting the forms and content of social support interventions according to different cultural backgrounds significantly enhances their effectiveness and acceptability[41].

The results of this study have important implications for clinical practice. Firstly, clinicians should recognize and incorporate comprehensive social support as part of depression treatment plans to improve therapeutic outcomes. Secondly, developing personalized social support plans that ensure comprehensive coverage of emotional, tangible, and informational support can meet the diverse needs of patients[43]. When designing and implementing social support interventions, a multidimensional integration approach should be considered, combining emotional, tangible, and informational support to formulate comprehensive intervention plans. Adjusting the forms and content of support according to patients' specific conditions and cultural backgrounds ensures the effectiveness and applicability of interventions[45]. Ensuring the continuity of support interventions through long-term follow-up and evaluation allows for timely adjustments and optimization of intervention strategies[42].

Despite existing research highlighting the significant role of social support in depression treatment, several limitations remain. Many studies are confined to specific regions or populations, lacking broad representativeness. Some research designs are cross-sectional, failing to reveal long-term causal relationships. The lack of standardization in the forms and content of social support interventions limits the comparability and reproducibility of research results[42]. Future research should consider increasing sample diversity, expanding the geographical and demographic diversity of study samples to enhance the generalizability of research findings.

4 CONCLUSION

Comprehensive social support has demonstrated significant positive effects in the treatment of depression, encompassing multidimensional interventions such as emotional support, tangible support, and informational support. These interventions not only enhance patients' treatment adherence and psychological health but also significantly improve their quality of life. The importance of social support in the treatment of depression cannot be overstated. It serves as a crucial supplement to psychotherapy and pharmacotherapy, improving treatment outcomes and quality of life through various forms of support.

Future research should further explore the effectiveness of multidimensional integration models of social support and culturally adaptive interventions. Policymakers should encourage and support the promotion and application of social support interventions, especially in multicultural and diverse communities, to ensure that every patient with depression receives adequate social support.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Berto P, D'Ilario D, Ruffo P, Di Virgilio R, Rizzo F. Depression: Cost-of-illness studies in the international literature, a review. Journal of Mental Health Policy and Economics, 2000, 3(1): 3-10.

[2] Luppa M, Heinrich S, Angermeyer MC, Konig HH, Riedel-Heller SG. Cost-of-illness studies of depression: A systematic review. Journal of Affective Disorders, 2007, 98(1-2): 29-43.

[3] Cai H, Qu Z, Li Z, Zhang Y, Hu X, Hu B. Feature-level fusion approaches based on multimodal EEG data for depression recognition. Information Fusion, 2020, 59: 127-138.

[4] Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). Journal of Clinical Psychiatry, 2015, 76(2): 155-162.

[5] Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychological Bulletin, 1985, 98(2): 310-357.

[6] Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. Journal of Personality Assessment,1988. 52(1): 30-41.

[7] Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry, 2014,13(1): 4-11.

[8] Berkman LF, Syme SL. Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. American Journal of Epidemiology, 1979, 109(2): 186-204.

[9] Cobb S. Social support as a moderator of life stress. Psychosomatic Medicine, 1976, 38(5): 300-314.

[10] House JS, Landis KR, Umberson D. Social relationships and health. Science, 1988, 241(4865): 540-545.

[11] Barrera M. Distinctions between social support concepts, measures, and models. American Journal of Community Psychology, 1986, 14(4): 413-445.

[12] Van der Poel MGM. Delineating personal support networks. Social Networks, 1993, 15(1): 49-70.

[13] Cutrona CE, Russell DW. Type of social support and specific stress: Toward a theory of optimal matching. In I. G. Sarason, B. R. Sarason, G. R. Pierce (Eds.), Social support: An interactional view.1990: 319-366.

[14] Shakespeare-Finch J, Obst PL. The development of the 2-Way Social Support Scale: A measure of giving and receiving emotional and instrumental support. Journal of Personality Assessment, 2011, 93(5): 483-490.

[15] Thoits PA. Stress, coping, and social support processes: Where are we? What next?. Journal of Health and Social Behavior, Extra Issue, 1995: 53-79.

[16] World Health Organization. Depression, 2020.

[17] Miller GE, Chen E, Cole SW. Health psychology: Developing biologically plausible models linking the social world and physical health. Annual Review of Psychology, 2009, 60: 501-524.

[18] Johnson JG, Cohen P, Kasen S, Brook JS. A longitudinal investigation of social causation and social selection processes involved in the association between socioeconomic status and psychiatric disorders. Journal of Abnormal Psychology, 2006, 115(3): 488-497.

[19] Clark AE, Diener E, Georgellis Y, Lucas RE. Lags and leads in life satisfaction: A test of the baseline hypothesis. The Economic Journal, 2003, 113(488): 998-1013.

[20] Thompson RA, Meyer S. Social support and resilience. In S. N. Goldstein R. B. Brooks (Eds.), Handbook of resilience in children. Springer. 2016: 27-41.

[21] Lakey B, Cohen S. Social support theory and measurement. In S. Cohen, L. G. Underwood, B. H. Gottlieb (Eds.), Social support measurement and intervention: A guide for health and social scientists. Oxford University Press, 2000: 29-52

[22] Thoits PA. Mechanisms linking social ties and support to physical and mental health. Journal of Health and Social Behavior, 2011, 52(2): 145-161.

[23] Wethington E, Kessler R C. Perceived support, received support, and adjustment to stressful life events. Journal of Health and Social Behavior, 1986, 27(1): 78-89.

[24] Durkheim E. Le Suicide: Étude de sociologie. Paris: Félix Alcan, 1897.

[25] Kawachi I, Berkman LF. Social ties and mental health. Journal of Urban Health, 2001, 78(3): 458-467.

[26] Bandura A. Self-efficacy: The exercise of control. W.H. Freeman, 1997.

[27] Uchino BN. Social support and health: A review of physiological processes potentially underlying links to disease outcomes. Journal of Behavioral Medicine, 2006, 29(4): 377-387.

[28] Kleiman EM, Liu RT. Social support as a protective factor in suicide: Findings from two nationally representative samples. Journal of Affective Disorders, 2013, 150(2): 540-545.

[29] Rutter M. Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. British Journal of Psychiatry, 1985, 147: 598-611.

[30] Lakey B, Orehek E. Relational regulation theory: A new approach to explain the link between perceived social support and mental health. Psychological Review, 2011, 118(3): 482-495.

[31] DiMatteo MR. Social support and patient adherence to medical treatment: A meta-analysis. Health Psychology, 2004, 23(2): 207-218.
[32] Pfeiffer PN, Heisler M, Piette JD, Rogers MA, Valenstein M. Efficacy of peer support interventions for depression: A meta-analysis. General Hospital Psychiatry, 2011, 33(1): 29-36.
[33] George LK, Blazer DG, Hughes DC, Fowler N. Social support and the outcome of major depression. The British Journal of Psychiatry, 1989, 154(4): 478-485.
[34] Cohen S. Social relationships and health. American Psychologist, 2004, 59(8): 676-684.
[35] Lin N, Ye X, Ensel WM. Social support and depressed mood: A structural analysis. Journal of Health and Social Behavior, 2012, 31(4): 344-359.
[36] Kim HS, Sherman DK, Taylor SE. Culture and social support. American Psychologist, 2008, 63(6): 518-526.
[37] Uchida Y, Kitayama S. Development and validation of a sympathy scale. Journal of Japanese Psychology, 2001, 72(2): 144-151.
[38] Markus HR, Kitayama S. Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review, 1991, 98(2): 224-253.
[39] Nakao M, Tamiya N. (). Role of culture in the link between social support and health. Journal of Epidemiology, 2013, 23(4): 243-250.
[40] Taylor SE, Sherman DK, Kim HS, Jarcho J, Takagi K, Dunagan MS. Culture and social support: Who seeks it and why? Journal of Personality and Social Psychology, 2004, 87(3): 354-362.
[41] Chu JP, Kim HS, Jeong YS, Hahm HC. The role of culture in managing mental health: An east Asian perspective. Social Work in Public Health, 2012, 27(4): 353-370.
[42] Heaney CA, Israel BA. Social networks and social support. In Glanz, K, Rimer, B. K, Viswanath, K. (Eds.), Health Behavior and Health Education: Theory, Research, and Practice (4th ed). Jossey-Bass, 2008: 189-210.
[43] Dennis CL, Dowswell T. Psychosocial and psychological interventions for preventing postpartum depression. Cochrane Database of Systematic Reviews, 2(CD001134), 2013.
[44] Bond GR, Drake RE, Mueser KT, Latimer E. Assertive community treatment for people with severe mental illness: Critical ingredients and impact on patients. Disease Management Health Outcomes, 2001, 9(3): 141-159.
[45] Alegria M, Atkins M, Farmer E, Slaton E, Stelk W. One size does not fit all: Taking diversity, culture and context seriously. Administration and Policy in Mental Health and Mental Health Services Research, 2008, 37(1-2): 48-60.
[46] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 2001, 16(9): 606-613.
[47] Denzin NK, Lincoln YS. The SAGE Handbook of Qualitative Research. SAGE Publications, 2011.
[48] Hovens JG, Giltay EJ, Wiersma JE, Spinhoven P, Penninx BW, Zitman FG. Impact of childhood life events and trauma on the course of depressive and anxiety disorders. Acta Psychiatrica Scandinavica, 2012, 126(3): 198-207.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3003

RESEARCH ON THE STRATEGY OF ORGANIC INTEGRATION OF CHINESE EXCELLENT TRADITIONAL CULTURE AND IDEOLOGICAL AND POLITICAL COURSES IN COLLEGES AND UNIVERSITIES IN THE CONTEXT OF NEW LIBERAL ARTS EDUCATION

Wei Zhang*, Jing Jing
Department of Literature and Arts, Southwest University of Science and Technology, Mianyang 621000, Sichuan, China.
Corresponding Author: Wei Zhang, Email: 2551939160@qq.com

Abstract: Under the background of new liberal arts, the integration of Chinese excellent traditional culture into the ideological and political courses of colleges and universities (hereinafter referred to as the "IPC") is a need to inherit the Chinese excellent traditional culture and improve the quality of ideological and political teaching. The article analyses the current realistic dilemma of the integration of the excellent traditional culture and the IPC by exploring and researching the fit of the integration of the two, combined with the questionnaire survey. The article puts forward specific measures from teaching materials, teaching mode, teaching platform and the discourse system of ideology and politics, which help to realize the organic integration of Chinese excellent traditional culture and IPC in colleges and universities.
Keywords: New liberal arts; Chinese culture; Ideological and political courses; Integration

1 INTRODUCTION

Carrying out the teaching of ideological and political education in different disciplines and courses is an inevitable requirement and development trend of ideological and political education in colleges and universities under the new situation [1]. Some scholars have explored the feasibility as well as effective ways of integrating college English teaching, network information technology [2-3] and psychology into ideological and political education [4].
"Excellent traditional Chinese culture is the cultural root of the Chinese nation, and the ideology, humanistic spirit and moral norms it contains are not only the kernel of our Chinese mind and spirit, but also of great value in solving human problems"[5]. The construction of new liberal arts lies in adapting to the development of philosophical and social sciences in the new era and cultivating liberal arts talents in the new era. The new liberal arts emphasize interdisciplinary integration and focus on the cultivation of humanistic literacy and critical thinking. With comprehensive, interdisciplinary and integrative characteristics, it provides a new opportunity for the integration of Chinese excellent traditional culture in the IPC in colleges and universities. In the context of new liberal arts education, the excellent traditional culture is an important value resource for the teaching of IPC in colleges and universities. The teaching of Chinese excellent traditional culture and the teaching of IPC in colleges and universities has the possibility and necessity of integration in teaching content and teaching objectives [6]. And favorable policy environment support is an effective guarantee for the integration of the two [7]. Numerous researchers believe that the development of socialist culture with Chinese characteristics must adhere to the position of excellent traditional culture, which is compatible with the basic theory of IPC [8]. They further put forward the practical path for the integration of Chinese excellent traditional culture into ideological and political education in colleges and universities from the aspects of teachers, teaching ability and teaching methods, including strengthening the integration of teachers, improving the traditional cultural literacy of teachers of IPC, and enhancing the ability of the majority of teachers to use traditional culture [8-11].
From the origin and development mode of ideological and political ideology, traditional culture is an important factor influencing political form [12]. The integration of Chinese excellent traditional culture into the IPC can effectively enhance the ideology and moral quality of college students [13]. One scholars believe that the essence of traditional culture should be explored to fully stimulate students' patriotic enthusiasm and national consciousness, guiding them to form correct moral and social values [14]. The teaching method of exploring the integration of the two can maximize the teaching quality of the ideological class in colleges and universities, improve the ideological education of students [15], and enable contemporary college students to form a correct self-cognition [16]. Moreover, the real integration of Chinese excellent traditional culture into the field of ideological education is of positive significance for improving college students' identification with Chinese excellent traditional culture, promoting cultural inheritance [17-18], and facilitating the development and prosperity of socialist culture.

2 THE NECESSITY OF ORGANIC INTEGRATION OF CHINESE EXCELLENT TRADITIONAL CULTURE AND IPC IN COLLEGES AND UNIVERSITIES

2.1 The Need to Pass on the Excellent Traditional Chinese Culture

The excellent traditional Chinese culture has a long and profound history and is the crystallization of the wisdom of Chinese civilization. It contains rich moral concepts such as benevolence, courtesy, righteousness and honesty, which are the cornerstones for building socialist core values. Cultivating and building cultural confidence of college students in the new era has to start from the excellent traditional Chinese culture as an entry point. However, due to the impact of foreign cultures and the insufficient protection and promotion of local cultures, many college students lack a deep understanding of Chinese excellent traditional culture. Based on this background, the task of integrating Chinese excellent traditional culture into the IPC is imminent. The organic integration of the two can enable college students to have a more comprehensive in-depth understanding of relevant national policies and the current development of Chinese traditional culture. As a result, a system of ideological and political education with Chinese characteristics is formed. The ideological quality and moral quality of college students are improved, contributing to the cultivation of college students' patriotism and sense of family and country, and the establishment of a correct worldview, outlook on life and values.

2.2 Needs of Ideological and Political Education and Teaching

General Secretary Xi stressed that "in the new era and new journey, the construction of the ideological and political course is facing a new situation and new tasks, and must have a new appearance and new action"[19]. IPC in colleges and universities is the main channel and main position of ideological and political education for college students, undertaking the important mission of cultivating qualified socialist builders. Yet, the traditional IPC elaborates too much on the theoretical knowledge of Marxism, having the problems of abstract content and single form, which leads to students' low interest in learning and poor learning effect. The excellent traditional Chinese culture emphasizes the cultivation of personal character and the enhancement of humanistic qualities. It contains many humanistic concepts, such as "cultivate oneself, unify the family, rule the country, and pacify the world", "the rise and fall of the world is the responsibility of every man", "what you don't want to be done to yourself, don't do it to others". These concepts are in line with the educational objectives of the IPC in colleges and universities, and have a positive impact on the overall development of college students. Under the background of the construction of new liberal arts and globalization, the appropriate integration of Chinese traditional culture into the IPC will improve the comprehensive quality and cultural heritage of college students and lay a solid foundation for their future growth and development. In addition, it is conducive to enhancing students' understanding of and respect for the world's multiculturalism, and cultivating them to become new-age talents with global vision and cultural tolerance. Meanwhile, to a great extent, the content of Civics teaching can be enriched, so as to improve the quality of Civics teaching and enhance the effectiveness of IPC.

3 THE REALISTIC DILEMMA OF ORGANIC INTEGRATION OF CHINESE EXCELLENT TRADITIONAL CULTURE AND IPC IN COLEGES AND UNIVERSITIES

Firstly, the teaching concept is the key to the integration of the two. At present, some teachers still adhere to the traditional concept of education. Teachers pay too much attention to knowledge instillation and ignore the subjectivity of students, resulting in a lack of interaction and vitality in the classroom. On the contrary, students are also used to passively accepting theoretical knowledge and lack of active thinking. Moreover, the contents of the courses taught by some teachers are traditional and outdated. Teachers' understanding of the excellent traditional culture remains on the surface. They do not dig deep into its essence, which makes the content of the lessons lack depth and breadth.

Next, teaching resources are the prerequisite for the organic integration of excellent traditional culture and the IPC. The current lack of teaching resources for the IPC of excellent traditional culture is mainly reflected in the shortage of teachers, outdated content of teaching materials and insufficient digital resources. On the one hand, the IPC in colleges and universities lack a team of dual-qualification teachers who understand both ideology and politics and traditional culture. Some teachers have limited ability to understand and teach traditional culture and insufficient ability to integrate cultural resources, making it difficult to achieve effective interdisciplinary integration. On the other hand, the existing Ideological and Political textbooks are of older versions and lack reflection of the values of the new era and the innovations of traditional culture. Furthermore, the complexity of cultural resources, the lack of appropriate technical personnel, and the insufficient investment of financial resources have made it possible to preserve fewer digitized cultural resources, to the detriment of online education and distance learning.

Additionally, the evaluation system for the organic integration of excellent traditional culture and the IPC programme is weak. The present evaluation system for students' learning outcomes is not yet improved. It is assessed only by means of written tests such as accompanying tests and final exams. This approach pays insufficient attention to students' mastery of traditional culture learning, as well as critical thinking and innovation ability, lacking a comprehensive examination of students' all-round development. Also, there is a lack of effective evaluation mechanisms to assess students' learning outcomes when they participate in practical activities related to traditional culture and interdisciplinary teaching.

Finally, the limited nature of cultural and practical activities is a major difficulty that restricts the in-depth combination of the two. Restricted by class scheduling and teaching resources, the IPC in colleges and universities tend to focus on the teaching of theoretical knowledge. In addition, some colleges and universities have limited resources to invest in traditional cultural practice activities, including infrastructure and funds for activities. As a result, the difficulty in

arranging rich and varied practical activities makes students' understanding and feeling of traditional culture stay on paper, making it difficult for them to experience and comprehend it in depth. Even when practical courses are offered or cultural practice activities are organized, the problems of low student participation, a single form of activity and superficial content exist.

4 THE PRACTICAL PATH OF ORGANIC INTEGRATION OF CHINESE EXCELLENT TRADITIONAL CULTURE AND IPC IN COLLEGES AND UNIVERSITIES

4.1 Digging Deep into the Chinese Excellent Traditional Cultural Resources Contained in the Teaching Materials of IPC to Enhance the Cultural Connotation of Ideological and Political Education

As an important carrier of ideological and political education, the teaching materials of IPC in colleges and universities contain rich resources of Chinese excellent traditional culture. Digging deep into these resources can not only promote the cultural connotation of the teaching materials, but also enhance students' understanding and love of Chinese excellent traditional culture. Above all, the existing teaching materials should be analyzed in depth and the elements of Chinese excellent traditional culture contained in them should be excavated. It can be combined with the essence of traditional Chinese philosophical thinking, like Confucianism and Taoism, to explore the similarities and complementarities between them and Marxism. Secondly, the content of the teaching materials can be optimized and adjusted to systematically sort out the practical value and significance of Chinese philosophy, literature, art, poetry and so on, so as to help college students to improve their knowledge and understanding of Chinese excellent traditional culture. Furthermore, the teaching content needs to be supplemented to keep up with the times. When supplementing and perfecting the integration of Chinese excellent traditional culture and new science and technology, and the innovation and reproduction of Chinese excellent traditional culture in the new era, the Chinese excellent traditional culture resources contained in the teaching materials of the IPC in colleges and universities should be deeply explored. It is also necessary to strengthen the traditional culture training of teachers of IPC, and to upgrade their traditional culture literacy and teaching ability.

4.2 Innovating the Teaching Mode of Ideology and Politics and Perfecting the Protection System of Human Training

On the one hand, colleges and universities should adjust the existing curriculum system of IPC and add course modules related to Chinese outstanding traditional culture. For example, Chinese philosophy, literature and art are added to enable students to learn and understand Chinese excellent traditional culture in a systematic way. When formulating teaching plans, knowledge points closely related to Chinese excellent traditional culture are consciously chosen as teaching contents. For instance, in the architecture course, the design concept of Suzhou Garden can be introduced to combine traditional architecture with Chinese aesthetics. Interdisciplinary curriculum design is encouraged to integrate elements of Chinese excellent traditional culture into the teaching of other disciplines.

On the other hand, teachers should adopt diversified teaching methods, such as role-playing, game interaction and group discussion. A platform for public sharing can also be created to encourage students to share their travelling experiences after class. In this way students can discuss together the attractive features of different cities and cultures. At the same time, colleges and universities should enrich practical activities on campus and organize students to participate in practical activities related to Chinese traditional culture. Students can better understand and feel the charm of Chinese traditional culture by experiencing the design of calligraphy and painting, poetry recitation competitions, and cultural seminars. Besides, cultural expeditions can be carried out during the summer and winter holidays. Students are organized to visit museums and historical monuments on the spot to enhance their intuitive feeling and knowledge of traditional culture.

4.3 Integrating and Innovating a Web-based Platform System for Integrating Chinese Traditional Culture into the IPC in Colleges and Universities

By integrating and innovating the network platform system of integrating Chinese excellent traditional culture into the IPC in colleges and universities, the teaching efficiency and quality can be greatly improved. It also provides a powerful teaching aid for teachers. Before all, the cultural resources should be classified and presented in many aspects through characters, pictures, videos and so on. Secondly, it is necessary to establish a network platform for teachers and students to share resources, so as to achieve the optimal allocation and efficient use of resources. A preference tracking system is set up to analyse students' learning records on the platform according to big data, so as to know students' preferences. From this, adjustments and enhancements can be made to the weak areas of the platform. The reasons for the wide audience of popular resources are then analyzed, so that better cultural resources can be pushed. Personalized recommendations are made according to students' learning habits and interests. Course content needs to be updated in a timely manner to ensure that the current educational resources are the latest versions. The interactive function between teachers and students can also be enhanced through forum discussions, voting and other ways to guarantee that the platform is rich in resources, the content is novel and interesting, and the learning effect feedback is timely.

4.4 Researching the Teaching Discourse of IPC and constructing the Ideological and Political Discourse System of Excellent Traditional Culture

Building a discourse system of excellent traditional culture for ideology and politics is a systematic project. It requires in-depth excavation of the essence of traditional culture, combining it with modern concepts of ideological and political education, and forming discourse expressions with characteristics and characteristics of the times. The key to studying the discourse of IPC teaching and constructing the discourse system of excellent traditional culture and Ideological and Political discourse lies in grasping the epochal, innovative and practical nature of the discourse. Firstly, the discourse of IPC teaching should be closely related to the context of the times, interpreting traditional values from a new perspective and reflecting the development achievements and social changes of contemporary China. New media on the Internet can be used to innovate communication methods and enhance the sense of the times and affinity of the discourse system. At the same time, paying attention to social hot spots and responding to the concerns of the times can give new vitality and vigour to traditional culture. Secondly, the discourse of IPC teaching should focus on innovative expression. The discourse expression of traditional culture is often historical and regional, and differs somewhat from the context of modern society. Therefore, when building the discourse system of the excellent traditional culture of ideology and politics, it is necessary to use modern discourse expression. That is, the ideological connotation of traditional culture is transformed into a form of discourse that is easy to understand, accept and spread. This can be achieved by adopting more vivid and imaginative expressions such as contemporary popular hot words and hot pictures. Last but not least, the practical application of the ideological and political discourse system should be strengthened. The ideological and political discourse system and the practice of ideological and political teaching should be combined. Through classroom teaching, campus cultural activities, and social practice in a variety of ways, the elements of traditional culture of ideology and politics should be integrated into the daily life of students. Students can feel the charm of traditional culture in practice and to strengthen the timeliness of the ideological and political education.

5 CONCLUSION

Chinese excellent traditional culture is the treasure of the Chinese nation, containing rich philosophical thoughts, moral concepts and humanistic spirit. IPC in colleges and universities, on the other hand, are an important position for cultivating students' worldview, outlook on life and values. In the tide of new liberal arts education, the organic fusion of Chinese excellent traditional culture and the ideological and political courses in colleges and universities is particularly crucial. This integration is not only a kind of inheritance and promotion of traditional culture, but also a kind of innovation and development of modern education concept. It helps to improve the cultural literacy and humanistic spirit of students, and can also better play the important role of the IPC in fostering virtue through education. The organic integration of Chinese excellent traditional culture and the IPC in colleges and universities is a continuous process. How to better balance the relationship between tradition and modernity, theory and practice, and how to ensure the quality and effectiveness of teaching. All these issues need to be explored and improved in our future practice. The study will continue to deepen the exploration and practice of this strategy, and keep exploring more effective teaching methods and means to cultivate new-age talents with a deep cultural heritage and a high sense of social responsibility in the context of the new liberal arts.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This article was supported by Postgraduate Innovation Fund Project by Southwest University of Science and Technology (24ycx1105); Southwest University of Science and Technology Postgraduate Education Teaching Reform - Ideological and Political Education and Management Program (24yjgs05, 24yjgs09).

REFERENCES

[1] Li F, Fu H. Study on college English teaching based on the concept of ideological and political education in all courses. Creative Education, 2020, 11(7): 997-1007.
[2] Luo J. Teaching reform of ideological and political courses based on “Internet+”. Journal of Physics: Conference Series. IOP Publishing, 2020, 1533(4): 042020.
[3] Cheng P, Yang L, Niu T, et al. On the ideological and political education of material specialty courses under the background of the internet. Journal of Higher Education Research, 2022, 3(1): 79-82.
[4] He X, Chen P, Wu J, et al. Deep learning-based teaching strategies of ideological and political courses under the background of educational psychology. Frontiers in Psychology, 2021, 12: 731166.
[5] China.gov.cn. Xi Jinping's speech at the National Conference on Propaganda and Ideology. 2018, 08.
[6] Chi Chengyong. On the integration of excellent Chinese traditional culture and the teaching of ideological and political theory courses in colleges and universities . Ideological and Theoretical Education, 2014 (12): 63-67.

[7] Liu Ze. Strategies for Integrating Chinese Excellent Traditional Culture into Civic and Political Science Courses in Colleges and Universities.Leadership Science Forum, 2024, (04):153-156.
[8] Zhang Cuifang. Analysis of the Path of Integrating Chinese Excellent Traditional Culture into the Teaching of Civics and Political Science Courses in Colleges and Universities. Industry and Technology Forum, 2022, 21(24): 144-146.
[9] Zhu Hao. The Path of Integrating Chinese Excellent Traditional Culture into Ideological and Political Education in Colleges and Universities. Western Quality Education, 2024,10(06):63-66.
[10] Wei Yuhang. Research on the integration path of excellent Chinese traditional culture and ideological and political courses in colleges and universities. Education World, 2024, 6(1).
[11] Li Q. Research on the integration of Excellent Chinese traditional culture into Ideological and Political education in Colleges and Universities. International Journal of Frontiers in Sociology, 2021, 3(21).
[12] Eisenstadt S N. Cultural Traditions and Political Dynamics: The Origins and Modes of Ideological Politics. Hobhouse Memorial Lecture. The British Journal of Sociology, 1981, 32(2): 155-181.
[13] Liao L. The Application of Chinese Excellent Traditional Culture in College Students' Ideological and Political Education. 2018 International Conference on Management and Education, Humanities and Social Sciences (MEHSS 2018). Atlantis Press, 2018: 217-223.
[14] Su J, Wei X, Wang Y, et al. Research on the Integration of Traditional Culture and Ideological and Political Education in Colleges and Universities. 2019 5th International Conference on Education Technology, Management and Humanities Science (ETMHS 2019). 2019: 1265-1269.
[15] Jiuyang W. Analysis on the Ways of Integrating Chinese Excellent Traditional Culture into the Teaching of Ideological and Political Theory Course in Colleges and Universities. Proceedings of 2019 5th International Conference on Education Technology, Management and Humanities Science (ETMHS 2019). Institute of Management Science and Industrial Engineering: Computer Science and Electronic Technology International Society. 2019: 212-216.
[16] Sun Z. Research on The Integration of Chinese Excellent Traditional Culture into Ideological and Political Course. International Journal of Social Sciences in Universities, 2020,208.
[17] Yingying W. Research on the path of integrating traditional culture into ideological and political education in colleges and universities. Frontiers in Educational Research, 2021, 4(10).
[18] Gan R. The Integration and Inheritance of Chinese Excellent Traditional Culture in the Teaching of Ideological and Political Courses in Colleges and Universities. International Journal of Social Science and Education Research, 2023, 6(3): 143-149.
[19] China.gov.cn. Xi Jinping gives important instructions on building school's ideological and political programme. 2024, 05.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3004

ANALYSIS AND PREDICTION OF AIR QUALITY INFLUENCE FACTORS IN CHANGSHA CITY

WenHui Zeng
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 2541290358@qq. com

Abstract: With the continuous advancement of modernization, the problem of air quality is becoming more and more serious. In this paper, LASSO regression analysis is used to screen out some variables that have little impact on the research object, so as to construct the variable system of the research question. Then the multivariate linear regression analysis is carried out according to the variable system. Finally, the main pollutants affecting the air quality of Changsha City are analyzed, which can effectively predict the air quality of Changsha City and take corresponding measures to improve the air quality.
Keywords: LASSO regression; Multiple linear regression; Air quality; Influence factor

1 INTRODUCTION

With the continuous advancement of China's industrialization process and rapid economic development, while bringing a lot of convenience to people's lives, it has also caused a lot of environmental problems, among which air pollution is especially prominent[1-2]. The economy of our country has entered a period of high-quality development, while ensuring the economic development, we must also pay attention to the protection of the environment[3-4]. Today's air quality assessment standards are too simple, only by calculating several air pollutant index to determine the quality of air quality is good or bad, in the current complex air pollution, it is obviously not enough[5]. By constructing a more comprehensive air quality evaluation system and improving analysis methods, this paper can better evaluate air quality, so as to grasp the key points affecting air quality and carry out atmospheric protection more quickly and efficiently[6].

2 DATA AND METHODS

2.1 Data Collection and Source

2.1.1 Data source

National bureau of statistics;
China statistics yearbook;
China environmental air monitoring analysis platform;
China meteorological administration.

2.1.2 Data content

1) Air quality monitoring data
The daily data set of air quality in changsha city is from September 2021 to November 2022. The data summary of the atmosphere daily data is also included in China's latest concentration data, which records the data of typical atmospheric pollutants such as $PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and volatile product O_3.
2) Historical weather forecast data
Changsha Weather Forecast Daily data set, the data span from September 2021 to November 2022. The daily data set includes meteorological elements such as air temperature (maximum and minimum temperatures), weather, wind direction, and wind power.

2.2 Data Processing

Data symbols are described in Table 1:

Table 1 Description of symbols

Symbol	Explain
y	AQI
x_1	$PM_{2.5}$
x_2	PM_{10}
x_3	SO_2

x_4	CO
x_5	NO_2
x_6	O_3
x_7	minat
x_8	wc
x_9	wd
x_{10}	wp

2.3 Variable Selection and Prediction Methods

2.3.1 Lasso filter variable method

The steps to preliminarily screen variables using Lasso are as follows:
1) Feature standardization.
2) Lasso linear regression model was established.
3) Select the best adjustment parameters.
4) Screen out important variables.

2.3.2 Multiple linear regression analysis

The so-called regression analysis is to analyze the relationship between the dependent variable and the independent variable according to some indicators, and compare the analyzed dependent variable with the true value, and judge the pros and cons of the regression according to the error or goodness of fit between them. It can also study the changes of the dependent variable according to the independent variable obtained from the provided data, which can be used for predictive analysis. There are many factors affecting air quality in this paper, so multiple linear regression is selected. Construct random variables y and independent variables $x_1, x_2, ..., x_p$, between the multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon$$

In the equation, $\beta_0, \beta_1, ..., \beta_p$ is unknown parameter, β_0 called the regression constant term, $\beta_1, ..., \beta_p$, called the regression coefficient.

3 RESULT AND ANALYSIS

3.1 Lasso Model Screening Variable Analysis

Table 2 Results of LASSO regression coefficients

The characteristics of LASSO regression	LASSO regression coefficient
$PM_{2.5}$	1. 194
PM_{10}	-0. 027
SO_2	-0. 817
CO	-0. 631
NO_2	-0. 060
O_3	0. 390
maxat	0
minat	0. 188
wc	-1. 475
wd	-0. 395
wp	-0. 940

As can be seen from the results in Table 2 above, the regression coefficient of LASSO regression model is returned. It can be seen that only maxat is compressed to 0 among the regression coefficients of various features. This suggests that maxat has no significant effect on air quality. Among the features that have a significant impact on air quality, the features that have a positive impact on air quality are: $PM_{2.5}$, O_3 and minat, totaling 3 features; the features that have a negative impact on air quality are: PM_{10}, SO_2, CO, NO_2, wc, wd and wp, totaling 7 features.

3.2 Multiple Linear Regression Modeling

3.2.1 Variable selection

The samples were selected by LASSO variables based on AQI data of time series and $PM_{2.5}$, PM_{10}, SO_2, CO, NO_2, O_3,

minat, wc and wp.

3.2.2 Model building

This study conducted computational modeling based on R language software, and selected data from September 2021 to November 2022. According to the output results of R language, the multiple linear regression model of air quality is as follows:

$$y = 2.981 + 1.201x_1 - 0.077x_2 - 0.967x_3 - 0.724x_4 + 0.006x_5 + 0.370x_6 + 0.265x_7 - 1.361x_8 - 0.188x_9 - 0.888x_{10}$$

3.2.3 Model test

For the multiple linear regression model of this equation, the linear goodness test of the linear fitting model, the sum of the significance test and (F test) of the equation's overall multilinear regression model, the sum of the significance test of the variable linear regression model and (t test) were respectively analyzed. The statistical significance of the results is as follows.

1) Goodness of fit test

From the goodness of fit of the equation, R=0. 9541, R^2=0. 9104, and nearly equal to 1, indicating that the fit is good, and the independent variable can effectively explain 91. 04% of the change of the dependent variable.

2) Significance test of the overall linearity of the equation (F test)

The probability value of F statistics is 0. 00, because 0. 00<0. 01, when the independent variable is introduced, the possibility of its significance is much smaller than 0. 01, so we can well exclude the original assumption of global regression coefficient 0. 0, indicating that there is an obvious linear relationship between the independent variable and the dependent variable.

3.4.3 Prediction of multiple linear regression model

The data from December 1 to December 5, 2022 are predicted according to the regression equation model, as shown in Table 3.

Table 3 Comparison of predicted and actual values

Date	True value	Multiple linear regression predicted value	Multivariate forecast relative error /%
2022/12/1	34	33. 0315	0. 97
2022/12/2	43	42. 5354	0. 99
2022/12/3	57	59. 1668	1. 04
2022/12/4	77	73. 1888	0. 95
2022/12/5	63	63. 7064	1. 01

As can be seen from Table 1, there is little difference between the predicted value of multiple linear regression and the true value, indicating that the prediction accuracy of multiple linear regression model is high.

4 DISCUSSION

In this paper, 10 variables with significant impact on air quality in Changsha City were selected through LASSO regression analysis, and an index system for evaluating air quality in Changsha City was established. Due to the accuracy of LASSO regression analysis, this evaluation system can provide a good basis for the next model construction. Then, through the introduction and establishment of multiple linear regression model, combined with the example of Changsha City air quality, the model is studied and analyzed in many aspects. The research results show that the multiple linear regression model is more accurate in the prediction of Changsha City air quality.

According to the established multiple linear regression equation, after data standardization, $PM_{2.5}$, NO_2, O_3 and minat are positively correlated with air quality. PM_{10}, SO_2, CO, wc, wd and wp were all negatively correlated with air quality. Among them, $PM_{2.5}$ has a greater and positive impact on air quality. $PM_{2.5}$ is an important indicator to evaluate air quality in the world today. The higher the concentration of $PM_{2.5}$ in the air, the worse the air quality and the more serious the pollution. Weather conditions have a large and negative impact on air quality, good weather is good air quality. Therefore, it can be explained that the multiple linear regression analysis is reasonable.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Kai Su, Zhongshan Peng, Dan Zhu, Ruiqian Liu, Qin Wang, Rong Cao, Jun He. Water quality evaluation based on water quality index and multiple linear regression: A research on Hanyuan Lake in southern Sichuan Province, China. Water environment research: a research publication of the Water Environment Federation, 2024, 96(6): e11055-e11055.

[2] Mahmuda Akter, Elias Khalil, Md. Haris Uddin, Md. Kamrul Hassan Chowdhury, Shah Md. Maruf Hasan.

Artificial neural network and multiple linear regression modeling for predicting thermal transmittance of plain-woven cotton fabric. Textile Research Journal, 2024, 94(11-12): 1279-1296.

[3] Mahmoudi Mohammadreza, Toufigh Vahab, Ghaemian Mohsen. A Novel Multiple Linear Regression Approach for Predicting the Unconfined Compressive Strength of Soil. International Journal of Geomechanics, 2024, 24(8).

[4] Sharaf AlKheder. Experimental road safety study of the actual driver reaction to the street ads using eye tracking, multiple linear regression and decision trees methods. Expert Systems With Applications, 2024, 252(PA): 124222.

[5] Zhong H, Hu H, Hou N, et al. Study on Abnormal Pattern Detection Method for In-Service Bridge Based on Lasso Regression. Applied Sciences, 2024, 14(7).

[6] Jiang J, Li Y, Kleeman M. Air quality and public health effects of dairy digesters in California. Atmospheric Environment, 2024, 331120588.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3005

AN ANALYSIS OF EARLY WARNING FOR CREDIT CARD CUSTOMER CHURN

ChangFu Yang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 625087024@qq.com

Abstract: This paper aims to explore the development of credit cards in the Chinese market and the challenges posed by the rise of internet finance, while also analyzing customer churn issues and their management strategies. Since the introduction of credit cards to China in 1986, despite the initial lack of supporting facilities such as POS machines, credit cards have served as a monetary credit voucher, facilitating the small loan business of commercial banks. Over time, the credit card market has experienced rapid growth, becoming a vital channel for personal consumer loans. However, the emergence of internet finance has had a profound impact on the traditional banking business model, with customer churn becoming an increasingly prominent issue.
Keywords: Credit cards; Internet finance; Customer churn; Data mining; Risk management

1 INTRODUCTION

Since the introduction of credit cards in China in 1986, they have not only transformed traditional payment habits but also provided financial support for the small credit services of commercial banks. The widespread use of credit cards has enabled consumers to enjoy the convenience of "spend now, pay later," which has in turn spurred the rapid development of the market. However, with the rise of internet finance, traditional banking services are facing unprecedented challenges. Internet finance companies, taking advantage of the internet's convenience, have quickly attracted a large number of customers, breaking the closed loop of banking services, especially in areas that banks find difficult to reach, such as personal loans for college students.

The development of internet finance is not simply a matter of internet plus finance,but rather it is a new form of finance that relies on new technologies such as cloud computing and mobile networks, and combines with new types of services like online payments and social media. This transformation has not only reduced the cost of financial services but also improved efficiency. However, it has also led to an increase in the rate of customer attrition from traditional banks. According to data from the central bank, the growth rate of credit card issuance has been declining year by year since 2017, and the issue of customer attrition is becoming increasingly serious.

The application of machine learning to predict credit card customer churn has seen significant advancements. Studies like those by Xu et al.[1] and Nie et al. [2] have shown that hybrid models and logistic regression can achieve high accuracy in predicting customer churn. Lin et al.[3] and Caigny et al.[4] have further enriched the field with their innovative approaches combining different techniques. Ensemble methods, as explored by de Bock and van den Poel[5], have also proven effective. Collectively, these studies highlight the growing sophistication of predictive analytics in customer retention strategies.

This paper utilizes a bank customer dataset obtained from the Kaggle website, employing text mining techniques to extract the main characteristics of churned customers and conducting Exploratory Data Analysis (EDA) to gain a deeper understanding of the data. Subsequently, a Random Forest model is applied for feature extraction to identify key characteristics of churned customers, and effective customer churn control and risk prevention recommendations are proposed.

2 MODELING ANALYSIS

2.1 Model Introduction

Random Forest (RF) is an ensemble learning method with decision trees as its core building block. The algorithm improves the accuracy and robustness of classification by constructing multiple decision trees and integrating their prediction results. In a random forest, each decision tree classifies the input samples independently.

For the Random Forest algorithm, it is an ensemble classifier composed of K decision trees $h(X,\theta_k),k=1,2,\ldots,K$ as the basic classifiers. When a sample to be classified is input, the classification result output by the Random Forest is determined by the voting of the classification results of each decision tree. The sequence of random variables $\theta_k,k=1,2,\ldots,K$ is determined by the two main randomization ideas of the Random Forest: Bagging and feature subspace. Let N denote the number of training samples, and M represent the number of features. The construction algorithm is as follows:

1) The number of input features m is used to determine the decision outcome at a node in the decision tree (m < M).
2) With replacement sampling is performed N times from the N training samples to form a training set

(Bootstrap sampling).

3) For each node, m features are randomly selected, and the best way to classify based on these m features is calculated.
4) The individual trees are combined to form the Random Forest.

The process of training each decision tree is to train the entire random forest, and because each decision tree is independent of each other, its training can be carried out together, which will greatly improve the efficiency of the model. Each decision tree is trained in the same way and then combined to obtain K decision trees, and the required random forest model is formed. The samples to be predicted are obtained by ranking the weights of the problem solving, and the average of the output results of each decision tree is taken as the result of random forest prediction.

Random Forest Regression (RFR) is formed by the growth of decision trees related to a random vector θ, where the dependent variable is a continuous variable, and it is assumed that the training set is independently drawn from the distribution of a random variable. Let $h_i(x)$ be the regression model result of a single decision tree, then the predicted value of the Random Forest Regression is obtained by averaging the regression results of k decision trees $\{h(X,\theta_i), i=1,2,\ldots,k\}$, that is:

$$H(x)=\frac{1}{k}\sum_{i=1}^{k}h_i(x)$$

Where $H(x)$ represents the result of the combined regression model.

The Random Forest method uses the bootstrap resampling method to obtain different sample sets, and constructs decision tree regression models using these sample sets, thereby increasing the differences between models and enhancing the ability of extrapolation prediction.

In summary, the steps of the Random Forest algorithm are as follows:

Step 1: A set of samples $\{T_k, k=1,2,\ldots,K\}$ is calculated by the weights selected by a decision tree for the problem.

Step 2: Classification results:

1) Individual decision trees are generated by randomly drawing K samples with replacement from the training set;
2) During the generation of the decision tree, a subset of attributes is randomly selected with equal probability;
3) The above two steps are repeated to generate K decision trees;
4) Finally, the optimal result of the final vote is output.

2.2 Indicator Selection

In order to identify the key factors influencing users' credit card satisfaction and the indicators that can reflect the trend of users' credit card use intention, this study used text mining technology to screen out 18 key indicators from the hot words related to credit card churn. These metrics include: user age, gender, household size, education level, marital status, income level, credit card tier, credit card registration time (months), number of products held, number of inactive months in the last 12 months, number of contacts in the last 12 months, total revolving credit card balance, average open purchase credit limit, change in transaction value between Q4 and Q1, total transactions in the last 12 months, number of transactions, and average frequency of credit card usage.

2.3 Model Solving

The random forest model is a powerful tool that is able to assess the impact of different factors on the churn rate of credit card users. In this study, we quantified the effect of each independent variable on the dependent variable by adjusting the model parameters and training a random forest model by adjusting the model parameters and using the churn rate of bank credit card users as the dependent variable. Once the model was trained, we got an importance score for each variable, which reflects the relative impact of each factor on the churn rate.

The relative error of the model is 0.0425, indicating that the model has high prediction accuracy. The results of the analysis showed that the total number of transactions and the number of transactions were the two most important factors influencing user churn, and their importance scores were significantly higher than those of other variables. This is followed by changes in the total amount frozen and the number of transactions, which constitute the second most important group of factors influencing user churn. The importance score of change in transaction amount and average frequency of use is low, but still higher than other factors, making up the third echelon. The importance score for the number of products held is in the fourth tier. The mean values for age and open purchase credit are classified as the fifth band while the importance scores for credit card time on book, months of inactivity, and number of contacts are in the sixth bracket. Factors such as educational attainment, number of family members, gender, marital status, income rating, and other factors have relatively low importance scores, while credit card ratings have the lowest degree of influence among all factors. Through the analysis of the random forest model, we can clearly identify the main factors affecting

the churn of bank credit card users, and provide targeted strategy suggestions for banks. Random forest model impact factor importance ranking can be seen in Figure 1.

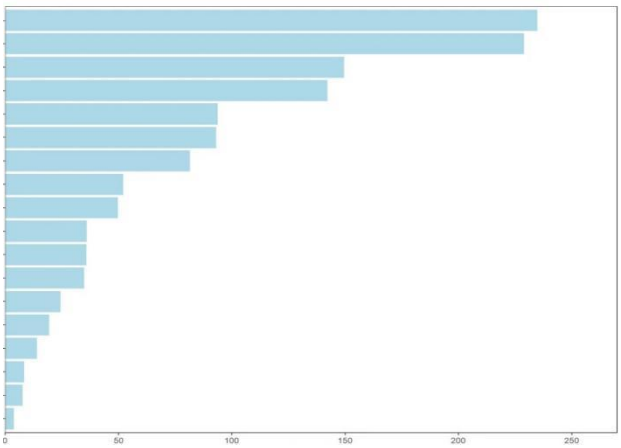

Figure 1 Random Forest Model Impact Factor Importance Ranking

According to the analysis results of the random forest model, we find that the main driving factors for the churn of bank credit card users include key economic indicators such as total transaction value and number of transactions. These findings point to the possibility that the credit card limits offered by banks for different levels of users may not be sufficient to meet their needs. Specifically, the growth of total transaction value is negatively correlated with the decrease in churn, suggesting that banks may be effective in reducing churn if they can provide credit lines that match users' spending power.In addition, an increase in the number of transactions is likewise associated with a lower churn rate. This means that frequent transaction activity may enhance users' dependence on and satisfaction with banking services. Therefore, banks can appropriately increase the frequency of credit card use on the premise of ensuring that users have the ability to repay, so as to stabilize and increase the user base.

Based on the importance of the influencing factors shown by the random forest model, banks can formulate targeted policies, such as adjusting credit limits and optimizing transaction experience, to reduce the churn rate of credit card users. In addition, by building a churn early warning model, banks can more effectively manage customers, identify high-risk customer groups, and take customer care measures to maximize the retention of these customers, thereby enhancing the ability of enterprises to resist customer churn risks. This data-driven strategy not only helps banks optimize customer relationship management, but also improves the personalization and competitiveness of banking services, ultimately leading to increased customer satisfaction and loyalty.

3 CONCLUSIONS AND RECOMMENDATIONS

The purpose of this paper is to use data mining technology to construct an early warning model for credit card customer churn for a bank. By quantitatively analyzing the churn problem of credit card customers and applying modern data mining methods such as random forest, an early warning model was successfully established, which can effectively monitor the risk of customer churn. This study combines statistical testing and data mining techniques to realize the integration of statistics and business practice, as well as the unification of qualitative and quantitative analysis methods. Through the test of the validation dataset, the model confirms its effectiveness in the early warning of credit card user churn, which provides strong support for solving the problem of user churn. The main results of this paper are summarized below:

1) Personalized marketing strategy: With the intensification of business competition, Internet financial enterprises need to implement personalized marketing for different credit card users. This study demonstrates the role of random forest model in quantifying user behavior characteristics and assisting marketing decision-making, and emphasizes the importance of data mining technology in banking business improvement and customer relationship management.
2) Feasibility of model establishment: Based on expert survey and statistical analysis, this study discusses the feasibility of establishing an early warning model for credit card customer churn and identifies the key factors affecting customer churn.
3) Empirical research: Through the case study of text data mining, a random forest prediction model is established, which provides an effective churn analysis framework for enterprises.

Credit card churn early warning analysis is an emerging field in credit card customer relationship management. The variability of customer needs and the expansion of choice increase the risk of customer churn. Therefore, understanding and maintaining valuable customers, improving the bank's competitiveness and reducing operational risks have become the key to the bank's business.

The user portrait system developed by the AI Lab provides the foundation for intelligent marketing in banks. Credit card churn warning models not only help target potential churned customers and improve customer retention, but can also be applied to all stages of the customer lifecycle, such as prospecting, nurturing high-value customers, increasing customer loyalty, and extending customer lifecycles. It is expected that more intelligent marketing model technologies and applications will be developed and applied in the future.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xu, Y., Rao, C., Xiao, X., Hu, F. Novel Early-Warning Model for Customer Churn of Credit Card Based on GSAIBAS-CatBoost. CMES-Computer Modeling in Engineering & Sciences, 2023, 137(3).

[2] Nie, G., Rowe, W.G., Zhang, L., Tian, Y., Shi, Y. Credit card churn forecasting by logistic regression and decision tree. Expert Syst. Appl., 2011, 38: 15273-15285.

[3] Lin, C. S., Tzeng, G. H., Chin, Y. C. Combined rough set theory and flow network graph to predict customer churn in credit card accounts. Expert Systems with Applications, 2011, 38(1): 8-15.

[4] De Caigny, A., Coussement, K., De Bock, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research, 2018, 269(2): 760-772.

[5] De Bock, K. W., & Van den Poel, D. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. Expert Systems with Applications, 2011, 38(10): 12293-12301.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3006

ANALYSIS OF GRASS-ROOTS BEHAVIOR AMONG YOUNG GROUPS

Chen Chen
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 2531086929@qq.com

Abstract: With the rise of the Internet and e-commerce, social platforms have integrated e-commerce into their ecosystems, giving rise to the "planting grass" culture. This paper examines the grass planting behavior of young users across different platforms. We conducted a literature review, designed a study using unequal probability three-stage sampling, and distributed 425 questionnaires, receiving 314 valid responses. Data was analyzed using descriptive statistics, cluster analysis, and a random forest model. Findings show that Little Red Book emphasizes practicality and authenticity but has inconsistencies between grass planting and purchasing. Tik Tok focuses on convenience and accuracy, with high consistency between planting and purchasing. Bilibili highlights professionalism, but users exhibit low purchasing frequency. Weibo users show inconsistent behavior and low usage frequency. Recommendations include enhancing consistency between planting and buying for Little Red Book, adopting user-specific strategies for Tik Tok, improving UP management for Bilibili, and diversifying methods for Weibo. Each platform has unique strengths and weaknesses. Long-term strategies and market research are essential for platforms to meet user needs and achieve rapid development.
Keywords: Grass platform; Unequal probability three-stage sampling; Random forest model; Cluster analysis

1 Introduction

According to the "50th Statistical Report on China's Internet Development," as of June 2022, the number of internet users in China reached 1.051 billion, with mobile internet users numbering 1.047 billion, accounting for 99.6%[1]. With the increasing number of internet users, the importance of mobile social networking in daily life has grown, and social media has rapidly risen[2]. In the era of social media, the media landscape has shifted from professional operations to global personal user participation, becoming a global sharing tool. Users utilize social media to establish virtual interpersonal relationships and achieve the integration of content production and consumption. Social media has surpassed traditional social functions, integrating business, social, and media resources to become a comprehensive platform.

Social platforms target the e-commerce market by attracting users through scenes, relationships, and content, making "planting grass" a crucial means of maintaining user growth. "Planting grass" content is widespread, influencing users' consumption decisions. Brands and merchants also use "planting grass" marketing to enhance user understanding and loyalty, thereby boosting sales. However, alongside "planting grass," there is also the phenomenon of "uprooting." Considering that young people are the main internet users with significant consumption potential, this paper studies the current situation of "planting grass" behavior among young users and provides targeted suggestions for social platforms.

Existing research primarily focuses on three aspects. First, research on "planting grass" marketing. Du Kangyi and Zhao Hongshan [3]pointed out issues in Xiaohongshu's marketing and provided suggestions; Jiang Jianguo and Chen Xiaoyu [4]discussed the problem of false marketing; Fu Donghan and Hu Li[5] suggested introducing industry KOLs, focusing on users' curiosity, and deepening user content; Wei Jiaxiao and Liang Yingtao[6], as well as Nie Luli[7], proposed strategies for optimizing short video KOL "planting grass" content; Li Zhongmei and Huang Min [8]studied Xiaohongshu's "planting grass" marketing strategies and provided countermeasures.

Secondly, consumer behavior research. Liu Siyu [9]used structural equation modeling to analyze survey data, finding that the affinity, professionalism, and reliability of "planters" positively influenced consumer interaction; Wang Shixin[10] used the Delphi method to extract and verify factors affecting consumers' "planting" and "uprooting"; Sun Yan[11] analyzed the transformation of user participation behavior from a participatory culture perspective; Fang Xuelan[12] studied the impact of the online "planting grass" landscape and called for rational treatment; Deng Sha[13] summarized the factors influencing consumers' purchase intentions through KOL "planting grass."

Lastly, research on "planting grass" culture. Hong Yu[14] explored the communication principles and issues behind the "planting grass economy" and proposed solutions; Lei Jinghui[15] studied the trust-building mechanisms in virtual sharing communities; Li Jiawen[16] analyzed the mechanisms behind "planting grass" culture; Dai Jiaqi and Xi Shizhen used the DART model and value co-creation theory to reveal the value co-creation mechanisms of "planting grass."

In summary, current research in China on "planting grass" mainly focuses on the content and optimization of "planting grass," consumer purchase intentions, and "planting grass" culture, with most studies taking Xiaohongshu as an example or focusing on KOLs. However, there is a lack of research on other mainstream "planting grass" platforms and user purchase decisions. Therefore, this paper attempts to explore the current situation of "planting grass" behavior among young users based on different "planting grass" platforms.

2 PRELIMINARY ANALYSIS OF THE SURVEY DATA

After completing the questionnaire design, we conducted a pilot distribution of the questionnaire to the young demographic based on the sampling frame. A total of 340 questionnaires were distributed and 340 were returned. After screening, 314 valid questionnaires were obtained. The collected data were then subjected to item analysis, as well as reliability and validity testing, to examine the discrimination of the questions, the reliability of the measurement results, and the validity of the questionnaire.

2.1 Item Discrimination Analysis

After collecting the questionnaires, we first analyzed the discrimination of each item on the scale. This was done to test the reliability of the scale and individual items. We used the high-low group mean difference test method. Items with low discrimination were excluded.

For the item analysis, we focused on Q22 (user satisfaction with TikTok). We summed the scores of all 9 items for each respondent, assigning the average score to any unanswered item. Respondents were then sorted by total score, with the top 27% as the high group and the bottom 27% as the low group. We performed an independent sample t-test on each item between the high and low groups. The t-values for all 9 items were significant ($p < 0.05$), indicating good discrimination.

Similarly, we analyzed Q23-Q25 for other platform users using the same method. The t-values for all items were significant ($p < 0.05$), indicating good discrimination.

In conclusion, the questionnaire passed the item analysis and is suitable for formal investigation.

2.2 Reliability Testing

Reliability testing of the questionnaire design refers to analyzing the accuracy of the measurement results, specifically evaluating the precision and consistency of data obtained from repeated use of the questionnaire. Reliability analysis reflects the true degree of the measured characteristics. We used Cronbach's alpha to measure the internal consistency of the questionnaire items, with the reliability coefficient ranging from [0,1]. The calculation formula for Cronbach's alpha is as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_T^2}\right) \quad (1)$$

In this context, k represents the total number of items in the scale, S_i^2 is the variance of the scores for item i, and S_T^2 is the variance of the total scores across all items. Cronbach's alpha evaluates the internal consistency of the scores across items in the scale. A higher Cronbach's alpha indicates greater reliability. Ideally, a well-designed questionnaire should have a reliability coefficient above 0.80. A coefficient between 0.70 and 0.80 is acceptable, but if the internal consistency of subscales is below 0.60 or the overall reliability is below 0.80, the questionnaire should be revised.

For this questionnaire, the Cronbach's alpha for each section exceeded 0.7, with an overall reliability of 0.726. Similarly, the Cronbach's alpha for the Q23-Q25 scales also exceeded 0.7, indicating the scientific and reasonable design of the questionnaire structure and items.

3 RELATED ANALYSIS

3.1 Content Validity

Content validity, also known as face validity or logical validity, refers to whether the designed items can represent the content or theme to be measured. The correlation between each subscale and the total scale is used as an indicator to assess the content validity of the scale, examining how well a scale represents the intended theme. The results show a significant correlation between each factor and the total scale score (p-values are all less than 0.01), indicating that the scale has good internal consistency.

3.2 Construct Validity

Construct validity refers to the degree of correspondence between the measurement results and the measured value. Common methods include correlation analysis, factor analysis, and structural equation modeling. Factor analysis can extract common factors to check if the questionnaire measures the assumed structure. Before factor analysis, we conduct the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. A KMO value close to 1 indicates strong variable correlation, suitable for factor analysis; a high Bartlett's test value indicates high variable independence, also suitable for factor analysis.

KMO>0.9 is very suitable for factor analysis; 0.8<KMO<0.9 is suitable; above 0.7 is acceptable, 0.6 is poor, and below 0.5 is unsuitable. Using SPSS, the KMO coefficient is 0.746, and the P-value is 0.000, indicating the questionnaire's structure is well-designed and suitable for factor analysis.

These results indicate that the pre-survey questionnaire meets the survey's objectives.

3.3 Test of Randomness

During the survey, ensuring the randomness of sampling was crucial, and we employed a run test to assess the randomness of categorical variables. The run test evaluates the randomness of a sample based on the number of consecutive occurrences of variable values. We sorted the sample observations in ascending order and identified the median (or mean), dividing the sample into two parts: above and below the median (mean). The run test examines the randomness of the sample based on the number of runs formed by alternating values. In this survey, we conducted a run test for single-sample variable values to determine if the occurrences of a particular variable were random. For instance, in the case of gender, where males were coded as 0 and females as 1, we observed 84 occurrences of 0 and 230 occurrences of 1, resulting in a total of 119 runs (R). We used a constructed Z statistic to perform the run test.

In the Z statistic, where n0 is the number of occurrences of 0, n1 is the number of occurrences of 1, and R is the total number of runs, we calculated Z = -0.73. At a significance level of α = 0.05, the value falls between -1.96 and 1.96, indicating no sufficient reason to reject the null hypothesis. Thus, the sequence of gender data in the survey exhibits a relatively high degree of randomness.

Using SPSS software, we conducted sample randomization tests on various categorical variables in the questionnaire. The results indicate that the sequence of data for most variables does not violate randomness, suggesting that the survey data achieved a high level of randomization and was successful.

4 ANALYSIS OF CHARACTERISTICS OF USERS

4.1 Model Selection and Establishment

In data mining, k-modes is considered a method suitable for clustering analysis of categorical data sets. It extends from k-means by adapting it to handle categorical data after discretizing numerical data. The k-modes clustering method measures distances between samples using the 0-1 matching dissimilarity measure and updates cluster centers using modes. Under this distance measure, dissimilarity between a sample variable and a cluster center is computed based on the count of features where they differ (1 for different values, 0 for identical values). In k-modes, smaller distances between two sample variables indicate greater similarity.

Let $D = \{x_1, x_2, ..., x_n\}$ be a dataset consisting of n sample variables. Each sample variable x_i has m categorical attributes, where $x_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$ composed of mmm attributes. Here, $x_{ij} \in A$, and A is the attribute set $A = \{A_1, A_2, ..., A_m\}$. The dissimilarity between two variables x_i, x_j is measured as:

$$d(x_i, x_j) = \sum_{l=1}^{m} \delta(x_{il}, x_{jl}) \tag{2}$$

If the n variables in D are divided into k clusters, with the i-th cluster centroid denoted as $q_i, i = (1, ..., k)$, and the distance measure from the i-th variable's data belonging to the r-th cluster is $d(x_i, q_r)$, then the objective function that defines the distances from all variables to all cluster centroids is:

$$P(W, Q) = \sum_{r}^{k} \sum_{i}^{n} w_{ir} d(x_i, q_r) \tag{3}$$

Here, $q_r = \{q_{r1}, q_{r2}, ..., q_{rm}\}$, where q_{rh} is the mode of the h-th attribute of the r-th cluster.

$$\sum_{r=1}^{k} w_{ir} = 1, \sum_{r=1}^{k} w_{ir} \in \{0,\ 1\},\ 0 < \sum_{i}^{n} w_{ir} < n, 1 \le i \le n, 1 \le r \le k \tag{4}$$

Due to the predominantly categorical nature of the collected data, we utilized the k-modes model for analysis in the following steps:

Specify the dataset D and the number of clusters k.

Arbitrarily select k sample variables as initial cluster centers, with each center representing a cluster.

Compute the distance measure $d(x_i, q_r)$ between all variables and the k initial cluster centers. Assign each variable to the closest cluster center to form new clusters.

For each attribute within each cluster, select the most frequent attribute value as the new ·

Regarding variable selection for modeling, we chose nine variables: gender, age, disposable income, platform engagement level, platform usage frequency, grass planting frequency, purchase frequency, preferred product categories, and preferred product attributes to classify user types.

4.2 Model Introduction

The random forest regression model uses the Bagging ensemble learning algorithm to create multiple samples by repeatedly sampling with replacement from the original dataset. Each sample set is used to build regression decision trees, where each tree selects feature variables as nodes and splits them based on the modeling sample data. Each split corresponds to an output value, allowing the model to classify feature variables and derive regression results. The final prediction in random forest regression is the average output of all decision trees. Like traditional linear regression models, variables are categorized into independent (feature) and dependent (target) variables, and feature selection is crucial before modeling. The process is outlined as follows:

1) Assuming there are M samples and N feature variables in the original dataset, the M samples are divided into training and testing sets in an 80:20 ratio. The training set is used to build the random forest model, while the testing set is used to evaluate the model's performance.
2) Using the Bagging ensemble learning algorithm, k iterations of random sampling with replacement are conducted on the training set, where each iteration samples approximately 2/3 of the training set data. These samples are used to form new sample sets, from which k regression decision trees are constructed. The remaining 1/3 of the training

set data, which is not sampled in each iteration, forms a set called Out-of-Bag (OOB) data. The OOB data is used for analyzing the importance of feature variables.

3) Selecting n (< N) feature variables from the total N variables to compare Root Mean Square Error (MSE) or Mean Absolute Error (MAE), dividing them into multiple units, with each unit corresponding to a fixed value. The selected n feature variables are used as the number of variables mtry randomly sampled each time a single decision tree grows.
4) The decision tree continuously splits and classifies feature variables based on sample data from the training set and the selected n feature variables, undergoing unpruned splitting growth until outputting the regression prediction result.
5) Taking the average of the output results from k decision trees as the final value for regression prediction using the random forest model.

4.3 Factor Analysis

After encoding and processing the information collected from the questionnaire, using whether to make a purchase as the dependent variable y, and gender, age, monthly disposable income, grass-planting platform, usage frequency, grass-planting format, grass-planting sharer, and grass-planting style as independent variables $x_i, i = (1,2,...,8)$, a random forest classification model is employed to compute feature importance. This is used to assess the influence of each factor on purchasing behavior.

$$y = \begin{cases} 1, & \text{purchase frequency} \neq 1 \\ 0, & \text{purchase frequency} = 1 \end{cases} \quad (5)$$

Among them, "content format" includes four formats: 'short videos', 'long videos', 'images and text', and 'live streaming'. "grass-root influencers" includes four types: 'celebrities', 'internet celebrities', 'professional bloggers', and 'amateurs'. "grass-root aesthetic" includes five styles: 'theme list-style', 'plot-style', 'unboxing-style', 'favorite items compilation-style', and 'experience review-style'.

Divide the data into training and test sets in a 70% to 30% ratio, and input the training set data into the random forest model for computation.

In terms of the Accuracy metric, factors such as usage frequency, grass-root style, and grass-root influencers have a significant impact on user purchase frequency. On the other hand, according to the Gini index, usage frequency, grass-root platform, and grass-root content style are the primary factors influencing user purchase frequency. Among these factors, usage frequency stands out as the most critical factor influencing whether users make purchasing decisions. Therefore, collectively, factors such as usage frequency, grass-root style, grass-root influencers, grass-root platform, and grass-root content style are crucial in influencing user purchasing behavior.

5 SUMMARY AND RECOMMENDATIONS

5.1 Summary

5.1.1 From the platform's perspective

Based on different content ecosystems, each platform has distinct grass-root characteristics:

1) Xiaohongshu (Little Red Book) emphasizes practicality and authenticity:

Practicality: Beauty and food content integrate product features through skill tutorials or practical user experiences, clearly showcasing user focus points.

Authenticity: Creators share post-product-experience notes from multiple dimensions, presenting product characteristics and usage experiences in a lifelike manner that attracts users.

2) Douyin (TikTok) focuses on convenience, closed-loop interaction, and precision:

Convenience: Features like live streaming and comment section links shorten the distance between users and brands, facilitating rapid conversion from interest to action, enhancing the effectiveness of grass-root recommendations.

Closed-loop interaction: Integration of public, commercial, and private domain traffic creates a closed-loop diversion enhancing the effectiveness of grass-root recommendations.

Precision: Douyin's centralized algorithm recommends content tailored to user preferences and interests, significantly increasing content consumption time and accurately meeting user needs, effectively stimulating grass-root interest.

3) Bilibili emphasizes creativity and professionalism:

Creativity: Diverse formats such as reviews, unboxing, and tutorials are presented through medium to long videos, captivating users with creative and entertaining content.

Professionalism: Gatherings of experts and diverse specialized content make Bilibili a unique hub for knowledge-based and technical content, establishing its distinctive content style.

4) Weibo focuses on longevity, practicality, and amplification:

Longevity: Multi-layered comments enhance product features and deepen user engagement, building a comprehensive scenario that sustains long-term grass-root influence.

Practicality: Shareable content featuring creators in various roles and everyday scenarios enhances practicality and penetration of grass-root content.
Amplification: Leveraging celebrity and trending topics attracts brands to promote and amplify product endorsements on Weibo, quickly expanding grass-root influence across various circles.

5.1.2. From the user's perspective

1) Xiaohongshu: Users show broad preferences but inconsistency between grass-root interest and purchasing behavior. Among young users aged 18-22 with disposable incomes of 1000-1500 RMB, there are two groups: frequent grass-root engagers who purchase occasionally. They exhibit diverse preferences across product categories and prioritize quality and price.
2) Douyin: Users generally align grass-root interest with purchasing behavior, but significant differences exist among user categories. The audience includes diverse income groups (below 1000 RMB to above 2000 RMB), with active daily users primarily interested in daily essentials. Higher-income male users prioritize daily essentials without specific product attribute preferences.
3) Bilibili: Users vary significantly in platform usage frequency, but overall, purchasing frequency is low. There are categories of users who rarely use the platform and those who log in daily, yet both groups show low purchasing behavior. Quality and price are important attributes across all user categories, focusing on practicality.
4) Weibo: Users demonstrate inconsistency between grass-root interest and purchasing frequency, with overall low platform usage. Among the two categories of female users, there are instances of alignment between usage and purchasing frequency, but discrepancies in grass-root engagement.
5) Users across platforms generally have broad preferences for product categories, particularly focusing on aesthetic appeal, with a majority falling within the 1000-1500 RMB income range.

5.2 Recommendations

5.2.1 Xiaohongshu

Xiaohongshu users have diverse preferences for products, but face challenges where grass-root interest does not always translate into purchase behavior, and where platforms for grass-root content and shopping differ. The key to addressing these issues lies in converting traffic into customers. While the platform is positioned as a sharing platform, influencers can potentially convert their followers into loyal customers.
Firstly, Xiaohongshu serves as a handy tool for users in daily life, offering solutions for various needs. To expand its market, highlighting advertisements where users can directly purchase products on the platform is essential.
Secondly, raising the bar for merchant entry and improving platform services, including pre-sales, after-sales, and user convenience, requires extensive market research to identify and address platform deficiencies.
Lastly, understanding user needs and boosting usage frequency is crucial. Adjusting the interface based on user preferences, inviting renowned brands to join the platform for increased exposure, and enhancing user engagement through tailored content and interactive activities are effective strategies.
These efforts aim to enhance Xiaohongshu's appeal and user engagement, ensuring that it meets user expectations while fostering a vibrant community of content creators and enthusiasts.

5.2.2 TikTok

TikTok users show consistent interest between grass-root content and purchases, yet there are significant differences across user categories. Male users, particularly those with higher disposable incomes, lean towards lifestyle products. TikTok's advantage lies in its convenience, with live streaming providing a direct product experience and comment links bridging the gap between users and brands. To enhance TikTok's grass-root strategy, addressing user categories is crucial.
Based on our segmentation, TikTok users are categorized into four groups. For females aged 23-25 with disposable incomes of 1000-1500 RMB, preferences include beauty, daily essentials, snacks, and clothing, focusing on product attractiveness, brand reputation, and pricing. However, their engagement in grass-root content and purchasing is relatively low. To address this, tailored pushes for beauty, daily essentials, snacks, and clothing categories can be based on user profiles and algorithmic analysis.
For males aged 23-25 with incomes above 2000 RMB, although they use the platform daily, they rarely engage in grass-root content or purchasing. There's untapped potential here due to their platform reliance. Market research is needed to explore products of interest to males, possibly including more male-oriented brands.

5.2.3 Bilibili

Bilibili users vary in how often they use the platform and generally have low purchase rates. Young males prefer educational supplies and value product reputation and pricing. Young females have a broader preference, including beauty products, daily essentials, snacks, beverages, clothing, and educational supplies, also focusing on product reputation and pricing.
Initially centered on anime and related content, Bilibili has evolved into a learning and lifestyle platform where UP creators share their favorites and Vlog records. Grass-root content is tailored to UP creator styles and segmented into different groups. Converting users into buyers effectively depends largely on managing UP creators on the platform.
Furthermore, Bilibili could strengthen partnerships with beauty and fashion brands, leveraging its professional creative style. Special incentive programs could encourage UP creators in these fields to produce high-quality content,

expanding user demographics and improving grass-root categories. Attention to pricing sensitivity among all users allows collaborations with brands to offer cost-effective grass-root experiences.

5.2.4 Weibo

Weibo users exhibit inconsistent rates of interest in products they endorse and purchase, alongside generally low usage frequency. The user base can be categorized into two main groups, both consisting of females. Users aged 18-22 with disposable incomes between 1000-1500 RMB frequently use the platform and make purchases, favoring beauty and skincare products while prioritizing brand reputation and pricing. Those aged 23-25 within the same income bracket show interest in lifestyle goods, clothing, and educational supplies, but their usage and purchasing frequency are comparatively lower. Weibo should enhance its appeal and increase user engagement, particularly by highlighting niche products.

For mature female users aged 23-25, Weibo serves primarily as a hub for celebrities and their fans, necessitating diverse content offerings beyond entertainment to cater to varying user needs, including increasing exposure for niche products. Regarding male users, Weibo could expand opportunities such as celebrity live-streamed product promotions and streamline purchasing processes to improve convenience.

Improving platform functionality, introducing shopping capabilities, simplifying the endorsement-to-purchase process, and enhancing user convenience are critical steps for Weibo. Strengthening community management, particularly through topic-specific forums and groups, and fostering closer interactions between fans and influencers will amplify the influence of key opinion leaders in product endorsements.

In conclusion, each grass-root marketing platform should leverage market research to optimize their strengths and address user needs to ensure sustained growth.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] China Internet Network Information Center. The 50th Statistical Report on Internet Development in China. (2022-08-31) [2022-09-20]. Available at: http://cnnic.cn/n4/2022/0914/c88-10226.html.
[2] Clay Shirky. Cognitive Surplus. Translated by Hu Yong and Haliss. Beijing: Renmin University of China Press, 2011: 204.
[3] Du Kangyi, Zhao Hongshan. Grass-Roots Marketing of Xiaohongshu App. National Circulation Economy, 2019(22): 34.
[4] Jiang Jianguo, Chen Xiaoyu. Network "planting grass": social marketing, consumption induction and aesthetic fatigue. Learning and Practice, 2019(12): 125131.
[5] Fu Donghan, Hu Li. A brief discussion on the phenomenon and development trends of mobile Internet grass planting marketing. Communication and Copyright, 2020(08): 142-144.
[6] Wei Jiaxiao, Liang Yingtao. Research on content optimization strategy of short video platform KOL "planting grass". Audiovisual World, 2021(05): 59-61.
[7] Nie Luli. Research on the content optimization strategy of social platform KOL "planting grass". News Frontier, 2022(13): 73-74.
[8] Li Zhongmei, Huang Min. Research on countermeasures of "planting grass" content marketing in the context of new media—taking Xiaohongshu as an example. Shopping Mall Modernization, 2022(21): 13.
[9] Liu Siyu. Research on women's impulse buying intention under social marketing. Xiamen University, 2020.
[10] Wang Shixin. "Planting grass"/"pulling grass": Research on fan consumption behavior in the context of social media. Henan University, 2020.
[11] Sun Yan. Research on the "grass planting" behavior of Xiaohongshu users from the perspective of participatory culture. Dongbei University of Finance and Economics, 2022.
[12] Fang Xuelan. Research on the "grass planting" landscape of the Internet from the perspective of consumerism. Zhejiang Gongshang University, 2022.
[13] Deng Sha. Research on the impact of KOL's "grass planting" on users' purchase intention. Yantai University, 2022.
[14] Hong Yu. A brief analysis of the "grass-planting economy" in the era of self-media. Audiovisual, 2020(02): 169-170.
[15] Lei Jinghui. Research on trust construction based on "grass planting" sharing in virtual sharing communities. Shenzhen University, 2020.
[16] Li Jiawen. A brief analysis of the culture of "planting grass" in the context of the Internet - Taking Xiaohongshu as an example. Sound Screen World, 2022(21): 117-119.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3007

STUDY ON THE INFLUENCING FACTORS OF GUANGXI'S TOTAL EXPORTS BASED ON RIDGE REGRESSION AND LASSO REGRESSION

YuHe Cheng
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: yuhecheng@stu.gxnu.edu.cn

Abstract: Since China's accession to the WTO, foreign trade has been developing rapidly, and as an important part of the national economy, the importance of foreign trade in China's economic development has been increasing. In recent years, the total amount of import and export of Guangxi Zhuang Autonomous Region has continued to increase, but it also faces some problems, so it is of practical significance to analyze the influencing factors of the total amount of import and export. Based on the goods import and export data of Guangxi Zhuang Autonomous Region from 2002 to 2021, this paper selects five factors, namely total retail sales of social consumer goods, fiscal expenditure, regional gross domestic product, per capita disposable income of urban permanent residents and the exchange rate of the RMB against the US dollar, and establishes a multiple regression model by using the R language software to analyze the influencing factors. The model was improved by statistical test, multiple covariance test and heteroskedasticity test, and random forest was used for prediction. Finally, according to the results of empirical analysis, corresponding countermeasure suggestions are put forward.
Keywords: Total import and export amount; Influencing factors; Ridge regression; LASSO regression

1 INTRODUCTION

In recent years, with the continuous advancement of global economic integration, studying the influencing factors of import and export trade has become an important topic in economics. Zhang Zhanpeng studied the impact of exchange rate fluctuations on China's import and export trade [1], pointing out that exchange rate fluctuations have a significant impact on trade volume, which provides important reference for analyzing the factors affecting Guangxi's total export volume. Xue Yushi explored the relationship between the four budgets and economic growth through empirical analysis in R language, providing technical reference[2].

In terms of regional economic research, Zhang Qingxiu and Li Hongmei used ridge regression to analyze the influencing factors of consumer demand in Hebei Province[3]. Li Jiacheng combined ridge regression and principal component regression methods to analyze the influencing factors of consumer level among residents in Hunan Province[4]. These studies indicate that ridge regression is effective in solving multicollinearity problems. Zhu Hailong and Li Pingping further used ridge regression and LASSO regression to study the influencing factors of fiscal revenue in Anhui Province, demonstrating the advantages of LASSO regression in variable selection and model simplification[5].

In addition, Li Duoduo practiced data visualization in multiple linear regression analysis through R Studio, providing a reference for data processing and result display in this study[6]. Sha Jing and Hu Deng respectively studied the influencing factors of total exports in Jiangsu Province and Shaanxi Province, using multiple regression analysis methods, providing a paradigm and empirical reference for this study[7-8].

This article aims to study the influencing factors of Guangxi's total export volume based on ridge regression and LASSO regression. By selecting relevant economic data from Guangxi region and using these two regression models to conduct quantitative analysis and variable selection on various influencing factors, key factors affecting Guangxi's total export volume are revealed, providing scientific basis for Guangxi's export trade policy and methodological support for related research.

2 MODELING AND VARIABLE SELECTION

2.1 Introducing Variable

The data in this paper comes from the Guangxi Statistical Yearbook, and the data of five variables of Guangxi Zhuang Autonomous Region from 2002 to 2021 are selected for research and analysis: the total amount of imports and exports (USD billion) y as the dependent variable, the total retail sales of consumer goods (RMB billion) x_1, the general budget expenditure of local finances (RMB billion) x_2, the Gross Regional Product (RMB billion) x_3, the per capita disposable urban income (RMB) x_4 and the RMB exchange rate (USD=1) x_5 as the independent variables. 4 and the exchange rate of RMB to USD (USD=1) x_5 are the independent variables. Analyse the factors affecting the total amount of import and export of Guangxi and construct a forecast model for the total amount of import and export.

2.2 Data Sources

This article selects data from the Guangxi Bureau of Statistics from 2002 to 2021 to analyze the impact of various explanatory variables on the total import and export volume of Guangxi.
Due to the different units of the selected indicators, the data was first standardized to eliminate the influence of different unit scales on different variables.

2.3 Multiple Regression Modelling

A multiple linear regression model was constructed using R software using the lm() function for parameter estimation as shown in Table 1.

Table 1 Least squares parameter estimation table

Coefficients: Estimate	Std.	Error	t-value	Pr(>|t|)	
(Intercept)	-9.049e-14	5.003e+00	0.000	1.00000	
x1	4.175e+01	7.282e+01	0.573	0.57544	
x2	2.143e+02	5.787e+01	3.703	0.00236	**
x3	-2.291e+02	1.169e+02	-1.960	0.07019	.
x4	2.333e+02	1.513e+02	1.541	0.14552	
x5	2.458e+01	1.006e+01	2.444	0.02840	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The standardised multiple regression model is:

$$y = -9.049 \times 10^{-14} + 41.75x_1 + 214.3x_2 - 229.1x_3 + 233.3x_4 + 24.58x_5 \quad (1)$$

$$R^2 = 0.9939 \quad \overline{R}^2 = 0.9917 \quad F = 456.6 \; p - value = 5.521e\text{-}15 \quad (2)$$

From the above regression results, it can be seen that the of the model is 0.9939 and the modified decidable coefficient is 0.9917, which indicates that the model fits the sample very well. The F-statistic of the model is 456.6, and the corresponding P-value is 5.521e-15, which is significantly less than 0.01, so the model passes the F-test. However, in this model, only the general budget expenditure of local finance (x_2) and the exchange rate (x_5) passed the t-test, while the other explanatory variables did not pass the t-test, suggesting that not all the explanatory variables have a significant effect. Therefore, it needs to be considered that the data selected may not fully meet the assumptions of the least squares estimation.

2.4 Hypothesis Testing for Linear Regression Models

Run the plot() function in R to perform residual analysis to test whether the model meets the corresponding assumptions, and output four model diagnostic plots, as shown in Fig. 1, in the order of residual-fit plots (top left), normal Q-Q plots (top right), scale-position plots (bottom left), and residual-leverage plots (top right).

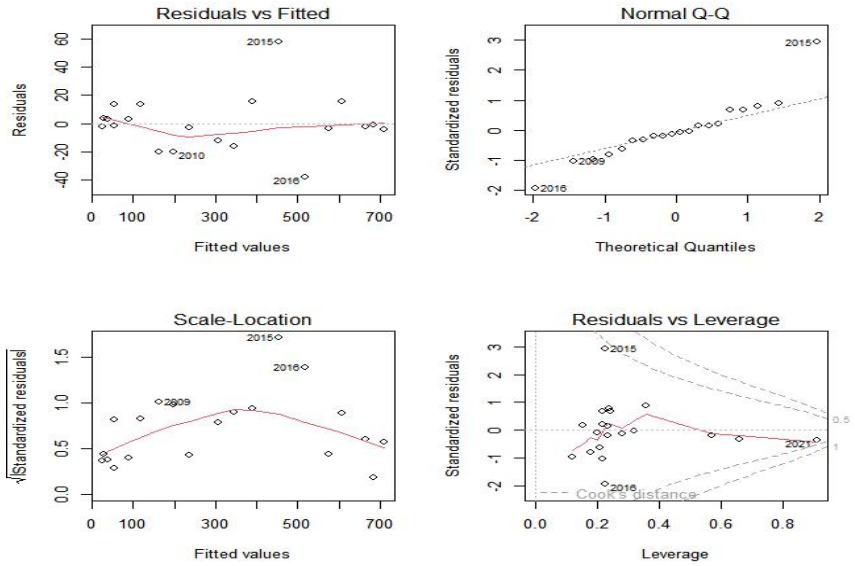

Figure 1 Model Diagnosis Diagram

As can be seen in Figure 1: The residual-fit plot (upper left) shows that the residual values are not systematically correlated with the fitted values, and the red line is basically smooth, indicating that the dependent variable is linearly

correlated with the independent variable, satisfying the linear correlation assumption. The data points in the normal Q-Q plot (top right) are arranged diagonally, which basically meets the normality assumption. The scale-position plot (bottom left) shows that the points around the horizontal line are randomly distributed, satisfying the assumption of homoscedasticity. In the residual-leverage plot (upper right), the model can be found to be free of strong influence points by Cook's Distance. In summary, the regression model meets the statistical assumptions and the model is valid and reasonable.

2.5 Multicollinearity Test

2.5.1 Correlation coefficient matrix

The correlation coefficient is used to determine the degree of correlation between two variables. The matrix of correlation coefficients is shown in Table 2:

Table 2 Correlation Coefficients Between Indicators, 2022-2021

	y	x_1	x_2	x_3	x_4	x_5
y	1	0.9890321	0.9955657	0.9875641	0.9906076	-0.6098053
x_1	0.9890321	1	0.994032	0.9925758	0.9963613	-0.6717087
x_2	0.9955657	0.994032	1	0.9928045	0.995308	-0.6360986
x_3	0.9875641	0.9925758	0.9928045	1	0.9981337	-0.6107133
x_4	0.9906076	0.9963613	0.995308	0.9981337	1	-0.6436158
x_5	-0.6098053	-0.6717087	-0.6360986	-0.6107133	-0.6436158	1

The above results show that there is a strong correlation between the independent variables, with most of the correlation coefficients reaching 0.9 close to one.

2.5.2 Variance inflation factor (VIF)

The empirical judgement method shows that when 0<VIF<10, there is no multicollinearity; when 10⩽VIF<100 , there is strong multicollinearity; when VIF⩾100, there is severe multicollinearity.

Table 3 Variance Inflation Factor (VIF)

x_1	x_2	x_3	x_4	x_5
201.247267	127.133335	518.562312	869.305744	3.841913

As can be seen from the results in Table 3, there is indeed a very serious multicollinearity directly in the independent variables of this data.

2.6 Modelling by Ridge Regression and LASSO Regression

From the above results, it can be seen that the model has serious multicollinearity, so the following paper uses ridge regression and LASSO regression to eliminate the multicollinearity and to build a model to study the problem.

2.6.1 Ridge regression

1) Selection of ridge parameters for ridge regression

The data were standardised for the independent variable and centred for the dependent variable respectively through R language to plot the ridge trace as shown in Figure 2. The ridge parameter estimates obtained from the generalised cross validation (GCV) method were selected to determine $\lambda = 0.02$.

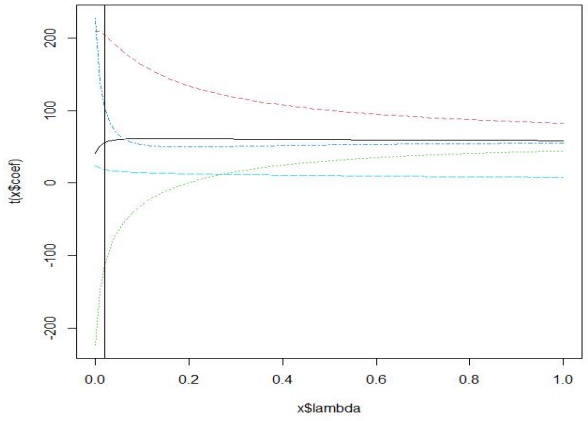

Figure 2 Ridge Trace Map

2) Parameter estimation to build a ridge regression model

The corresponding parameter estimates were obtained using the linearRidge() function to establish the ridge regression model, and the results were analysed based on the results given by the ridge regression model, and the results are listed in Table 4.

Table 4 Results of ridge regression analysis

	Estimate	Scaled estimate	Std.Error (scaled)	T-value (scaled)	Pr(>|t|)	
(Intercept)	5.929e-15	NA	NA	NA	NA	
x1	6.176e+01	2.692e+02	6.143e+01	4.383	1.17e-05	***
x2	1.104e+02	4.810e+02	7.062e+01	6.812	9.65e-12	***
x3	2.541e+01	1.108e+02	5.511e+01	2.010	0.0445	*
x4	5.303e+01	2.312e+02	3.976e+01	5.814	6.09e-09	***
x5	1.110e+01	4.839e+01	3.360e+01	1.440	0.1498	

The standardised ridge regression equation is:

$$y = 269.2x_1 + 481x_2 + 110.8x_3 + 231.2x_4 + 48.39x_5 \quad (3)$$

After standardised treatment, the intercept term of this model has no null value, and from (3), it can be seen that: x_1, x_2, x_3, x_4, x_5 are positively correlated with the total amount of Guangxi's import and export, and the changes of the above variables will cause the total amount of import and export to change in the same direction. Although the ridge regression method significantly improves the variable coefficients, there are still insignificant regression coefficients. The disadvantage of ridge regression is that it cannot perform variable selection and still includes all independent variables, thus failing to completely solve the problem of multicollinearity. Therefore, Lasso regression will be used in the following section to remedy this shortcoming.

2.6.2 LASSO

1) Select the variables in order

A LASSO regression model is built and variables are selected sequentially. Table 5 shows the variable selection of each explanatory variable sequentially as the parameter t increases into the regression model. Figure 3 shows the results of variable selection for the LASSO regression model, where the bottom horizontal axis indicates the ratio of the model regression coefficients, the data on the right vertical axis indicates the corresponding explanatory variables, and the data on the left vertical axis indicates the standardised parameters; the dotted lines represent the variables, and the vertical solid lines represent the penalty values.

Table 5 LASSO Regression Variable Selection

Call:					
lars(x = trainx, y = trainy)					
R-squared: 0.992					
Sequence of LASSO moves:					
	x2	x4	x5	x1	x3
Var	2	4	5	1	3
Step	1	2	3	4	5

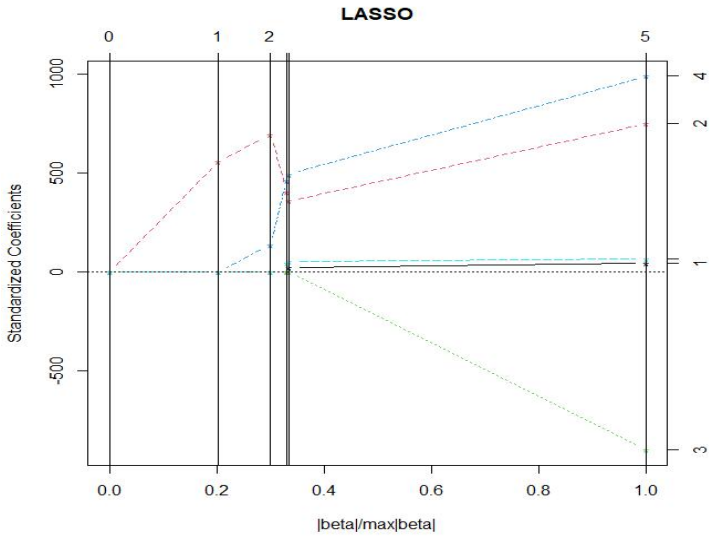

Figure 3 LASSO Plot of Total Imports and Exports Data

As can be seen from Table 5 and Figure 3, the variables selected sequentially for the LASSO regression are

$x_1, x_2, x_3, x_4, x_5$and the judgment coefficient is 0.992, indicating a very good fit.

2) Principle of minimum value of C_p

The statistic is used as a measure of multicollinearity between the variables and the smaller the value of C_p, the better the number of subsets it selects. The variation of values in LASSO solution is shown in Table 6. Where Step represents the number of steps and Rss represents the residual sum of squares.

Table 6 Variation of Values in LASSO Solving

Step	Rss	Cp
1	88033	94.8418
2	8109	2.5657
3	7031	3.2937
4	6994	5.2504
5	5934	6.0000

As can be seen from Table 6, the $C_p$value reaches the minimum value of 2.5657 in the second step, corresponding to the variable selection step is the second step. Therefore, after the screening of variables, and are selected as the two independent variables, and the expression of LASSO regression can be obtained:

$$y = 4901.764 + 203.49028x_2 + 41.52979x_4 \quad (4)$$

Through LASSO regression, two influential explanatory variables were selected, i.e., the main factors that significantly affect the total amount of import and export of foreign trade in Guangxi Province are: general budget expenditure of local finances (x_2), and per capita disposable income of urban residents (x_4). And both variables are positively correlated with the total amount of imports and exports.

1 CONCLUDE

By comparing the above analyses, it is found that the model obtained by the LASSO regression method is better, so it can be seen through model (3): local financial general budget expenditure (x_2) and urban residents' disposable income per capita (x_4) are the most significant factors affecting the total import and export trade of Guangxi in 2002-2021. Therefore, Guangxi can increase its total import and export trade by increasing the general budget expenditure of local finance and raising the disposable income of urban residents, so as to promote the development of Guangxi's import and export trade.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Zhang Zhanpeng. Research on the impact of exchange rate changes on China's import and export trade. North University of Technology, 2022.
[2] Xue Yushi. The relationship between the four budgets and economic growth - an empirical analysis based on R language. Journal of Economic Research, 2022(29): 109-112.
[3] Zhang Qingxiu, LI Hongmei. Analysis of factors influencing consumer demand in Hebei Province based on ridge regression. China Market, 2022(23): 23-27.
[4] Li JC. Analysis of factors affecting consumption level of Hunan residents based on ridge regression and principal component regression. China Collective Economy, 2022(21): 10-12.
[5] Zhu Hailong, LI Pingping. Analysis of factors affecting fiscal revenue in Anhui Province based on ridge regression and LASSO regression. Jiangxi University of Science and Technology, 2022, 43(01): 59-65.
[6] Dodo Li, Xing Yu, Sheng Han, He Zhu, Yi Yuan, Jie Shen, Jingfeng Lin, Xia Li, Yena Gan, Jianping Liu. Practice of R Studio software for data visualisation of multiple linear regression analysis. Chinese Journal of Evidence-Based Medicine, 2021, 21(04): 482-490.
[7] Sha Jing, Yang Yang, Zeng Gongli. A study on multiple regression analysis of factors affecting total exports of Jiangsu Province. Software, 2020, 41(10): 256-259.
[8] Hu Deng. Research on the influencing factors of total import and export amount in Shaanxi--Based on multiple linear regression model. Contemporary Economy, 2018(14): 88-89.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3008

ANALYSIS AND FORECAST OF THE AVERAGE SALES PRICE OF RESIDENTIAL COMMERCIAL HOUSING

XinYue Dang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 845387601@qq.com

Abstract: In recent years, China's residential commercial housing market has shown pronounced regional disparities, especially between large cities and smaller urban areas. This study utilizes cluster analysis to compare market conditions in 2020 and 2002. In 2020, regions were categorized into three tiers, with Beijing and Shanghai exhibiting the most significant differences. In contrast, the 2002 data divided regions into four tiers, with Guangdong Province standing alone in the first tier, indicating a more balanced market. From 2012 to 2020, Beijing's average housing prices increased substantially, whereas changes in the Guangxi Zhuang Autonomous Region were more moderate, reflecting significant regional economic disparities. Neural network models are used to make predictions on the data. This empirical analysis underscores the diversity and economic factors influencing China's residential commercial housing market. By comparing different regions and examining temporal sequences, the study provides theoretical insights into regional development imbalances, emphasizing the need for precise regional policies to promote coordinated market development.
Keywords: Average sales price of residential commercial housing; Cluster analysis; Neural networks; Forecast

1 INTRODUCTION

Commercial residential housing refers to properties planned and developed by state real estate development institutions and various enterprises, available for sale, rental, or mortgage to private homeowners. Purchasing commercial residential housing is the primary means through which individuals and families acquire housing[1].

Real estate investment, as one of the three major sectors of fixed asset investment, holds considerable significance. Real estate is not only a key component of the social consumption market but also a pillar industry driving economic development[2]. A critical variable reflecting the development of real estate is property prices. The year 2008 marked a turning point in national policy; in response to rising property prices, the government began adjusting real estate policies to regulate the market. In 2010, to curb the rapid increase in housing prices and reduce speculative purchases, the government issued the "New Ten National Articles." This policy aimed to counteract the rapid rise in prices and speculation. Subsequently, in 2011, the "New Eight National Articles" were introduced to further control the excessive increase in housing prices[3]. After 2013, following two to three years of emergency suppression policies, the government would relax policies to adapt to a slowdown in economic growth or to address inventory reduction needs. This means that when the market declines or demand decreases, the government may appropriately relax policies to promote real estate sales[4]. In 2018, despite multiple rounds of policy regulation, housing prices continued to show a persistent upward trend[5].

In recent years, influenced by significant pandemic risks and policies in certain countries, housing prices have started to decline significantly. By October 2022, the number of cities experiencing a decrease in commercial residential property prices had markedly increased among 70 major and medium-sized cities, with prices falling month-on-month across various city tiers[6]. The 2016 Central Economic Work Conference provided a more detailed description of efforts to reduce real estate inventory, emphasizing the need for all regions to adhere to the national development strategy that "houses are for living in, not for speculation." The conference also stressed the importance of researching and accelerating the establishment of a series of foundational systems and long-term mechanisms tailored to local conditions and market dynamics.

This paper, incorporating both temporal and spatial dimensions, offers valuable theoretical insights for understanding and addressing regional development imbalances. The study's conclusions underscore the necessity of adopting more precise regional policies to promote the coordinated development of the national commercial residential property market.

2 DATA SOURCES AND DESCRIPTIONS

The research data was sourced from the National Bureau of Statistics, showcasing data from certain provinces in 2020. The prices of residential commercial properties are influenced by multiple factors, which are complex and include both quantitative and qualitative data. For the convenience of this study, the impact indicators on housing prices are primarily represented by quantifiable variables.

After preliminary data processing and initial field visits, six variables have been selected as explanatory variables for subsequent research(Table 1 and Table 2).

Table 1 Variable Description Table

variable	illustrate
Y	Average sales(Yuan)
X_1	Sales area(ten thousand/m2)
X_2	Investment in real estate development(Yuan)
X_3	Gross Domestic Product (GDP)(100 million yuan)
X_4	Year-end resident population(10,000 people)
X_5	All the people Disposable income per cianapita(Yuan)
X_6	Consumer Price Index

Table 2 Raw Data Tables

region	Y	X_1	X_2	X_3	X_4	X_5	X_6
Beijing	42684	733.59	2317.08	35943.3	2189	69434	101.7
Tianjin	16391	1220.74	2084.8	14008	1387	43854	102
Hebei	8251	5572.25	3746.74	36013.8	7464	27136	102.1
Shanxi	6877	2549.49	1431.8	17835.6	3490	25214	102.9
……	……	……	……	……	……	……	……
Hainan	16751	626.19	946.41	5566.2	1012	27904	102.3
Chongqing	8917	4814.49	3189.05	25041.4	3209	30824	102.3
Sichuan	8041	10902.37	5330.14	48501.6	8371	26522	103.2
Guizhou	5600	4929.93	2572.34	17860.4	3858	21795	102.6
Yunnan	8267	4175.88	3317.55	24555.7	4722	23295	103.6
Tibet	8824	81.62	118.47	1902.7	366	21744	102.2
Shaanxi	9624	3902.4	3225.45	26014.1	3955	26226	102.5
Gansu	6467	1863.81	1010.28	8979.7	2501	20335	102
Qinghai	8164	420.64	292.32	3009.8	593	24037	102.6

Overall, due to regional and economic development factors, the average sales prices of residential commercial housing in Beijing and Guangxi Zhuang Autonomous Region exhibited a substantial disparity around 2010. This gap widened further after 2014, and by now, the difference has more than doubled compared to 2012.

Table 3 Describe the Statistical Table

Average sales price of residential commercial housing(Yuan)	Beijing	Guangxi
average value	28483.3	5172.484
maximum	42684	6440
minimum	16553.48	3909.83
range	26130.52	2530.17

Just like Table 3, between 2012 and 2020, the average price of residential commercial housing in Beijing surged from 16,553.48 yuan to 42,684 yuan, an increase of 26,130.52 yuan, which is more than ten times the increase of 2,530 yuan in Guangxi Zhuang Autonomous Region and almost fourteen times the increase in Guizhou Province.

3 ANALYSIS OF THE CURRENT SITUATION OF THE AVERAGE SALES PRICE OF COMMERCIAL HOUSING

In recent years, the overall trend of housing prices in China has exhibited a pattern of being higher in the south than in the north, and higher in the east than in the west. The continuous and significant increase in residential property prices is particularly evident in the developed regions of eastern China, as well as in the central, western, and southern parts of northern China. Even within the same region, the disparity in housing prices is considerable due to uneven economic development. Additionally, policies and the impact of the recent pandemic have contributed to significant fluctuations in housing prices. This paper will conduct a detailed analysis of the average sales prices over the past few years from both horizontal and vertical perspectives.

3.1 Analysis of Regional Differences (Based on Cluster Analysis)

In order to analyze the development of residential commercial housing across provinces, cities, and autonomous regions in mainland China, relevant variables were selected. The initial analysis was conducted using 2020 data(Figure 1).

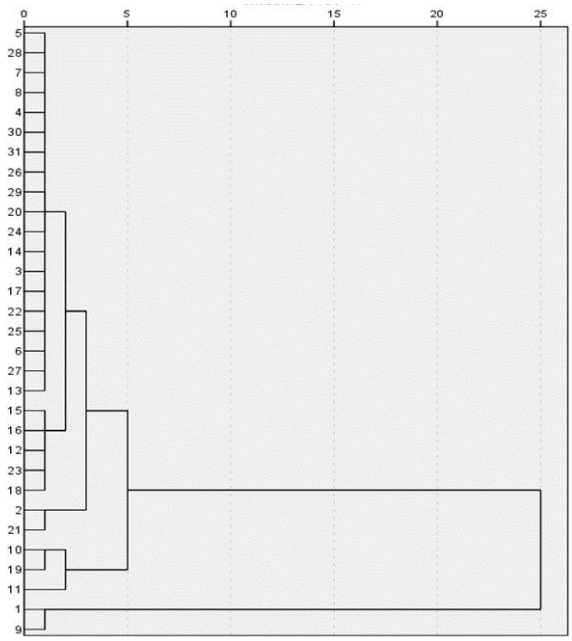

Figure 1 Phylogenetic Clustering Taxonomic Pedigree Diagram(2020)

These regions can be categorized into three groups based on housing conditions, as follows:
First Tier: Beijing, Shanghai.
Second Tier: Zhejiang, Guangdong, Jiangsu.
Third Tier: Tianjin, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Anhui, Fujian, Jiangxi, Hebei, Shandong, Henan, Hunan, Hubei, Guangxi, etc.
Among these, the two cities in the first tier exhibit the most substantial differences from other regions.
The following results of the analysis using the 2002 data(Figure 2):

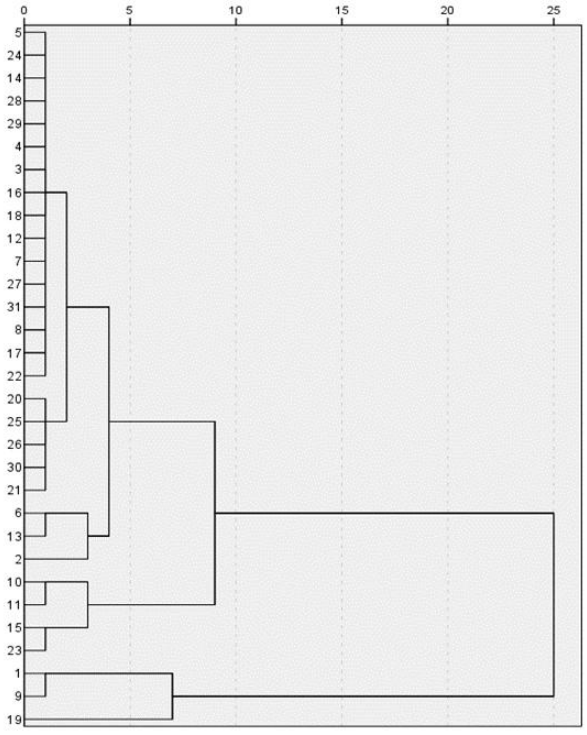

Figure 2 Phylogenetic Clustering Taxonomic Pedigree Diagram(2002)

These regions can be classified into four categories based on housing conditions, as follows:
First Tier: Guangdong Province.
Second Tier: Beijing Municipality, Shanghai Municipality.
Third Tier: Sichuan Province, Shandong Province, Zhejiang Province, Jiangsu Province.

Fourth Tier: Tianjin Municipality, Hebei Province, Shanxi Province, Inner Mongolia Autonomous Region, Liaoning Province, Jilin Province, Heilongjiang Province, Anhui Province, Fujian Province, Jiangxi Province, Henan Province, Hubei Province, Hunan Province, Guangxi Zhuang Autonomous Region, and others.

The housing market conditions were relatively more balanced in 2002. China's vast territory, large population, and uneven resource distribution generally result in high-quality resources being concentrated in major cities, while small and medium-sized cities lag behind. This development disparity has become increasingly pronounced in recent years. The rapid economic development in regions like Beijing and Shanghai has led to a corresponding swift advancement in the housing market, widening the gap from other regions.

4 NEURAL NETWORK MODEL PREDICTION

4.1 Model Building

From the previous discussion, it is evident that constructing an appropriate regression model for a practical problem is very challenging. Apart from building a regression model, one can also view the complex intermediate process of modeling as a black box and use neural network models to predict the data.

First, preliminary data processing is performed. This involves ensuring the relative completeness of the data information and checking for any missing values that may still exist. It is also hoped that the data information can be normalized.

Subsequently, the data is divided into training and testing sets, with 70% of the data allocated for training the neural network model and the remaining 30% reserved for testing it.

Generally, the number of hidden layers lies between the input and output layers, typically amounting to two-thirds of the input layer. Here, a feedforward network model with a single hidden layer containing two neurons is constructed, resulting in the model.

To evaluate the model's fitting performance, the root mean square error (RMSE) is used, which measures the closeness of the model's predictions to the actual labels by calculating the distance between predicted and actual values. The smaller the RMSE, the closer the predictions are to the true values.

First, the RMSE for the training set is calculated, yielding a result of 0.0427. Next, predictions are made on the testing dataset, resulting in the predicted values(Table 4).

Table 4 The Test Set Predicts the Outcome

region	Dependent variable
Tianjin	0.38258
Shanxi	0.07793
Inner Mongolia	0.14331
Heilongjiang	0.07033
Zhejiang	0.43241
Henan	-0.04120
Guangxi	0.00838
Hainan	0.12312
Guizhou	-0.01197
Tibet	0.04445
Xinjiang	0.02924

The correlation coefficient between the predicted and actual values, which indicates the strength of their linear relationship, is 0.832. A correlation coefficient close to 1 suggests a strong linear relationship, implying that the predictions are very close to the actual values and the predictive performance is satisfactory.

4.2 Model Modifications

Based on the original model, we will proceed to modify it by adding an additional hidden layer. We will construct a neural network model with the first and second hidden layers containing four and two hidden units, respectively. The specific model is illustrated below:

The root mean square error (RMSE) for the training set of this model is 0.0386, slightly lower than the previous model's 0.0427, but the difference is minimal. Similarly, the RMSE for the test set is 0.0782, also marginally lower than the previous model's 0.0826. Compared to the previous model, this model does not show significant improvement in predictive performance.

5 CONCLUSION

From the empirical analysis presented earlier, the following conclusions are drawn:

Regarding the development status of residential commercial housing, it is evident that the disparity in development between medium-sized and large cities is becoming increasingly apparent. In economically rapidly developing regions such as Beijing and Shanghai, the development of residential commercial housing is also accelerating, widening the gap with other areas.

There is a viewpoint that post-pandemic, the suspension of business operations and production, coupled with a decline in residents' expected income, will change the consumption patterns of some people, leading to reduced unnecessary expenditures. Consequently, residents' housing consumption and investment behavior will become more cautious and rational, causing a decline in the real estate market's sales level. However, there is also a perspective that once the pandemic stabilizes, people will engage in retaliatory consumption, and the suppression of consumption by the pandemic will be temporary.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Bai Y. Research on Influencing Factors of Average Sales Price of Residential Commercial Housing in Major Cities of China. IOP Publishing Ltd.IOP Publishing Ltd, 2020:012020.
[2] Na C, Chen-Xu T, Wei L, et al. Analysis and Forecast of the Residential Commercial Housing Price in Zhoukou. Mathematics in Practice and Theory, 2019.
[3] Wanchuan W. Prediction Model and Application of BP Neural Network. Journal of Yangzhou Polytechnic Institute, 2013.
[4] G Peter, Zhang, et al. Neural network forecasting for seasonal and trend time series. Operations Research, 2005.
[5] Marzban C, Stumpf G J. A Neural Network for Damaging Wind Prediction. Weather & Forecasting, 1998, 13(1): 151-163.
[6] Chen Y, Yang B, Dong J. Time-series prediction using a local linear wavelet neural network. Neurocomputing, 2006, 69(4/6): 449-465.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3009

FINANCIAL CREDIT RISK ASSESSMENT BASED ON MACHINE LEARNING

MingYue Gao
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 925483956@qq.com

Abstract: In the era of big data, the financial industry is facing new challenges and opportunities. Through big data and artificial intelligence technology, we can more accurately assess and manage various types of financial risks, including credit risk, market risk, fraud risk, etc. In this paper, the decision tree model is used to model and analyze the credit risk in financial risk, and the financial risk prevention and control system is established. For credit risk, according to the bank customer information data, after data processing, the decision tree classification method is used to judge whether the customer may default in the future through the customer's basic information, and finally the main discriminant basis is age, expected income, balance and number of credit cards. Then, from the five dimensions of establishing a sound risk assessment and early warning mechanism, improving citizens' financial risk awareness, promoting the construction of financial stability guarantee fund, strengthening industry supervision and self-discipline, and improving the legal and regulatory system, feasible suggestions are put forward for financial risks.
Keywords: Financial credit risk; Machine learning; Data mining; Decision tree model

1 INTRODUCTION

In the context of globalization and economic integration, financial risks have become an important factor affecting national economic security and social stability. Especially under the impetus of scientific and technological progress, the financial industry is undergoing unprecedented changes, which not only brings development opportunities, but also brings unprecedented challenges.

In recent years, China's financial industry has undergone tremendous changes driven by science and technology. With its unique advantages and broad application prospects, financial technology has become an important force to promote financial innovation and development. However, the development of financial technology has also brought a series of new financial risks. For example, the rise of Internet finance has made the boundaries of financial services more blurred, and the participation of non-traditional financial institutions has increased the complexity and uncertainty of the market. At the same time, the application of big data, artificial intelligence and other technologies in the financial field has also brought new risks such as data leakage and algorithm discrimination. In addition to technological factors, macroeconomic environment, policy changes, market volatility and other factors also have an important impact on financial risks. In the context of globalization, the unstable factors of the international financial market may be transmitted to the domestic financial market through channels such as trade and capital flows, posing a threat to China's financial stability. Therefore, we need to pay close attention to the changes in the economic and financial situation at home and abroad, and strengthen the construction of risk monitoring and early warning mechanisms.

In order to cope with these challenges, it is particularly important to promote the construction of financial stability guarantee fund. As a special financial system arrangement, the financial stability guarantee fund aims to enhance the ability of the financial system to cope with sudden risks by raising funds in advance and effectively dealing with risks afterwards. The sources of funds such as the payment of financial institutions, the return of disposal funding and the injection of financial funds provide a solid material basis for the operation of the financial stability guarantee fund. However, in the face of systemic and cross-industry financial risks, the existing guarantee fund may be difficult to respond effectively. Therefore, it is of practical significance and urgency to establish a financial stability guarantee fund. By improving the operation mechanism of the financial stability guarantee fund, it can better play its important role in preventing and defusing financial risks.

In order to monitor and respond to financial risks in a timely manner, the construction of early warning systems is particularly important. By integrating various types of risk information and establishing a scientific risk assessment model, the early warning system can issue an alarm before the risk breaks out, providing sufficient time and space for decision makers to respond. In the process of dealing with financial risks, enterprises need to constantly innovate. Through technological innovation and business transformation, we can improve our competitiveness and ability to resist risks. By improving the company's governance structure and strengthening internal control, we can reduce risks and improve the company's market reputation and brand value.

2 RESEARCH STATUS AND METHODS

2.1 Research Status

Credit score is the core evaluation index in customer default prediction. Fisher first proposed to provide a quantitative

risk judgment basis for lenders by comprehensively evaluating the borrower's credit history, financial status, repayment ability and other factors[1]. Through credit scoring, lenders can more accurately assess the default risk of borrowers, so as to formulate more reasonable loan policies and risk management strategies. In addition to credit scoring, a variety of classification prediction models are widely used abroad to predict customer defaults. These models include but are not limited to logistic regression decision trees, random forests, neural networks, etc.

Wiginton used Logistic regression model to explore the credit score in his study[2]. He emphasized that there is no need to impose special restrictions on the distribution of explanatory variables when making judgment analysis, which increases the flexibility and universality of the model in practical applications. Subsequently, Makowski introduced the decision tree model into personal credit evaluation for the first time[3]. This innovation has brought new perspectives and methods to the field of credit evaluation. The credit scoring model constructed by Coats and Fant is based on the neural network method, and widely collects the actual loan data of various countries for empirical research, thus verifying the validity and practicability of the model. These studies not only enrich the theoretical framework in the field of credit scoring, but also provide lenders with more diversified and more accurate assessment tools[4].

In recent years, with the progress of science and technology and the continuous innovation of data analysis methods, the research in the field of customer default prediction has also made significant progress. More and more studies have confirmed that machine learning algorithms show better performance than traditional methods in predicting customer default risk. Obare et al. fully proved this point. They deeply explored the application of machine learning in customer default prediction and achieved remarkable results. Machine learning algorithms can process a large amount of data and extract complex patterns and rules that are difficult to identify manually[5]. Chopra and Bhilare conducted in-depth research on bank loan data sets and found that the gradient boosting model showed significant advantages over the decision tree in prediction performance[6]. Ampountolas et al. conducted empirical research using actual data to compare the performance of various machine learning models in classification prediction[7]. The results show that the random forest model has certain advantages in classification prediction, which provides a new reference for loan institutions in selecting and applying risk prediction models.

The above research not only promotes the application of machine learning in the field of loan risk prediction, but also provides a more accurate and efficient tool for the risk management of financial institutions. By training and optimizing these algorithms, the default risk of customers can be predicted more accurately, so as to provide more reliable decision support for loan institutions. Compared with traditional credit scoring methods and classification prediction models, machine learning algorithms have stronger flexibility and adaptability. They can automatically adjust parameters to adapt to the characteristics of different data sets and maintain high prediction accuracy in the face of complex and changeable market environment and customer behavior.

In addition, with the continuous development of big data and artificial intelligence technology, the customer default prediction system is also constantly upgrading and improving. By using more abundant data sources and more advanced algorithm models, lenders can achieve a more comprehensive and in-depth analysis of the borrower's credit status, and further improve the accuracy and reliability of loan default prediction.

2.2 Decision Tree Algorithm

The decision tree is a tree structure in the form of a flow chart. It is used to calculate the probability that the expected value of the net present value is greater than or equal to zero by constructing a decision tree based on the known probability of occurrence of various situations, so as to evaluate the project risk and judge its feasibility. This decision branch is drawn into a graph like the branch of a tree, so it is named decision tree. In machine learning, decision tree is a prediction model, which represents a mapping relationship between object attributes and object values.

The decision tree contains three types of nodes: decision nodes (usually represented by rectangular boxes), opportunity nodes (usually represented by circles), and endpoints (usually represented by triangles). Each internal node represents a test on an attribute, each branch represents a test output, and each leaf node represents a category or prediction result.

The decision tree classification in this paper is based on the Gini value in the decision tree, that is, the Gini coefficient, which is an important indicator for evaluating the uniformity of data distribution or the purity of nodes. In the decision tree algorithm, the Gini coefficient is used for the decision-making process of feature selection and node division. The smaller the value is, the more uniform the data distribution is. The larger the value, the more uneven the data distribution.

The classification calculation formula of Gini coefficient is:

$$\mathrm{gini(T)} = \frac{S_1}{S_1+S_2}\mathrm{gini(T_1)} + \frac{S_2}{S_1+S_2}\mathrm{gini(T_1)} \tag{1}$$

Among them, S_1 and S_2 are the respective sample sizes of the two types after division.

In the process of building a decision tree, we first need to draw a decision tree according to the actual situation, which includes predicting various events that may occur in the future and expressing these situations in a tree diagram. Then, the expected value of each node is calculated, and the scheme is optimized by comparing the expected values of different schemes. Finally, pruning may be required to further simplify the decision tree and improve the prediction accuracy. Decision tree is divided into classification tree and regression tree. The classification tree is used to deal with discrete variables, while the regression tree is used to deal with continuous variables.

In machine learning and data mining, decision tree is a very common technology, which can be used for classification, regression, feature selection and other tasks. It has been widely used in the financial field because of its intuitive, easy

to implement and strong explanatory. The decision tree divides the data into different subsets by constructing a tree structure, so as to realize the classification and prediction of the data. In the risk management of customer default, the decision tree can help financial institutions identify the key factors affecting customer default, establish a customer default prediction model, and then provide decision support for risk management. In general, decision tree is an intuitive and powerful decision analysis method, which can help people make wise decisions in complex situations. By constructing decision trees, people can better understand various possible results and probabilities, so as to make more reasonable choices.

3 ESTABLISHMENT OF DECISION TREE MODEL

This study aims to use decision tree technology to judge and predict the default risk of bank customers, which has important theoretical and practical significance. From a theoretical point of view, this study can further improve and enrich the theoretical system of risk management. By applying decision tree technology to customer default risk management, we can deeply explore the inherent laws and characteristics of customer default risk, and reveal the key factors affecting customer default and its mechanism of action. This will help to promote the innovation and development of risk management theory and provide financial institutions with more scientific and effective risk management methods and tools. From a practical point of view, this study has important application value. First of all, by constructing a customer default prediction model, financial institutions can achieve accurate identification and evaluation of customer default risks, so as to formulate more reasonable credit policies and risk management strategies. Secondly, the interpretability of the decision tree model is strong, so that risk managers can intuitively understand the working principle and prediction results of the model, which is convenient for application and adjustment in practice. In addition, with the development of big data technology, financial institutions can obtain more abundant customer data, which provides a broader space and possibility for the application of decision tree model.

3.1 Data Source

In foreign countries, especially in Western countries, the long-standing credit culture and consumption concept have prompted people to be more inclined to achieve their life goals through borrowing. In this environment, borrowing is regarded as a normal economic behavior, which helps to improve the quality of life. In China, due to the different historical, cultural and economic development stages, people pay more attention to savings and stability, and hold a more cautious attitude towards borrowing. In addition, the policy and social environment have also affected people's consumption and borrowing habits to a certain extent. In recent years, China has also promoted the development of consumer finance to encourage reasonable consumer credit to meet people's growing consumer demand.
In this paper, the data are selected from 10000 customer data of banks in three regions abroad, and ten relevant information about customers are selected from them, namely: credit score, geographical location, gender, age, tenure, balance, number of products, whether to have a credit card, whether to be an active user, and estimated income.

3.2 Descriptive Statistics

Figure 1 shows that boys account for 45.43 % and girls account for 54.57 %, indicating that the gender ratio of the subjects surveyed is relatively balanced. At the same time, from the geographical point of view, France has a larger number, Germany and Spain have the same number. Secondly, from the perspective of age distribution, most people are between 20 and 80 years old, which meets the age limit requirements of bank customers. Finally, in terms of tenure, the number of people in office for 1 to 9 years is relatively uniform, indicating that most people work relatively stable.

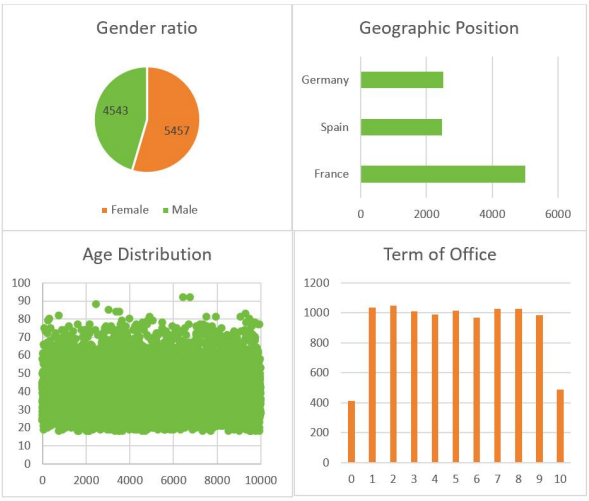

Figure 1 Descriptive Statistic

3.3 Data Preprocessing

Through observation, it is found that the data does not have missing values and outliers, so it does not need to be processed. However, there are two non-numeric characteristic variables. For the subsequent modeling analysis, the two characteristic variables of country and agenda are converted into values through the relevant code in python. At the same time, in order to facilitate the examination of whether the customer defaults, the credit score variable is divided into two categories according to the average value. The data below the average value is recorded as 0, representing default, and the data above the average value is recorded as 1, representing trustworthiness. In order to construct a decision tree model to predict bank customer defaults, the processed data is divided into a training set and a test set according to the ratio of 8:2, as shown in the Table 1.

Table 1 Dataset Classification

	Number of trustworthy people	Number of defaults	Grand total
Training sets	4109	3891	8000
Testing set	1027	973	2000
Grand total	5136	4864	10000

3.4 Basic Evaluation

ROC curve, also known as receiver operating characteristic curve or receiver operating characteristic curve, is to draw the relationship curve between True Positive Rate and False Positive Rate under different thresholds by changing the decision threshold of the two classifiers. On the ROC curve, the point closest to the upper left of the coordinate map is the critical value with high sensitivity and specificity. When the ROC curve of the model is closer to the critical value, the classification result of the method is more effective. The area under the curve is a quantitative index used to evaluate the overall performance of the diagnostic method. The value range is 0 to 1, and the larger the value is, the better the classification effect is.

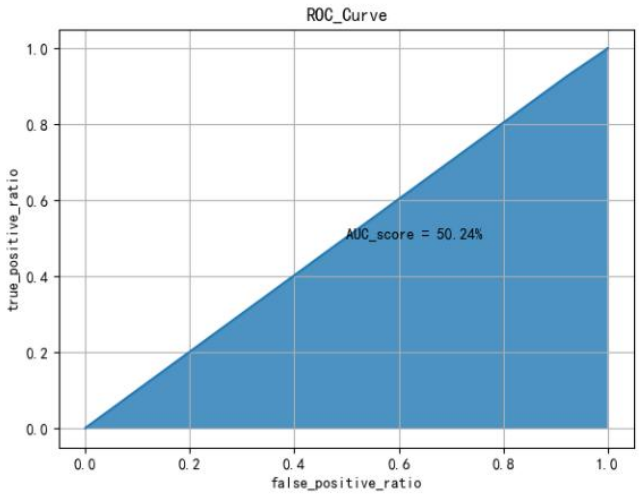

Figure 2 ROC Curve

As shown in the Figure 2, the AUC value is 50.24 %, indicating that the gap between the two samples of the data is obvious, and the classification result is better using the decision tree.

3.5 The Results of Customer Default Prediction Based on Decision Tree Model

First, the initial decision tree model will be trained using the training set of the divided bank customer credit score data. By inputting the training data into the model, the model will learn how to predict whether customers default based on their credit score characteristics. After the training, we get an initial decision tree model with predictive ability.

After obtaining the trained decision tree model, our next step is to import the test set of bank customer credit score data into this model. The role of the test set is to evaluate the predictive performance of the model, rather than participating in the training process of the model. We will use the model to predict the data in the test set to determine whether the customer defaults. In order to evaluate the prediction performance of the model, we need to calculate some evaluation indicators. These indicators will help us understand the accuracy, recall rate and other key information of the model in predicting whether customers default. By comparing these indicators, we can have a comprehensive understanding of the performance of the model.

Finally, in order to improve the prediction effect of the model on customer default, we need to optimize the parameters

of the decision tree model. This can be achieved by modifying the parameter values of the model, such as adjusting the maximum depth of the tree, the minimum number of samples required for leaf nodes, etc. In the process of adjusting the parameters, we evaluate the influence of different parameter settings on the performance of the model according to the change of the evaluation index, so as to find the optimal parameter combination. Through this process, a more accurate and reliable decision tree model can be obtained to predict the default risk of bank customers.
According to the above process, the final decision tree is obtained in the Figure 3.

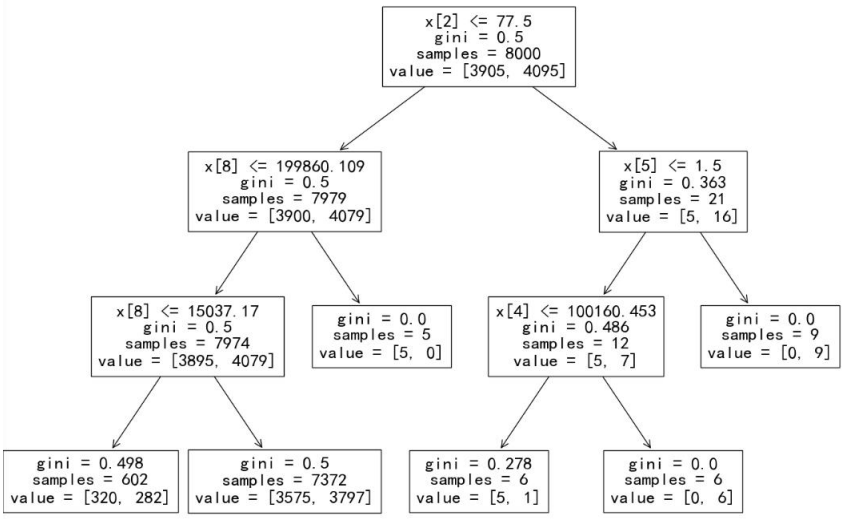

Figure 3 Decision Tree

According to the results, judging whether a bank customer defaults has nothing to do with geographical location and gender. The overall decision tree is first divided into two categories based on age. For those younger than 77.5 years old (left half branch), the estimated income is used as the classification point, and then divided downward. For those who are older than 77.5 years old (the right half branch), first take whether they have a credit card as the classification point, and then examine the balance status of the customers who have a credit card, and finally judge whether they default according to their term of office.

4 CONCLUSIONS AND RECOMMENDATIONS

4.1 Research Conclusion

Based on the bank customer data, the decision tree is established to determine whether the customer defaults. Finally, the customer will be judged to be in default in two cases. One is when the customer is younger than 77.5 years old and the expected income is lower than 15037.17, and the other is older than 77.5 years old. At the same time, the number of credit cards is small and the balance is less than 100160.453. This result is in line with expectations, and by analyzing the various attributes and historical data of customers, it can help banks assess customer credit risk, help banks understand customer repayment ability more accurately, and thus manage loan risk more effectively.
At the same time, banks can understand the differences between different customer groups, and adjust the customer service strategy accordingly, which can provide decision support for banks and help them formulate more intelligent loan approval strategies. For example, stricter risk control measures can be adopted for high-risk customers, while more favorable loan conditions can be provided for low-risk customers, thereby improving customer satisfaction and loyalty. Finally, banks can conduct customer default risk assessment based on decision trees, and banks can reduce the losses caused by loan defaults. Early detection of customers who may default, you can take appropriate measures, such as raising interest rates, reducing the amount of loans or requiring guarantees to reduce the risk.
In summary, the customer default prediction model based on decision tree is of great significance to banks, which can help banks better manage risks, optimize customer service, and improve the accuracy and efficiency of loan approval.

4.2 Suggestions

With the continuous development of the global economy, financial risks have increasingly become an important issue that we cannot ignore. Especially in the financial industry reform driven by scientific and technological progress, the complexity and diversity of financial risks are becoming more and more prominent. According to the conclusion of this paper and the current financial risk situation, the following suggestions are put forward:

4.2.1 Establish a sound risk assessment and early warning mechanism

Financial institutions should establish a sound risk assessment system to regularly assess and monitor various financial risks. Through big data analysis, machine learning and other technical means, a risk early warning model is constructed to detect and warn potential risks in time. At the same time, we should strengthen the comprehensive monitoring of

market, credit, liquidity and other risks to ensure that the risks are measurable, controllable and bearable.

4.2.2 Improve citizens' awareness of financial risks

Through media publicity and educational activities, citizens' awareness and awareness of financial risks should be improved. Enable citizens to rationally view financial market fluctuations and avoid blind investment and excessive borrowing. At the same time, we should strengthen the protection of financial consumers' rights and interests and maintain the fairness, transparency and stability of financial markets.

4.2.3 Promote the construction of financial stability guarantee fund

Drawing on international experience and combining with China's actual situation, we will accelerate the construction of financial stability guarantee funds. The source of funds should be diversified, including financial institutions to pay, financial capital injection and so on. Through the establishment of a financial stability guarantee fund, it can provide financial support for the disposal of systemic financial risks and reduce the risk exposure of financial institutions and the entire financial system.

4.2.4 Strengthen industry regulation and self-discipline

The regulatory authorities should strengthen the supervision of financial institutions to ensure that their business operations are compliant and robust. At the same time, promote financial institutions to strengthen self-discipline, establish and improve the internal risk control system, improve risk management and disposal capacity. In addition, information sharing and collaboration between industries should be strengthened to jointly cope with cross-industry financial risks.

4.2.5 Improve the legal system

In view of the new problems and challenges in the financial field, the relevant laws and regulations system should be improved in time. Through legislative means to clarify the power and responsibility relationship between financial institutions and regulatory authorities, and regulate the order of financial markets. At the same time, increase the punishment of illegal acts, to form an effective deterrent and restraint.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Fisher R A. The use of multiple measurements in taxonomic problems. Annals of eugenics, 1936, 7(2): 179-188.
[2] Wiginton J C. A note on the comparison of logit and discriminant models of consumer credit behavior. Journal of Financial and Quantitative Analysis, 1980, 15(3): 757-770.
[3] Makowski P. Credit scoring branches out. Credit World, 1985, 75(1): 30-37.
[4] Coats P K, Fant L F. Recognizing financial distress patterns using a neural network tool. Financial management, 1993: 142-155.
[5] Obare D, Murary M. Comparison of Accuracy of Support Vector Machine Model and Logistic Regression Model in Predicting Individual Loan Defaults. Am. J. Appl. Math. Stat, 2018, 6(6): 266-271.
[6] Chopra A, Bhilare P. Application of ensemble models in credit scoring models. Business Perspectives and Research, 2018, 6(2): 129-141.
[7] Ampountolas A, Nyarko Nde T, Date P, et al. A machine learning approach for micro-credit scoring. Risks, 2021, 9(3): 50.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3010

AN EMPIRICAL ANALYSIS OF THE INDUSTRIAL STRUCTURE AND EMPLOYMENT STRUCTURE IN JIANGSU PROVINCE

Ling Li
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 3561186176@qq.com

Abstract: Based on the development context of Jiangsu Province, this study uses data from 2002 to 2020 to examine the interaction between industrial structure and employment structure. It analyzes the output value and employment share of the three industries in Jiangsu, calculates their structural deviations, and explores the causal relationships between industrial and employment structures. Additionally, it identifies problems in their development and proposes optimization suggestions. The results indicate that there is no causal relationship between output value and employment in the primary industry. Increased labor productivity has reduced labor demand, making it difficult to reallocate surplus labor. In the secondary industry, output growth far exceeds employment growth, indicating no causal relationship, but there remains significant employment potential. The tertiary industry shows a unidirectional causal relationship where increased output value promotes employment. This industry has a strong capacity to absorb labor and significant development potential.
Keywords: Industrial structure; Employment structure; Grey relational analysis; Granger causality test

1 INTRODUCTION

Since the reform and opening-up policy, Jiangsu Province's economy has grown rapidly, achieving remarkable accomplishments. In 2013, its GDP was 217 times that of 1978, with an average growth rate of 17.14%. However, Jiangsu still faces issues related to both the quantity and quality of its labor force, and the inconsistency in the adjustment and upgrading of its industrial and employment structures has led to wasted labor resources and difficulties in overcoming traditional employment concepts. During the "13th Five-Year Plan" period, Jiangsu continued to adjust the structure of its three major industries. The proportion of the tertiary industry increased, while the proportions of the primary and secondary industries decreased, although the secondary industry still dominates. Compared to economically leading provinces like Guangdong and Shandong, Jiangsu's industrial structure remains biased towards the secondary industry. In comparison with Beijing and Shanghai, Jiangsu has higher proportions of the primary and secondary industries but lags in the development of the tertiary industry. In terms of employment, during the "13th Five-Year Plan" period, the employment proportions of the primary and secondary industries in Jiangsu decreased, while the tertiary industry's proportion continued to rise. By the end of 2020, they accounted for 13.8%, 39.7%, and 46.5% respectively. The growth momentum of traditional service industries weakened, and the growth rate of industrial added value slowed. Reasonable adjustments to the industrial and employment structures are crucial for future economic development and labor allocation.

Many scholars have studied the relationship between industrial structure and employment structure. Xiao Ma, Hongjuan Li, and Xiuli Sang analyzed the correlation between industrial structure and employment structure in Yunnan Province. Their research found that the primary industry can absorb a certain amount of labor but does not significantly increase its output value. Increased employment in the secondary industry notably promotes its output growth, while the output and employment in the tertiary industry reach equilibrium in the short term[1]. Wenqian Luo studied the dynamic relationship between industrial structure and employment structure, revealing that industrial structural adjustments have a lagged effect on employment structure[2]. The primary industry temporarily enhances its employment absorption capacity in the short term but diminishes in the long term, whereas the absorption capacity of the secondary and tertiary industries weakens in the short term but strengthens in the long term.

Other scholars have researched the factors influencing the coordination between industrial structure and employment structure. Yan He, Changyao Zhang, Zibao Sun used gray relational analysis to analyze industry and employment data, measuring the factors causing deviations between industrial structure and employment structure[3]. Their study found that urbanization levels can promote short-term adjustment of industrial structure in Tibet, while over the long term, industrial and employment structures tend to become more coordinated. Xingji Diao employed SVR and gray prediction models to explore factors influencing the coordination between industrial structure and employment structure in Hunan Province[4]. The results identified four major influencing factors: regional economic development level, quality of regional workforce, degree of government intervention, and level of technological development.

Some scholars have analyzed industrial and employment structures using methods such as structural deviation. Lijuan Wu and Ke Zhu examined the main difficulties faced by Guangxi in adjusting its industrial structure and proposed pathways for improvement[5]. Their study suggests that the primary industry should explore new advantages and promote the external transfer of surplus labor. The secondary and tertiary industries, which dominate economic development, should accelerate industrial upgrading. Qing Wen analyzed the coordination between industrial and employment structures using structural deviation and overall coordination degrees, concluding that both are at a low

level, indicating significant potential for improvement[6]. Xinhao Qin compared indicators such as structural deviation, correlation coefficients, and employment elasticity, finding weak correlation in the primary industry's coordination with employment, a mismatch between jobs and talent in the secondary industry, and significant potential for labor absorption in the tertiary industry[7]. Fenglin Liu studied the structural deviation and coordination between industrial and employment structures in He tian, finding that the primary industry has improved economically, rural surplus labor is shifting to the tertiary industry, while the secondary industry is stagnant, and the tertiary industry significantly drives economic growth and labor absorption[8]. Yadi Ren analyzed the coupling effect between industrial and employment structures in the Beijing-Tianjin-Hebei region using economic data and indicators like structural deviation[9]. Shu-xi Yan quantitatively examined the relationship between industrial and employment structures in Yulin, using correlation analysis and structural deviation for the primary and secondary industries, providing recommendations while noting the lack of analysis for the tertiary industry[10].

Based on the above literature, most existing studies mainly conduct structural deviation analysis of industrial and employment structures. This paper, however, conducts research from two aspects. Firstly, it refers to previous studies and selects the following indicators to analyze the relationship between industrial structure and employment structure: employment elasticity, structural deviation, and gray relational analysis. Secondly, it uses the Granger causality test to conduct a dynamic analysis of the relationship between industrial structure and employment structure and provides relevant recommendations based on the changes in their relationship.

2 ANALYSIS OF THE RELATIONSHIP BETWEEN INDUSTRIAL STRUCTURE AND EMPLOYMENT STRUCTURE IN JIANGSU PROVINCE

2.1 Analysis of Employment Elasticity

Employment elasticity refers to the ratio of employment growth rate to economic growth rate. It is generally used to describe the capacity of economic growth to absorb labor resources. When the elasticity coefficient is greater than zero, it indicates that the industry promotes employment, and the capacity of economic growth to drive employment is determined by the magnitude of the elasticity. The larger this value, the stronger the ability to absorb labor. When the elasticity coefficient is less than zero, it indicates that the industry has a crowding-out effect on employment. The larger the absolute value of the elasticity coefficient, the stronger the crowding-out effect. If economic growth is positive and employment growth is negative, this situation is called a crowding-out effect. If economic growth is negative and employment growth is positive, this situation is called an absorption effect. The formula for employment elasticity is as follows:

$$\text{Employment elasticity} = \frac{\text{Employment growth rate}}{\text{GDP growth rate}} \quad (1)$$

By substituting the data from 2002 to 2020 into the formula, it can be seen that the employment elasticity coefficients of the primary industry are all negative. This is due to the increase in labor productivity in Jiangsu Province, which has led to a decrease in demand for labor. The secondary industry had positive elasticity coefficients from 2003 to 2014, indicating that economic growth drove employment growth. This was caused by the initial low technology levels, which required more labor. During these nine years, economic growth had a driving effect on employment, with an average driving capacity of 0.17. However, from 2015 to 2020, the elasticity coefficients of the secondary industry were all negative. This was due to technological innovation and increased labor productivity, leading to a reduction in the required labor force. The tertiary industry had positive elasticity coefficients from 2003 to 2020, indicating that economic growth significantly drove employment growth. It is evident that since its inception, the tertiary industry has consistently had a driving effect on employment.

Overall, the employment elasticity of the primary industry has always been negative, indicating that the primary industry's ability to absorb labor is weak and it does not promote employment. The employment elasticity of the secondary industry was greater than zero before 2014, indicating that the development of the secondary industry could increase labor, with a strong ability to absorb labor. However, from 2015 to 2020, the employment elasticity of the secondary industry was less than zero, indicating a weak ability to absorb labor. The employment elasticity of the tertiary industry has always been greater than zero and relatively stable, indicating that the development of the tertiary industry can increase employment and has a strong ability to absorb labor.

2.2 Structural Deviation Degree Analysis

In this article, the structural deviation degree used is the difference between the ratio of the added value proportion to the employment proportion of each industry . Structural deviation degree is used to determine whether the industrial structure and employment structure are coordinated, which can influence the arrangement of the three industries and employment numbers in Jiangsu Province. The judgment criteria are as follows: if the structural deviation degree tends to zero, it indicates that the two are in a balanced state; if the structural deviation degree is greater than zero, it indicates that the employment proportion is smaller than the industry value-added proportion, suggesting that there is significant employment potential in this industry; if the structural deviation degree is less than zero, it indicates that the employment proportion is greater than the industry value-added proportion, implying that there is pressure to transfer labor out of this industry, meaning there is surplus labor. A structural deviation degree with an absolute value greater than zero indicates an imbalance between the two, and the larger the absolute value, the more likely structural

unemployment may occur in future development. Therefore, we will use the structural deviation degree to determine whether the industrial structure and employment structure of the three industries in Jiangsu Province are coordinated. The formula for structural deviation degree is as follows:

$$\text{Structural Deviation Degree} = \frac{\text{Proportion of Industry Value Added}}{\text{Proportion of Industry Employment}} - 1) \quad (2)$$

By processing the relevant data from Jiangsu Province from 2002 to 2020 and substituting it into the above formula, we can observe the following: The structural deviation degree of the primary industry was consistently less than zero from 2002 to 2020, hovering around -0.7. This indicates that the value-added proportion of this industry lags behind its employment proportion. In other words, over these 18 years, there has been pressure to transfer labor out of the primary industry, resulting in a significant surplus of labor. The structural deviation degree of the secondary industry has always been greater than zero. Between 2004 and 2014, the deviation hovered between 0.1 and 0.2. During this period, both the value added and the number of employed persons in the secondary industry increased, indicating that the secondary industry's capacity to absorb labor was moderate. From 2015 to 2020, the structural deviation tended towards 0.1, suggesting that there is still potential for employment growth in the secondary industry. For the tertiary industry, the structural deviation degree gradually approached zero between 2002 and 2016, indicating that economic growth in the tertiary industry had a strong positive impact on employment, achieving a coordinated development between industrial structure and employment structure. However, between 2016 and 2020, the structural deviation ranged from 0.1 to 0.2, suggesting that there is still considerable potential for employment growth in the tertiary industry.

Based on the analysis of the structural deviation degree results, it can be concluded that, compared to the secondary and tertiary industries, the primary industry is more likely to experience structural unemployment. The proportion of employment and the value-added proportion in the primary industry are both declining, but the proportion of employment is decreasing much faster than the value-added proportion. Meanwhile, the value added in the secondary industry has been consistently increasing, and although the value added in the primary industry is also increasing, it is doing so at a much slower rate. This indicates that labor productivity in the primary industry is relatively low. The imbalance in development suggests that there will be unemployed workers in the primary industry.

2.3 Grey Relational Analysis of Employment Structure and Industrial Structure

Grey relational analysis is a quantitative method used to describe and compare the development trends of a system. Grey relational degree serves as a measure of the degree of correlation between factors. In this study, data from Jiangsu Province on industrial structure and employment structure from 2002 to 2020 are selected for relational analysis. Let X_0 represent the sum of absolute deviations of various industries, X_1 represent employment in the primary industry, X_2 represent employment in the secondary industry, and X_3 represent employment in the tertiary industry. The modeling steps are as follows:

Step 1: Define the analysis sequences: Establish the reference sequence as X_0, where $X_0(t)$ represents the deviation degree in year t, with t denoting the year. Set the comparison sequences as X_1, X_2, and X_3, where $X_1(t)$ represents the employment in the primary industry in year t, $X_2(t)$ represents the employment in the secondary industry in year t, and $X_3(t)$ represents the employment in the tertiary industry in year t, with t representing the respective year.

Step 2: To begin with the data, it's necessary to normalize the variables, as the data in each column may vary in scale, making direct comparison difficult or impractical. In this study, the method employed is initial value normalization. This involves dividing each column of data by its first value to derive new normalized data. Specifically, the formula for initial value normalization of the reference column is as follows:

$$X_0 * (t) = \frac{X_0(t)}{X_0(1)} \quad (3)$$

The formula for initial value normalization of the comparison columns is:

$$Xi(t) = \frac{Xi(t)}{Xi(1)} \quad (4)$$

Here, where iii represents the primary, secondary, and tertiary industries (i = 1, 2, 3).

Step 3: Calculate the correlation coefficient. First, calculate the difference sequence, where the absolute difference between the mother sequence and the child sequence at different times is:

$$\Delta_i(t) = X_i(t) - X_0 * (t) \quad (5)$$

In which, when t takes a certain moment, Δmax and Δmin are respectively the maximum and minimum values, and the formula for calculating the correlation coefficient is:

$$\xi_i(t) = \frac{\min(\Delta_i(\min)) + \rho \max(\Delta_i(\max))}{| X_0(t) - X_i(t) | + \rho \max(\Delta_i(\max))} \quad (6)$$

Here, the resolution parameter ρ is a constant, with values ranging between 0 and 1. In this study, ρ is equal to 0.5

Step 4: Calculate the correlation degree.

Because the correlation coefficient is the comparison of the correlation degree values between the comparison sequence and the reference sequence at each moment, and it has more than one number, so the correlation values are averaged as the final correlation degree.

Step 5: Correlation degree ranking.

Using MATLAB to process Jiangsu Province's data on the three major industries from 2002 to 2020, the results are shown in the table below.

Table 1 A List of Grey Relational Degrees for the Three Industries

Item	The primary industry	The second industry	the third industry
Grey relational degree	0.8335	0.6511	0.6080

Based on Table 1, it can be concluded that the correlation between employment numbers and industrial structure is greatest for the primary industry, followed by the secondary industry, and then the tertiary industry. Further comparison reveals:

The employment correlation of the primary industry is the highest among the industrial structure adjustments, indicating that over the 18 years, the employment dynamics in Jiangsu Province's primary industry have played the most significant role in industrial upgrading. Its grey relational degree of 0.8335 indicates a high correlation. Analyzing the period from 2002 to 2020, the proportion of the primary industry's value added showed a declining trend overall, decreasing from 10.47% to 4.42%, a drop of 6.05 percentage points. Concurrently, during this period, the employment proportion in the industry also exhibited a declining trend, decreasing from 39% to 13.8%, a decline of 25.2 percentage points. This decrease is nearly four times that of the decline in the value added proportion, indicating that the primary industry, facing saturated employment, has been transferring employment to the secondary and tertiary industries.

The correlation coefficient between employment in the secondary industry and industrial structure is 0.6511, indicating a moderate correlation, suggesting that employment in the secondary industry plays a certain role in industrial upgrading. Based on the data above, we understand that Jiangsu Province's proportion of value added in the secondary industry has generally declined, decreasing by 9.78 percentage points. However, during the same period, the employment proportion in the industry increased by 7.2%. This indicates that despite absorbing some labor, the industrial structure has not changed significantly or has even declined, highlighting the need for adjustments in Jiangsu Province's secondary industry to enhance value added.

Although employment in the tertiary industry exhibits the weakest correlation among the three sectors, it still shows a relatively high correlation coefficient, indicating that employment in the tertiary industry also plays a positive role in industrial upgrading. The value added in the tertiary industry has consistently increased from 2002 to 2020, growing by 24.17%. Simultaneously, the employment proportion in this industry has also increased by 18%. This demonstrates that the tertiary industry, after absorbing labor, significantly enhances industrial value added, making a notable contribution to the Gross Domestic Product (GDP).

Based on the analysis of employment elasticity, structural deviation, and grey relational degree, the following conclusions can be drawn: The primary industry exhibits a displacement effect on employment, indicating surplus labor. Its employment situation significantly affects industrial upgrading, continuously shifting labor towards the secondary and tertiary industries as its employment numbers saturate. Employment in the secondary industry plays a moderate role in industrial upgrading by absorbing labor. However, despite absorbing labor, the industrial structure has not changed significantly and may even have declined. The tertiary industry's economic growth drives employment growth with strong labor absorption capabilities. After absorbing labor, it enhances industrial value added, making the largest impact on economic development.

3 EMPIRICAL ANALYSIS OF INDUSTRIAL STRUCTURE AND EMPLOYMENT STRUCTURE

3.1 Introduction of Variables and Models

The paper selects the value added of the three industries (denoted as G1, G2, G3) and the employment numbers in these industries (denoted as J1, J2, J3) in Jiangsu Province from 2002 to 2020 as variables. To mitigate the impact of heteroscedasticity, the data are logarithmically transformed, represented as LNG1, LNG2, LNG3, LNJ1, LNJ2, LNJ3.

The paper further analyzes the relationship between industrial structure and employment structure using the Granger causality test, and conducts tests for stationarity and cointegration of the data. The Granger test method is primarily used to analyze causal relationships between variables. In the context of time series, the Granger causality between two economic variables X and Y is defined as: if, with the inclusion of past information of both X and Y, the prediction of X based solely on the past information of Y is significantly better than predicting X without considering Y, then X is considered to be a Granger cause of Y. When conducting the Granger causality test, the first consideration is whether the time series are stationary, as non-stationary series may lead to spurious regression issues. Therefore, before performing the Granger causality test, it is essential to test the stationarity of each indicator's time series using the Augmented Dickey-Fuller (ADF) test.

The Granger causality test assumes that all predictive information about Y and X is contained within their respective time series. The test requires estimating the following regressions:

$$Y_t = \sum \alpha_i X_{t-i} + \Sigma \beta_j Y_{t-j} + u_{1t} \tag{7}$$

$$X_t = \sum \lambda_i X_{t-i} + \Sigma \delta_j Y_{t-j} + u_{2t} \tag{8}$$

Assuming that the white noises u1t and u2t are uncorrelated. Equation (7) posits that the current Y is related to itself and to the past values of X, and Equation (8) assumes a similar scenario.

For Equation (7), the null hypothesis is: $H0: \alpha_1=\alpha_2=\alpha_3=\ldots=\alpha_q=0$

For Equation (8), the null hypothesis is: $H0: \delta_1=\delta_2=\delta_3=\ldots=\delta_s=0$

Discuss in four scenarios:

The first scenario is that X causes changes in Y, indicating a unidirectional causal relationship from X to Y. If the coefficients of lagged X in Equation (7) are significantly different from zero and the coefficients of lagged Y in Equation (8) are not significant, then X is considered to cause changes in Y.

The second scenario is that Y causes changes in X, indicating a unidirectional causal relationship from Y to X. If the coefficients of lagged Y in Equation (8) are significantly different from zero and the coefficients of lagged X in Equation (7) are not significant, then Y is considered to cause changes in X.

The third scenario is that X and Y mutually cause each other's changes, indicating bidirectional causality. If the coefficients of lagged Y in Equation (8) and the coefficients of lagged X in Equation (7) are both significantly different from zero, then X and Y are considered to have bidirectional causality.

The fourth scenario is that X and Y are independent or there is no causal relationship between them. If the coefficients of lagged Y in Equation (8) and the coefficients of lagged X in Equation (7) are both not significant, then X and Y are considered to have no causal relationship between them.

3.2 Data Stationarity Test

Based on the data of Jiangsu Province's primary, secondary, and tertiary industry output and employment from 2002 to 2020, the correlation coefficients between the structure of the three industries and employment structure were computed using EViews. The results are shown in Table 2.

Table 2 The Correlation Coefficients between the Output and Employment Numbers of the Industries

Item	LNJ1	LNJ2	LNJ3
LNG1	-0.98	0.87	0.97
LNG2	-0.99	0.91	0.98
LNG3	-0.99	0.86	0.98

From the table, it can be seen that Jiangsu Province's industrial output is negatively correlated with the primary industry and has some correlation with the secondary and tertiary industries. While these variables indeed exhibit correlation, it does not imply consistency or causality among them. Therefore, Granger causality tests are necessary to determine causality. Next, this paper conducts tests for stationarity, cointegration, and causality on the sequences to further elucidate the relationship between Jiangsu Province's industrial structure and employment structure.

Since non-stationary sequences can result in spurious regression, this study uses the Augmented Dickey-Fuller (ADF) test to assess stationarity. If the sequences are non-stationary, they will be differenced to achieve stationarity before conducting Granger causality tests. The ADF test results obtained using EViews are shown in Table 3.

Table 3 The Correlation Coefficient between Industrial Output and Employment Numbers

Item	LNG1	LNG2	LNG3	LNJ1	LNJ2	LNJ3
Prob.	0.0823	0.000	0.0068	0.0242	0.5118	0.7552

The unit root tests indicate that the time series LNG1, LNJ2, and LNJ3 are non-stationary at the 5% critical value level. Therefore, next steps involve differencing these series. First-order (or second-order) differencing is applied to achieve stationarity. The results of the differencing are shown in Tables 4 and 5.

Table 4 The first-Order Difference of Industrial Output and Employment Numbers

Item	D(LNG1)	D(LNG2)	D(LNG3)	D(LNJ1)	D(LNJ2)	D(LNJ3)
Prob.	0.2778	0.454	0.0161	0.0390	0.9959	0.5453

Table 5 The Second-Order Difference of Industrial Output and Employment Numbers

Item	D(LNG1,2)	D(LNG2,2)	D(LNG3,2)	D(LNJ1,2)	D(LNJ2,2)	D(LNJ3,2)
Prob.	0.0012	0.0045	0.0020	0.0053	0.0079	0.0121

Therefore, based on the first-order differencing results in Table 4, LNJ1 is stationary at the 5% critical value level. Additionally, from the second-order differencing results in Table 5, LNG1 and LNJ3 are stationary at the 5% critical

value level. Thus, LNJ1 is first-order stationary, while LNG1 and LNJ3 are second-order stationary.

3.3 Analysis of the Relationship Between Industrial Structure and Employment Structure

3.3.1 Cointegration analysis of the development of the three industries

Since in the previous section we have already analyzed the structural deviation between industrial structure and employment structure, the next step in exploring the relationship between the two will focus on examining the coordination of the development of Jiangsu Province's three industries. From the stationary tests mentioned earlier, it was found that LNG1 is a first-order differenced stationary series, while LNG1 and LNJ3 are second-order differenced stationary series. Next, a Granger causality test will be conducted on the three industries to examine whether there is consistency in the development of industrial structures and whether the development of the three industries is coordinated. As mentioned earlier, correlation coefficients alone cannot fully reflect their relationships. Therefore, the next step involves verifying the development relationship between the three industries through cointegration tests and Granger causality tests to examine the causality relationship.

According to the cointegration test results, it can be concluded that there are three cointegration relationships among the time series at a 5% significance level. This indicates that the development of the three industries in Jiangsu Province over the past 18 years exhibits a certain degree of balance. Next, a Granger causality test will be conducted on the three industries. The results are presented in Table 6.

Table 6 Granger Causality Test Results for the Three Industries

Null hypothesis	Sample size	F-value	P-value
LNG2 is not the Granger cause of LNG1	18	9.50	0.00
LNG1is not the Granger cause of LNG2	18	2.06	0.17
LNG3 is not the Granger cause of LNG1	18	1.77	0.20
LNG1 is not the Granger cause of LNG3	18	0.40	0.53
LNG3 is not the Granger cause of LNG2	18	0.97	0.34
LNG2 is not the Granger cause of LNG3	18	14.71	0.00

Based on the analysis from Table 5-6, under a significance level of 10%, it can be concluded that the second industry not being accepted as a Granger cause of the first industry implies that the second industry can promote the development of the first industry and has a stimulating effect on it.It can also be observed that the first industry cannot promote the development of the second industry, and there is a 17% probability of acceptance. Therefore, there is a unidirectional causality test between the first industry and the second industry, indicating that the first industry does not drive the development of the second industry, while the second industry does have a driving effect on the first industry. The third industry does not reject the null hypothesis at a 10% significance level, suggesting that the development of the third industry does not promote the first industry, with a 20% probability of acceptance. Similarly, from the Granger causality test, it is also found that the development of the first industry does not promote the third industry, with a 53% probability. Therefore, there is no causal relationship between the first industry and the third industry. The third industry is not the Granger cause of the second industry, indicating that the third industry does not drive the development of the second industry. At the same time, the second industry can promote the development of the third industry, showing a unidirectional causal relationship between the second industry and the third industry.

3.3.2 Granger causality test

Next, we will continue to conduct Granger causality tests on the three industries and employment. The results are as shown in the table 7 below.

Table 7 Granger Causality Test between the Three Industries and Employment

Null hypothesis	Sample size	F-value	P-value
LNJ1 is not the Granger cause of LNG1	18	1.79	0.20
LNG1is not the Granger cause of LNJ1	18	0.96	0.34
LNJ2 is not the Granger cause of LNG2	18	2.14	0.16
LNG2 is not the Granger cause of LNJ2	18	50.29	0.24
LNJ3 is not the Granger cause of LNG3	18	0.06	0.27
LNG3 is not the Granger cause of LNJ3	18	0.03	0.02

According to the table, at a significance level of 5%, it is not rejected that the employment in the primary industry is not

the Granger cause of its output increase, indicating that its employment is not the reason for the growth in output. Furthermore, the increase in output of the primary industry from 111.044 billion yuan to 453.672 billion yuan over 18 years occurred alongside a decline in employment numbers. This suggests that during this period, Jiangsu Province invested in automation technology in the primary industry and overall workforce skills improved. Simultaneously, the labor force in the primary industry shifted towards other industries.

The Granger causality test accepts that the increase in output in the secondary industry is not caused by its employment, with a 16% acceptance probability. Additionally, the test accepts that the increase in output does not cause growth in employment. Therefore, there is no causal relationship between employment and output growth in the secondary industry. From the above data, we can also analyze that over these 18 years, the output in 2020 was 7.89 times that of 2002, while employment in 2020 was only 1.33 times that of 2002. This indicates that the increase in output in the secondary industry did not significantly increase its employment numbers. Similarly, the increase in employment did not have a substantial impact on the development of its output. This corroborates the findings of the Granger causality test.

The hypothesis that the increase in employment in the tertiary industry is not caused by its output growth is accepted with a 27% probability. However, it is accepted that the increase in output in the tertiary industry affects the growth in its employment. This situation is related to the development of the tertiary industry in Jiangsu Province. According to the development situation in Jiangsu Province, the movement of personnel contributes to the development of the tertiary industry, leading to an increase in employment numbers. The rapid development of the industry has also resulted in an increase in employment numbers. From the data, it can be seen that employment in the tertiary industry has increased by more than 10 million people. Among the three major industries, the tertiary industry has developed the fastest, with the most prominent changes in employment numbers.

In summary, there is no causal relationship between output and employment numbers in the primary and secondary industries. The increase in output in the primary industry coincides with a decrease in employment, indicating an improvement in labor productivity rather than relying on increased employment for economic development. In contrast, both output and employment in the secondary industry are increasing, but the growth in output far exceeds the increase in employment, suggesting that the secondary industry is not solely dependent on labor to drive economic growth.

On the other hand, there is a unidirectional causal relationship between output and employment in the tertiary industry. An increase in output in the tertiary industry promotes an increase in employment numbers, highlighting its strong ability to absorb labor and significant potential for development.

4 ADVICE

This article analyzes the data of the three major industries' output and employment in Jiangsu Province from 2002 to 2020. Using methods such as structural deviation and Granger causality test models, it investigates the relationship between industrial structure and employment structure. The specific conclusion is:

The output of the primary industry shows no causal relationship with employment. With slow growth in output, surplus labor emerges. Concurrently, workers from the primary industry transition to the secondary and tertiary sectors, though their capacity in these sectors is limited. Therefore, recommendations for the primary industry include optimizing it, enhancing specialization in agricultural production, improving labor quality, and guiding surplus labor towards the secondary and tertiary sectors.

There is no causal relationship between output and employment in the secondary industry. Employment lags significantly behind output growth, indicating a need for internal restructuring despite labor absorption. Recommendations for the secondary industry thus focus on stabilizing its development, elevating skill levels, optimizing industry structure, and phasing out outdated capacities.

The tertiary industry exhibits a unidirectional causal relationship between output and employment. Increased output promotes employment growth, effectively absorbing skilled labor and making substantial contributions to economic development. Consequently, suggestions for the tertiary industry involve intensifying its development efforts, expanding employment opportunities, nurturing technical expertise, particularly in modern service sectors.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ma Xiao, Li Hongjuan, Sang Xiuli. Analysis and Forecast of Coordinated Development between Industrial Structure and Employment Structure in Yunnan Province. Reform and Opening-up, 2020(15): 76-85.

[2] Luo Wenqian. Study on the Dynamic Relationship between Industrial Structure Adjustment and Employment Structure in Yunnan Province. China Collective Economy, 2022(06): 27-30.

[3] He Yan, Zhang Changyao, Sun Zibao. Research on the Coordinated Development of Industrial Structure and Employment Structure in Tibet. Journal of Tibet University for Nationalities (Philosophy and Social Sciences Edition), 2021, 42(02): 112-117.

[4] Diao Xingwen. Evaluation of the Coordination between Industrial Structure and Employment Structure and Development Forecast: An Analysis of Provincial Data Based on the SVR Model. Journal of Social Sciences of Jiamusi University, 2021, 39(03): 65-68.
[5] Wu Lijun, Zhu Ke. Research on the Coordination of Industrial Structure and Employment Structure in Guangxi. Marketing Circle, 2021(26): 87-89.
[6] Wen Qing. Research on the Coordination of Industrial Structure and Employment Structure in Urumqi. Cooperative Economy and Science, 2021(14): 7-10.
[7] Qin Xinhao, Chen Mengdi, Xu Nazii, Bian Fei, Wang Qian, Mao Yanxin. Analysis of the Coordination between Forestry Industry Structure and Employment Structure in China. Chinese Forestry Economics, 2022(01): 60-65.
[8] Liu Fenglin, Yin Cong, Lu Xusheng. Research on the Coordinated Development of Industrial Structure and Employment Structure in Southern Xinjiang. Cooperative Economy and Science, 2022(04): 52-53.
[9] Ren Yadi, Tang Enbin. Research on the Relationship between Industrial Structure and Employment Structure in Beijing-Tianjin-Hebei Region. IOP Conference Series: Earth and Environmental Science, 2019, 330(2).
[10] Yan Shu-xi. A Quantitative Study on the Relationship between Industrial Structure and Employment Structure in Yulin. 3rd International Conference on Social Science and Technology Education (ICSSTE 2017), 2017.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3011

ANALYZING REGIONAL ECONOMIC INFLUENCING FACTORS BASED ON DIFFERENT CONTRACTION METHODS

CaiYun Peng
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email:pcyjying@163.com

Abstract: Currently, the biggest issue facing social development in China is the imbalance and lack of coordination in regional economic development, with multiple factors influencing regional economies. In order to scientifically measure these influencing factors, we established a multiple linear regression model based on data from 31 provinces nationwide in 2020. We fitted the model using four different solution methods – traditional OLS estimation, ridge regression, Lasso, and elastic net estimation. By comparing the fitting effects of different models, we further analyzed the merits and demerits of various contraction methods, and found that elastic net exhibits excellent performance in terms of prediction accuracy and model simplicity. Lastly, based on the results from elastic net estimation, we conducted an analysis of the factors affecting regional economic development.
Keywords: Regional Economic Development; OLS; Ridge Regression; Lasso; Elastic Net

1 INTRODUCTION

In recent years, China has synergistically leveraged technological innovation and market value realization, providing robust support for economic development. Studying the relationship between technology and regional economic growth holds significant practical importance[1]. Factors influencing regional economic disparities in China, as identified by Wang Anqi[2] and Cao Haibo[3], include geographic location, capital factors, technological progress, degree of economic openness, and policy institutions. Zhu Xuexin et al. utilized a broad Cobb-Douglas production function to examine the contributing factors to national economic growth and the contribution rates of different indicators of technological progress[4]. Wang Wei et al. conducted empirical research on the impact of technological innovation capability on regional economic development using a comprehensive evaluation method[5]. Wang Xu et al. focused on the quantitative factors of the input-output process of technological innovation in their research[6]. Song Meizhe and Li Mengsu found spatial gradient distribution patterns such as "low in the west, high in the east" and "low inland, high coastal" in their study of the coupling coordination of higher education, technological innovation, and economic development[7]. Zhang Zhiruo et al. studied the spatiotemporal coupling characteristics of technology finance and regional economic development in different regions[8]. Hong Mingyong employed econometric analysis to investigate the relationship between technological innovation and regional economic growth nationwide[9]. Foreign scholar Lukas Hogenschurz researched the impact of science, technology, and innovation on economic and social benefits, emphasizing the importance of converting knowledge and innovation outcomes into societal and economic entities[10].

In above, building upon existing research, this study selects ten relevant factors influencing regional economic development—namely, education, technology, policy institutions, trade openness, urbanization rate, labor input, geographical location, environment, among others—to construct a multiple linear regression model. Various methods for solving multiple linear regression models are employed to select the optimal model. Subsequently, a detailed analysis of each factor is conducted, and corresponding recommendations are proposed to relevant departments.

Common estimation methods for multiple regression models after regularization include ridge regression and Lasso. Both of them reduce the risk of overfitting by adding a penalty term to the loss function and reduce the information redundancy caused by multicollinearity between independent variables[11]. Elastic net regression is a combination of ridge regression and Lasso regression, which can solve some limitations of ridge regression and Lasso regression[12].

Using traditional Ordinary Least Squares (OLS) estimation initially to fit, test, correct, and evaluate the research object, followed by separate applications of Ridge Regression, Lasso, and Elastic Net, is advantageous for finding the best model and assessing the research object. Improving methods for solving multiple linear regression models like OLS, Ridge Regression, and Lasso can effectively enhance model prediction accuracy, prediction efficiency, and model simplicity[13]. Scaling the original Elastic Net further allows achieving optimal results. For each fixed parameter, the Elastic Net problem can be efficiently solved using algorithms designed for solving the Lasso problem. Compared to solving the Lasso problem under equivalent circumstances, Elastic Net also reduces computational speed [12].

2 PREPARATORY KNOWLEDGE

Before establishing the model, it is necessary to introduce relevant data and models.

2.1 Selection of Indicators

Based on references, most scholars agree that factors such as human capital, education, resource and environmental quality, geographic location, technological progress, degree of economic openness, and policy institutions significantly influence regional economic development. Different scholars select various indicators to measure regional economic development levels across different aspects. Sturm Thomas argues that no single rational theory can provide a satisfactory explanation for every aspect [10]. Therefore, this paper synthesizes several references and selects 10 factors from the following aspects to analyze the factors influencing regional economic disparities. The selected indicators are denoted as $x_1, x_2, \cdots, x_{10}$ respectively, and are summarized in tabular form as follows(Table 1).

Table 1 Summary of Indicators and their Symbol Explanations

Primary indicator	Secondary indicator	Symbol
Economic development level	Gross Regional Product (GRP)	y
Education	Local fiscal expenditure on education	x_1
technology	Number of domestic patent applications accepted	x_2
policy systems	Local government general budget expenditure	x_3
	Proportion of fiscal revenue to GDP	x_4
trade openness	Total import and export volume divided by GDP	x_5
urbanization level	Urbanization rate	x_6
labor input	Urban employed population	x_7
geographical location	Longitude	x_8
	Latitude	x_9
environment	Total wastewater discharge (ten thousand tons)	x_{10}

The data for this study is sourced from the "China Statistical Yearbook 2021" [14].

2.2 Multiple Regression Model

The factors that typically influence the dependent variable are more than one, so it is generally necessary to establish a multiple regression model. Its general model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i, \tag{1}$$

Where Y is the dependent variable, $X_1, X_2, \cdots, X_k$ is the independent variable, $\beta_0, \beta_2, \cdots, \beta_k$ represent the parameters of the independent variables, and ε_i is the error term.

In a typical linear regression model, p predictor variables $x_1, x_2, \cdots, x_p$ are first specified, and the response Y is generated by model(1). The ultimate goal of linear regression is to minimize the gap between predicted values of the model and actual values. The loss function for a typical multiple linear regression model is:

$$L(\beta) = \|y - X\beta\|^2 \tag{2}$$

The loss function is typically solved using OLS (Ordinary Least Squares) for parameter estimation. This involves finding values that minimize the sum of squares of differences between actual values and model-predicted values, serving as estimates. The objective function of the loss function at this point is:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2 \tag{3}$$

However, OLS often performs poorly in prediction and explanation. Therefore, many scholars have proposed adding penalties to the loss function based on OLS.

2.3 Contraction Methods

2.3.1 Ridge regression

Ridge regression minimizes the sum of squared residuals while constrained by the L2 norm of the coefficients, which reduces the residual sum of squares and prevents coefficients from becoming too large. As a form of continuous shrinkage method, ridge regression achieves improved predictive performance through a bias-variance trade-off. However, ridge regression does not produce a parsimonious model because it always retains all predictor variables in the model.

The loss function for ridge regression estimation, which includes an L2 norm penalty, is:

$$\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{4}$$

2.3.2 Lasso

Lasso is another estimation method that adds a penalty to least squares estimation by applying an L1 penalty to regression coefficients. Lasso performs both continuous shrinkage and automatic variable selection, enabling model simplification. The loss function of Lasso is as follows:

$$\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{5}$$

Ridge regression and Lasso have certain limitations when addressing certain problems. For example:

1) In the case where p>>n, due to the nature of convex optimization problems, Lasso can select at most n variables before saturating, hence it cannot handle p>>n scenarios.
2) If there is a group of variables with very high pairwise correlations, Lasso tends to arbitrarily select one variable from this group without specifically considering which one to choose.
3) In the typical scenario where n>p, if there are highly correlated predictor variables, empirical observations suggest that Ridge regression generally outperforms Lasso in terms of predictive performance.

2.3.3 Original Elastic Net

To address these issues, this paper considers a new regularization technique—Elastic Net. Similar to Lasso, Elastic Net can select groups of correlated variables. Simulation studies and practical data examples indicate that Elastic Net often outperforms Lasso in terms of prediction accuracy. Elastic Net combines the features of Lasso regression and Ridge regression by adding two penalties to the loss function of Ordinary Least Squares (OLS) estimation, namely:

$$L(\lambda_1, \lambda_2, \beta) = \|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 |\beta|_1, \tag{6}$$

The paper transforms the Elastic Net problem, which involves two penalties, into a Lasso-type problem with a single penalty by a simple transformation. This transformation extends the sample size from n to n+p, overcoming the first limitation of Lasso. The transformed penalty function of Elastic Net after this transformation is:

$$L(\gamma, \beta^*) = \|y^* - X^*\beta^*\|^2 + \gamma |\beta^*|_1, \tag{7}$$

The Elastic Net estimation also overcomes some limitations of both Lasso regression and Ridge regression. However, for each fixed λ, it appears to induce twice the amount of shrinkage, which may not effectively reduce substantial variance and introduces unnecessary additional bias. Therefore, the paper aims to improve the predictive performance of Elastic Net by correcting this double shrinkage issue.

2.3.4 The Elastic Net

After scaling the coefficients of the original Elastic Net estimation through a proportional transformation, the corrected Elastic Net estimation maintains the variable selection properties of the original Elastic Net. This scaling is the simplest method to eliminate excessive shrinkage. The scaled Elastic Net estimator is given by:

$$\hat{\beta}(elastic \quad net) = (1+\lambda_2)\hat{\beta}(naive \quad elastic \quad net), \tag{8}$$

After scaling by $1+\lambda_2$, Elastic Net automatically achieves optimal balance between maximum and minimum and resolves the issue of excessive shrinkage tendencies in the original Elastic Net in regression problems.

3 SIMULATION EXAMPLE OF REGIONAL ECONOMIC INFLUENCING FACTORS ANALYSIS IN CHINA

Due to the inconsistent units among the research data, we standardized the data to remove the dimensional effects, and then established the following multiple linear regression model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{10} x_{10i}, \tag{9}$$

3.1 Traditional OLS Regression

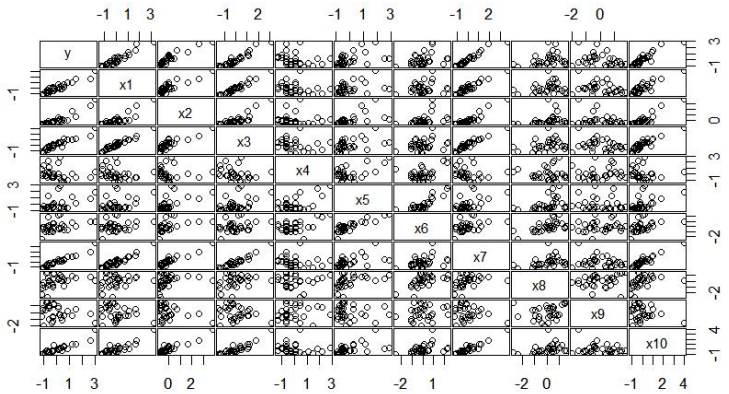

Figure 1 Scatter Plot of Variables in Regional Economics

Examining the correlations between variables reveals(Figure 1) that the dependent variable shows a clear linear relationship with some independent variables, while the linear relationship with other independent variables is not clear. There are also significant correlations among some independent variables. Therefore, the model we established may suffer from multicollinearity. To further assess whether the model exhibits multicollinearity, we calculated the variance inflation factors (VIFs) for each independent variable(Table 2).

Table 2 VIF Values for Each Independent Variable

Coefficients	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
VIF	45.127	13	38.524	2.837	9.385	6.606	48.195	2.051	1.5	14

From Table 2, it is evident that the VIF values for most variables are well above 10, indicating the presence of multicollinearity in the model. To assess the fitting and prediction effectiveness of the model, this study divides the observations into two parts, with 70% used as the training set and 30% as the test set.

3.1.1 Model estimation

Firstly, the fit using all variables in the ordinary least squares estimation was examined. The results indicate that only one independent variable is significant, with a model fit of 98.27%. However, the model shows signs of overfitting. It is necessary to address multicollinearity issues and revise the model accordingly.

3.1.2 Addressing multicollinearity

This study employs the most commonly used stepwise regression method to address the model. The results of stepwise regression analysis indicate that when using $x_2, x_3, x_4, x_6, x_{10}$ as coefficients in the regression equation, the minimum AIC value is -65.86. The remaining variables are examined for their impact on the model's fit. The fitting results indicate that all variables except x_{10} are significant. Therefore, variable x_{10} is excluded to optimize the model, and the model's fit is re-evaluated(Table 3).

Table 3 Refining the OLS Estimation Results

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.06210	0.04537	-1.369	0.191247
x_2	0.36036	0.08442	4.269	0.000673***
x_3	0.56689	0.08084	7.012	4.19e-06 ***
x_4	-0.32878	0.06378	-5.155	0.000118***
x_6	0.29429	0.07632	3.856	0.001555**

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

At this point, the t-tests for the parameters of each selected independent variable in the model have all passed, indicating their significance. The model's F-test has also been successful, confirming its overall significance. The standard error of residuals is 0.1768, with an R-squared of 97.86%, indicating a good fit of the model.

3.1.3 Model validation

Next, the model is tested. Using the Breusch-Pagan test to check for heteroscedasticity, the calculation shows that the model does not exhibit heteroscedasticity. Further, plotting the residual plot and regression scatter plot(Figure 2) both indicate that the model fits well.

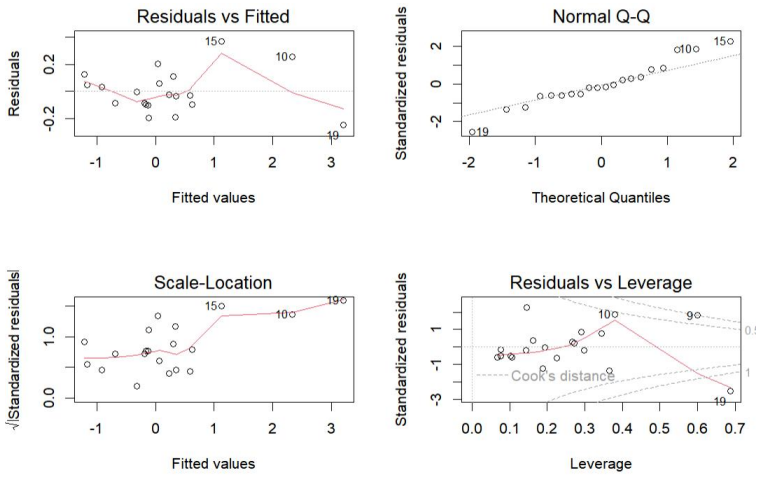

Figure 2 Residuals and Regression Plot of OLS Estimation

At this point, the root mean square error (RMSE) of the prediction sample is 0.377, with a coefficient of determination R-squared of 88.87%. The prediction errors for the sample are relatively large, but the overall fit is good. Therefore, under traditional OLS estimation, the regression equation obtained using standard methods is:

$$y = -0.062 + 0.36x_2 + 0.567x_3 - 0.329x_4 + 0.29x_6 \quad (10)$$

Among these, variables representing technological capability, policy institutions, and urbanization level remain in the model. The coefficient of the variable representing the ratio of fiscal revenue to GDP is negative, indicating an inverse relationship with the dependent variable, contrary to reality. The omitted education expenditure forms the basis for

cultivating highly skilled technology personnel, and the labor force is a crucial factor influencing economic development. These variables are indispensable factors in measuring regional economic development. Therefore, despite the good model fit and prediction error of the multiple linear regression model based on traditional OLS estimation methods, its effectiveness and reliability are low for sparse models with small sample sizes.

3.2 Ridge Regression

Ridge regression is a biased estimation regression method specifically used for analyzing collinear data. Its essence lies in shrinking the coefficients of variables with less importance to the dependent variable towards zero, rather than completely excluding them from the model. This trade-off sacrifices some information and precision to obtain regression coefficients that are more realistic and reliable, thus improving upon least squares estimation.

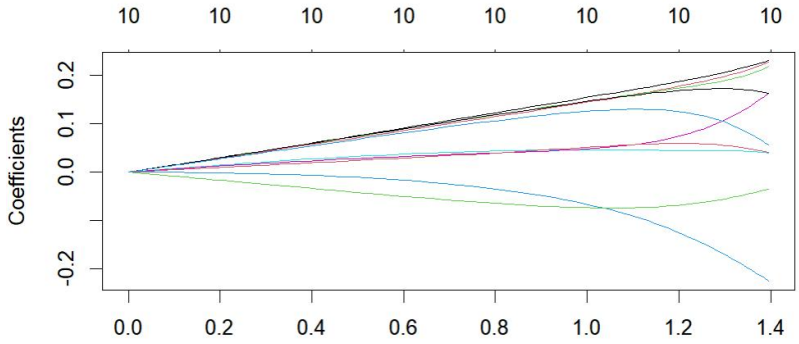

Figure 3 Ridge Trace Plot

From the ridge trace plot(Figure 3), it is observed that the magnitude of curve changes increases with the change in λ, and the curve tends to become unstable. It is challenging to directly determine the appropriate ridge parameter λ value from this plot. Therefore, we opt for ten-fold cross-validation to select the optimal ridge parameter value(Figure 4).

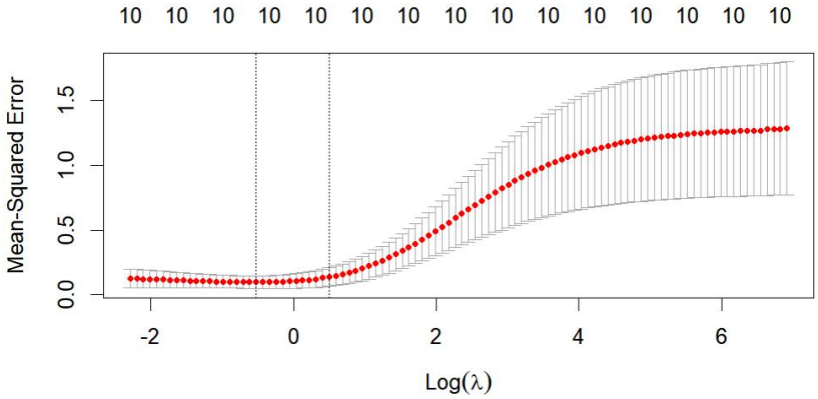

Figure 4 Ridge Regression Ten-fold Cross-Validation Plot

The vertical lines (in Figure 4) correspond to two λ values of 0.5924 and 1.6485. Under the condition of choosing the minimum λ value, the coefficients of the model variables are as follows(Table 4):

Table 4 Ridge Regression Estimation Results

Coefficients	Intercept	x_1	x_2	x_3	x_4	x_5
Estimate	0.01572446	0.18040776	0.1703356	0.16814676	-0.11132384	0.04487903
Coefficients	x_6	x_7	x_8	x_9	x_{10}	
Estimate	0.06605877	0.16508513	0.05887699	-0.071032	0.12825863	

Upon calculation, the root mean square error (RMSE) and coefficient of determination R-squared for predicting sample data are 0.1707 and 94.12%, respectively. After testing, the significance of each parameter in the ridge regression has been confirmed. Therefore, the fitted regression equation under ridge regression is:

$$y = 0.02 + 0.18x_1 + 0.17x_2 + 0.17x_3 - 0.11x_4 + 0.04x_5 + 0.07x_6 + 0.17x_7 + 0.06x_8 - 0.07x_9 + 0.13x_{10} \quad (11)$$

where except for x_4 and x_9, the coefficients of the remaining independent variables are positive. Based on the coefficients of the model's independent variables, the root mean square error of prediction, and the coefficient of

determination, ridge regression demonstrates higher predictive accuracy compared to traditional OLS estimation. However, when all independent variables are included, the coefficient of determination is lower than that of OLS estimation.

3.3 Lasso

Lasso can shrink some regression coefficients, meaning it forces the sum of the absolute values of the coefficients to be less than a fixed value and directly sets some coefficients of independent variables to zero. Therefore, it possesses the characteristic of promoting model parsimony.

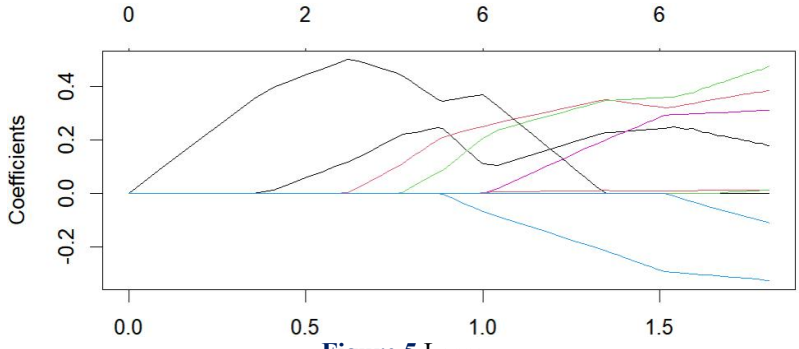

Figure 5 Lasso

The above figure(Figure 5) shows the coefficients changing with the parameter λ under Lasso estimation. As λ increases, the variables gradually enter the model. Similarly, ten-fold cross-validation is used to select the λ value with the minimum cross-validation error.Lasso Ten-fold Cross-Validation Plot can be seen in Figure 6.

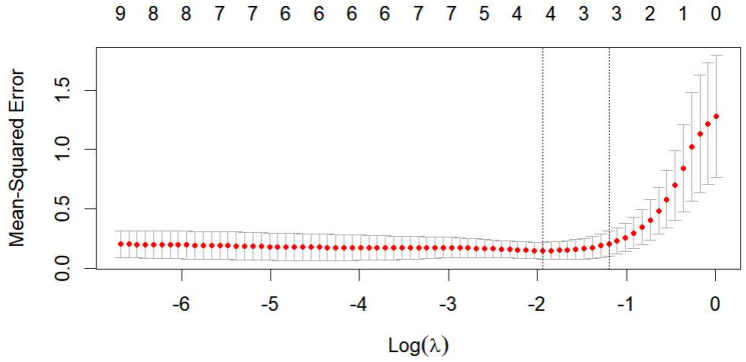

Figure 6 Lasso Ten-fold Cross-Validation Plot

Under the condition of selecting the minimum λ value 0.1434, the coefficients of the model's independent variables are as follows:

Table 5 Lasso Regression Estimation Results

Coefficients	Intercept	x_1	x_2	x_3	x_4	x_5
Estimate	0.06432890	0.23477557	0.16895035	0.05188976	—	—
Coefficients	x_6	x_7	x_8	x_9	x_{10}	
Estimate	—	0.38232146	—	—	—	

After calculation, the root mean square error (RMSE) and coefficient of determination R-squared for predicting the sample by the model are 0.248 and 92.56%, respectively. Following validation, the significance of selected parameters in Lasso regression has been confirmed.

Therefore, the regression equation fitted under Lasso regression is:

$$y = 0.06 + 0.23x_1 + 0.17x_2 + 0.05x_3 + 0.38x_7 \quad (12)$$

where all coefficients of the selected independent variables are positive. The included variables represent education, technology, policy institutions, and labor input, while variables representing the proportion of fiscal revenue to GDP for policy institutions, the ratio of import and export volume to GDP for trade openness, urbanization rate for urbanization level, longitude and latitude for geographical location, and total wastewater discharge for environmental conditions were excluded. The selection and exclusion of these variables align with practical considerations.

In terms of model effectiveness, Lasso regression outperforms OLS estimation; in terms of model simplicity, Lasso outperforms ridge regression. However, in terms of predictive accuracy and model fit, Lasso estimation is inferior to

both OLS and ridge regression. This phenomenon can be explained by the fact that OLS estimation minimizes the difference between model predictions and actual values, thereby achieving the highest R2 in terms of model fit. Ridge regression includes all independent variables in the fitting model, retaining more information than Lasso estimation, and thus outperforms Lasso in terms of predictive accuracy and model fit.

3.4 Elastic Net

Elastic Net is an estimation method that combines the regularization techniques of Ridge Regression and Lasso, leveraging the strengths of both methods.

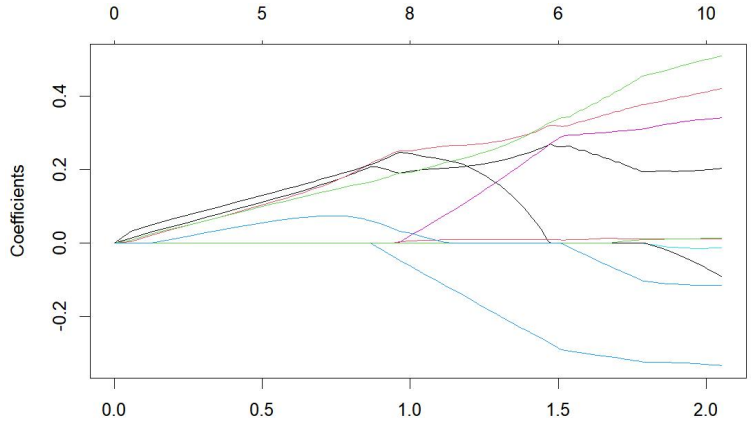

Figure 7 Elastic Net

Similar to Lasso estimation, individual variables (in Figure 7) gradually enter the model as their coefficients increase, thereby validating Elastic Net's advantage in variable selection. The selection of the optimal value is based on ten-fold cross-validation, choosing the value with the smallest cross-validation error(Figure 8). Following this, under the condition of selecting the minimum value, calculate the coefficients of the variables in the model(Table 5). Elastic Net Regression Estimation Results can be seen in Table 6.

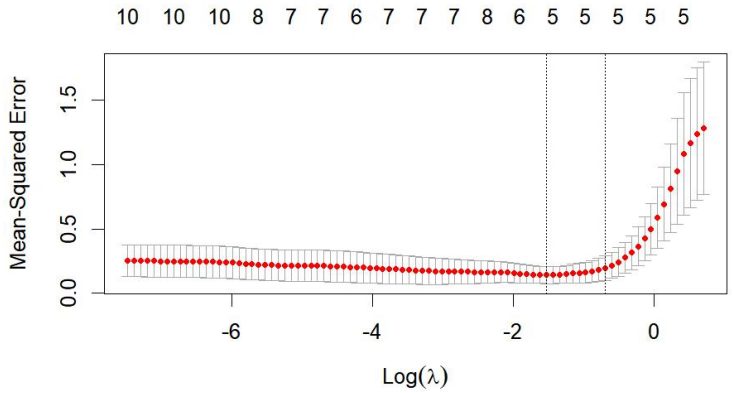

Figure 8 Elastic Net Ten-fold Cross-Validation Plot

Table 6 Elastic Net Regression Estimation Results

Coefficients	Intercept	x_1	x_2	x_3	x_4	x_5
Estimate	0.06028641	0.20452320	0.21156278	0.16460067	—	—
Coefficients	x_6	x_7	x_8	x_9	x_{10}	
Estimate	—	0.21113873	—	—	0.06343535	

Based on the calculation, the root mean squared error (RMSE) and the coefficient of determination R-squared for predicting the sample data are 0.248 and 92.56%, respectively. Upon inspection, the significance of the selected parameters under Lasso has been verified. Therefore, the regression equation fitted under Elastic Net is:

$$y = 0.06 + 0.2x_1 + 0.21x_2 + 0.16x_3 + 0.21x_7 + 0.06x_{10}, \quad (13)$$

where all coefficients of the selected predictor variables are positive. The selected variables are the same as those chosen in the Lasso estimation, representing education, technology, policy institutions, labor input, and additionally, total wastewater emissions representing environmental conditions. Other variables have been excluded. The selection

and exclusion of these variables are consistent with practical considerations. From the perspective of model effectiveness, Elastic Net outperforms OLS estimation; in terms of model parsimony, Elastic Net is superior to Ridge Regression; and regarding prediction accuracy and model fit, Elastic Net surpasses several other models.
Summary of simulated data from different estimation methods, including their model fitting and parameter selection, along with computed prediction mean square errors on test data, are as follows:

Table 7 Summary of Regression Results under Different Methods

	R^2	Parameter selection	RMSE	Variable selection
OLS	97.86%	—	0.1768	(2,3,4,6)
Ridge Regression	95.12%	$\lambda = 0.5924$	0. 1707	All
Lasso	92.56%	$\lambda = 0.1434$	0.2480	(1,2,3,7)
Elastic Net	95.02%	$\lambda = 0.2169$	0.1631	(1,2,3,7,10)

From Table 7, it is evident that Elastic Net demonstrates superior predictive accuracy compared to several other models. It exhibits greater sparsity than Ridge Regression, better effectiveness than OLS, and superior model fit compared to both Ridge Regression and Lasso, achieving an R-squared of 95.02%.
Specifically, OLS achieves an R-squared of 97.86% with a root mean square error of 0.1768. OLS estimation does not perform variable selection and may suffer from multicollinearity when not adjusted in the model. Ridge Regression and Lasso are improvements upon OLS, each with their own strengths, while Elastic Net combines the advantages of both Ridge Regression and Lasso. Therefore, this paper further analyzes the regression results under Elastic Net estimation.

3.5 Analysis of Elastic Net Regression Results

From equation (15), it is evident that the regression equation retains several significant factors influencing regional economic impact: education, technology, policy institutions, labor input, and environment. In theory, high levels of education, enhanced technological innovation capabilities, robust human capital, policy support, increased investment in labor assets, and environmental improvement all contribute to regional economic development. However, given the differences in regional capabilities in technological innovation, human capital, overall social fixed asset investment, and regional trade openness across the country, the actual impact of technological innovation, human capital, fixed asset investment, and trade openness on economic promotion varies.
Different regions have varying realities. In the production process of using technology as a factor of production, the benefits and contribution rates of each factor to economic development are different. Since all variables have been standardized during the data processing and analysis stages, the coefficients of each variable in the regression equation can to a considerable extent represent their contribution to the economy. Since the coefficients of each variable are positive, it indicates that these variables promote economic development. The contributions of education investment, technological strength, policy institutions, labor input, and environment to China's regional economic development are 20%, 21%, 16%, 21%, and 6%, respectively.

4 CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

This study established a multiple linear regression model to analyze factors influencing regional economic impact. The fitting, testing, model refinement, and evaluation were initially conducted using traditional least squares estimation. Subsequently, Ridge Regression, Lasso Regression, and Elastic Net were employed to identify the best fitting model. Finally, the regression equation derived from Elastic Net estimation was selected to analyze factors affecting regional economic development.
Comparing the four different model solving methods, Elastic Net was found to significantly enhance model fit, prediction accuracy, efficiency, and simplicity. Traditional OLS estimation often performs poorly when the number of predictors is much larger than the sample size or in scenarios of sparse models with many variables and small samples. Elastic Net effectively addresses these challenges by expanding the dimensionality of the sample and automating variable selection, transforming scenarios where the number of predictors greatly exceeds the sample size into standard multiple regression models. Scaling the original Elastic Net further optimizes results towards extremum solutions. Efficient algorithms solving the Lasso problem are utilized for each fixed parameter in the Elastic Net problem. Compared to Lasso under similar conditions, Elastic Net also reduces computational speed.
The final selected regression equation retained several factors with significant impact on regional economic development: education, technology, policy institutions, labor input, and environment. Less relevant factors were excluded from the model. The respective contributions of the selected variables to regional economic impact were determined as follows: education 20%, technology 21%, policy institutions 16%, labor input 21%, and environment 6%.

4.2 Recommendations

Based on the empirical findings of this study, the following recommendations can be provided:

4.2.1 Developing precision education-driven strategies for underdeveloped regions

For educationally disadvantaged regions like Tibet and Qinghai, it is essential to systematically and incrementally enhance technological innovation capabilities and develop precision education-driven strategies. Additionally, strategies aimed at increasing educational contributions in these areas should not solely focus on increasing educational investment. Instead, they should leverage opportunities for collaboration with other regions and universities, strengthen the sharing of scientific and research information among regions, enhance the transfer of technology and research outcomes to enterprises, and ensure that these achievements are applied in production and daily life [15].

4.2.2 Comprehensive Promotion of Regional Technological Innovation and Coordinated Economic Development

The government should comprehensively consider various situations regarding the strength of technological innovation capabilities and the level of economic development. It should accelerate achieving "prosperous regions helping less developed regions" in terms of technological innovation capabilities among different provinces, thereby avoiding further widening of spatial disparities nationwide[16]. Meanwhile, regions with strong technological innovation capabilities such as Guangdong Province should pay more attention to sustainable development of the environment and resources. Regions with weaker technological innovation capabilities can focus on integrating distinctive resources with technological innovation to stimulate the development of tourism, agriculture, and aquaculture industries, thereby promoting synchronized development of technology and the economy.

4.2.3 Increase Policy Support to Promote Labor Mobility

Labor mobility often depends on policy incentives and support. Workers move between regions, sectors, and employment statuses seeking higher wages. Increasing policy support can encourage individuals with higher and moderate educational levels to move to underdeveloped areas, thereby supporting regional development. Similarly, retaining low-skilled labor locally also contributes significantly to regional development.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Liu Qiongfang, Liu Chao. Research on the Dynamic Impact of Scientific and Technological Innovation and Industrial Structure on Regional Economy: An Empirical Study Based on GVAR Model. Times Finance, 2021(13): 33-35.

[2] Wang Anqi. Analysis of differences and influencing factors of high-quality regional economic development in China. Zhongnan University of Economics and Law, 2021.

[3] Cao Haibo. Analysis of regional economic growth differences and their influencing factors in China. Jilin:Jilin University, 2012.

[4] Zhu Xuexin, Fang Jianwen, Zhang Bin. The Impact of Scientific and Technological Innovation on China's Economic Development: An Empirical Study Based on Panel Data. Journal of Soochow University (Philosophy and Social Science), 2007(4): 18-21.

[5] Wang Wei, Xu Yilun, Yang Wenjuan, Tao Xu. The impact of scientific and technological innovation capability on regional economic development. Journal of Shandong Normal University (Natural Science Edition), 2017, 32(4): 126-136.

[6] Wang Xu, Chen Rong, Li Mingbao. Research on the Impact of Scientific and Technological Innovation on Regional Economy: An Empirical Analysis Based on Inter-provincial Panel Data. Industrial Technology and Economics, 2018, 37(9): 39-44.

[7] Song Meizhe, Li Mengsu. Measurement of the coupling and coordination relationship between higher education, scientific and technological innovation and economic development and its influencing factors. Modern Education Management, 2019(3): 19-25.

[8] Zhang Zhiruo. Research on the coupling relationship between science and technology finance and regional economic development in China. SCIENTIA GEOGRAPHICA SINICA, 2020, 40(5): 751-759.

[9] Hogenschurz Lukas, Acs Z. J. Strategies for the Development of Science, Technology and Innovation as a Public and Social Good. Journal of Entrepreneurship and Innovation in Emerging Economies, 2021, 7(1): 1-7.

[10] Sturm Thomas. Scientific innovation. Theoria: An International Journal for Theory, History and Foundations of Science, 2019, 34(3): 321-341.

[11] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1): 49-67.

[12] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 2005, 67(2): 301-320.

[13] Zou H. The adaptive lasso and its oracle properties. Journal of the American statistical association, 2006, 101(476): 1418-1429.

[14] National Bureau of Statistics of the People's Republic of China. China Statistical Yearbook from 2017 to 2021. Beijing:China Statistics Press.

[15] Chen Yong. Research on the influencing factors and solutions of regional economic development imbalance under the background of regional integration development in the Yangtze River Delta. Journal of Business Economics, 2022(02): 14-16.
[16] Fan Jie, Liu Hanchu. The Influence and Adaptation of Scientific and Technological Innovation Driven on the Change of China's Regional Development Pattern during the 13th Five-Year Plan Period. Economic Geography, 2016, 36(1): 1-9.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3012

ANALYSIS OF FACTORS INFLUENCING THE ENGEL INDEX BASED ON REGRESSION MODELS

GaoBo Peng
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 1062041360@qq.com

Abstract: This paper explores the macroeconomic indicators affecting the Engel Index. Initially, a multicollinearity test reveals the presence of multicollinearity among the independent variables. To eliminate multicollinearity, ridge regression modeling is employed. The analysis identifies six independent variables, with only one having an insignificant regression coefficient. The results show that labor force, urbanization rate, and trade balance are negatively correlated with the Engel Index, while the added value of the primary industry and the gross national income index are positively correlated with the Engel Index.
Keywords: Engel Index; Multicollinearity; Ridge regression; Principal component regression

1 INTRODUCTION

The Engel coefficient is an economic indicator introduced by German economist and statistician Ernst Engel in the 19th century. It is used to measure the standard of living and level of wealth of a household or country. The Engel coefficient is calculated as the ratio of food expenditure to total consumption expenditure. It reflects the proportion of total expenditure that a household or country spends on food. According to Engel's findings, when a household or country has lower income, the proportion of food expenditure is higher; conversely, as income increases, the proportion of food expenditure decreases. Therefore, a higher Engel coefficient indicates a lower standard of living and greater poverty, whereas a lower Engel coefficient indicates a higher standard of living and greater wealth.

The Engel coefficient is widely used not only at the household level but also in national-level analyses. Research by Lancaster et al. shows that the Engel coefficient is widely used due to its simplicity and ability to intuitively reflect living standards [1]. Kaus studied the consumption tendencies of residents in over 50 countries across 12 categories, demonstrating that Engel's law can reveal patterns of consumption [2]. Lewbel's research indicates that the basic meaning of Engel's law is that the rate of increase in food expenditure is slower than the rate of increase in income, showing that the proportion of food expenditure decreases as income increases [3]. Chai et al. found that Engel's law can reveal the relationship between household income and consumption structure [4]. Wang et al. discovered that the Engel coefficient varies among different income groups, indicating a relationship between income levels and the Engel coefficient [5]. Dorothy Brady's research showed that families without children have lower Engel coefficients, and the coefficient increases with the number of children in the family [6]. Angus Deaton and Anne Case analyzed the impact of various cultural customs on residents' dietary activities and found that different cultural habits significantly affect the Engel coefficient [7]. Crawford et al. showed that the Engel coefficient is primarily influenced by the number of women in the household; the fewer the women, the higher the Engel coefficient [8]. John found that the household head's gender, age, and family size also affect the Engel coefficient [9].

2 MAIN THEORIES

2.1 Collinearity

Collinearity is an important concept in statistics and regression analysis, especially in multiple regression models. Collinearity refers to the phenomenon where there is a high correlation among the independent variables. This can lead to instability in the regression coefficients, thereby affecting the explanatory power and predictive ability of the model. Collinearity is classified into perfect collinearity and near-perfect collinearity. Perfect collinearity occurs when an independent variable can be exactly represented by other independent variables, making regression analysis infeasible. Near-perfect collinearity occurs when there is an approximate linear relationship among independent variables. Although regression analysis can still be performed, the results may be unstable.

Collinearity primarily affects regression analysis in the following ways: firstly, it causes instability in regression coefficients, leading to large fluctuations in estimates, and even sign reversal. Secondly, it increases the standard errors of the regression coefficients, affecting the significance tests. Lastly, collinearity makes it difficult to distinguish the independent effects of the variables, reducing the explanatory power of the model. Methods to detect collinearity include the variance inflation factor (VIF), eigenvalue decomposition, and the condition number. A VIF value exceeding 10 typically indicates strong collinearity. Eigenvalues close to zero or a condition number greater than 30 also suggest severe collinearity.

Methods to mitigate collinearity include: removing highly correlated independent variables, using principal component analysis (PCA) to transform the variables and adding penalty terms through ridge regression to reduce the regression

coefficients.

2.2 Ridge Regression

Arthur E. Hoerl and Robert W. Kennard proposed ridge regression in 1970. It is a biased estimation method that improves upon ordinary least squares (OLS). The basic idea is: for a linear regression model:

$$Y = X\beta + \varepsilon \tag{1}$$

the least squares estimate of the parameters is: $\widehat{\beta} = (X^T X)^{-1} X^T Y$ If there is strong multicollinearity among the independent variables, meaning the determinant of $X^T X$is close to zero (nearly singular), the estimation results may be highly biased. When the number of effective equations is less than the number of unknowns, there is no unique solution, leading to infinitely many solutions. Using the least squares method in this scenario results in instability and unreliability. By adding a normal matrixλI, where is the ridge parameter and is the identity matrix, we obtain the ridge regression estimate:

$$\widehat{\beta} = (X^T X + \lambda I)^{-1} X^T Y \tag{2}$$

The cost function modifies the residual sum of squares (RSS) by adding a penalty term $\sum_{j=1}^{p} \beta_j^2$ for the coefficients, ensuring that while minimizing the RSS, the coefficients do not become excessively large:

$$J_\beta(\beta) = RSS + \lambda \sum_{j=1}^{p} \beta_j^2 = RSS + \lambda \|\beta\|^2 \tag{3}$$

Typically, the results of ridge regression models are slightly lower than those of ordinary regression models, but their significance is much higher. Ridge regression is particularly useful in studies with collinearity problems and large amounts of ill-conditioned data.The fonts, their sizes, and styles can be seen in the table & figure below.

3 DATA SOURCES AND VARIABLE DESCRIPTIONS

3.1 Data Sources

The data used in this paper comes from the National Bureau of Statistics, covering a period of seventeen years from 2004 to 2020. The data spans seven dimensions: Gross National Income Index (last year=100), labor force (in millions), urbanization rate (%), trade balance (in billions of RMB), added value of the primary industry (%), Consumer Price Index (CPI, last year=100), and the Engel Index (%). Among these, the urbanization rate is not direct data but is calculated according to the National Bureau of Statistics' definition: urbanization rate = urban population / total population (both based on the permanent population, not the registered population). The urban population and total population data are sourced from the National Bureau of Statistics.

3.2 Variable Descriptions

This paper selects the Engel Index (%) as the dependent variable, denoted as. The independent variables are labor force (in millions), urbanization rate (%), trade balance (in billions of RMB), added value of the primary industry (%), Consumer Price Index (CPI, last year=100), and Gross National Income Index (last year=100), denoted as,x_2, x_3, x_4, x_5, x_6 respectively.

4 EMPIRICAL ANALYSIS

4.1 Collinearity Analysis

First, the pairs function in R was used to create scatter plots, providing an overall view of the multidimensional data. See Figure 1.

From the scatter plot matrix, it is evident that some variables are correlated. For instance, the dependent variable shows a strong negative correlation with the independent variables and, and a strong positive correlation with. Additionally, and are also strongly positively correlated.

The strong correlations among the independent variables can be confirmed by examining their correlation coefficients, as shown in Table 1.

Table 1 Correlation Coefficients

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.0000000	0.8071398	0.6484744	0.8420276	0.1139538	0.6551198
x_2	0.8071398	1.0000000	0.8002923	0.9474073	0.5091849	0.8482647
x_3	0.6484744	0.8002923	1.0000000	0.7670221	0.4033238	0.6397549
x_4	0.8420276	0.9474073	0.7670221	1.0000000	0.3173843	0.7060632
x_5	0.1139538	0.5091849	0.4033238	0.3173843	1.0000000	0.6955236
x_6	0.6551198	0.8482647	0.6397549	0.7060632	0.6955236	1.0000000

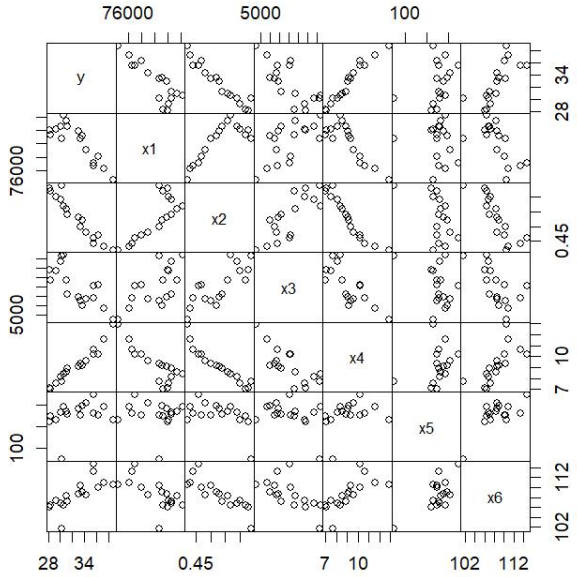

Figure 1 Scatter Plot Matrix

From the table, we see that the correlation coefficients between and,,, are 0.8071398, 0.6484744, 0.8420276, and 0.6551198, respectively, indicating strong positive correlations. The correlation coefficients between and,, are 0.8002923, 0.9474073, and 0.8482647, respectively, indicating very strong positive correlations. The correlation coefficients between x_3 and x_4, are 0.7670221 and 0.6397549, respectively, indicating strong positive correlations. The correlation coefficient between and is 0.7060632, indicating a strong positive correlation. The correlation coefficient between and x_6 is 0.6955236, indicating a strong positive correlation.

Using the kappa function in R, the condition number of the independent variable matrix was found to be 2610187. Using the vif function in R, the variance inflation factors (VIF) are obtained as shown in Table 2.

Table 2 Variance Inflation Factors

x_1	x_2	x_3	x_4	x_5	x_6
5.407842	31.326773	2.852933	18.563677	3.668128	7.660773

A condition number greater than 100 or a VIF value greater than 10 indicates severe collinearity among the independent variables. In summary, the six factors selected as potential influencers of the Engel Index exhibit severe collinearity.

4.2 Ridge Regression

As required by ridge regression, the data needs to be standardized. To determine the ridge parameter, ridge trace plots need to be created, which can be done using the lm.ridge function and the matplot function in R. See Figure 2 for the ridge trace plots.

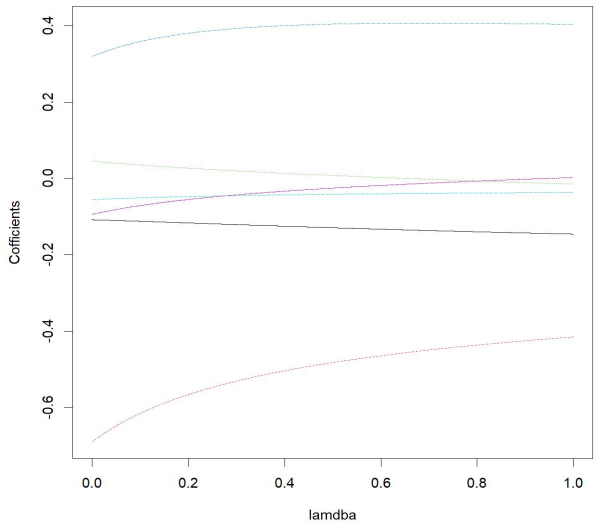

Figure 2 Ridge Trace Plots

From the plots, it can be seen that finding the value of the ridge parameter where all the curves stabilize is difficult, as most curves stabilize very slowly. Thus, it is hard to directly determine an appropriate ridge parameter. However, the select function in R can be used to calculate the value of based on several statistical criteria, or generalized cross-validation can be used to determine the ridge parameter kkk directly. In this section, generalized cross-validation is employed to determine the value of.
The lm.ridge function can then be used to provide the parameter estimates, but the summary function cannot be used to view the results. Therefore, the linearRidge function is used in this section to obtain the parameter estimates. The results are shown in Table 3.

Table 3 Regression Coefficients

(Intercept)	x_1	x_2	x_3	x_4	x_5	x_6
6.5e-16	-6.4e-04	-7.2e-04	-5.7e-04	7.2e-04	2.5e-04	5.5e-04

From the table, it can be seen that labor force (in millions), urbanization rate (%), and trade balance (in billions of RMB) are negatively correlated with the Engel Index (%). In contrast, the added value of the primary industry (%), Consumer Price Index (CPI, last year=100), and Gross National Income Index (last year=100) are positively correlated with the Engel Index (%). The magnitudes of the regression coefficients are also similar.
Next, the summary function in R can be used to output the results of the linear ridge regression model test. The specific results are shown in Table 4.

Table 4 Model Test Results

Coefficient	Estimate	Scaled Estimate	Std. Error (scaled)	t value (scaled)	Pr(>|t|)
(Intercept)	6.54E-16	NA	NA	NA	NA
x1	-0.00065	-0.00258	0.000726	3.559	0.000372
x2	-0.00072	-0.00289	0.000725	3.981	6.85e-05
x3	-0.00057	-0.00228	0.000726	3.143	0.001671
x4	0.00073	0.002919	0.000725	4.023	5.74e-05
x5	0.000251	0.001002	0.000726	1.38	0.167575
x6	0.000555	0.00222	0.000726	3.06	0.002213

From the figure, it can be seen that only the regression coefficient of did not pass the significance t-test, indicating that five variables are significant. Therefore, using ridge regression to analyze the relationship between the independent variables and the dependent variable is appropriate.

4.3 Principal Component Regression

In this section, PCR is used for modeling. Six independent variables must undergo principal component analysis to reduce the data's dimensionality and simplify the model. The princomp function in R is employed for principal component regression modeling, and the summary function is used to obtain the specific results. See table 5 for the detailed results.

Table 5 Model Test

Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
Comp.1	2.0807435	0.7215823	0.7215823
Comp.2	1.0133088	0.1711324	0.8927147
Comp.3	0.60044424	0.06008888	0.95280358
Comp.4	0.41187548	0.02827357	0.98107715
Comp.5	0.30507618	0.01551191	0.99658906
Comp.6	0.143058208	0.003410942	1.00000000

The results show that 72.2% of the data variance can be explained by the first principal component, 17.1% of the variance by the second principal component, and 6.0% by the third principal component. The variance explained by the first principal component is about four times that of the second principal component and approximately twelve times that of the third principal component. The contributions of the second and third principal components decline sharply relative to the first principal component, and the last few principal components explain very little variance. This indicates that most of the variations in the independent variables can be explained by a few dimensions or principal components.
Using the summary function, the linear combinations of the vectors constructing each principal component can also be obtained. See table 6 for the specific results.

Table 6 Composition of Principal Components

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
x1	0.399	0.443	0.375	0.371	0.605	0.0
x2	0.471	0.0	-0.341	-0.203	0.0	0.785
x3	0.408	0.0	-0.835	0.356	0.0	0.0
x4	-0.445	-0.254	0.0	0.637	0.0	0.573
x5	-0.267	0.804	0.0	0.141	-0.510	0.0
x6	-0.427	0.292	-0.399	-0.441	0.574	0.222

The results show that all six independent variables collectively determine the first principal component. However, the first three independent variables are positively correlated with the first principal component, while the last three are negatively correlated. The second principal component is constructed only by the labor force (in millions), added value of the primary industry (%), Consumer Price Index (CPI, last year=100), and Gross National Income Index (last year=100).

Based on the contributions of each principal component, this section uses only the first two principal components for the linear regression model. The lm function in R is used to output the least squares estimates, and the summary function is used to display the results of the multiple linear model test. See table 7 for the specific results.

Table 7 Model Test

|Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.157e-16	6.781e-02	0.000	1.0000
z1	-4.364e-01	3.259e-02	-13.391	2.26e-09
z2	-2.253e-01	6.692e-02	-3.367	0.0046
Residual Standard Error		0.2796 on 14 degrees of freedom		
Multiple R-squared		0.9316		
Adjusted R-squared		0.9218		
F-statistic		95.33 on 2 and 14 DF		
p-value		7.009e-09		

It can be seen that the regression coefficients of the first and second principal components pass the significance t-test, indicating that all variables are significant. Additionally, the model's significance test is also significant, as the F-statistic is 95.33, and the corresponding p-value is 7.009e-09, which is less than 0.01. The R-squared and adjusted R-squared values are 0.9316 and 0.9218, respectively, showing that the model can explain about 92% of the data variance. This collectively indicates that using PCR modeling is appropriate. It can also be seen that both the first and second principal components negatively impact the Engel Index. Combining the results of the linear combinations constructing each principal component, it can be concluded that the labor force, urbanization rate, trade balance, and Consumer Price Index are negatively correlated with the Engel Index, while the added value of the primary industry and Gross National Income Index are positively correlated with the Engel Index.

5 RESULTS AND DISCUSSION

Factors such as the labor force, urbanization rate, trade balance, added value of the primary industry, and Gross National Income Index all have significant impacts on the Engel Index. Specifically, the labor force, urbanization rate, and trade balance are negatively correlated with the Engel Index, while the added value of the primary industry and Gross National Income Index are positively correlated with the Engel Index.

From the data in Table 1, it can be observed that the Engel Index in China has been decreasing year by year, indicating an improvement in the living standards of Chinese residents. However, in 2022, the Engel Index increased compared to the previous year, likely due to the impact of the COVID-19 pandemic.

Based on the conclusions drawn from the regression model, to continue improving the living standards of Chinese residents and further reduce the Engel Index, it is necessary to continue increasing the values of the labor force, urbanization rate, and trade balance while decreasing the values of the added value of the primary industry and Gross National Income Index.

In terms of the labor force, the number of workers in China has been decreasing in recent years due to the aging population, which has led the country to gradually relax its family planning policies.

Regarding the urbanization rate, this value has been on an upward trend and should continue to be maintained.

Concerning the trade balance, it has generally been steadily increasing. However, there have been some years of decline since 2016, which may be related to the sanctions imposed by the United States on Chinese enterprises. Therefore, to maintain an increasing trade balance, China needs to develop its core competitiveness.

Regarding the added value of the primary industry, the added value of agriculture is relatively low. An excessive proportion of agriculture can lead to an increase in the Engel Index. However, agriculture is fundamental and its total amount should not be reduced. Therefore, China should vigorously develop the secondary and tertiary industries while ensuring a stable agricultural foundation.

As for the Gross National Income Index, it is influenced by inflation, which is inevitable. Hence, China should aim to keep the inflation rate under control.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Lancaster G, Ray R, Valenzuela M R. A Cross-Country Study of Equivalence Scales and Expenditure Inequality on Unit Record Household Budget Data. Review of Income & Wealth, 1999, 45(4): 455-482.
[2] Kaus W. Beyond Engel's law - A cross-country analysis. The Journal of Socio-Economics, 2013.
[3] Benjamin, S., Loeb. The Use of Engel's Laws as a Basis for Predicting Consumer Expenditures. Journal of Marketing, 1955, 20(1): 20-27.
[4] Chai A, Moneta A. Retrospectives: Engel Curves.The journal of economic perspectives, 2010, 24(1): 225-240.
[5] Wang X, Woo WT. The size and distribution of hidden household income in China. Asian Economic Papers. 2011, 10(1): 1-26.
[6] Brady DS, Barber HA. The pattern of food expenditures. The Review of Economics and Statistics. 1948, 30(3): 198-206.
[7] Angus Deaton, Anne Case. Analysis of Household Expenditure. LSMS working paper, 1992(4).
[8] Crawford I, Laisney F, Preston I. Estimation of household demand systems with theoretically compatible Engel curves and unit value specifications. Journal of econometrics. 2003, 114(2): 221-241.
[9] Gibson J. Why Does the Engel Method Work? Food Demand, Economies of Size and Household Survey Methods.Oxford Bulletin of Economics & Statistics, 2002(4): 341-359.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3001

APPLICATION AND EFFECTIVENESS ASSESSMENT OF QUALITY CERTIFICATION AND STANDARDISATION IN AGRICULTURAL PRODUCT E-COMMERCE

YuanBo Jia[1,2*], Rui Yan[3], Khorloo Yundendorj[4]
[1]*Graduate School, University of Finance and Economics, Ulaanbaatar 13381, Mongolia.*
[2]*School of Computer Information Management, Inner Mongolia University of Finance and Economics, Hohhot010070, China.*
[3]*Department of Finance, The Second Affiliated Hospital of Inner Mongolia Medical University, Hohhot010030, China.*
[4]*Department of Interdisciplinary Studies, University of Finance and Economics, Ulaanbaatar 13381, Mongolia.*
Corresponding Author: YuanBo Jia, Email: ncdjyb@163.com

Abstract: As a key strategy to promote China's agroecological progress, rural revitalization and farmers' income growth, agricultural product e-commerce is facing unprecedented market potential thanks to the extensive penetration of Internet technology and the rapid rise of new transaction modes. However, there are still some bottlenecks in this field in China, such as the lack of a product quality monitoring system, inconsistency in logistics and distribution services, and lagging regulations and supervision, which undoubtedly pose challenges to the high-quality development of agricultural e-commerce. Therefore, the establishment of a perfect quality certification and standardization system for agricultural e-commerce is a key initiative to break the status quo and guide the sustainable and sound development of e-commerce agriculture. Based on this, this paper first clarifies the connotation of quality certification and standardization of agricultural e-commerce, then analyses the application model of quality certification and standardization of agricultural e-commerce from several aspects, including practitioners, logistics and distribution, sales channels and policies and regulations, and finally introduces specific implementation methods.
Keywords: Quality certification; Standardisation; Agricultural e-commerce

1 INTRODUCTION

Driven by a series of national policies and measures to actively support the rural economy, agricultural e-commerce has shown a strong momentum of development, played a key role in the rural revitalization project, effectively promoted the trade circulation of agricultural products, increased the economic income of farmers' groups, and played a driving role in the process of poverty alleviation in impoverished areas. However, the development of agricultural e-commerce has not been smooth, accompanied by significant challenges of imperfect normative and standardization systems. Problems such as mixed quality of agricultural products, low brand recognition, and redundant construction of sales channels are increasingly highlighted, which is a significant gap with the long-term goal of agricultural e-commerce set by the government. Therefore, there is an urgent need to accelerate the process of upgrading the quality certification and standardization of agricultural e-commerce to achieve sustainable and high-quality development.

2 CONNOTATION OF QUALITY CERTIFICATION AND STANDARDIZATION OF AGRICULTURAL PRODUCTS E-COMMERCE

The agricultural products e-commerce quality certification mechanism is a rigorous procedure implemented within the framework of electronic business transactions, which mainly focuses on the characteristics of agricultural products, such as quality, safety, origin, and production process, etc., and carries out detailed assessment and confirmation through specific standards and methods. Its core objective is to strengthen consumer confidence in agricultural products and promote the sustainable development of e-commerce platforms. This process involves the following key links: first, the environmental quality certification, which mainly emphasizes the ecological context in which the agricultural products are grown and ensures that they are grown in an environment free of pollution and potentially hazardous factors, to safeguard the purity and health of the agricultural food products [1]. The second is the quality certification of the production process, which mainly focuses on the compliance of the production technology of agricultural products to ensure that the entire production process is carried out in a monitorable standardized process, thus enhancing the stability of the quality of agricultural products. The last is the setting and verification of quality certification scales, such as organic certification, green labeling, or geographical indications, which aim to ensure that agricultural products meet preset high quality or specific standards.

The standardization process of agricultural e-commerce is essentially an orderly integration of the entire chain of online transactions, to enhance the standardization level of the entire industry, thereby optimizing product quality, enhancing the competitiveness of agricultural products in the marketplace, and promoting the sustainable evolution of the e-commerce model. This concept covers several core areas: firstly, standardization in the production chain emphasizes following nationally and industry-recognized best practices for fine-tuned management of agricultural operations, such

as efficient use of land, scientific planting strategies, precise fertilization and irrigation, effective pest prevention and control, and standardized harvesting and packaging processes. Such standardization not only improves the efficiency of agricultural production and reduces the waste of resources, but also contributes to environmental protection and ensures the high quality and safety of agricultural products [2]. Second, product standardization focuses on the unity and consistency of products, covering clear product specifications, quality requirements, and standardized packaging design. For example, specific produce sizes, sweetness indicators, and strict pesticide residue control standards are set to meet the diversified needs and trust of consumers. Finally, standardization in the distribution chain involves standardization of logistics, warehousing, and marketing processes. For example, by implementing a comprehensive agricultural product tracking system, consumers can clearly understand the complete path of the product from field to table, increasing transparency. At the same time, strict control of transportation and storage environments effectively prevents products from being damaged or having quality problems in circulation, ensuring the rights and interests of consumers. Overall, the standardization of agricultural product e-commerce is an all-round, multi-level process aimed at creating an efficient, safe, and reliable trading environment for agricultural products.

3 THE APPLICATION MODEL OF QUALITY CERTIFICATION AND STANDARDIZATION IN AGRICULTURAL E-COMMERCE

3.1 Standardization of Practitioners

At present, the reserve of talent in the field of agricultural e-commerce in rural areas of China has not reached saturation level, and agricultural e-commerce has not yet got rid of the concept of "opening a store" at the beginning of the development process, and there is a serious lack of composite talents who can comprehensively manage the business process as well as experts with professional skills. The standardization process of the industry calls for the emergence of high-quality talent. Based on this, the application of quality certification and standardization of agricultural e-commerce needs to focus on the standardization of practitioners. On the one hand, the government should actively guide the flow of talent to the agricultural industry and rural areas, and motivate the industry's best talents to join the field of agricultural e-commerce through reasonable and reliable policy means, such as employment subsidies and preferential policies. This requires close collaboration between government departments at all levels and various units, and at the same time, based on the means of publicity and promotion to gradually change the concept of social cognition for the flow of talent [3]. On the other hand, in agricultural e-commerce transactions the information distortion problem occurs frequently, the consumer's trust is often due to the false or misleading description of the goods damaged, resulting in a gradual decline in subsequent purchase intentions. In addition, it is also necessary to carry out quality certification and standardized management of professionalism for practitioners, who need to operate pragmatically, rigorously manage marketing means, eliminate false and exaggerated publicity, and commit to improving consumer satisfaction, to form a positive cycle in the field of agricultural e-commerce, and set up a high-quality brand image and industry integrity culture.

3.2 Standardization of Logistics and Distribution

From the perspective of supply chain management, a mature logistics system for agricultural products can significantly reduce the cost and loss of the circulation link and enhance the market competitiveness of agricultural products. At present, agricultural product trading mainly relies on the traditional B2B2C model, which leads to scattered orders and is difficult to integrate, reducing the operational efficiency of agricultural enterprises and transportation efficiency. Using the intensive advantages of e-commerce, with the help of big data analysis, a large number of orders can be consolidated and processed, batch delivery, to avoid individual agricultural products in the mixed transport suffering contamination or deterioration, thus saving transportation costs. For the characteristics of agricultural products such as perishable and storage difficulties, logistics companies should give priority to freshness and anti-pollution, the use of targeted transportation strategies, such as special transport vehicles, shorten the transportation cycle of agricultural products from the origin to the market to ensure the quality of goods. In addition, the establishment of a perfect tracking system for agricultural products is a key link to ensure the quality of agricultural products and food safety. The tracking system can effectively eliminate the information barriers between consumers, distributors, and producers, and provide powerful evidence for resolving subsequent disputes. In the e-commerce environment, especially for special agricultural products such as beef, fruits and vegetables, and aquatic products, the application of tracking technology is more convenient. China has already implemented the construction of traceability systems in many places and set up several traceability production bases. To comprehensively improve the efficiency and quality of logistics and distribution, it is necessary to implement quality certification and standardized operations in the sales model, product quality tracking, transport processes, and distribution services, which not only reduces costs but also enhances the profitability of producers and suppliers.

3.3 Standardization of Marketing Channels

Traditional marketing strategies for agricultural products are relatively outdated, generally relying on large-scale wholesaling, bazaar retailing, or word-of-mouth building confined to a single channel to boost sales. In recent years, the 2020 epidemic has accelerated the development of changes in agricultural sales channels, with online shopping

platforms and virtual market browsing becoming increasingly popular. With the influence of online media and the boom of live sales, agricultural e-commerce has gained huge attention, promoting the construction of its quality certification and standardization system [4]. Under the background of the national strategy of "three products and one standard", those poor areas where the sales of agricultural products are the main economic driver, should take the initiative to adopt this strategy, utilize the "three products and one standard" system to promote the local agricultural products, shape a unique e-commerce brand, and establish a sound quality certification and standardization system for agricultural products. The company should also establish a sound quality and safety monitoring system for agricultural products and a standardized marketing and sales management system.

3.4 Standardization of Policies and Regulations

The construction of a perfect quality certification and standardization system for the e-commerce of agricultural products is inseparable from the effective implementation of policies and regulations and the strengthening of the supervision mechanism. The role of regulations at the macro level is to drive the orderly development of the entire agricultural e-commerce industry chain through precise formulation and jointly promote economic prosperity. Government departments need to deepen the improvement of relevant laws in key areas such as quality assurance of agricultural products, network security maintenance, and IT compliance. Based on the Law on the Protection of Consumer Rights and Interests and the Contract Law, the protection of consumer rights and interests should be strengthened, and their right to obtain traceability information when purchasing agricultural products online should be expanded to ensure fair trade. At the same time, based on the Law on Quality and Safety of Agricultural Products, corresponding regulations have been formulated to promote China's agricultural product quality standards to be in line with international standards. In response to problems in the deep processing of agricultural products in rural areas, such as outdated technology, illegal additives, and jerry-building, policies, and regulations should establish strict quality and safety thresholds to prevent these problems from affecting product quality and consumer health. In warehousing and transportation, there is an urgent need to clarify standardized operating procedures to prevent food spoilage due to mismanagement and ensure food safety. In the special environment of e-commerce of agricultural products, quality, and safety standards should focus on minimizing the loss of commodities in the distribution process and how to adapt to the diversified business conditions of farmers. Therefore, more comprehensive and practical standards need to be developed to adapt to this complex and changing trading environment.

4 THE IMPLEMENTATION METHOD OF QUALITY CERTIFICATION AND STANDARDIZATION OF AGRICULTURAL E-COMMERCE

4.1 The Practical Combination of Agricultural E-Commerce and Quality Certification and Standardization

The process of rural revitalization calls for a strategy that goes beyond a single system of agricultural products and e-commerce, and requires the close integration of the two to build a perfect and efficient framework for the standardization of agricultural e-commerce to activate the entire industrial chain and promote rural revitalization.

First of all, on the technical level, there is an urgent need to strengthen the construction of informationization in rural areas. In the process of building rural network facilities and Internet coverage, we can draw on the experience of advanced regions at home and abroad, such as Japan, the United States, and other developed countries that designed "highly informative rural programs" and "online agricultural technology service network" and other initiatives for the development of China's agricultural e-commerce. Provides a certain reference for the development of China's agricultural e-commerce. Drawing on the excellent elements, we can establish an advanced information service system for agricultural products, deliver market information in real-time, ensure that the information released is true and accurate, solve the problem of information asymmetry, and encourage technological innovation and the introduction of talents through financial support [5].

Secondly, from the overall consideration of the industrial chain, it is crucial to build a professional and trustworthy agricultural e-commerce website, which can play a significant role in helping the construction of quality certification and standardization of different links in the practical application, and make up for the gap in the quality certification and standardization of agricultural e-commerce. Developed countries have a rich variety of models, including the O2O model combining brick-and-mortar stores and online, B2B professional online trading, and various types of online markets and community agricultural stores. Although the number of domestic e-commerce platforms has grown, quality certification and standardized regulation still need to be improved. Due to the seasonal and geographical characteristics of agricultural products, as well as China's vast territory and uneven development, it is extremely challenging to realize the closed loop of an efficient supply chain. Therefore, it is urgent to build a high-quality, comprehensive, and professional e-commerce platform, and at the same time, strengthen the cooperation between e-commerce platforms and producers to ensure the quality of products and smooth sales channels.

Finally, from the perspective of industry practitioners, improving the skills and training quality of all industry participants is the key to promoting the standardization and sustainable development of agricultural e-commerce. At present, the shortage of talent in the fields of information technology, logistics management, and marketing in China has seriously constrained the development of e-commerce standardization. Therefore, increasing the training of talents and balancing supply and demand is a key path to achieving this goal.

4.2 Evaluation and Improvement of Quality Certification and Standardization of Agricultural Products E-Commerce

After the standards are established and put into practice, it is crucial to ensure their effective implementation and grounding, and to this end, a series of comprehensive strategies need to be adopted to promote the standardization process of the agricultural e-commerce industry and related subjects, to promote the sound and sustainable growth of the industry.

The first task is to sort out and widely disseminate the standards in an orderly manner. First of all, the newly introduced standards should be sorted out in depth to form a clear and easy-to-understand operation manual containing detailed explanations of the standards, practical case analyses, and answers to frequently asked questions, so that all kinds of participants can quickly grasp the core content and key points of implementation. At the same time, through online and offline educational activities, such as training courses, seminars, and lectures, the concept of standardization is vigorously popularized to enhance the awareness and acceptance of standards within and outside the industry. The government, industry associations, and e-commerce platforms should work together to promote knowledge popularization and ensure that all participants fully understand and voluntarily comply with the standards.

Second, strengthen the implementation and supervision mechanism. The core of the implementation phase is to put the standards into practice. E-commerce companies need to integrate the standards into their daily operations, control product quality at the source, optimize supply chain management, and improve service quality. Government departments and industry associations need to set up a strict monitoring system to regularly assess the implementation of enterprise standards, correct and punish violations, and build an effective incentive and constraint mechanism [6]. At the same time, third-party certification organizations are encouraged to intervene and certify the quality of agricultural products to protect consumer trust.

Third, continuous improvement and correction. The vitality of the standard stems from its flexibility. During the implementation process, all parties should actively feedback on problems and challenges, especially the factors affecting the implementation of standards. The government, industry associations, and standard-setting bodies should set up smooth feedback channels, according to the feedback timely adjustment and update standards to ensure that it always keeps pace with industry development, to maintain the quality certification and standardization system of advanced and practical.

Fourth, drive innovation and standards integration. The development of agricultural e-commerce cannot be separated from the promotion of technological innovation. At the same time, the implementation of standards encourages enterprises to increase investment in research and development and promote the close integration of technological innovation and standards. Through the development and promotion of new technology standards, enterprises are guided to adopt efficient production methods and management methods to enhance the value and market competitiveness of agricultural products, while strengthening the convergence with international standards to enhance the internationalization of the industry.

Finally, strengthen the collaborative quality certification and standardization of the industrial chain. Agricultural product e-commerce involves many links and multiple participants, so when implementing standards, attention should be paid to the synergistic effect of the industrial chain. Through the establishment of information sharing, cooperation, and benefit distribution mechanisms, it promotes close collaboration and resource complementarity in each link, thus enhancing the standardization level and operational efficiency of the whole industry chain.

Therefore, the quality certification and standardization implementation phase of agricultural e-commerce involves a complicated and closely interrelated process system, which requires the active participation and coordination of multiple actors such as government departments, industry organizations, merchants, and customers [7]. By integrating and promoting the knowledge of standards, strengthening the implementation and supervision, constructing a feedback and revision system, promoting technological innovation and standardization, as well as promoting industrial chain collaboration and standardization structure, the effective implementation of the specifications can be ensured, thus promoting the healthy development of the e-commerce industry of agricultural products.

5 CONCLUSION

With the acceleration of globalization and the trend of consumption upgrading, consumers' pursuit of quality of life has prompted a continuous rise in the market demand for agricultural products, and the requirements for product quality are also rising simultaneously. This trend provides unprecedented opportunities for the rise of agricultural e-commerce in China. However, China's development in this field is still in its infancy and still faces a series of challenges and risks. The construction of a perfect information standard system for agricultural products and the effective use of network technology is crucial for promoting the transformation and upgrading of China's agricultural industry. It can be said that strengthening the standardization process of agricultural production, cultivating specialized agricultural information management talents, and strengthening the quality certification and standardization of agricultural products in the context of e-commerce are the key paths to driving the prosperity of the agricultural and rural economy and moving towards agricultural modernization.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yang Yang. Discussion on the development path of agricultural products e-commerce platform based on big data background. Shanghai Business, 2023(1): 38-40.
[2] Chen Lu, Peng Chengyuan, Zhao Jianwei. Research on the Reasons and Countermeasures Constraining the Development of Live Streaming of Agricultural Product E-commerce. Journal of Shanxi Finance and Taxation College, 2022, 24(1): 5.
[3] Ren Xin, Wu Liancui." Internet +" Background of agricultural e-commerce development status quo, problems and countermeasures. China Business Theory, 2024(4): 34-37.
[4] Ou Danli, Wei Man. Research on value co-creation of digital-enabled agricultural e-commerce. Agricultural Outlook, 2022, 18(11): 3-11.
[5] Peng Chao, Zhai Shixian. The development of high-quality agricultural e-commerce from the perspective of new agricultural management subjects. China Farmers' Cooperatives, 2022(9): 12-15.
[6] Wang Ying. Exploration of high-quality development and governance path of China's agricultural e-commerce in the era of digitalization. Management Science and Research: Chinese and English Edition, 2023, 12(6): 122-127.
[7] Liu Aiying. Dilemma of agricultural e-commerce development and cracking--Taking Xingtang County as an example. Journal of Shijiazhuang Institute of Vocational Technology, 2022, 34(6): 39-42.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3013

ANALYSIS AND PREDICTION OF INFLUENCING FACTORS ON THE PROBABILITY OF STROKE

XinChun Wang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 2545083901@qq.com

Abstract: With the continuous development of big data, statistical model analysis has permeated various fields of life, particularly in clinical medicine. In addressing the clinical prediction of patients' stroke probability, modeling methods such as ANOVA, support vector machines, binomial logistic regression analysis and random forest models are essential. Given that the dependent variable in this study is a binary classification problem, logistic regression analysis and random forest models were selected for the analysis. This paper elaborates on the principles of logistic regression analysis and random forest models and random forest models, providing the regression equation for the regression model and the importance scores of each variable in the random forest model. Additionally, the predictive capabilities of these two models were evaluated, including an assessment of prediction accuracy.Through the application of regression and random forest models, we aim to enhance the clinical prediction accuracy of patients' stroke probability, thereby providing a more reliable basis for clinical decision-making.
Keywords: Stroke; Influencing factors; Logistic Regression Analysis; Random forest regression model; Prediction accuracyt

1 INTRODUCTION

With the increasing development of data information and data technology, the application field of statistics has become more extensive, especially in the field of clinical medicine. Clinical prediction models refer to establishing models using various relevant factors to calculate the probability of the occurrence of a certain disease or the future disease status[1]. Clinical risk prediction models mainly involve doctors using established predictive models based on various medical influencing factors to analyze and calculate the overall probability of the future occurrence of a specific disease or the risk of future disease for patients[2]. Diagnostic probability models focus on analyzing the probability of a particular clinical disease diagnosis in an individual patient, determining the likelihood of diagnosing a disease based on clinical signs and characteristics of the patient or a specific group, predicting the current status of the patient (whether they are ill), which is more common in cross-sectional studies[3]; prognostic models are used to predict the likelihood of an individual patient experiencing a specific event in the future, mainly referring to recurrence, death, disability, and future complications, generally in cohort studies[4].

Logistic regression models and random forest models are widely used in the field of clinical prediction. For example, factors influencing the occurrence of a certain disease are explored, and the probability of disease occurrence is predicted based on these factors[5]. In this study, taking the analysis of the probability of stroke as an example, two groups of people are selected, one group with stroke and the other without stroke, with different disease characteristics and daily lifestyles. Therefore, the dependent variable is whether the person has a stroke, with values of "yes" or "no," and there can be many independent variables, such as age, gender, whether they have heart disease, whether they have hypertension, etc[6]. Independent variables can be either continuous or categorical.

2 DATA SOURCES AND DESCRIPTIONS

The data in this article is from the stroke prediction dataset published on the website. The analysis of this study involved a total of 1110 research subjects. The original dataset had 11 variables, with the response variable being whether or not the individual had a stroke, which is a binary variable. The occurrence of stroke is influenced by multiple factors, including both quantitative and qualitative data. For the convenience of this study, the indicators affecting the occurrence of stroke are mainly represented by quantifiable variables(Table 1).

Table1 Variable Description Table

Symbols	Variable Name	Illustrate
Y	Stroke	Y=1:Yes; Y=0:No
A	Age	A=0.25:age in (0,31]; A=0.5:age in (31,52] A=0.75:age in (52,68]; A=1:age in (68,100]
G	Gender	G=1:Male; G=0:Female
X_1	Hypertension	X_1=1:Yes; X_1=0:No
X_2	Heart disease	X_2=1:Yes; X_2=0:No
X_3	Married	X_3=1:Yes; X_3=0:No
X_4	Work Type	X_4=1:employed; X_4=unemployed
X_5	Residence type	X_5=1:Urban; X_5=0:Rural

X_6	Avg glucose level	Numeric
X_7	Body Mass Index	Numeric

Before modeling, conduct descriptive statistics on the data. Among 1110 samples, 249 individuals suffered from stroke, with 108 males and 141 females(Figure 1). The likelihood of women suffering from a stroke is slightly higher than that of men.

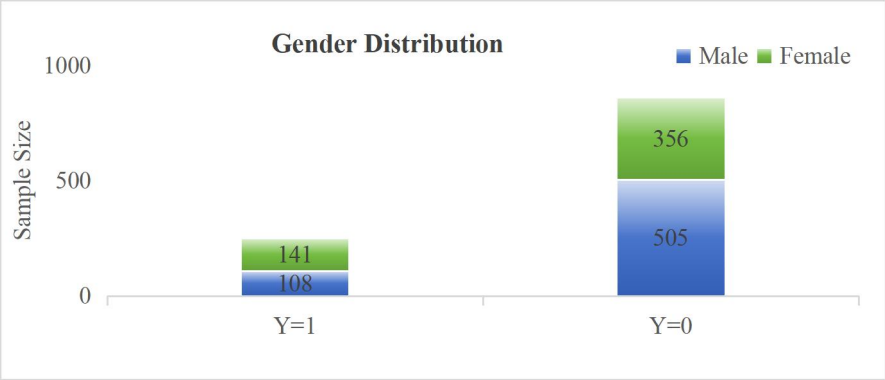

Figure 1 Gender Distribution

From the age distribution chart, it is evident that the largest number of samples falls within the older age group(Figure 2). The probability of stroke is highest among middle-aged individuals. However, it should not be overlooked that young adults also exhibit a certain incidence of stroke, indicating that advanced age is not the sole factor for stroke occurrence. This underscores the objectivity and importance of data analysis.

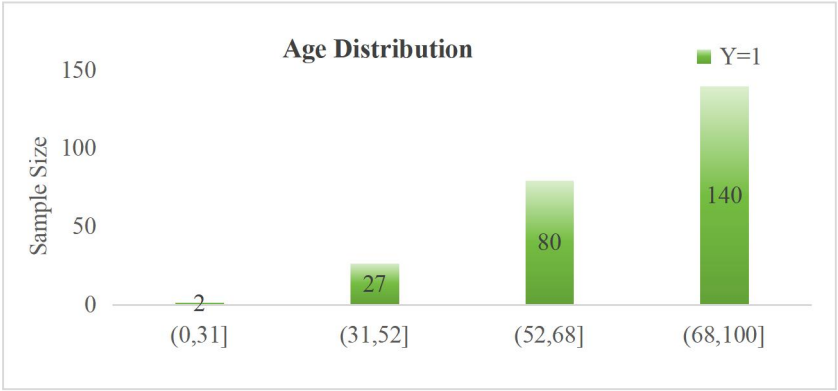

Figure 2 Age Distribution

Additionally, among other phenomena, the highest numbers of afflicted individuals are those who are married and employed, while the numbers of those with hypertension and heart disease are the lowest(Figure 3). However, it is important to note that this does not imply that stroke is unrelated to hypertension and heart disease. This is not only because descriptive statistics do not allow for direct conclusions, but also because hypertension and heart disease are merely factors that can trigger strokes; they are related but not prominently. Our study focuses on risk prediction for patients who have not yet had a stroke, analyzing the characteristics of relevant phenomena during the latency period.

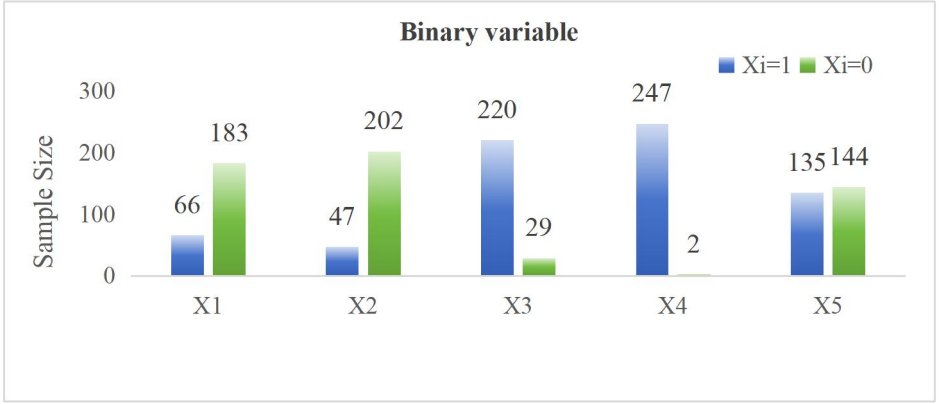

Figure 3 Binary variable Distribution

Next, we further explore the relationship between variables and stroke(Figure 4). By calculating the correlation coefficients between the independent variables and the dependent variable, as well as the correlation coefficients between each pair of variables, we can draw some conclusions. From the graph, we can see that the correlation coefficient between age and stroke is the highest, indicating that, the older the age, the higher the future risk of stroke. On the other hand, the correlation coefficients for gender, residential type, and body mass index are almost zero, suggesting that these three variables may not be related factors for stroke.

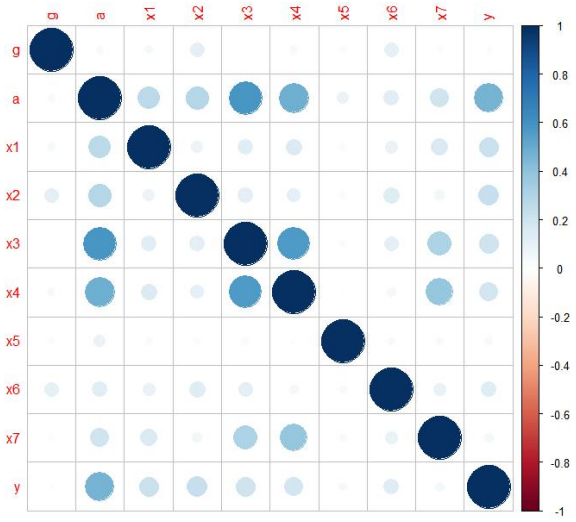

Figure 4 Correlation Coefficients

3 ANALYSIS OF FACTORS INFLUENCING STROKE INCIDENCE

There are many factors that contribute to stroke, such as heart disease and hypertension. Most of these indicators are categorical variables that do not follow a normal distribution and have high collinearity among explanatory variables. Therefore, traditional regression models do not meet the necessary assumptions. Consequently, traditional regression models are abandoned. In this section, a random forest model is constructed to evaluate the probability of stroke occurrence.

3.1 Implementation of Logistic Regression Models

The number of decision trees that corresponds to the minimum out-of-bag error is 328, determined by building a random forest model.

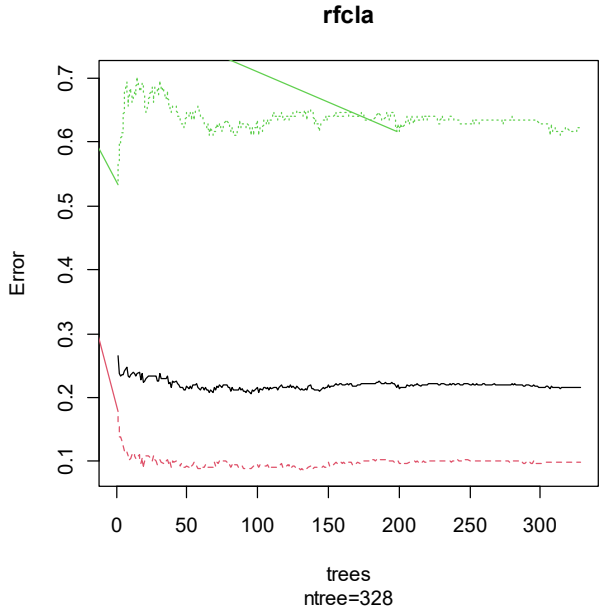

Figure 5 Out-of-bag Error/ntree=328

As shown in the above figure 5, this chart depicts the decreasing trend of out-of-bag data obtained from the random forest model. When the number of decision trees increases, the out-of-bag data tends to stabilize.From the diagram, it

can be roughly estimated that the out-of-bag error rate is around 60%, with the Type I error rate and Type II error rate being approximately 60% and 10% respectively. Below, the specific values for the out-of-bag error rate, Type I error rate, and Type II error rate are calculated.

Next, we will visualize the out-of-bag error rate, the first type error rate, and the second type error rate, and calculate their specific values(Table 2).

Table2 Out-of-bag Error Rate and **Confusion matrix**

OOB estimate of error rate: 21.62%			
Confusion matrix:			
	0	1	class.error
0	543	59	0.09800664
1	109	66	0.62285714

The Type I error rate and Type II error rate are respectively at 62.3% and 9.8%, with the Type I error rate being somewhat high. However, the out-of-bag error rate is at 21.62% , indicating that the prediction accuracy of this random model is at 78.38%, which shows that the overall predictive performance of this random model is relatively good.

3.2 Implementation of Logistic Regression Models

In a random forest model, it is not possible to obtain the average regression coefficient for each independent variable individually. Instead, the importance of each variable should be determined using the comprehensive scoring method, which involves the average decrease in mean squared error and the average decrease in the accuracy error of the forest model. These scores evaluate the average impact of each variable on the weights of other dependent variables. Within the research framework of this paper, the influence and extent of each independent variable on the primary dependent variable are illustrated in the following diagram.

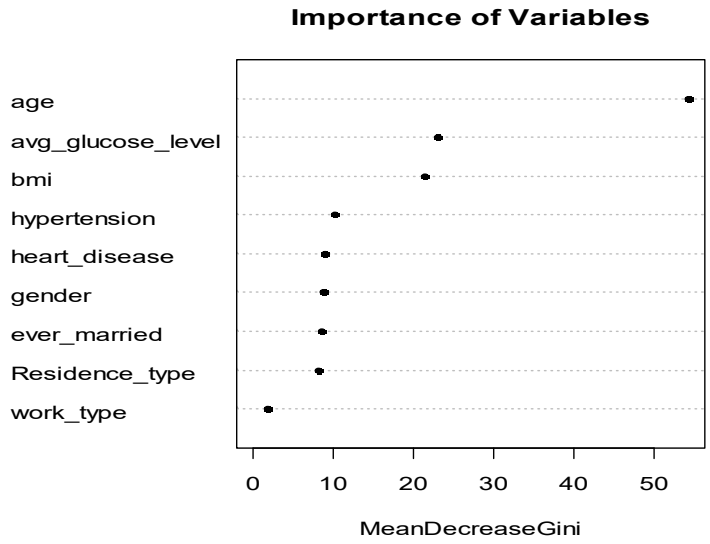

Figure 6 Importamce of Variables

In Figure 6, it can be observed that, according to the mean squared error values, the top five significant factors influencing stroke are age, average glucose level, body mass index, hypertension, and heart disease. Among these, age is the most significant factor affecting stroke, indicating that the older a person is, the higher the probability of having a stroke. This highlights that elderly individuals are particularly at risk of stroke. The second most significant factors are average glucose level and body mass index, suggesting that to reduce the probability of stroke, it is crucial to control the increase in average blood glucose level and body mass index by managing blood sugar levels and preventing obesity in daily life. Additionally, attention must be paid to the prevention and control of hypertension and heart disease. Although these two factors are not as critical as the first three, they are still contributors to stroke and should not be ignored. Patients with hypertension or heart disease must control the progression of these conditions to prevent inducing a stroke. Among the factors, although indicates that married individuals have a higher incidence rate than unmarried individuals, this might be because married individuals are generally older than unmarried ones, so it is not considered for now. Residence type has a certain impact on stroke, but its influence is not significant. Occupation work type is the least important factor affecting stroke in the random forest model, so having or not having a job has the least impact on stroke.

3.3 The Confusion Matrix of the Random Forest Model and Prediction Accuracy

Table3 The Confusion Matrix of the Random Forest Model

Actual Predict	0	1
0	230	28
1	43	31

Just like Table 3, using this confusion matrix, we can determine that the prediction accuracy of the random forest model is 78.6%, which is greater than 75%, indicating that this model has good prediction accuracy and strong predictive ability.

4 CONCLUSION

We use a random forest model to determine the number of decision trees that result in the smallest error. The first type of error rate and the second type of error rate are respectively, with the first type of error rate being a bit high, but the out-of-bag error rate is, indicating that the prediction accuracy of this random forest model is 78.38%, which means the overall prediction performance of this random forest model is quite good. Among the nine influencing factors, based on the mean squared error, the top five factors significantly impacting stroke are age, average glucose level, body mass index, hypertension, and heart disease. Among these, age is the most significant factor affecting stroke; the older the age, the higher the probability of having a stroke. This indicates that elderly people are at a particularly high risk of stroke. Secondly, average glucose level and body mass index have the highest impact on stroke, implying that controlling average glucose levels and body mass index is crucial in reducing the probability of stroke. It is important to manage blood sugar levels and prevent obesity in daily life. Lastly, attention should also be paid to the prevention and control of hypertension and heart disease. Although these two factors are not as critical as the previous three, they are still factors that can induce strokes and should not be ignored. Patients with hypertension or heart disease must control the progression of their conditions to prevent strokes. The prediction accuracy of this random forest model is 78.6%.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Breiman L. Random forests. Machine learning, 2001, 45(1): 5-32.
[2] Liaw A., Wiener, M. Classification and regression by randomForest. R news, 2(3): 18-22.
[3] Diaz-Uriarte, R., Alvarez de Andres, S. Gene selection and classification of microarray data using random forest. BMC bioinformatics, 2006, 7(1): 3.
[4] Ishwaran H., Kogalur UB. Random forests for survival, regression and classification (RF-SRC). R packag e version 1, 2007, 4.
[5] Ishwaran H., Kogalur UB., Blackstone EH., Lauer MS. Random survivalforests. The annals of applied sta tistics, 2008, 2(3): 841-860.
[6] Genuer Robin, Jean-Michel Poggi, Christine Tuleau-Malot. Variable selection using random forests. Pattern Recognition Letters, 2010, 31(14): 2225-2236.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3014

RESEARCH ON THE MARKETING STRATEGY OF DOMESTIC BEAUTY INDUSTRY FROM THE PERSPECTIVE OF DIGITAL ECONOMY

ZhaoShuo Wu
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: wuzhaoshuo0520@qq.com

Abstract: With the rapid development of digital technology, digital economy has become one of the main driving forces for global economic growth. In China, the digital economy has made remarkable achievements. Among them, digital technologies such as e-commerce, social media, big data and artificial intelligence have been widely used in various industries. In this context, the domestic beauty industry has also ushered in unprecedented development opportunities and challenges. From the perspective of digital economy, this study aims to deeply analyze the strategies and practices of domestic beauty brands represented by PROYA in digital marketing, explore their successful factors and development paths, and provide theoretical reference and practical guidance for the future development of domestic beauty brands. Logistic regression model is used to analyze the influencing factors of customer purchase decision, and chi-square test is used to analyze and study PROYA's product categories, so as to study the current situation and future development trend of domestic cosmetics industry under the digital economy. Finally, combined with the conclusions drawn from the model analysis, suggestions and prospects are put forward.
Keywords: Digital economy; Cosmetic industry; Logistic regression analysis; Chi-square test

1 INTRODUCTION

With the vigorous development of the digital economy and the transformation of consumers' lifestyles, the domestic cosmetics industry is undergoing an unprecedented transformation. The popularity of digital technology and the rise of the Internet have completely changed the pattern and operation mode of the traditional cosmetics industry, and promoted the rise and growth of domestic cosmetics brands in the market.

This paper aims to explore the marketing strategy of domestic cosmetics from the perspective of digital economy. In foreign countries, the theory of marketing and brand marketing is relatively perfect and mature. With the development of the Internet, foreign scholars have begun to study network marketing. Philip Kott, a famous marketing management expert, has made a detailed analysis of the traditional 4P marketing theory, and further developed and improved it, and put forward the concept of enterprise marketing management to adapt to the Internet era. Hyeon-Joo[1] studied the impact of marketing 4P on consumers' purchase motivation, satisfaction and loyalty, aiming to analyze the impact of cosmetics marketing mix on consumers' purchase behavior. Syawaluddin[2] used multiple linear regression model and coefficient of determination to study the influence of social media, e-marketing and product quality on the process of purchasing natural cosmetics. Lee[3] analyzed the relationship between consumers' desire to consume cosmetics and the influence of purchase motivation that promotes consumption desire on purchase intention. In recent years, China's domestic makeup industry has flourished and the market scale has continued to expand. Many local brands have successfully won the trust of consumers by continuously improving product quality, strengthening R & D innovation and clever marketing. The rise of online channels and the gradual development of the international market have also given domestic cosmetics a broader space for development. Technological progress, the trend of consumer upgrading and the brand's international development strategy have jointly promoted the prosperity of the domestic cosmetics industry, making it gradually emerge in the global market. However, in the traditional beauty market, the beauty brands of the United States, Japan, South Korea, France and other countries have long monopolized the domestic market. Many foreign beauty brands have established a long history and strong brand awareness in the Chinese market, and have rich experience in market promotion and channel expansion. Through ingenious marketing strategies and sales channels adapted to the Chinese market, they better meet the diversified needs of Chinese consumers. At the same time, the international vision and cross-cultural design of foreign brands have also played an active role in attracting young consumers, keeping them competitive in the Chinese market. Therefore, in the face of the new pattern of digital economy, domestic beauty brands are facing many challenges and competitive pressures. In the field of digital marketing, how to effectively use big data analysis, social media communication, content marketing and other means to enhance brand exposure and influence has become a key issue that domestic beauty brands need to think about and solve.

As a well-known local beauty brand in China, PROYA was founded in 1931. After nearly a century of development, it has become one of the leading brands in China's beauty industry. With a wide range of product lines in the fields of skin care, make-up, personal care and so on, PROYA is committed to providing high-quality, safe and reliable products, which are favored by consumers. At present, PROYA has a total of thousands of products, and has set up business networks in major department stores, cosmetics stores and large supermarket chains. It has successfully established a

multi-brand, multi-type, multi-channel and multi-mode operation system, and its comprehensive strength is in the leading position in the cosmetics industry.

From the perspective of digital economy, this study aims to deeply analyze the strategies and practices of domestic beauty brands represented by PROYA in digital marketing, explore their successful factors and development paths, and provide theoretical reference and practical guidance for the future development of domestic beauty brands. Through the in-depth study of domestic cosmetics marketing strategy, it is expected to inject new vitality and impetus into the innovation and development of the industry.

2 MODEL CONSTRUCTION

2.1 Analysis of Influencing Factors of Customer Purchase Decision -- Logistic Regression Model

Logistic regression, also known as Logistic regression analysis, is a generalized linear regression analysis model. Binary Logistic regression analysis is a kind of Logistic regression analysis, which is often used in data mining and economic forecasting. Sometimes it is also used to explore the risk factors of disease occurrence, and then calculate the probability of disease occurrence. In order to more effectively quantify and calculate the influencing factors of consumers' purchase of PROYA products, we applied binary logistic regression analysis to quantify the variables.

2.1.1 Single factor analysis of variance

Based on the existing data, six variables that are most likely to affect consumers' purchase of PROYA products are selected: gender, age, city, occupation, average monthly income, and monthly cosmetics expenses. Before analyzing the influencing factors of consumers' purchase of PROYA products, we first need to process the city variables, divide the cities into five city levels according to different regions, and then carry out Logistic single factor regression analysis on the six indicators that affect consumers' purchase of PROYA products.

Table 1 Single Factor Analysis

Factor		Purchase group (n=304)	Non-purchase group (n=108)	χ^2	p-value
Gender	Male	76 (25%)	42 (38.89%)	3.961	0.041
	Female	228 (75%)	66 (61.11%)		
Age	Less than 18 years old	8 (2.63%)	4 (3.7%)	0.851	0.837
	18 to 25 years old	182 (59.87%)	70 (64.81)		
	26 to 35 years old	92 (30.26%)	26 (24.07%)		
	36 to 45 years old	22 (7.24%)	8 (7.41%)		
Occupation	Students	134 (44.08%)	50 (46.3%)	0.640	0.958
	On-the-job personnel	120 (39.47%)	40 (37.04%)		
	Freelancers	40 (13.16%)	16 (14.81%)		
	Self-employed	8 (2.63%)	2 (1.85%)		
	Entrepreneurs	2 (0.66%)	0 (0%)		
Urban hierarchy	First-tier cities	76（25%）	26（24.07%）	4.165	0.384
	New first-tier cities	94（30.92%）	28（25.93%）		
	Second-tier cities	74（24.34%）	28（25.93%）		
	Third-tier cities	40（13.16%）	24（22.22%）		
	Fourth-tier city	20（6.58%）	2（1.85%）		
Monthly average income	Under 3000	114（37.5%）	114（37.5%）	8.985	0.032
	3000-5000	80（26.32%）	80（26.32%）		
	5000-7000	64（21.05%）	64（21.05%）		
	More than 7000	46（15.13%）	46（15.13%）		
Monthly cosmetics expenses	Under 200	104（34.21%）	104（34.21%）	8.566	0.046
	200-500	116（38.16%）	116（38.16%）		
	500-1000	66（21.71%）	66（21.71%）		
	More than 1000	18（5.92%）	18（5.92%）		

Table 1 shows the results of single factor analysis showed that 304 of the 412 consumers who purchased PROYA products were set as the purchase group, and the remaining 108 were set as the non-purchase group. The results of the above table show that there is no significant correlation between age, urban hierarchy, occupation and whether to buy PROYA products ($P > 0.05$). Consumers with different genders, monthly average income, and monthly cosmetics

expenses had statistically significant differences in purchases (P < 0.05). The statistical results showed that the proportion of female consumers, monthly average income of 3000-5000 yuan, 5000-7000 yuan, more than 7000 yuan, monthly cosmetics expenditure of 200-500 yuan and monthly cosmetics expenditure of 500-1000 yuan in the purchase group was significantly higher than that in the non-purchase group, and the difference was statistically significant (P < 0.05).

2.1.2 Multiple logistic regression

Sex, monthly average income and monthly cosmetics expenses were selected as the influencing factors for purchasing PROYA products. The factors affecting customers' purchase of PROYA products in the single factor analysis were set as independent variables, and the following definitions and assignments were included in the Logistic multivariate regression analysis.

It is assigned as follows Table 2:

Table 2 Variable Assignment

Variable	Influencing Factors	Variable Assignment
X_1	Gender	Male = 1; Female = 2
X_2	Monthly average income	'Under 3000' = 1;'3000-5000' = 2;'5000-7000' = 3;'more than 7000' = 4
X_3	Monthly cosmetics expenses	'Under 200' = 1;'200-500' = 2;'500-1000' = 3;'more than 1000' = 4
Y	Purchase experience	No = 0; yes = 1

Multivariate Logistic regression analysis was performed to screen out the influencing factors that affect consumers' purchase of PROYA products, and the following Table 3:

Table 3 The Recommended Fonts

Variable		β	SE	χ^2	P	OR	95% CI Value	
							Lower Limit	Upper Limit
Gender	Female	0.952	0.624	5.328	0.027	2.592	1.763	8.810
Monthly average income	3000-5000	0.318	0.618	4.265	0.047	1.088	1.017	2.442
	5000-7000	0.082	0.649	4.016	0.038	1.786	1.304	3.878
	More than 7000	0.007	0.658	5.094	0.049	1.993	1.274	3.603
Monthly cosmetics expenses	200-500	0.062	0.789	4.036	0.037	1.064	1.027	5.001
	500-1000	0.631	0.777	4.659	0.017	1.880	1.410	8.625
	More than 1000	0.882	0.816	5.169	0.028	2.417	1.488	11.965

According to the further analysis of the above table, it is concluded that women's purchase of PROYA products is about 2.6 times that of men (OR = 2.592). The main reason is that PROYA products are mainly for women's makeup products. In our daily life, men rarely buy cosmetics, and many men buy cosmetics for their girlfriends or mothers. Consumers with an average monthly income of 3000-5000 yuan buy PROYA products about 1.1 times that of less than 3000 yuan (OR = 1.088). Consumers with an average monthly income of 5000-7000 yuan per month to buy PROYA products is about 1.8 times that of 3000 yuan or less (OR = 1.786); consumers with an average monthly income of more than 7000 yuan buy PROYA products about 2.0 times that of less than 3000 yuan (OR = 1.993). Consumers who spend 200-500 yuan per month on cosmetics are about 1.1 times more likely to buy PROYA products than those who spend less than 200 yuan (OR = 1.064); consumers who spend 500-1000 yuan per month on cosmetics to buy PROYA products is about 1.9 times that of 200 yuan or less (OR = 1.880); consumers who spend more than 1,000 yuan per month on cosmetics buy PROYA products about 2.4 times as much as those who spend less than 200 yuan (OR = 2.417). The main reason is that PROYA products are mid-to-high-end products. Consumers with high monthly average income will be more inclined to buy PROYA, while consumers with low monthly average income will tend to choose brands with low prices.

2.2 Product Category Analysis--Chi Square Test

Different products have different effects, and different ages have different skin states. Consumers will choose products according to their own skin states when buying PROYA products. Therefore, age has a certain relationship with the types of products purchased by consumers. At the same time, at different ages, we are engaged in different occupations,

so the occupation and the types of products purchased by consumers also have a certain relationship. Next, we will explore the relationship between age, occupation and PROYA's product categories.

2.2.1 Age and product categories

Before exploring the relationship between age and PROYA product categories, we first need to conduct a chi-square test to determine the relationship between the two, and get the following Table 4:

Table 4 Chi-square Test

	Value	Degree of freedom	Asymptotically significant
Pearson chi-square	58.579	18	0.000
Number of effective cases	898		

From the above chi-square test analysis table, the chi-square value is 58.579, and the p value is 0.000 less than 0.05, indicating that there is a significant correlation between age and product type variables[4]. The following correspondence analysis is performed to obtain the corresponding correspondence diagram as follows:

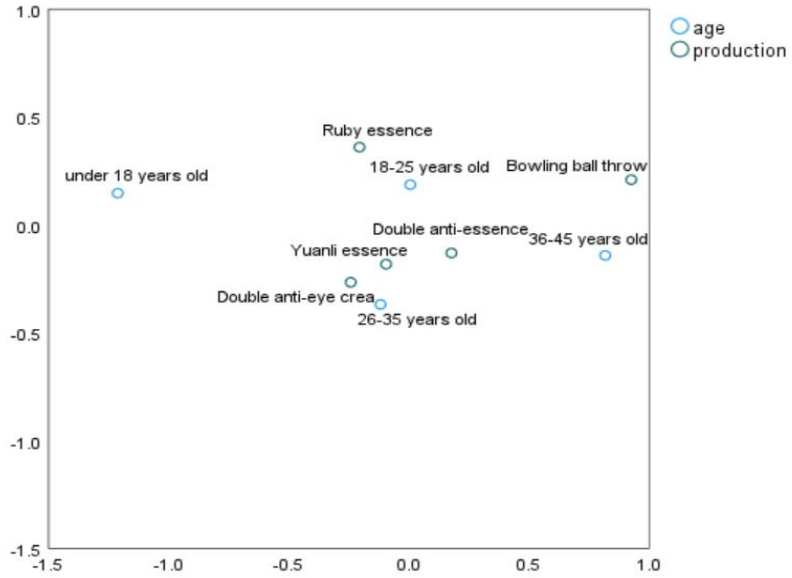

Figure 1 Correspondence Diagram of Age and Product Category

Figure 1 shows that consumers under the age of 18 hardly buy PROYA products and are not very interested in them. The main reason is that the people under the age of 18 are basically in high school. For students, the college entrance examination is the most important. All the time spent on learning, there is little time to understand cosmetics and dress up yourself. Consumers aged 18-25 and 26-35 mainly buy ruby essence, Yuanli essence, double-antibody essence, double-antibody eye cream and have great interest in these types. The main reason is that most of these people are in the era of university, graduate or work. More and more consumers will pay attention to their skin condition. Makeup can improve skin appearance, modify facial features, make people feel more confident, and help to cope with various social occasions and challenges. Consumers aged 36-45 tend to buy bowling throws or are very interested in such products.

2.2.2 Occupation and product categories

Before exploring the relationship between occupation and PROYA's product categories, we first need to conduct a chi-square test to determine the relationship between the two, and get the following:

Table 5 Chi-square Test

	Value	Degree of freedom	Asymptotically significant
Pearson chi-square	52.157	24	0.000
Number of effective cases	898		

Table 5 shows that the chi-square value is 52.157, and the p value is 0.000 less than 0.05, indicating that there is a significant correlation between occupation and product type variables. The following correspondence analysis is carried out, and the corresponding correspondence diagram is as follows:

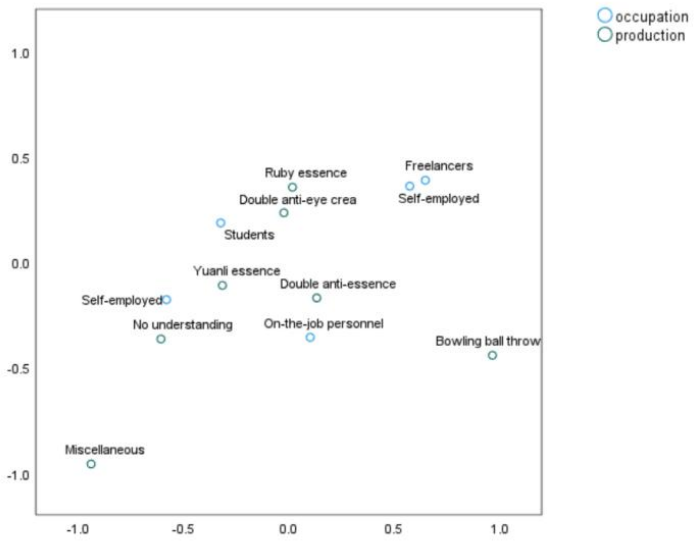

Figure 2 Occupation and Product Category Corresponding Diagram

Figure 2 shows that students mainly buy double-anti-eye cream and have great interest in it. The main reason is that students are heavy in their studies and often study late into the night. The double-antibody eye cream mainly improves the dark yellow around the eyes, contains the same formula of the double-antibody series, and adds ingredients to brighten the skin around the eyes. It helps to improve many eye problems such as dark eyes, blister eyes, dry lines and fine lines. Therefore, double-anti-eye cream is very popular among students. For on-the-job personnel, the working environment of different occupations is different, which will affect our skin condition. The indoor office crowd will be better than the outdoor office crowd in the skin condition. When choosing the product category, the main purchase is the double-antibody essence. Of course, some people choose to buy the source essence and the ruby essence and have interest. For entrepreneurs, the entrepreneurial stage is very hard, there is little time to understand cosmetics, and most of the entrepreneurial male population, they often choose to repair the source of the series of essence. Bowling throw is mainly used for deep moisturizing, repairing skin, improving skin discomfort symptoms such as redness. Long-term use can improve skin lock water, repair damaged lipid barrier, and strengthen and toughen skin. Although this product is also suitable for some of our people, it is not very popular due to the better effect and more efficacy of other essences.

3 CONCLUSION

3.1 Research Conclusions

3.1.1 Influencing factors of customer purchase decision

Through Logistic regression analysis, the risk factors that affect consumers' purchase of PROYA products are gender, average monthly income, and monthly cosmetics expenses. It is found that female consumers buy PROYA products about 2.6 times that of men; consumers with an average monthly income of 3000-5000 yuan, 5000-7000 yuan and more than 7000 yuan purchase PROYA products at 1.1 times, 1.8 times, and 2.0 times, respectively, below 3000 yuan. Consumers who spend 200-500 yuan, 500-1000 yuan and more than 1,000 yuan per month on cosmetics buy PROYA products 1.1 times, 1.9 times, and 2.4 times as much as those who spend less than 200 yuan.

3.1.2 The product range of PROYA

Through the analysis of product types, it can be seen that there is a correlation between age, occupation and PROYA product types. Consumers under the age of 18 almost do not buy PROYA products and are not very interested in them. 18-25 years old and 26-35 years old students and in-service personnel mainly buy ruby essence, source power essence, double anti essence, double anti eye cream these products and have great interest in these kinds; consumers aged 36-45 tend to buy bowling throws; for entrepreneurs, they often choose the essence of the source of the repair series.

3.2 Suggestion and Prospect

3.2.1 Suggestion

Strategies to cope with market changes. PROYA Company should continue to pay attention to market dynamics, understand changes in consumer demand, and grasp the development trend of the industry. Through market research, the company can better understand market demand and provide a basis for product development and market strategy formulation[5]. According to market changes, PROYA should adjust its product strategy in time, including product line update, product quality improvement and product pricing optimization. At the same time, the company should also pay attention to the application of new technologies, through technological innovation to enhance the competitiveness of

products. In the face of market changes, PROYA needs to continue to expand its sales channels, including online and offline channels. Through diversified sales channels, the company can cover a wider range of consumer groups and increase market share.

Optimize customer service experience. Improve the professional ability of the customer service team, and regularly provide product knowledge, communication skills and customer service attitude training for the customer service team to ensure that they have professional and efficient service capabilities. In order to meet the needs of different consumers, PROYA can provide customer service support through various channels such as telephone, email, online customer service, and social media to ensure that consumers can get help quickly and easily, establish an efficient customer service process, and ensure that consumers' consultations and problems can be quickly responded to and resolved. Encourage consumers to provide feedback, understand consumers' evaluation of customer service through satisfaction surveys, online evaluations, etc., and make corresponding improvements based on feedback.

Improve brand awareness. Use social media platforms for extensive and accurate brand promotion, which can include publishing new developments about products, sharing skin care knowledge and skills, interacting with consumers, holding online activities, etc. In cooperation with influential online celebrities, bloggers, stars, etc., they can endorse or recommend PROYA's products. Through their fan effect, they can expand brand exposure and improve brand awareness. Advertisements are placed on major media platforms, including television, the Internet, outdoors, etc., so that more potential consumers can understand the PROYA brand. At the same time, we should pay attention to the innovation and attractiveness of advertising content to attract target consumers.

Optimize sales channel. For offline sales channels, optimize offline experience, and provide high-quality shopping experience in physical stores, such as professional product consulting, skin testing, makeup service, etc., so that consumers can feel the professionalism and intimate service of the PROYA brand. For online sales channels, PROYA can optimize online stores and improve user experience[6]. For example, optimizing store page design, improving page loading speed, and improving shopping process. In addition, it can also strengthen the construction of customer service team, improve the response speed and service quality of customer service, so that consumers can feel better service in the shopping process.

3.2.2 Prospect

The domestic cosmetics brand has made significant development and progress in the past few years, but the road ahead is still full of challenges and opportunities. Domestic cosmetics brands need to continue to strengthen brand building and positioning, enhance brand awareness and reputation. Through the unique brand story, high-quality product experience and innovative marketing strategy, establish their own brand image and establish a deep emotional connection with consumers. With the increasing diversification of consumer demand for cosmetics, domestic cosmetics brands need to continuously innovate and optimize products to meet the personalized needs of consumers. Strengthen product research and development, improve product quality and technical content, is the key to enhance brand competitiveness. At the same time, it is necessary to actively expand online and offline channels to improve the coverage and accessibility of products. It is also necessary to continuously innovate marketing strategies and use new marketing methods such as social media and live delivery to interact and communicate with consumers more deeply[7].

In short, domestic cosmetics brands need to constantly innovate, optimize products, expand channels, practice the concept of green environmental protection and actively expand the international market in the future development, so as to enhance brand competitiveness and market position. At the same time, it is also necessary to maintain a keen insight into market changes and consumer demand, adjust strategies in a timely manner, seize opportunities, and meet challenges.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Seo H J, Sheen Y S, Oh S Y. The Effects of Cosmetic Industry Marketing-mix Strategies on Consumers' Purchasing Behaviors. J Korea Soc Beauty Art, 2016, 17(3): 131-147.

[2] Syawaluddin S, Joni J, Erwin E. Influence of social media advertising, e-marketing and product quality on the process of purchasing nature cosmetics. International Journal of Research in Business and Social Science, 2019, 8(5): 316-321.

[3] Lee J, Choi E J. A study on the effect of cosmetic consumption desire on purchase intention of customized cosmetics: purchasing motivation as a mediating effect. Journal of Wellness, 2020, 15: 43-58

[4] McHugh M L. The chi-square test of independence. Biochemia medica, 2013, 23(2): 143-149.

[5] Rezvani M, Ghahramani S, Haddadi R. Network marketing strategies in sale and marketing products based on advanced technology in micro-enterprises. International Journal of Trade, Economics and Finance, 2017, 8(1): 32-37.

[6] Tian J. Research on the Marketing Strategy of Beauty Brands in the Background of Social Media. International Conference on Economic Management and Cultural Industry. Atlantis Press, 2022: 1267-1277.

[7] Zuhria K H, Ratnaningtyas S. Integrated Communication Strategy for Awareness of Emotional Marketing Campaign for Beauty Brand. Asian Journal of Research in Business and Management, 2023, 5(2): 78-90.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3015

LSTM MODEL ENHANCED BY KOLMOGOROV-ARNOLD NETWORK: IMPROVING STOCK PRICE PREDICTION ACCURACY

XiaoXuan Yao
School of Mathematics and Statistics,Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: xiaoxuan.yao@qq.com

Abstract: This study addresses the accuracy limitations of traditional LSTM models in stock price prediction by proposing an innovative hybrid model, the LSTM-KAN model. Combining the classical Long Short-Term Memory (LSTM) network with the Kolmogorov-Arnold Network (KAN), this model aims to enhance the performance of the LSTM model in predicting complex financial time series by leveraging the highly nonlinear expressive power of KAN. Through empirical analysis of historical stock data, a comparative study is conducted to examine the differences between the LSTM-KAN model and the basic LSTM model in terms of prediction error, stability, and generalization capability. The results demonstrate that the LSTM-KAN model significantly reduces prediction errors in most cases, improving prediction accuracy and providing new perspectives and tools for stock market analysis.
Keywords: LSTM; Kolmogorov-Arnold Network; Stock price prediction; Time series analysis; Nonlinear models

1 INTRODUCTION

As global financial markets become increasingly complex and volatile, accurate stock price prediction has become a critical factor for investor decision-making. Traditional prediction methods, such as technical analysis and fundamental analysis, often fall short when faced with high-dimensional, nonlinear financial time series data that contain significant noise. In recent years, the rapid development of machine learning and deep learning technologies has brought new breakthroughs to this field. Among these, Long Short-Term Memory (LSTM) networks, a special form of recurrent neural networks, have been successfully applied to various prediction problems, including stock price prediction, due to their excellent performance in handling sequential data.

However, while LSTM excels at capturing long-term dependencies, its prediction accuracy and generalization ability still have room for improvement in the highly nonlinear financial market environment. The Kolmogorov-Arnold Network (KAN), a novel neural network based on the Kolmogorov-Arnold representation theorem, offers strong nonlinear expressive power with relatively low model complexity, providing a new approach to solving such complex problems.

Currently, scholars at home and abroad have extensively explored the application of LSTM in stock price prediction and have achieved certain results. Some studies have attempted to combine LSTM with other machine learning algorithms or deep learning models to improve prediction accuracy, such as ensemble learning and convolutional neural networks (CNN)[1]. Despite this, research on combining KAN with LSTM for stock price prediction remains relatively sparse, especially in terms of model construction, parameter optimization, and performance evaluation, leaving considerable room for further exploration.

Therefore, this study aims to design and implement an innovative hybrid model, the LSTM-KAN model, to integrate the powerful memory capability of LSTM with the nonlinear expressive advantages of KAN, overcoming the limitations of a single model in handling complex financial time series data. Specific research objectives include: constructing the LSTM-KAN hybrid model framework, exploring the effective integration mechanism of the two models; validating the improvement in prediction accuracy, stability, and generalization capability of the LSTM-KAN model using historical stock data; analyzing the specific impact of KAN's introduction on the performance of the LSTM model, and exploring the underlying theoretical basis.

2 THEORETICAL FOUNDATIONS

2.1 Long Short-Term Memory Networks (LSTM)

LSTM is a special type of recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber in 1997. It aims to address the vanishing and exploding gradient problems commonly encountered by traditional RNNs when dealing with long-term dependencies. LSTM introduces memory cells (cell states), input gates, forget gates, and output gates to control the flow of information, allowing the model to effectively learn patterns in long sequence data. Memory cells can accumulate and retain important information over long periods, while the gate mechanisms control the reading, writing, and forgetting of information, thus enhancing the model's ability to handle sequential data[2].

2.2 Kolmogorov-Arnold Representation Theorem

The Kolmogorov-Arnold representation theorem is a significant result in mathematics, proposed by Andrey Kolmogorov and Vladimir Arnold. It asserts that any continuous multivariable function can be composed of a series of simple functions, formally represented as nested single-variable functions. Equation (1) represents the mathematical formulation of the Kolmogorov-Arnold representation theorem. This theory inspired the design of the Kolmogorov-Arnold Network (KAN), where each node can be regarded as a highly nonlinear mapping. The network approximates complex function relationships by combining these mappings. The advantage of KAN lies in its potential for efficient expression and the compactness of its model structure. KAN is based on a supervised learning task aimed at approximating a function f, which maps the inputs x of all data points to their outputs y. This method uses the Kolmogorov-Arnold theorem to decompose any multivariate function into a series of single-variable functions and summations. The equation indicates that for each input dimension x_p, there is a univariate function h_p that aggregates the outputs of these univariate functions, as expanded in Equation (2).

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}\left(x_p\right) \right) \tag{1}$$

$$f(x) = \sum_{i_{L-1}=1}^{n_{L-1}} \phi_{L-1,i_L,i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \cdots \left(\sum_{i_2=1}^{n_2} \phi_{2,i_3,i_2} \left(\sum_{i_1=1}^{n_1} \phi_{1,i_2,i_1} \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0}\left(x_{i_0}\right) \right) \right) \right) \cdots \right) \tag{2}$$

2.3 Structure and Characteristics of KAN

KAN's core lies in its structural design, which differs from traditional neural networks by treating the activation function as a part of the model for learning. This means that each connection in the network not only has weight parameters but also has an activiation function that is trainable, allowing the network to automatically discover the most suitable non-linear transformation. This provides greater flexibility in handling non-linear problems, especially in scenarios with complex data distributions or highly non-linear relationships. The left side of Figure 1 shows the activation symbols flowing through the network, while the right side illustrates the activation function parameterized as B-splines, enabling switching between coarse and fine-grained grids[3].

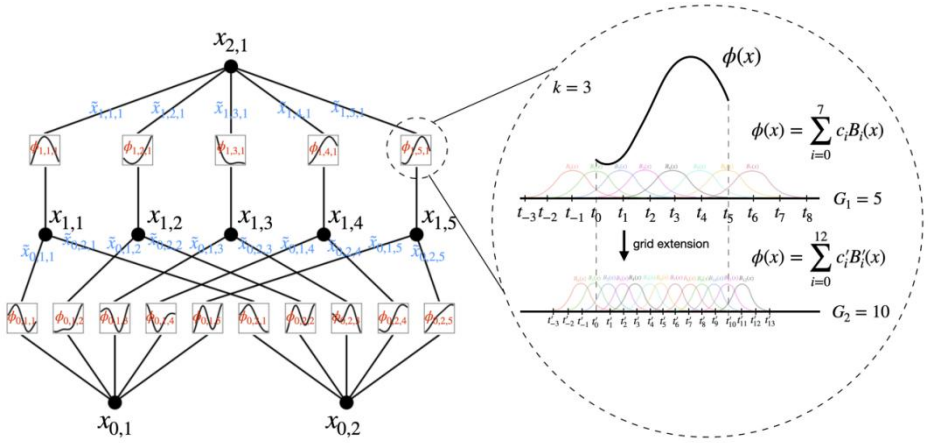

Figure 1 KAN Structure Diagram

2.4 Overview of Stock Price Prediction Methods

Stock price prediction is a significant topic in financial engineering and econometrics, involving various statistical and machine learning methods. Early methods primarily included time series analysis (such as ARIMA models), moving averages, and exponential smoothing. With the increase in computational power and data volume, machine learning methods have become increasingly popular, including support vector machines (SVM), random forests, and neural networks. In recent years, deep learning techniques, particularly RNNs and their variants (such as LSTM and GRU), have gained attention for their ability to better capture sequential data patterns[4].

3 CONSTRUCTION OF THE LSTM-KAN MODEL

3.1 Model Design and Parameter Configuration

The core design of the LSTM-KAN model is to combine the memory capabilities of LSTM in handling sequential data with the advantages of KAN in expressing complex nonlinear relationships. The LSTM layers are responsible for capturing long-term dependencies in time series, while the KAN layers further refine this information by utilizing flexible basis function activation and piecewise polynomial weights to adapt to the highly nonlinear patterns in stock price prediction. The model's parameter configuration includes input size, hidden size, number of layers, output size, dropout ratio, and whether to use batch-first mode for the LSTM. Additionally, the KAN layer's configuration involves grid size, the order of piecewise polynomials, scaling noise, and scaling of activation functions. These parameters collectively determine the model's complexity and adaptability.

3.2 Code Implementation

In the code implementation, the LSTM-KAN model class is first defined, inheriting from nn.Module. During model initialization, the LSTM layer is set up, and the KAN model (via the KAN class) is innovatively included as a subsequent processing layer for the LSTM output, replacing the conventional fully connected layer. In the forward propagation function, the final hidden state of the LSTM is used as input to the KAN layer, designed to leverage KAN's characteristics to further enhance nonlinear expression in predictions[5].

3.3 KAN Layer Implementation

The KAN layer's implementation is based on piecewise polynomial weights and optional independent scaling factors. It internally manages grid generation, weight initialization, and regularization loss calculation. The computation of piecewise polynomial weights considers grid steps, scaling noise, and the influence of the base activation function, achieved through methods like curve2coeff. Additionally, the KAN layer supports the computation of regularization losses, including L1 regularization and entropy regularization terms, to prevent overfitting and promote the model's generalization capabilities.

3.4 Model Flexibility and Future Research

By combining LSTM with KAN, the LSTM-KAN model can handle both the long-term dependencies of time series data and the nonlinear complexities inherent in financial data, potentially achieving lower prediction errors in stock price forecasting tasks. The model's flexibility and customizability, such as adjusting grid size and selecting activation functions, provide extensive room for adaptation to different market characteristics and datasets. Future research can further explore model performance under different parameter settings and how the depth and complexity of the KAN layer affect prediction accuracy and computational efficiency.

4 EXPERIMENT DESIGN AND DATA PROCESSING

4.1 Experiment Objectives

The aim of this chapter is to evaluate the performance of the LSTM-KAN model compared to the pure LSTM model in stock price prediction tasks. The effectiveness of the KAN layer in enhancing nonlinear expression capabilities will be assessed. The study uses stock data from the Shanghai Stock Exchange Composite Index (SSE) spanning from January 5, 1998, to June 2, 2020, to explore the models' potential in predicting actual financial data.

4.2 Data Preprocessing

First, daily trading data of the SSE within the specified date range were collected, including eight variables: opening price, closing price, highest price, lowest price, price change, price change percentage, trading volume, and trading amount. Outliers and missing data were removed to ensure data integrity. All numerical features were normalized to the same scale to facilitate model training. A fixed random seed was used to ensure reproducibility of the experiments. The dataset was split into a training set (80%) and a test set (20%). The input for the models consists of data from the previous five days (sequence_length=5), predicting the closing price on the sixth day.

4.3 Model Construction and Training

A basic LSTM model was constructed with an input dimension (input_size) of 8 and an output dimension (output_size) of 1 (i.e., the predicted closing price), using a five-day data sequence as input. Based on this, an LSTM model with an integrated KAN layer was constructed. The parameter configuration for the KAN layer, as described in previous chapters, optimizes its flexibility and efficiency in nonlinear expression. Both models were trained for 100 epochs, recording the loss for each epoch, and optimized using the Adam optimizer.

4.4 Model Evaluation

The performance of the models was evaluated primarily through Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics directly reflect the deviation between the predicted and actual values. The RMSE and

MAE of the LSTM and LSTM-KAN models on the test set were compared to analyze the performance improvement brought by the KAN layer. Additionally, the trend of loss during the training process was observed to assess the models' convergence and the risk of overfitting.

5 RESULTS ANALYSIS AND DISCUSSION

5.1 Summary of Experiment Results

In this study, we trained both a basic LSTM model and an LSTM-KAN model, using trading data from the Shanghai Stock Exchange Composite Index (SSE) from January 1998 to June 2020 to evaluate their predictive performance. Data preprocessing included normalization and splitting the dataset into 80% training and 20% testing sets. The models used the previous five days of trading data to predict the closing price on the sixth day. During training, we recorded the loss for each epoch. After 100 epochs, we assessed the models' prediction accuracy by calculating the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on the test set.

5.2 Comparison of Results

The loss variations over epochs for both LSTM and LSTM-KAN models are plotted, with blue representing the LSTM model and yellow representing the LSTM-KAN model, as shown in Figure 2. Initially, the LSTM model's loss decreases rapidly, but from the fourth epoch onwards, the LSTM-KAN model's loss decreases faster, showing better performance. A comparison of the LSTM and LSTM-KAN predictions against actual values on the test set is illustrated in Figure 3. The blue line represents the LSTM model, and the yellow line represents the LSTM-KAN model. When zoomed in, it is evident that the LSTM-KAN model predictions are closer to the actual values. As shown in Table 1, the basic LSTM model stabilizes during training, with a final RMSE of 0.065467 and an MAE of 0.052623 on the test set. This indicates that the LSTM model can learn the patterns in historical data relatively well but may have limitations when dealing with complex nonlinear relationships. The LSTM model combined with the KAN layer also shows good convergence characteristics during training. However, compared to the basic LSTM model, the LSTM-KAN model reduces the test set RMSE to 0.008226 and the MAE to 0.005742, demonstrating a significant performance improvement. This confirms the KAN layer's ability to handle nonlinear financial data, particularly in capturing complex price fluctuation patterns.

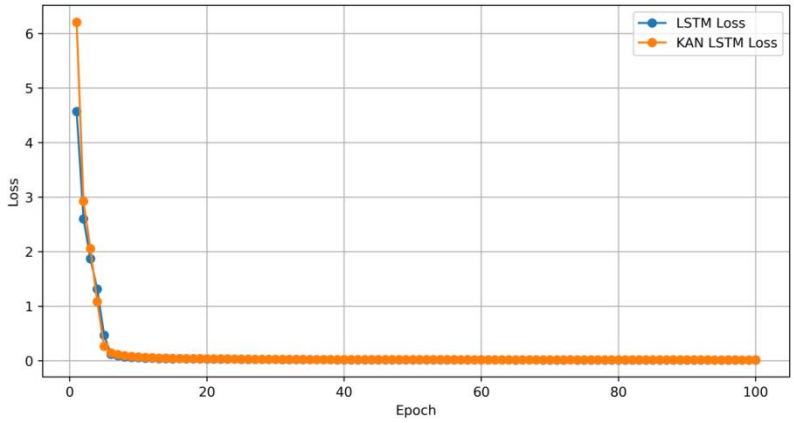

Figure 2 Loss Comparison

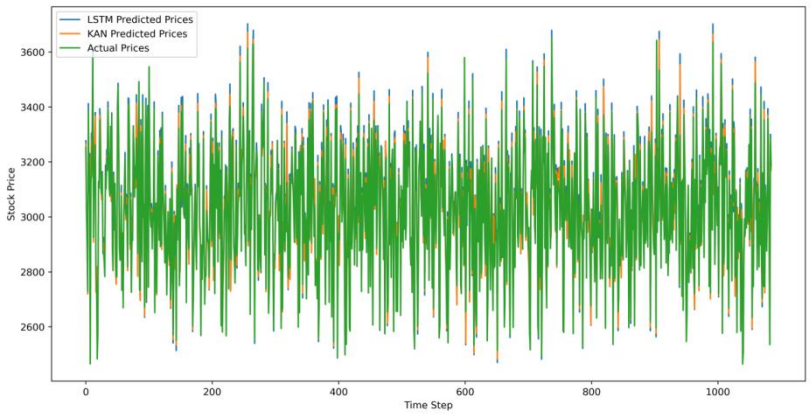

Figure 3 Predicted vs Actual Stock Prices

Table 1 Metrics Comparison

Metrics	LSTM	LSTM-KAN
RMSE	0.065467	0.008226
MAE	0.052623	0.005742

5.3 Discussion

In analyzing the results, we found that the LSTM-KAN model significantly reduced prediction errors on the test set, indicating that the introduction of the KAN layer indeed enhanced the model's nonlinear expression capability. However, it is important to note that the increased training complexity of the model might lead to a risk of overfitting. By observing the loss curves during training, we found that the LSTM-KAN model exhibited strong generalization ability, maintaining good performance on the test set despite the increased model complexity. Additionally, the parameter configurations of the KAN layer, such as grid size and the order of piecewise polynomials, significantly impact the model's performance, requiring further research and optimization. Finally, while the experimental results performed well on SSE data, the effectiveness might be influenced by the dataset, necessitating validation on other markets or datasets.

6 CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

This study explored the potential of the Kolmogorov-Arnold Network (KAN) to enhance prediction accuracy by comparing the performance of a basic LSTM model and an LSTM-KAN model in the task of stock price prediction. Through empirical analysis of historical data from the Shanghai Stock Exchange Composite Index (SSE) from January 1998 to June 2020, we concluded the following:

1) The LSTM-KAN model outperformed the traditional LSTM model in stock price prediction, significantly reducing prediction errors and improving prediction accuracy.
2) The introduction of the KAN layer enhanced the model's nonlinear expression capability, allowing it to better capture complex stock price fluctuation patterns, especially in handling nonlinear and complex financial time series data.
3) Although the training complexity of the LSTM-KAN model increased, reasonable parameter configurations and training strategies enabled the model to maintain good generalization ability, avoiding the risk of overfitting.

6.2 Future Prospects

Despite the positive results of this study, several future research directions deserve further exploration:

1) Application to Different Markets and Financial Products: Apply the LSTM-KAN model to different markets or various types of financial products to test its universality and robustness across different datasets.
2) Optimization of KAN Layer Parameters: Further study the optimization strategies for KAN layer parameters to find the best parameter combinations, enhancing the model's efficiency and prediction accuracy.
3) Incorporation of External Information: Explore the inclusion of additional macroeconomic indicators, news sentiment analysis, and other external information to improve the model's prediction accuracy and practical applicability.
4) Development of Real-time Prediction Systems: Develop real-time prediction systems to enable the model to adapt promptly to market dynamics, which is crucial for high-frequency trading and real-time risk management.
5) Improving Model Interpretability: Work on enhancing the interpretability of the model to make it more understandable and actionable for end-users.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

FUNDING

This article was supported by the project titled "Identification, Measurement, and Governance of Relative Poverty in Rural Guangxi from the Perspective of Rural Revitalization" (Project No. 23BTJ001) funded by the Guangxi Philosophy and Social Science Office. The project titled "Reform of the Evaluation Model of Physical Education Entrance Examination in Guangxi Based on Data Mining" (Project No. 2022ZJY2311) funded by the Guangxi Zhuang Autonomous Region Admissions Examination Institute.

REFERENCES

[1] Chen S, Ge L. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 2019, 19(9): 1507-1515.
[2] Kim T, Cho S. Predicting residential energy consumption using CNN-LSTM neural networks. Energy, 2019, 18272-81.
[3] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, Max Tegmark. KAN: Kolmogorov-Arnold Networks. arXiv preprint arXiv: 2024. 19756.
[4] Wenchao G, Zhigang L, Chuang G, et al. Stock price forecasting based on improved time convolution network. Computational Intelligence, 2022, 38(4): 1474-1491.
[5] Zhang, Zhiping, Wang, et al. Design of financial big data audit model based on artificial neural network. International Journal of System Assurance Engineering and Management, 2021, (prepublish): 1-10.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3016

ANALYSIS AND DECISION-MAKING OF REGIONAL ECONOMIC VITALITY AND ITS INFLUENCING FACTORS

ShaSha Zhang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: sszhangq@qq.com

Abstract: This paper takes Henan Province as the research object, selects GDP as the index of regional economic vitality, and selects the index that affects GDP from the aspects of economic benefit, opening up, population, government regulation, residents' quality of life and enterprise vitality. Firstly, a multiple linear regression model is established to test the multicollinearity of the independent variables. Then, the ridge regression and LASSO regression models are established to correct them, and the multicollinearity problem between independent variables is solved. By comparing and analyzing the two models, the LASSO regression model is the optimal regression model. Secondly, the Holt exponential smoothing model is used to predict the time series of the five variables showing a linear trend in the LASSO regression equation. The simple exponential smoothing model is used to predict the four variables of random fluctuation, and the data of each variable in 2024 are predicted. The predicted value of GDP in Henan Province in 2024 is 61441.55 billion yuan. Finally, some suggestions are put forward for Henan Province from several aspects, so that the economic development of the region can form a virtuous circle, so as to enhance the competitiveness of its regional economy and promote the sustainable development of the economy.
Keywords: Multivariate linear regression model; Ridge regression model; LASSO regression model; Time series prediction; Regional economic vitality

1 INTRODUCTION

1.1 Research Background

China implemented reform and opening up in 1978. In the early stage, the development model of "first rich and then rich" showed great superiority, which made the economy of the eastern coastal areas of China develop rapidly, broke the obstacles brought by human resources and production resources in the process of economic development, and achieved great results.

However, in this development process, the regional economic development gap between the eastern, central, western and northeastern regions of China is increasing, which is not conducive to the rational allocation of various resources, nor to the long-term stable and healthy development of China's economy. The development of China's regional economy is unbalanced. This phenomenon not only hinders the development of China's economic level, but also has a negative impact on social stability and prosperity. After the implementation of the reform and opening up, China has formulated many corresponding development strategies to shorten the imbalance of regional economic development in China, such as the strategy of "western development" and "central rise". Although various development strategies have been formulated, this phenomenon in China has not been well improved, and this phenomenon has become a challenge that China must deal with in economic development in the new era[1].

1.2 Research Meaning

China is a large developing country with numerous economic regions, and the conditions of each region are very different. Therefore, the level and status of regional economic development are very different. How to make the economy of all regions of our country develop, the regional economy becomes more active, and the overall efficiency becomes higher is a difficult problem that China's current development economics needs to explore and overcome[2]. The region and industry are closely linked, they are inseparable, regional economic development will inevitably lead to the development of some industries. Only by building a unified, open, competitive and orderly Chinese market, uniting all regions, complementing each other's advantages and cooperating with each other, can the whole national economy be put on the agenda.

Industrialization, marketization and socialization are not only the historical task that China has to complete, but also the great historical mission entrusted by the era of realizing China's modernization. The task is very difficult. Economic development plays a very important role in this task, which is a very important link that we must go through to complete this dual important historical mission. To realize modernization development requires a complete process, which should be phased, planned and have complete steps. Reform and opening up is the most fundamental feature of China in the new era. China is currently undergoing a transition from a traditional planned economic system to a modern socialist market economy. Therefore, it is of great significance to develop China's regional economy and shorten the gap between regions, not only to realize China's modernization development, but also to maintain the stable development of the motherland, maintain the unity of the motherland, and establish a great image of a powerful country.

2 RESEARCH THOUGHT

2.1 The Overall Framework

Taking a region (or city or province) as an example, this paper determines the indicators that affect the regional economic vitality by consulting the statistical yearbook, establishes the relationship model of the factors affecting the economic vitality, compares the models, selects the optimal model, and puts forward some suggestions to improve the economic vitality of Henan Province according to the model results, and analyzes the influence and function of the regional economic vitality from various angles. Some policy suggestions are put forward to improve the economic vitality of the region (or city or province) discussed in the analysis, so as to improve the economic vitality of the region and strengthen the sustainable development of the economy. The specific framework is shown in Figure 1.

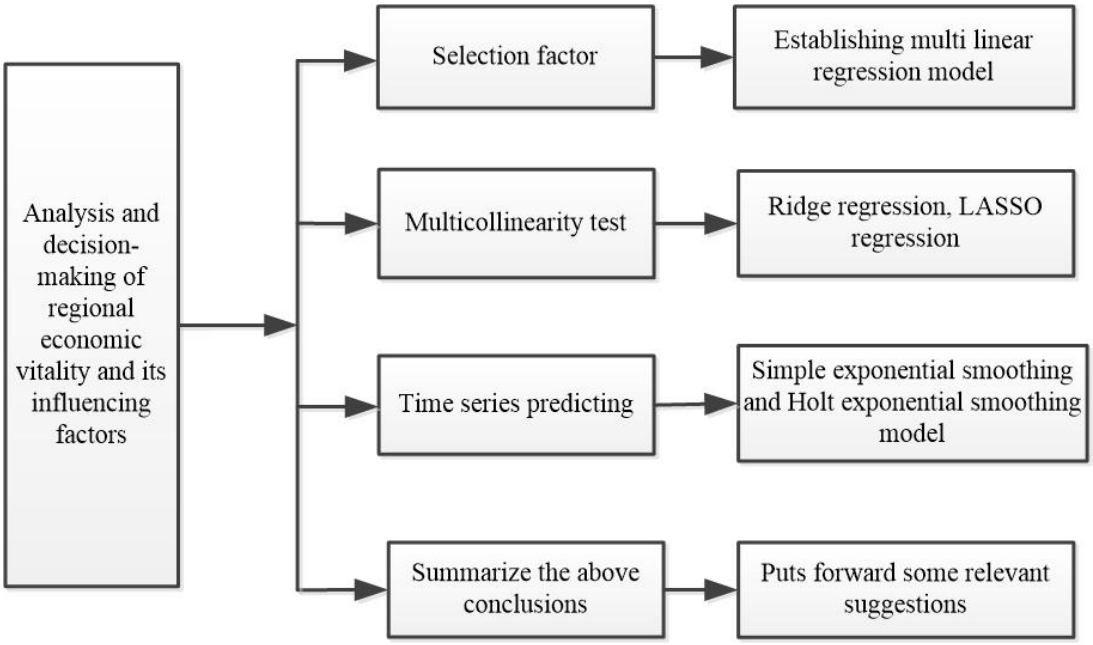

Figure 1 Research Idea Diagram

2.2 Research Specific Ideas

With the continuous advancement of China's modernization drive, the regional economy has developed rapidly under the impetus of various driving forces. Under the stimulation of various preferential policies issued by various regions, effectively improving regional economic vitality has become an irreplaceable part of regional comprehensive competitiveness.

Taking Henan Province as an example, this paper analyzes and studies the economic vitality of Henan Province. Firstly, by consulting the statistical yearbooks of Henan Province over the years, it determines the indicators that affect the regional economic vitality of Henan Province, and explains the rationality and correctness of selecting these indicators. Through the website of the National Bureau of Statistics, the data corresponding to each index of Henan Province from 2009 to 2023 were searched and sorted out. According to the selected indicators, a multiple linear regression model is established. According to the solution results, suggestions are put forward to help improve the economic vitality of Henan Province, and its impact on regional economic vitality is analyzed from various aspects.

From the results of the multiple linear regression model, there is a great correlation between the independent variables. Then the multicollinearity test is carried out on the independent variables, and the variables with multicollinearity are obtained. Then the ridge regression model and the LASSO regression model are established, and the regression equations of each regression model are obtained. According to the regression equation, the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, is calculated, and the comparison diagram between the fitted value and the actual value is obtained. By comparing the two models, the optimal regression model is obtained. Through the time series diagram of the independent variables in the optimal regression equation, the time series prediction of each index is judged by establishing what model, and the standardized data of each index in 2024 is predicted, and the data is substituted into the optimal regression equation, and the predicted value of GDP in Henan Province in 2024 is obtained.

Finally, combined with the analysis results, starting from the actual economic development of Henan Province, comply with the basic principles of regional economic development, break through various constraints that hinder regional economic development, and put forward some constructive policy recommendations for Henan Province, so that the economic vitality of Henan Province forms a virtuous cycle, economic sustainable development, and constantly strengthen the comprehensive competitiveness of the region.

3 MODEL ASSUMPTIONS AND SYMBOLIC DESCRIPTION

3.1 Model Assumptions

It is assumed that the data in the statistical yearbooks of Henan Province over the years are true and effective; assume that the data in the National Bureau of Statistics are true and effective; it is assumed that the selected indicators can fully reflect the changes in regional economic vitality; assume that human error is not considered. All the above assumptions are true.

3.2 Symbol Description

The symbols used in this article, the relevant instructions are shown in Table 1.

Table 1 Symbol Definition

Sign	Definition
β_0	Regression constants
$\beta_j(j=1,2,\cdots k)$	Regression coefficient
y	Dependent variable
ZX_i	Represents the standardized variable of X_i
$E(X_i)$	Represents the mean value of variable X_i
$Var(X_i)$	Represents the variance of variable X_i
r_{ij}	Correlation coefficient between the variables X_j representing the variable X_i
P_{ij}	Denotes partial correlation coefficient

From the above table, we can clearly see the relevant definitions of the symbols used in this article.

4 ESTABLISHMENT AND SOLUTION OF MODEL

4.1 The Establishment and Solution of Multiple Linear Regression Model

4.1.1 The construction of economic vitality index system

Henan Province is the largest grain production base in China, and its agricultural food and animal products rank first in the national ranking. In addition, Henan Province is not only outstanding, but also has a very rich mineral tourism resources and its transportation is also very convenient. However, there are other problems in Henan Province, such as the lack of excellent talents and the attraction of outstanding talents outside the province, which leads to the weak economic strength of Henan Province. Compared with other neighboring provinces, its regional comprehensive competitiveness is relatively small. Therefore, this paper selects Henan Province as the research object, and systematically and deeply studies the regional economic vitality of Henan Province.

First of all, in the selection of the impact indicators of economic vitality in Henan Province, it must be based on the principles of scientificity and integrity, and the selected indicators can accurately and truly show the impact of various factors on the economic vitality of Henan Province. Because the regional economic vitality is a complex and pluralistic system, the indicators selected in this paper should reflect the development status and trend of regional economic vitality from multiple levels, multiple perspectives and multiple dimensions. Then, when establishing an index system that affects economic vitality, it is mainly based on theoretical analysis, but due to the limitations of data sources and data availability when constructing indicators. Therefore, the constructed indicators should be as easy to understand as possible, the evaluation methods should be combined with the old and the new, and the selected indicators should ensure the authenticity and reliability of the data sources. Finally, regional economic vitality is an abstract phenomenon that is difficult to present in detail. When selecting indicators, it is necessary to pay attention to whether there is an inevitable connection with it. Based on the research of other scholars, the index system constructed for different periods is not exactly the same, so it is necessary to select indicators according to the actual situation of the research object of this topic[3].

The purpose of this paper is to focus on the economic vitality of Henan Province, and to establish an index system based on the relevant data and the actual economic development of Henan Province. In this paper, GDP (billion yuan) is selected as the representative index of economic vitality in Henan Province, and the relevant indexes affecting regional economic vitality are selected from six aspects: economic benefit, opening up, population, government regulation, residents' quality of life and enterprise vitality. They are per capita GDP (yuan / time), the annual cumulative number of tourists (millions of people), the total amount of foreign-funded enterprises in and out (thousands of dollars), the natural population growth rate (%), the consumer price index (last year = 100), the urban unemployment rate (%), fiscal

revenue (billions of yuan), fiscal expenditure (billions of yuan), per capita disposable income (yuan), total retail sales of social consumer goods (billions of yuan), R&D funds for industrial enterprises above designated size (ten thousand yuan) and the number of industrial enterprises above designated size (units)[4]. The details are shown in Table 2.

Table 2 The Selection of Economic Vitality Indicators in Henan Province

Influencing factor	Index
economic benefit	GDP per capita (yuan / person)
opening to the outside world	Cumulative annual number of visitors (millions)
	Total import and export volume of foreign-invested enterprises (USD 1000)
population	Natural population growth rate (%)
	Consumer Price Index (last year = 100)
	Urban unemployment rate (%)
government regulation	Fiscal revenue (billion yuan)
	Fiscal expenditure (billion yuan)
residents' quality of life	Per capita disposable income (yuan)
	Total retail sales of social consumer goods (billion yuan)
enterprise vitality	R&D expenditure of industrial enterprises above designated size (ten thousand yuan)
	Number of industrial enterprises above designated size (units)

4.1.2 Based on the establishment of multiple linear regression model

Let the multivariate linear regression model of random variable y and variable $x_1, x_2, \cdots, x_k$ ($k \geq 2$) be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{1}$$

Among them, β_0 is the regression constant, $\beta_j \left(j = 1, 2, \cdots, k \right)$ is the regression coefficient, y is the dependent variable, $x_1, x_2, \cdots, x_k$ is k independent variables that have a significant impact on y , and ε is a random term that shows the comprehensive impact of various errors on the dependent variable[5].

If n sets of observation data $x_{i1}, x_{i2}, \cdots, x_{ik}, y_i \left(i = 1, 2, \cdots, n \right)$ are obtained, the above equation can be expressed as the following equation.

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n \end{cases} \tag{2}$$

4.1.3 Based on the solution of multiple linear regression model

Through the website of the National Bureau of Statistics, the data corresponding to each index of Henan Province from 2009 to 2023 are searched and sorted out. The 12 indexes selected from the six aspects of economic benefits, opening up, population, government regulation, residents' quality of life and enterprise vitality are expressed in turn by A, B,...,L. Firstly, the correlation between GDP, the representative index of economic vitality, and these 12 indexes is analyzed[6]. It is concluded that the correlation coefficients between GDP and per capita GDP, fiscal revenue, fiscal expenditure, per capita disposable income, total retail sales of social consumer goods and R&D funds of industrial enterprises above designated size are all higher than 0.9. There is a strong positive correlation; the correlation coefficients between GDP and the total number of tourists received throughout the year, the total amount of foreign-invested enterprises in and out, the natural population growth rate and the number of industrial enterprises above designated size are all higher than 0.6, showing a strong positive correlation. However, the correlation between GDP and consumer price index and urban unemployment rate is not very high, and the correlation coefficients are -0.27 and -0.04 respectively.

Therefore, this paper considers GDP, the representative index of regional economic vitality in Henan Province, as the dependent variable (y)[7]. From the six aspects of economic benefits, opening up, population, government regulation, residents' quality of life and enterprise vitality, 10 indicators affecting economic vitality are selected as independent variables, which are set as per capita GDP (x_1), total number of tourists received throughout the year (x_2), total inflow and outflow of foreign-invested enterprises (x_3), natural population growth rate (x_4), fiscal revenue (x_5), fiscal

expenditure (x_6), per capita disposable income (x_7), total retail sales of social consumer goods (x_8), R&D funds of industrial enterprises above designated size (x_9), and number of industrial enterprises above designated size (x_{10}). Establish a multiple linear regression model, use R software to write code, and use the least squares estimation method to obtain the OLS regression model as follows:

$$\begin{aligned} y = {} & 0.837x_1 - 130.1x_2 + 0.000027x_3 - 309.7x_4 - 0.586x_5 + 0.277x_6 \\ & -0.727x_7 + 0.773x_8 + 0.0006x_9 + 0.0056x_{10} + 3625 \end{aligned} \tag{3}$$

Through the above equation, we get that the indicators that have the greatest impact on the economic vitality of Henan Province are: the annual cumulative number of tourists and the natural population growth rate, and these independent variables are negatively correlated with GDP. The total amount of foreign-invested enterprises, the R&D funds of industrial enterprises above designated size and the number of industrial enterprises above designated size have the least impact on the economic vitality of Henan Province.

We use R software to write code to test the OLS regression model. The correlation coefficient of the OLS regression model is 0.999, indicating that the fitting effect of the model is very good. And the corresponding P value is less than 0.01, indicating that the OLS regression model is highly significant and the overall fitting effect is good.

It can be seen from the t-test results that the per capita GDP and the total retail sales of social consumer goods pass the t -test, while other variables do not pass the t-test, and the impact on the dependent variables is not significant[8-12]. Combined with the above analysis, it can be seen that there is a great correlation between these independent variables that do not pass the t test, so this paper should consider that the variable does not pass the t test may be the reason for the multicollinearity of these variables.

4.1.4 Analysis of effect

According to the results of the multiple linear regression model, this paper gives the action plan to improve the economic vitality: from the perspective of opening up, Henan Province can improve the vitality of economic growth by reducing the number of tourists. From the perspective of population, Henan Province can appropriately reduce the number of unemployed people and increase the consumer price index of residents by controlling the natural growth rate of the population, reducing the population base, increasing the total per capita GDP and increasing the setting of posts, so as to improve the vitality of economic growth. From the perspective of government regulation and control, Henan Province can appropriately promote a virtuous cycle of economic vitality by reducing fiscal revenue and increasing fiscal expenditure; from the perspective of residents' quality of life, Henan Province can appropriately reduce per capita disposable income and increase the sales of social consumer goods to improve the vitality of economic growth.

4.1.5 Multiple collinearity diagnosis

Considering that the independent variables in the OLS regression model may have multicollinearity, this paper uses the most common eigenvalue determination method to diagnose the multicollinearity of the independent variables.

Using R software to write the code, the condition number is 32151, indicating that there is a serious multicollinearity between the independent variables. In order to determine which independent variables have multicollinearity, the eigenvalue of each variable is obtained by using R software to write code.

$$\zeta = (7.97, 1.07, 0.62, 0.24, 0.09, 0.006, 0.003, 0.0009, 0.0008, 0.0002) \tag{4}$$

It can be seen that the eigenvalues of x_6 , x_7 , x_8 , x_9 and x_{10} are close to 0, so there is multicollinearity between x_1 , x_2 , x_3 , x_4 and x_5 . When there is multicollinearity between independent variables, the accuracy of parameter estimation of the model will decline sharply, so that the estimated value cannot be explained from the perspective of economy and society, so that the application value of the model will decline sharply. In the following, we will use ridge regression and LASSO regression to modify the multiple linear regression model respectively, and at the same time solve the problem of multicollinearity between independent variables.

4.2 The Establishment and Solution of Ridge Regression Model

4.2.1 The establishment of ridge regression model

When there is multicollinearity between independent variables, ridge regression is a coefficient estimation method proposed by Hoerl and Kennard. Although ridge regression will lose some information and reduce the accuracy of fitting, the obtained regression coefficient is more realistic and reliable, and the fitting of ill-conditioned data is better than the least square method[13].

The linear regression model is established as follows:

$$Y = Z\theta + \varepsilon \tag{5}$$

In the above formula, Y is the dependent variable, Z is the independent variable, θ is the standard regression coefficient, ε is a random error, P is the number of independent variables, n is the number of samples.

Where,

$$Y=\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix},\quad Z=\begin{bmatrix} Z_{11} & \cdots & Z_{p1} \\ \vdots & \ddots & \vdots \\ Z_{1n} & \cdots & Z_{pn} \end{bmatrix},\quad \theta=\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix},\quad \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{6}$$

Due to the different dimensions or the large numerical gap, the solution results will cause great errors. It is necessary to use the following formula to centralize and unit the length of Y and Z respectively.

$$Z_{ij}=\frac{X_{ij}-\overline{X}_j}{\sqrt{\sum\left(X_{ij}-\bar{x}_j\right)^2}} \tag{7}$$

The least squares estimate of θ can be expressed as:

$$\hat{\theta}=\left(Z^TZ\right)^{-1}Z^TY \tag{8}$$

When there is multicollinearity between independent variables, matrix Z^TZ is a singular matrix, and its corresponding eigenvalues are very small. The elements on the diagonal of matrix $\left(Z^TZ\right)^{-1}$ are very large, which will lead to very unstable parameter estimation. The slight change of data will make the estimated value of parameters change greatly, and the regression coefficient cannot accurately and objectively reflect the influence of independent variables on dependent variables. Ridge regression is to correct the shortcomings of the least squares method. A diagonal matrix kI is added on the basis of matrix Z^TZ, so that the eigenvalues of the matrix become larger, and the singular matrix becomes a non-singular matrix, so as to improve the stability of parameter estimation and make the regression coefficient more accurately reflect the objective reality. The standardized coefficient $\Theta_{(k)}$ of ridge regression estimation is:

$$\hat{\Theta}_{(k)}=\left(Z^TZ+kI\right)^{-1}Z^TY=\left(Z^TZ+kI\right)^{-1}Z^T\hat{\Theta} \tag{9}$$

Where, k is the ridge regression parameter, and the value range is 0~2. When k=0, it is the least squares estimation; when the $k\neq 0$ and k values are larger, the predicted variance is larger, and the influence of multicollinearity on the stability of the regression coefficient becomes smaller.

4.2.2 The solution of ridge regression model

Ridge regression is a biased estimation regression method used to solve the multicollinearity between independent variables. By giving up the unbiasedness of least squares estimation, the regression coefficient obtained at the cost of losing some information and reducing the accuracy of fitting is more realistic and more reliable. The regression method is suitable for the fitting of ill-conditioned data.

Considering the different dimensions of each variable, the R software is used to write the code. Firstly, the original data is standardized to eliminate the influence of the dimension on the accuracy of the model. Then, the ridge regression analysis is carried out. The ridge parameter k is 0-2, and the step size is 0.2. The ridge trace diagram corresponding to the 10 ridge parameter values is shown in Figure 2.

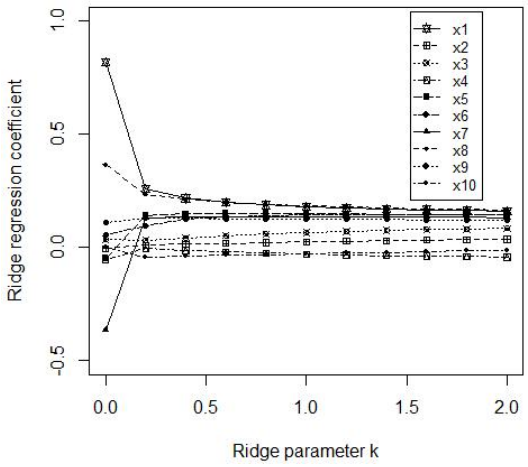

Figure 2 Ridge Trace Map

It can be seen from the above figure that according to the general principle of ridge regression k value selection, the independent variables x_2, x_3, x_4, and x_{10} with relatively stable coefficients and small absolute values are eliminated, and the new ridge trace is obtained, as shown in Figure 3.

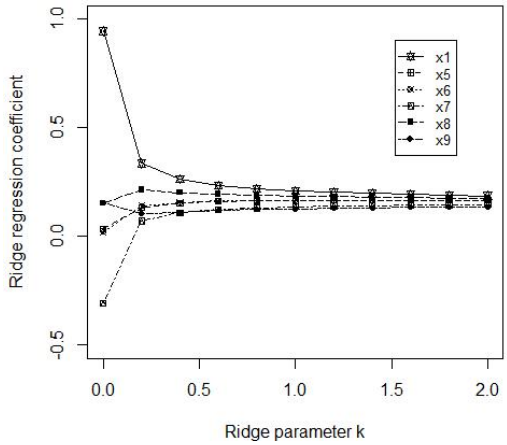

Figure 3 Ridge Trace Map

It can be seen from the above figure that the ridge regression coefficient changes more and more slowly after removing some variables. In summary, when $k > 0.8$, the value of the ridge regression coefficient basically reaches a stable state. Therefore, this paper uses R software to write code and select the ridge regression coefficient corresponding to each other when the ridge parameter $k=0.8$ is selected. The standardized ridge regression equation is:

$$\hat{y} = 0.220x_1 + 0.165x_5 + 0.166x_6 + 0.132x_7 + 0.189x_8 + 0.124x_9 \quad (10)$$

The F value of the ridge regression model is 10090, and the p value is less than 0.01 when the t test is performed, indicating that the model is very effective. The correlation coefficient is 0.999, the fitting effect is good, and the goodness is high. This model can be used for research and analysis.

From the standardized ridge regression equation, it can be seen that per capita GDP, fiscal revenue, fiscal expenditure, per capita disposable income, total retail sales of social consumer goods and R&D expenditure of industrial enterprises above designated size are positively correlated with the economic vitality of Henan Province. At the same time, the importance of the main six factors affecting the economic vitality of Henan Province is ranked from large to small: per capita GDP, total retail sales of social consumer goods, fiscal expenditure, fiscal revenue, per capita disposable income, and R&D expenditure of industrial enterprises above designated size.

Using R software to write the code, the unstandardized ridge regression equation is:

$$\hat{y} = 0.967x_1 + 0.429x_5 + 0.1x_6 - 0.609x_7 + 0.324x_8 + 0.0008x_9 + 199.9 \quad (11)$$

In order to further better observe the fitting effect of the ridge regression model, this paper calculates the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, according to the above equation, and uses R software to write the code to draw a comparison diagram, as shown in Figure 4.

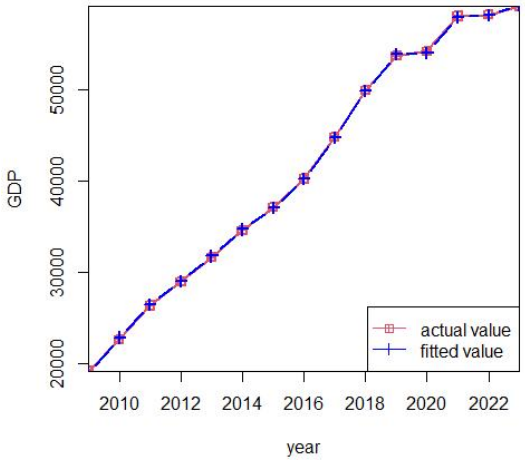

Figure 4 The Comparison Chart Between the Actual Value of GDP and the Fitting Value of Ridge Regression

From the above figure, we can see the intuitive effect of the comparison between the actual value of GDP and the ridge regression fitting value, which shows that the fitting effect of the model is very good.

4.3 The Establishment and Solution of LASSO Regression Model

4.3.1 The establishment of LASSO regression model

LASSO model is designed to give priority to the importance of independent variables. LASSO regression determines the best model by compressing the estimation. Its fundamental property is to obtain a more accurate model by establishing a penalty function. Under the constraint that the sum of the absolute values of the regression coefficients is less than a constant, the sum of the residual squares is minimized, and the variables whose regression coefficients are close to or equal to 0 are eliminated, so as to solve the problem of multicollinearity[14].

LASSO regression is based on ordinary linear regression with additional penalty terms, and its estimation is:

$$\hat{\beta}_{Lasso} = \arg\min_{\beta \in R^p} \|Y - X\beta\|^2 \tag{12}$$

$$s.t. \sum_{j=1}^{p} |\beta_j| \le t, t \ge 0 \tag{13}$$

Equivalent to:

$$\hat{\beta}_{Lasso} = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{14}$$

In the above formula, $\arg\min(\cdot)$ is the function of finding the minimum value of the parameter, and $\hat{\beta}_{Lasso}$ is the objective function of minimization; λ is the adjustment parameter, and $\lambda \ge 0$; $\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$ represents the effect of model fitting, and $\lambda \sum_{j=1}^{p} |\beta_j|$ represents the penalty of parameters[15].

The smaller the value of λ, the smaller the punishment will be, and the more variables will be retained in the model; the greater the value of λ is, the greater the punishment will be, and the fewer variables are retained in the model. Through the selection of parameter λ, variable selection can be achieved[16].

4.3.2 The solution of LASSO regression model

Considering the different dimensions of each variable, R software is used to write code. Firstly, the original data is standardized to eliminate the influence of dimension on the accuracy of the model. Then, LASSO regression analysis is carried out to obtain Figure 5.

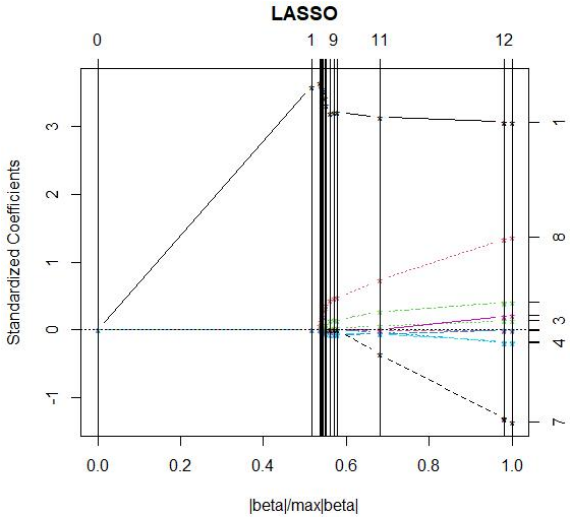

Figure 5 LASSO Regression Diagram

The coefficient selection methods include the cv method of k-fold cross-validation and the method of using Cp statistics to evaluate regression for coefficient selection. In this paper, the Mallows Cp statistic selection coefficient is selected, and the standardized LASSO regression equation is obtained by using R software to write code.

$$\hat{y} = 0.840x_1 - 0.003x_2 + 0.015x_3 - 0.009x_4 - 0.017x_5 - 0.096x_7 + 0.196x_8 + 0.071x_9 - 0.015x_{10} \tag{15}$$

From the above equation, it can be seen that the coefficient of the fiscal expenditure variable is zero, and this variable is eliminated. The importance of the main nine independent variables affecting the economic vitality of Henan Province is ranked from large to small: per capita GDP, total retail sales of social consumer goods, per capita disposable income, R&D expenditure of industrial enterprises above designated size, fiscal revenue, total inflow and outflow of foreign-invested enterprises, number of units of industrial enterprises above designated size, natural population growth rate and cumulative number of tourists received throughout the year.

In order to further observe the fitting effect of the LASSO regression model, this paper calculates the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, according to the above equation, and uses R software to write the code to draw a comparison chart, as shown in Figure 6.

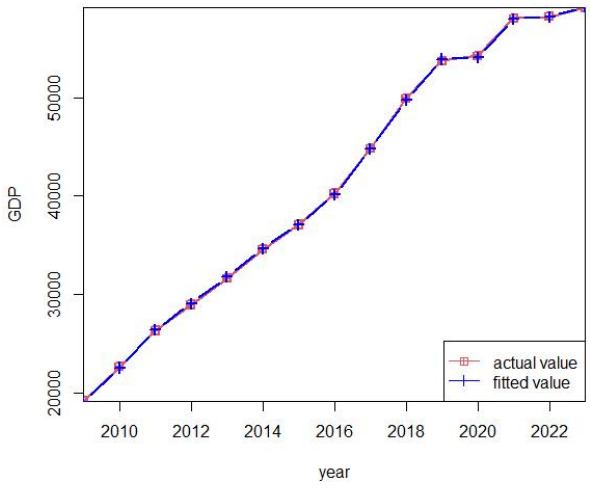

Figure 6 The Comparison Between the Actual Value of GDP and the Fitting Value of LASSO Regression

From the above diagram, we can see the intuitive effect of the comparison between the actual value of GDP and the fitting value of LASSO regression, and the fitting effect of the model is very good.

4.4 Model Analysis

In this paper, ridge regression and LASSO regression model are used to eliminate the effect of multicollinearity between variables. By comparing the results of the test and parameter test of each model, the optimal regression model is selected. Using R software to write code, the comparative analysis of ridge regression and LASSO regression model is shown in Table 3.

Table 3 Model Analysis

Model	RMSE	R^2
Ridge regression	0.011106	0.999
LASSO regression	0.008112	0.999

From the above table, it can be seen that the RMSE of the LASSO regression model is small, indicating that the deviation between the fitting value of the LASSO regression model and the actual value is smaller, and the effect is optimal; by comparing the goodness of fit, it can be concluded that the values of the ridge regression and the LASSO regression model are the same, and the fitting effect is very good. In summary, the LASSO regression model is a relatively better model.

4.5 Time Series Predicting

From the comparative analysis results of the model, it can be seen that the LASSO regression model is a relatively better model, and in the LASSO regression equation, the fiscal expenditure variable is eliminated. Therefore, we first use R software to write the code, and get the time series diagram of other independent variables in the equation, as shown in Figure 7, Figure 8, Figure 9.

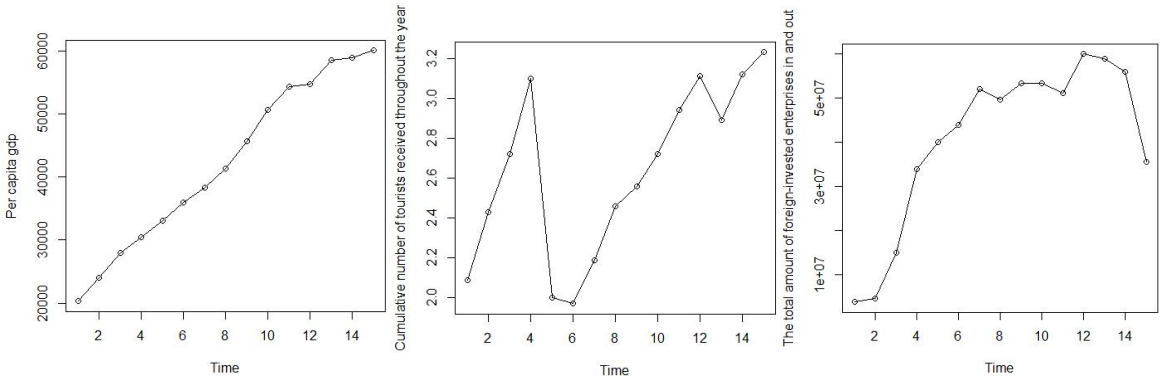

Figure 7 x_1, x_2, x_3 Time Series Diagram

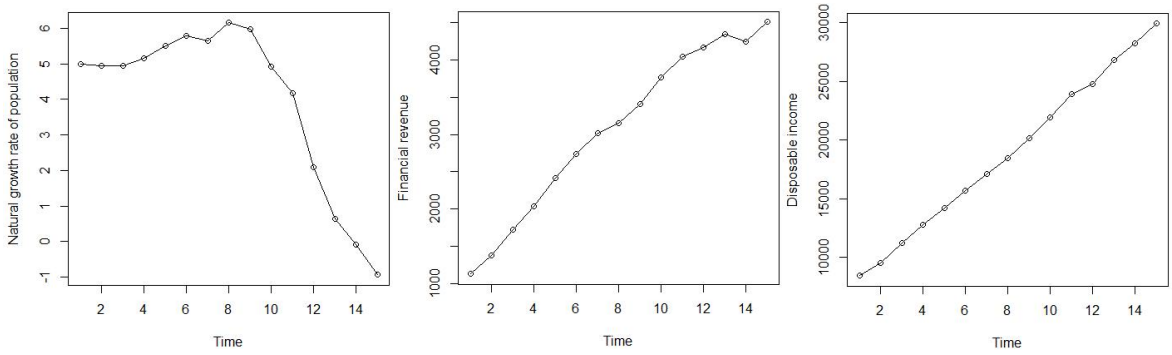

Figure 8 x_4, x_5, x_7 Time Series Diagram

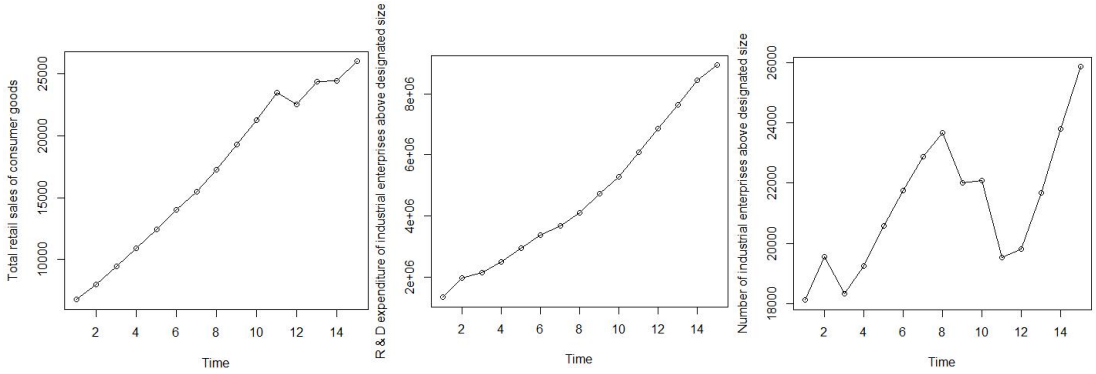

Figure 9 x_8, x_9, x_{10} Time Series Diagram

From the time series diagram of the above variables, it can be seen that the five variables of per capita GDP, fiscal revenue, per capita disposable income, total retail sales of social consumer goods, and R&D expenditure of industrial enterprises above designated size show a linear trend, and the remaining four variables show random fluctuations without significant changes.

For the five variables showing a linear trend, this paper uses the Holt exponential smoothing model for time series prediction. For these four variables that show random fluctuations without significant changes, a simple exponential smoothing model is used for time series prediction. The standardized data of each index in 2024 are shown in Table 4.

Table 4 The Predictive Value of Independent Variables in 2024

Year	x_1	x_2	x_3	x_4	x_5	x_7	x_8	x_9	x_{10}
2024	1.472	1.355	-0.276	-2.095	1.497	1.822	1.560	1.985	2.076

From the results of the LASSO model, the standardized LASSO regression equation is:

$$\hat{y} = 0.840x_1 - 0.003x_2 + 0.015x_3 - 0.009x_4 - 0.017x_5 - 0.096x_7 + 0.196x_8 + 0.071x_9 - 0.015x_{10} \quad (16)$$

The standardized data of each index predicted in the above table in 2024 are substituted into the above regression model, and then the predicted value of GDP in Henan Province in 2024 is calculated to be 61441.55 billion yuan.

5 SUGGESTIONS ON PROMOTING THE ECONOMIC DEVELOPMENT OF HENAN PROVINCE

5.1 Increase Investment in Domestic Enterprises to Form a Sustainable Industrial Chain

The government of Henan Province should increase the investment attraction and investment attraction of enterprises, speed up the implementation of excellent projects as soon as possible, especially introduce and train some leading enterprises and leading industries, so as to improve the scientific and technological technology of Henan Province, make the industrial chain longer, improve the comprehensive benefits of Henan Province, and improve the vitality of regional economy in an all-round way. It can also start from the regional financial field, through continuous improvement and broadening the financing channels of private enterprises and state-owned enterprises, in order to create a good financial environment and achieve a sustainable financial chain.

5.2 Improve the Efficiency of Government Work

The government of Henan Province can integrate the resources within the scope of Henan Province through policy intervention, which can provide more excellent resource services for enterprises, so that enterprises have a broader space for development. The government can also relatively reduce the more complicated procedures and reduce

unnecessary operations for the establishment process of enterprises. The government should improve work efficiency, because its work efficiency has a direct impact on the cost of enterprises. Low efficiency work will bring bad results to the competitiveness of enterprises, and enterprises play a very important role in regional economic competitiveness. The competition among regions is the competition among enterprises in different regions.

5.3 Pay Attention to System and Mechanism Innovation, Establish Market Competition Mechanism

The development trend of the economy is influenced by the system of our country. The quality of the system plays a decisive role in the economic growth of various places, especially in the degree of modern informationization, the development degree of the factor market and the quality of the market environment in our country. This shows that a good market environment can attract many foreign excellent resources to the region, and the more it can promote the rapid growth of the local economy.

In addition, by improving the environment of the market, elements can be attracted to take root locally, so that these excellent resources can be retained locally, forming the local core strength and improving the local economic development level. Competition plays a vital role in economic growth. Competition with other enterprises can make enterprises improve their fighting spirit, constantly improve themselves, improve their competitiveness, and make themselves better. Two-way competition can enable enterprises to continuously innovate. In general, the economy can be improved as a whole through mutual competition.

5.4 Strengthen Regional Cooperation and Achieve Coordinated Development of Regional Economy

Regional cooperation plays a vital role in the coordinated development of economy. Good regional cooperation not only helps to improve the status and role of all regions in their respective division of labor, but also helps market players to obtain the required factor supply and product demand from a wider range. Henan Province is close to Anhui Province and Shandong Province in the east, and is closely linked to Shaanxi Province in the west. Anhui Province is rich in labor resources, tourism resources and tourism resources. Shandong Province not only has very large reserves of coal and oil, but also has very convenient transportation. Shaanxi Province has a wide variety of mineral resources. Strengthening cooperation with neighboring provinces can not only effectively improve the advantages of each region, but also effectively play the special capabilities of each region, forming a highly competitive regional economy with strong competitiveness according to local conditions, division of labor and cooperation, complementary advantages and common development.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Hu Jinyana, Cai Yongfanga, Cheng Jiaminga. Evaluation Model of the Regional Economic Vitality. Journal of Physics: Conference Series, 2020(3).
[2] Zhao Dan, Zhao Zhi, Chen Zening. Construction of Regional Economic Vitality Model. Journal of Economic Science Research, 2020(2): 19-23.
[3] Xin Siqian, Xin Sihan, Hu Buen, Wang Penghui, Fang Kang. Analysis and Decision of Regional Economic Vitality and its Influencing Factors. Frontiers in Economics and Management, 2020(1): 76-85.
[4] Pang Jinfeng, He Jintao, Luo Lan, Li Yao, Mao Yiren, Zhang Hanrui. Analysis of China's Regional Economic Vitality and Its Influential Factors Based on System Cluster Model and Computer Multiple Regression Model. Journal of Physics: Conference Series, 2020.
[5] Zhong Zhenfang. Factors Influencing Regional Economic Vitality Based on Regression Analysis. E3S Web of Conferences, 2020.
[6] Liu Yi, You Xiaoyu, Zhang Chunshuo. Regional Economic Vitality Based on Weighted Grey Relational Analysis. Journal of Economic Science Research, 2020(2): 12-18.
[7] Gan Mingli. Empirical Analysis on Influencing Factors of Financial Revenue in Sichuan Province. Journal of Economics and Public Finance, 2022(8): 68-74.
[8] Zhang Enquan. Empirical analysis of the influencing factors of fiscal revenue in Xinjiang. Economic Forum, 2015(5): 26-30.
[9] Guo Lin, Guo Li, Zhang Na, Tang Biao. Empirical Analysis on the Influencing Factors of Fiscal Revenue in China. Modern Business Trade Industry, 2016(27): 135-136.
[10] Li Lvxiu, Dong Xiyuan. Empirical analysis based on Eviews on China’s fiscal revenue factors. China Collective Economy, 2021(16): 92-94.
[11] Xie Liying. Empirical Analysis on the Influencing Factors of Fiscal Revenue in Anhui Province. Modern Business Trade Industry, 2019(9): 115-116.
[12] You Jiawei. Empirical analysis of the influencing factors of China’s fiscal revenue. China Collective Economy, 2019(7): 100-103.
[13] McDonald Gary C. Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 2009(1): 93-100.

[14] Tibshirani Rob. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B: Methodological, 1996(58): 267-288.
[15] Zou H. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association, 2006(101): 1418-1429.
[16] Nie Ruichao. Analysis of influencing factors of Fiscal revenue in Beijing based on Ridge regression and Lasso regression model. International Journal of New Developments in Engineering and Society, 2022(6): 1-5.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3017

TREND ANALYSIS AND FORECAST OF HOUSEHOLD APPLIANCE OWNERSHIP AND ELECTRICITY CONSUMPTION IN XIANGYANG

MinShi Zheng
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 1229564969@qq.com

Abstract: With the development of the times, the types of household appliances are also diversified, and then people's demand for electricity is also growing. So it is worth thinking about whether there is a relationship between the amount of household durable goods and the amount of electricity used. Whether the energy-saving and emission-reduction implemented by the state is carried out by the manufacturers or not, the relationship between the two can also be reflected. In this paper, we focus on Xiangyang, the results show that household electricity consumption is closely related to household appliances. It can be inferred that the total electricity consumption will continue to increase in the next few years, but it will fall slightly each year. It is also a bold bet that when all appliances reach saturation, there may be a negative increase in electricity consumption driven by some policies.
Keywords: Electricity consumption; Principal component analysis; Multiple regression analysis; Energy saving and emission reduction

1 INTRODUCTION

1.1 Research Significance

With the development of science and Technology and People's pursuit of high quality of life, the purchasing power of durable goods in China's households has increased significantly, thus directly contributing to the increase in the ownership of household appliances per 100 households, taking the quantity of household appliances as the starting point of electricity consumption of urban and rural residents can be relied on. In theory, the quantity of electricity consumption and the quantity of household appliances are increasing in direct proportion. However, the national policy of energy conservation and emission reduction is being vigorously implemented, whether the implementation of this model can also be reflected. This paper analyzes the relationship between the changes in the ownership of typical household appliances in Xiangyang and the average electricity consumption per 100 households in the corresponding years, to analyse the implementation of energy saving and emission reduction measures for household appliances in Xiangyang in recent years. At the same time, the trend of electricity consumption research is also helpful to the power industry, as follows: first, electricity can not be stored, power generation and use must be completed in an instant, therefore, the amount of electricity generated by the power plant should be consistent with the load required by the city. Second, the power industry is an industry that requires very high capital and technology requirements. The construction time of any power plant is more than three to five years. Only accurate forecasting can make reasonable arrangements for the use of human, material and financial resources, there will be no shortage of electricity or excess generating equipment.

1.2 Research Status

With the increase of electricity consumption year by year, the development of electricity consumption has become the most potential growth point in the electricity market. The analysis of residential electricity demand has gradually become one of the hot spots of people from all walks of life.
The scholar Yang Zhengde, compared the energy consumption data of various home appliances in China with that of other countries, such as the European Union, which has established energy-saving standards, analysis of our country in the standardization of energy-saving home appliances there are still some gaps[1]; Gu Xinling, a scholar, used the factor analysis method to study the different factors of the amount of durable goods owned by urban residents in various cities and towns in the country. In the end, the score factor tells us that the amount of durable goods owned by the coastal areas is the most, second, the average ownership of durable goods is relatively low in the central region and other inland cities, and in the northwest and southwest regions, where the economy is relatively low. Therefore, it can be considered that the difference in the ownership of durable goods is closely related to the local economic level, this is especially true between urban and rural areas[2]; Liang Huifang, a scholar, has divided household electricity into two parts: electrical appliances and electric lamps. The amount of electricity used for electric lamps is related to the number of people and the size of houses, while the amount of electricity used for electrical appliances is related to the amount of various household appliances owned, the power consumption level is analyzed and predicted under the above factors[3];There are various approaches to the study of electricity consumption and household appliances and the relationship between them, and different approaches will draw conclusions from different levels[4].

1 DATA SITUATION

1.1 Data Pre-processing

In the process of data collection, there is a small amount of missing data, so the mean interpolation is used to deal with the missing values. The mean interpolation can be calculated to estimate the missing value. The mean interpolation method is to estimate the value of the middle point (X_0, Y_0) through two points (X_1, Y_1) and (X_2, Y_2) .Assuming $y = f(x)$ is a straight line, calculate the function $f(x)$ from two known points, you can find y as long as you know x .When only one value is missing, the unknowns can be estimated by means of the average of two known values to fill in the missing values.

1.2 Descriptive Analysis

2.2.1 Analysis of domestic electricity consumption data in Xiangyang city

The annual total electricity consumption of the city is a group of very complex and lengthy data. In order to make the data more intuitive and facilitate the follow-up analysis, we need to use statistical software to make descriptive statistics of electricity consumption. This paper uses R software to make a time series diagram of the total living electricity consumption in Xiangyang city in the past ten years. The timing diagram is a horizontal axis for time, and the horizontal axis for a two-bit plane coordinate diagram of the total living electricity consumption in Xiangyang city every year. It can intuitively help us master the changing trend of electricity consumption over time.

As can be seen from Figure 1, the annual electricity consumption in Xiangyang city has an obvious increasing trend, which is conceivable. With the development of society, our demand for electricity is more and more, and the domestic electricity consumption is naturally increasing trend. And can be seen by figure 1, during the period of 2008-2011, the growth trend of electricity consumption is relatively slow, after 2011-2013 and 2016, the rapid growth of electricity consumption, and the popularity of household appliances in life, and the emergence of a wide variety of household appliances, of course, may be with global warming, the world temperature, hot summer, colder winter more environmental factors.

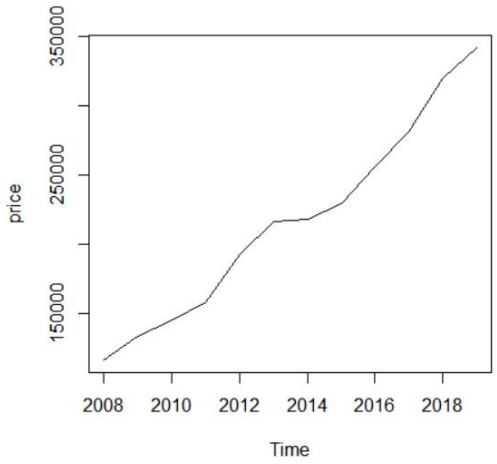

Figure 1 Electricity Consumption Sequence Diagram of Residents in Xiangyang City

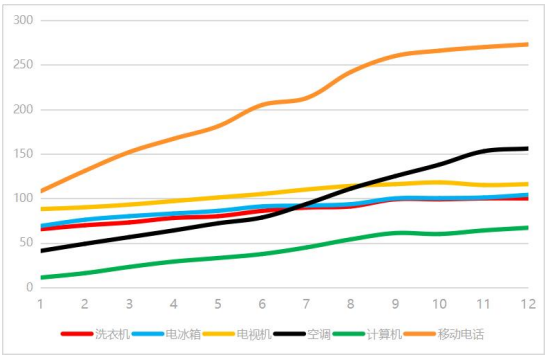

Figure 2 Statistical Chart of Household Electrical Appliances Ownership

2.2.2 Descriptive statistics of the ownership data of various household appliances

As can be seen from Figure 2, any kind of home appliance shows a trend of growth over time; The growth rate of mobile phone is the fastest, but after 2016 (the abscissa is 9), the upward trend is obviously flat, the peak is around 275, indicating that in recent years per capita mobile phone is about 2.5-3; Second faster growth rate is the air conditioning

and computer, and the annual growth rate is almost a value, also can be seen that air conditioning and computer in 2008, their ownership is very little, every hundred less than half of the people have, with the improvement of the quality of life, the pace of rapid development and social development is consistent; Relative TV and refrigerator starting point is very high, in 2008 both hundred household ownership is close to 100, with 12 years of growth, refrigerators reached 100%, can think every household has a refrigerator, and television in 2012 (5) reached 100, in the later growth has more than 100%, the average every family has more than one TV.

This result is also just like the common phenomenon in our life, air conditioning and computers are step by step towards our life, but in some rural areas, the ownership of these durable goods is still relatively low; Washing machines and refrigerators are almost one unit in both urban and rural areas; most households have two or more.

2 MODELING AND ANALYSIS

2.1 Regression Analysis of Electricity Consumption and Household Appliances

2.1.1 Plot scatters

Before using SPSS software to do regression analysis on electricity consumption and household appliances, it is necessary to understand the correlation between electricity consumption and each influencing factor, and judge whether there is a linear relationship between the two, whether the regression analysis can be carried out. So we take the electricity as a dependent variable on the y-axis, and then take each household appliance as an independent variable on the x-axis, making scatter plots.Too many graphics, no longer display one by one.

From the results of the graph, we can see that there is an obvious linear relationship between each type of household appliances and their electricity consumption, washing machines have not had a strong linear relationship with electricity consumption in the last four years, and you can also see that computers and mobile phones tend to follow the same pattern. To sum up, it is possible to set up linear regression equation with these six factors as independent variables.

3.1.2 Establish a regression model

Import data into SPSS, select Regression analysis from the analysis menu, enter dependent variables and multiple independent variables, and select [stepwise] regression, the stepwise regression method selects the entry regression equation that conforms to the criterion and contributes the most to the dependent variable, and removes the model that conforms to the elimination criterion from the model, it is repeated until the independent variables in the equation meet the criteria for entering the model. The results of the analysis are as follows:

Table 1 Variable was Entered / Removed

Models	Variables have been entered	The variable was removed means
1	Air-conditioning	Step (criterion: F-to-enter probability <=. 050, the probability of F-to-remove> =. 100)

Table 2 Variance Analysis Table of Dependent Variable (Electricity Consumption)

Model	Sum of squares	degree freedom	ofmean square	F Value	significance
Regression	5.607E+10	1	5.607E+10	287.375	.000
Residuals	1951285548	10	195128554.8		
Total	5.803E+10	11			

As can be seen from Table 1, only one variable of air conditioning entered the equation, and the remaining five variables were excluded, indicating the existence of serious collinearity among the six independent variables.

Table 2 is the analysis of variance table, which shows the results of the analysis of variance in the process of regression fitting. Sig. indicates the probability that the F value is greater than the F critical value. The results show that when the regression equation includes air conditioning as an independent variable, the probability of significance is less than 0.001, which rejects the original hypothesis that the regression coefficients are all 0. Therefore, it is considered that the regression model included only one independent variable, namely, air conditioning. If forced to build a model in this way, it is contrary to the original intention of our research. Therefore, we need to do collinearity analysis between the excluded five independent variables and the included six independent variables.

In order to judge the collinearity between the six independent variables, we change the stepwise regression method to the input regression method when we build the model, that is, we force the six independent variables into the model, in the output statistics to add [collinearity diagnosis] , collinearity diagnosis, the output results are as follows. The "Eigenvalues" column in table 3 has five eigenvalues of 0.004, 0.000, 0.000, 0.000, 3.691E-5, all very close to 0, the corresponding five conditional indexes are 40.285, 131.862, 175.527, 232.522, 430.702, respectively, all of which are greater than 30. The two indexes indicate that there must be serious collinearity among these five variables.

Table 3 Diagnosis of collinearity of independent variables

model	Dimensions	Eigenvalues	Conditional Index	Proportion of variance						
				Constant	Washing Machine	Refrigerator	Television	Air conditioning	Computer	Mobile phones
1	1	6.847	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.148	6.801	.00	.00	.00	.00	.00	.00	.00
	3	.004	40.285	.00	.00	.00	.00	.44	.03	.01
	4	.000	131.862	.03	.00	.09	.07	.02	.28	.09
	5	.000	175.527	.05	.00	.09	.10	.18	.66	.27
	6	.000	232.522	.22	.52	.00	.01	.01	.01	.24
	7	3.691E-5	430.702	.70	.48	.81	.82	.36	.01	.40

2.2 Factor Analysis of Electricity Consumption and Household Appliances

Using SPSS software, factor analysis of 6 variables including washing machine, refrigerator, TV, air conditioner, computer and mobile phone was carried out. One of the ways to determine the principal component is to take the component with an eigenvalue greater than 1 as the principal component, that is, table 4. The second column of the total variance interpretation table shows that only the eigenvalue of the first factor is greater than 1 and the total column shows that only the eigenvalue of the first factor is 5.897, the variance percentage of the first factor is 98.279% , which means that the variance explained by this component accounts for 98.3% of the total variance. Therefore, the determination to extract a principal component greatly reduces the complexity of the original data and only loses 1.7% of the information.

Table 4 Explanation of Total Variance

Component	Initial eigenvalue			Extract the sum of squares of loads		
	Total	Percentage of variance	Cumulative	Total	Percentage of variance	Cumulative
1	5.897	98.279	98.279	5.897	98.279	98.279
2	.065	1.089	99.368			
3	.027	.444	99.812			
4	.006	.100	99.912			
5	.003	.054	99.967			
6	.002	.033	100.00			

The factor score coefficient matrix is shown in table 5. According to the factor score coefficient and the standardized value of the original variable, the scores of each factor of each observation can be calculated. Finally, the principal component expression can be written as:

$$FAC1 = 0.169 \times xyj + 0.168 \times dbx + 0.167 \times dsj + 0.166 \times kt + 0.169 \times jsj + 0.169 \times yddh \quad (1)$$

Table 5 Table of Component Score Coefficients

	Components
	1
Washing Machine（xyj）	.169
Refrigerator（dbx）	.168
Television（dsj）	.167
Air conditioning（kt）	.166
Computer（jsj）	.169
Mobile phone（yddh）	.169

2.3 Regression Analysis of Power Consumption and Factor

In SPSS Software, the results of factor analysis are saved as variables. The factors after dimensionality reduction and electricity consumption were linear regression. R-square and modified r-square can reflect the goodness of fit. The output result shows that the adjusted r-square is 0.924, so the goodness of fit of the model is better. The results of analysis of variance table show that the probability of significance is less than 0.001 when the regression equation contains factor scores, so the equation fitting is better. According to the regression coefficient table, the regression model can be written as:

$$\text{Power} = 217453.592 + 70063.367 \times FAC1\_1 \quad (2)$$

Using R software to forecast the quantity of six kinds of household appliances per 100 households, the result is as table 6. The predicted factor scores are calculated by taking the data into the factor scores. Based on the obtained score factor, the power consumption in the next three years is predicted by introducing the functional expression between the score factor and the power consumption. The factor score formula is:

$$FAC1 = 0.169 \times xyj + 0.168 \times dbx + 0.167 \times dsj + 0.166 \times kt + 0.169 \times jsj + 0.169 \times yddh \tag{3}$$

Table 6 Each Electrical Appliances per 100 Households have Three-phase Forecast Table

	Washing Machine	Refrigerator	Television	Air conditioning	Computer	Mobile phone	FAC1_1
2020	105	105	116.8	154.6	70	274	138.6902
2021	108.5	108	117.3	158	76	276	141.7856
2022	112	111	117.8	163	82	279	145.3156

3 CONCLUSION

Whether we use computer ownership or car ownership to represent electricity consumption, we can see from the trend chart, the trend of electricity consumption has not significantly reduced, so energy-saving efforts still have to be carried out vigorously. At the same time, we used SPSS software to analyze the six household appliances studied, and found that there may be collinearity among them, and there is a large relationship between the amount of household cars and the amount of electricity used, the second is the amount of computers, so I think we can judge the intensity of energy saving and emission reduction and the implementation results mainly by the amount of household cars and computers, follow-up efforts to increase energy conservation and emission reduction can also focus on cars and computers to start.
We have found that domestic electricity consumption is closely related to household appliances. Therefore, it can be inferred that although Xiangyang's total electricity consumption will continue to increase with the increase of the year, but the annual increase will be slightly lower than the previous year. At the same time, it can be boldly predicted that when all electrical appliances reach saturation, under the promotion of "Energy conservation and emission reduction" policy, there may be negative growth in electricity consumption. This shows that with the development of social economy, home appliances in the family life more and more heavy proportion, people more and more enjoy the convenience of science and technology to people's lives. But Xiangyang's annual electricity consumption has not risen much from the previous year's, indicating that household appliances are becoming more environmentally friendly and energy-efficient. This is also the latest achievement of the country to vigorously promote energy saving and emission reduction[5]. It also proves that the energy-saving and emission-reduction policies have been implemented in every household. With the introduction of the 14th five-year plan, which aims to reach“Peak carbon” by 2030 and“Carbon neutrality” by 2060, policy after policy shows how determined the country is to save energy and reduce emissions[6]. The most direct manifestation of the effect of energy saving and emission reduction is the change in electricity consumption. We can boldly predict that after 10-15 years, the total electricity consumption of our country will reach its peak, after 50 years, the total electricity consumption of our country tends to a stable value and may even fall back.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yang Zhengde. Speed up the energy-saving standardization of household appliances.Standardization in Shanghai, 2004(08): 31-35.
[2] Gu Xinling. Statistical application of factor analysis in durable goods ownership of urban residents. The Journal of Tongling University, 2016, 15(04): 29-33+48.
[3] Liang Huifang, Su Ming, Tian Lei. Is the China's residential electricity consumption mode consistent with the requirements of the new urbanization?—— Empirical study based on the electrical equipment and electricity consumption status of urban households . The Economic Journal, 2014, 1(01): 163-180.
[4] Meng Ming, Zhao Ningning. Analysis of the current situation and influencing factors of residential electricity consumption. Think Tank Era, 2019, (37): 246-247.
[5] Guo Songliang, Yan Pengjun, E Haokun. Short-term prediction of total social electricity consumption in Beijing based on ARIMA model. Journal of Beijing Information Science and Technology University (Natural Science Edition), 2020, 35 (05): 93-96.
[6] Hongyun Han, Radwan Amira. Economic and social structure and electricity consumption in Egypt. Energy, 2021, 231.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3018

ANALYSIS OF FACTORS INFLUENCING HUNAN PROVINCE'S GDP TOTAL

QiBin Zhu
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 1138398110@qq.com

Abstract: This article employs multiple linear regression, ridge regression, and LASSO regression methods to analyze the total GDP of Hunan Province from 2002 to 2021, focusing on eight influencing factors including individual employment, total value of goods imports and exports by foreign-invested enterprises, and local fiscal expenditure. The results indicate a significant positive correlation between the total value of goods imports and exports by foreign-invested enterprises, local fiscal expenditure, and Hunan's GDP. Conversely, factors such as research and development (R&D) activities of industrial enterprises above designated size show a significant negative correlation with GDP. The experimental analysis results suggest positive implications for the healthy and stable growth of Hunan Province's GDP.
Keywords: Multiple linear regression; Ridge Regression; LASSO regression; Factors influencing GDP; Multicollinearity

1 INTRODUCTION

GDP, also known as Gross Domestic Product, refers to the final output of production activities by all resident units in a country or region during a specified period. GDP is a core indicator of national economic accounting and a crucial measure of a country or region's economic condition and development level. It reflects the scale of economic development, assesses overall economic strength, and evaluates the pace of economic development of a country or region, demonstrating its comprehensive national power[1].

Additionally, GDP is used for economic structure analysis, such as industrial, demand, and regional structure analysis, providing essential foundations for macroeconomic decision-making. GDP growth is vital for all regions and countries, as it meets the needs of economic and social development[2]. During GDP growth, each region enhances its status to varying degrees by expanding its economic size locally. This is because regions contribute more to their development as their economic strength increases. When a region's GDP reaches a certain level, its influence on its resident population within the region grows significantly[3]. This influence also extends to the entire regional population. GDP, when combined with related indicators, helps calculate other significant metrics of importance[4].

Since the beginning of the 21st century, especially since China's accession to the World Trade Organization, China's GDP has experienced rapid development[5]. By 2010, it surpassed Japan's GDP to become the world's second largest, trailing only the United States[6]. Since 2010, particularly following the 18th National Congress of the Communist Party of China, high-quality development has become a defining characteristic of Hunan Province. The economic aggregate has consistently surged forward. Hunan Province's Gross Regional Product (GRP) exceeded 2 trillion yuan in 2012 and surpassed 4 trillion yuan in 2020, marking a rapid ascent across three trillion-level thresholds in just eight years. It is projected to reach another milestone by surpassing 5 trillion yuan for the entire year[7].

Hunan Province's per capita GRP has exceeded $10,000, doubling compared to 2012. Given the impact of the COVID-19 pandemic and the current complex international situation, it is essential to study the factors influencing Hunan Province's GDP to sustain healthy development, avoid economic crises, steadily enhance residents' income, improve their quality of life, and provide relevant recommendations for future development[8].

This article utilizes data related to Hunan Province's GDP from 2002 to 2021, employing multiple linear regression, ridge regression, and LASSO regression methods combined with relevant literature to quantitatively and qualitatively analyze the factors influencing GDP, aiming to provide recommendations for future development.

2 PRELIMINARY KNOWLEDGE

2.1 Variable Selection and Explanation

To comprehensively consider the factors influencing GDP and based on the actual situation in Hunan Province, this paper selects Hunan Province's GDP as the dependent variable. The independent variables chosen are individual employment, total value of goods imports and exports by foreign-invested enterprises, local fiscal expenditure, per capita consumer expenditure of residents, total water supply, total retail sales of social consumer goods, electricity generation, and research and development (R&D) activities of industrial enterprises above designated size. To facilitate subsequent research, alphabetic symbols are used to represent these nine variables as detailed in the table 1 below:

Table 1 Variable Description

φ	Total GDP
x_1	Individual employment
x_2	Total value of goods imports and exports by foreign-invested enterprises
x_3	Local fiscal expenditure
x_4	Per capita consumer expenditure of residents
x_5	Total water supply
x_6	Total retail sales of social consumer goods
x_7	Electricity generation
x_8	Research and development (R&D) activities of industrial enterprises above designated size

The units for the variables in the table are respectively: 100 million yuan, persons, billion US dollars, 100 million yuan, yuan, billion cubic meters, 100 million yuan, billion kilowatt-hours, and 100 million yuan.

2.2 Data Source

The purpose of this article is to study the factors influencing the total GDP of Hunan Province and to fit a regression model to explain the linear relationship between explanatory variables and the dependent variable. The data used in this study were sourced from the official websites of the National Bureau of Statistics (http://www.stats.gov.cn/) and the Hunan Provincial Bureau of Statistics (http://www.tjj.hunan.gov.cn/), covering nine economic indicators from the years 2002 to 2021, spanning a period of 20 years. The downloaded data were processed to convert them into a standard data format. Due to the age of some indicators, early data points were missing. To ensure the integrity of the analysis, this article employed simple non-random methods to impute missing values, using techniques such as mean, median, and mode.

3 MULTIPLE LINEAR REGRESSION

3.1 Model Establishment

In many real-life problems, there are often multiple factors influencing the dependent variable. When there exists a linear relationship between the dependent variable and several independent variables, this modeling problem is referred to as multiple linear regression.

The basic model of multiple linear regression is as follows:

$\varphi = \beta_0 + \beta_i x_i + \varepsilon$, and $1 \le i \le 9$.

In the equation, β_i represents the regression parameters of each factor in the model.Obtained through the method of least squares or maximum likelihood estimation;$\beta_0$is the regression intercept, $x_i$is the independent variable, also known as the explanatory variable or predictor variable; ε represents the error term, which, similar to the simple linear regression model, εhas a mean of 0 and a variance of σ^2.

The regression equation satisfies the following basic assumptions:

A1. The explanatory variable x_i is a constant variable.Furthermore, it is not a random variable, and the independent variables in the design matrixXare mutually independent, meaning XX^T is nonsingular with a nonzero determinant.The number of samples n should be greater than the number of independent variables k, and X is a full-rank matrix.

A2. The random error term of the regression equation has the characteristics of a mean of 0, homoscedasticity, and independence.

A3. The random errors must follow a normal distribution.

3.2 Results and Analysis

3.2.1 Multiple linear regression modeling

Due to the lack of uniformity in data units and their large magnitudes, standardizing the independent variables allows for more accurate comparison of their effects on the dependent variable. The scale function in R language is used for data standardization. Based on the method of least squares, the linear regression equation is established using the lm function. From Table 2, the multiple linear regression equation is obtained as follows:

$$\varphi = 7.98 \times 10^3 - 1.418 \times 10^{-4} x_1 + 0.1055 x_2 + 0.9085 x_3 + 1.406 x_4 \quad (1)$$

$$+ 5.518 \times 10^{-4} x_5 - 26.25 x_6 + 4.844 x_7 - 5.192 x_8 \quad (2)$$

Table 2 Linear Regression Parameter Estimates

Variables	Parameter Estimates	Standard Error	T value	Pr>|t|
Intercept	7.980e+03	1.025e+04	0.778	0.45291
x_1	-1.418e-04	1.878e-04	-0.755	0.46605
x_2	1.055e-01	5.298e-01	0.199	0.84576
x_3	9.085e-01	6.433e-01	1.412	0.18553

x_4	1.406e+00	7.863e-01	1.788	0.10127
x_5	5.518e-04	5.154e-04	1.071	0.30722
x_6	-2.625e+01	3.045e+01	-0.862	0.40714
x_7	4.844e+00	1.197e+00	4.045	0.00193
x_8	-5.192e+00	5.995e+00	-0.866	0.40492

The fitted coefficient R^2 reached 0.9995, adjusted to 0.9991, indicating that the explanatory variables effectively explain the incidence of hypertension.This statement that explains the incidence of hypertension very well. However, at a significance level of 0.05, from Table 2, it is observed that all independent variables except x_7 are not significant. Additionally, individual employment (persons), total retail sales of social consumer goods (100 million yuan), and R&D activities (100 million yuan) show a negative correlation with total GDP, which is contrary to common sense. Therefore, it is necessary to test the model for heteroscedasticity and autocorrelation of the error term. Check whether there is a linear relationship between variables and whether there is multicollinearity among explanatory variables.

3.2.2 Test for linearity of error terms

We used the plot function on the results obtained from the lm function, and then used the crPlots function to test for linearity. The evaluation of model fit based on Figure 1 is shown below.

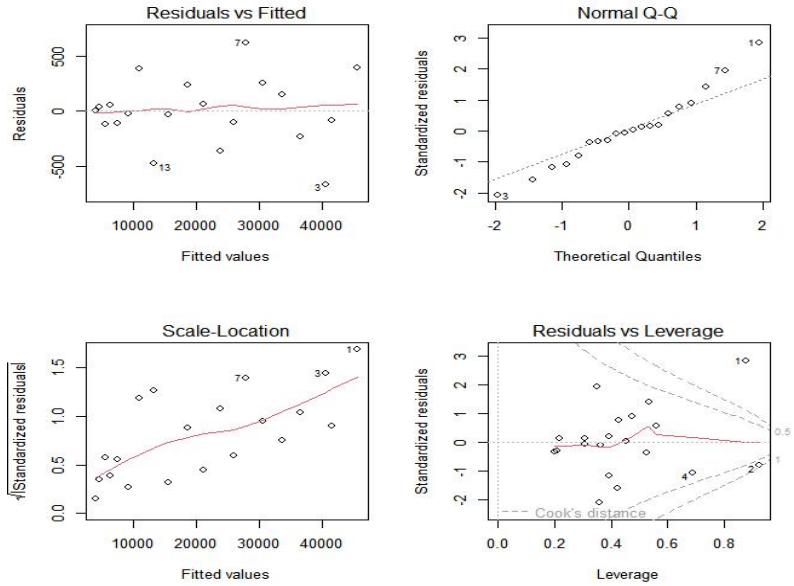

Figure 1 Linear Relationship Diagnostic Plot

3.2.3 Test for multicollinearity

The heteroscedasticity test and autocorrelation test passed, indicating no issues with the error terms. Considering the multicollinearity between variables, the variance inflation factor (VIF) was computed using the vif function from the car package. VIF measures the extent of variance inflation among explanatory variables. Generally, a VIF > 10 indicates severe multicollinearity. The diagnostic results are shown in Table 3:

Table 3 Results of Multicollinearity Test

	vif	Vif > 10
x_1	38.029722	TRUE
x_2	1292.832916	TRUE
x_3	402.769319	TRUE
x_4	2314.114228	TRUE
x_5	184.884440	TRUE
x_6	5.056438	FALSE
x_7	40.460161	TRUE
x_8	82.761979	TRUE

According to the data in the table, except for variable x_6, the variance inflation factor (VIF) values of the other independent variables are very high, clearly exceeding VIF > 10. This indicates significant multicollinearity among the explanatory variables that cannot be ignored.

4 RIDGE REGRESSION AND LASSO REGRESSION

4.1 Ridge Regression Modeling

From the earlier scatter plot matrix, it is evident that there is linear correlation among multiple sets of independent variables. Additionally, the VIF also indicates a significant multicollinearity issue among the independent variables. Therefore, the credibility of the multiple linear regression model built from these data is not high. The analysis of the model coefficients further confirms this point. Ridge regression is essentially an improved least squares estimation method that sacrifices some unbiasedness of least squares estimation to obtain regression coefficients that are more practical and reliable, albeit at the cost of losing some information and reducing precision.
Using the linearRidge function from the ridge package in R for ridge regression, simultaneously selecting ridge regression parameters. Utilizing the lm.ridge function from the MASS package to set parameter ranges, the ridge trace plot is obtained as shown in Figure 2, and the parameter estimates are shown in Table 4.

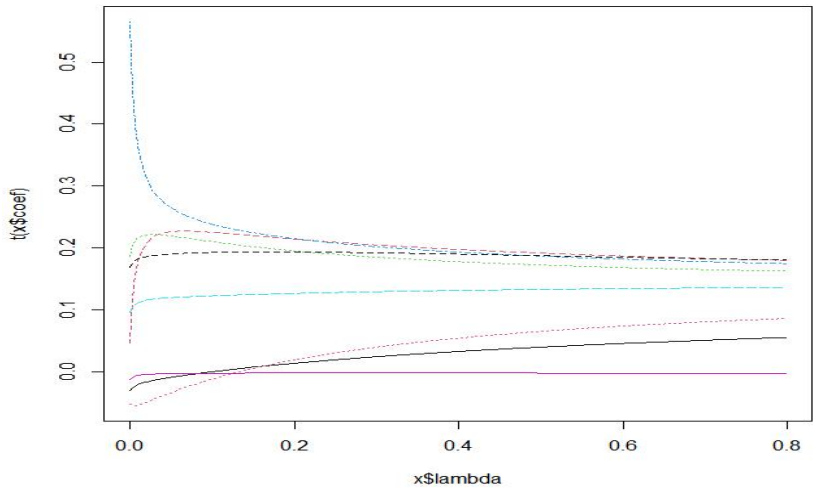

Figure 2 Ridge Trace Plot

Next, the regression equation obtained will be tested by randomly splitting the data into two parts: 70% of the data will be selected as the training set to train the neural network model, while the remaining 30% will serve as the test data to validate the model. The root mean square error (RMSE) will be used as the evaluation criterion. The RMSE for the training set and test set, as well as for the training set and test set after ridge regression, are shown in the table below:

Table 4 RMSE of Training Set and Test Set

DATA TYPE	RMSE
traindata	4.003×10^{-18}
testdata	0.147
Traindata(after ridge regression)	4.644×10^{-17}
Testdata(after ridge regression)	0.034

Based on the above data, the RMSE for the training samples and the test samples obtained from ridge regression are essentially consistent.

4.2 Lasso Regression

LASSO, first proposed by Robert Tibshirani in 1996, stands for Least Absolute Shrinkage and Selection Operator. It is a method based on the principle of shrinkage estimation. By constructing a penalty function, LASSO aims to obtain a more refined model that compresses certain regression coefficients, enforcing the sum of their absolute values to be less than a specific value. Additionally, it sets some regression coefficients to zero, thereby retaining the benefits of subset shrinkage. LASSO is particularly effective for biased estimation in data with complex collinearity.
Similar to ridge regression, LASSO transforms a constrained optimization problem into an unconstrained penalty function optimization problem by adding a penalty term. However, unlike ridge regression, LASSO does not yield an analytical solution. Nevertheless, its regression results assist in appropriate feature selection, making it advantageous compared to ridge regression.
Figure 3 illustrates the results of coefficient changes with parameter variations. The horizontal axis represents the ratio of model coefficients, the vertical axis represents the corresponding explanatory variables, dashed lines represent variables, and vertical lines denote penalty values.

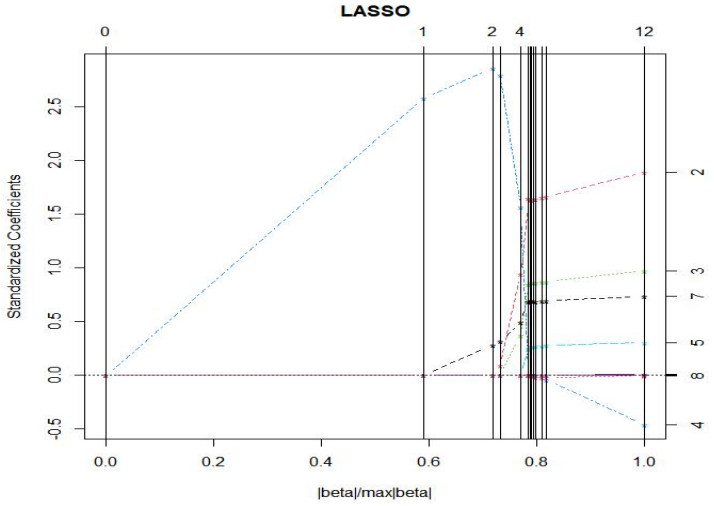

Figure 3 Lasso Regression Plot

Use cross-validation to select T value,Get the left picture in Figure 4.At this point we know that the recommended T value is around 0.8. At this time, it can be seen from the above figure: when the value of T is 0.84848, a model containing 7 independent variables is generated. The right figure in Figure 4 shows the estimated coefficients. We can clearly see that the 7 independent variables used to predict the dependent variable.
The model obtained using LASSO regression is as follows:

$$\varphi = 0.50x_2 + 0.26x_3 - 0.03x_4 + 0.008x_5 + 0.20x_7 - 0.01x_8 \quad (3)$$

Clearly, LASSO regression has identified x_2、x_3、x_4、x_5、x_7、x_8 to predict the dependent variable. These six variables are respectively the total value of goods imports and exports by foreign-invested enterprises, local fiscal expenditure, per capita consumer expenditure of residents, total water supply, electricity generation, and R&D activities. Among them, per capita consumer expenditure and R&D activities have a negative impact on the GDP of Hunan Province, but their coefficients are close to zero in absolute value, which aligns with the actual situation.

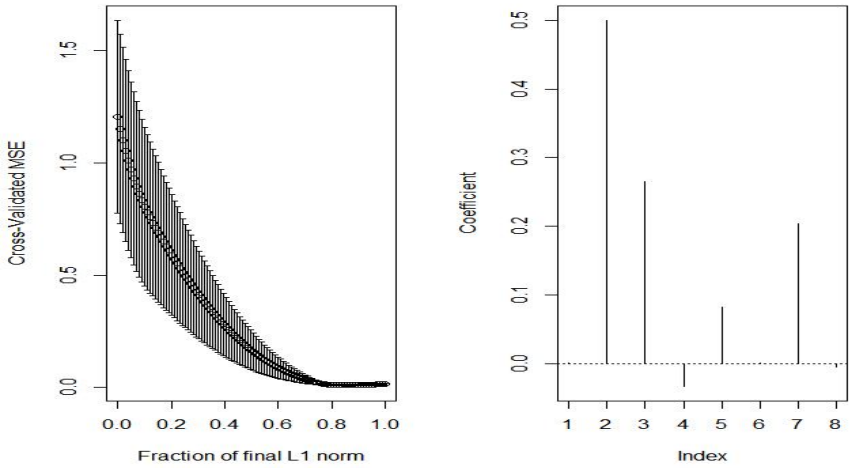

Figure 4 Cross-Validation Performance Plot

5 CONCLUSION AND OUTLOOK

5.1 Problems with this Article

5.1.1 Error term test subjective

In the heteroscedasticity test of the error term, only the method of observing the normal distribution chart is used to make judgments, and the image is relatively subjective and not very convincing. The model may have some effects of heteroskedasticity without improvement. Although the p value satisfies the condition of non-significant, it is initially believed that the error terms are independent. However, the DW value is 1.824. Although it is close to 2, it still shows that there is a weak, that is, part of the negative correlation. There is no moderate improvement here.

5.1.2 Exception points are not processed

On the one hand, outliers may be due to different sources of data collection, resulting in the data obtained in a certain year not being collected under the same standard. On the other hand, it would be too hasty to directly eliminate outliers. This article has not found a more suitable method to deal with these outliers.

5.1.3 Missing value imputation may be suboptimal

This paper uses the mean and median to replace missing data on prevalence. The advantage is that the number of samples is not reduced and random errors are not introduced. The disadvantage is that the variance of the variables calculated for non-random data is reduced. If the number to be supplemented is large, the standard deviation will be underestimated, leading to errors in the judgment of variable correlation. Need to be used with caution.

5.2 Conclusion

The article analyzes the influencing factors of Hunan Province's GDP through multiple linear regression, ridge regression and LASSO regression methods. By comprehensively comparing the results of the three analysis methods, we obtain the total import and export of goods by foreign-invested enterprises, local fiscal expenditures, and per capita consumption of residents. Expenditure, total water supply, and power generation all have a non-negligible impact on the total GDP of Hunan Province, and the total import and export of goods by foreign-invested enterprises, local fiscal expenditures, per capita consumption expenditure of residents, and power generation are all positively related to the GDP of Hunan Province, but R&D is negatively related to Hunan Province's GDP.

Therefore, in order to promote the healthy and sustainable development of GDP in Hunan Province, the following suggestions can be put forward:

(1) The total import and export of goods by foreign-invested enterprises has a significant positive correlation with the GDP of Hunan Province. Therefore, it is necessary to vigorously attract foreign enterprises to invest in Hunan and increase the total import and export volume.

(2) There is a significant positive correlation between local fiscal expenditure and Hunan Province's GDP. Local fiscal expenditures have a significant impact on promoting local employment, economic development, and infrastructure development.Therefore, we must find ways to increase local fiscal revenue so that we can more effectively increase GDP development.

(3) R&D can have a positive impact on the high-tech industry in Hunan Province. Educational experiments are investments that must be invested in costs. They are the key to truly solving the bottleneck problem Therefore, even if R&D will have a negative impact on total GDP, we cannot stop investing in it.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Li Bingxiao, Zhang Shiwei, Zheng Shuyu, Zhao Zhifan. Research on hydropower prediction based on a combined model of multivariate linear regression and ARIMA. Science and Technology Innovation, 2022(33): 71-74.

[2] Liang Haonan. Empirical study on factors affecting GDP in Anhui Province - Based on multiple regression analysis . Times Finance, 2021(24): 73-75.

[3] Ma Liyun. Ridge regression analysis of life insurance demand factors in my country. Modern Commerce and Industry, 2019 (05): 117-118.

[4] Shu Shuhua. Analysis of factors affecting household paper consumption in my country based on ridge regression . China Paper, 2022, 43(14): 48-51.

[5] Zhang Lei, Wu Hao. Multiple linear regression analysis of factors influencing employment numbers. Fujian Computer, 2022, 38(10): 12-16.

[6] Li Bingxiao, Zhang Shiwei, Zheng Shuyu, Zhao Zhifan. Research on hydropower prediction based on a combined model of multivariate linear regression and ARIMA. Science and Technology Innovation, 2022(33): 71-74

[7] Wang Peng, Cheng Wenshi. Research on the intensive land use and driving factors of cultivated land in Jiuquan City based on ridge regression model. Land and Natural Resources Research, 2022(04): 1-5.

[8] Zhang Qingxiu, Li Hongmei. Analysis of influencing factors of consumption demand in Hebei Province based on ridge regression. China Market, 2022(23): 23-27.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3037

A REFLECTIVE INQUIRY INTO LANGUAGE LARGE-UNIT TEACHING BASED ON CORE LITERACY

XiaoYu Liu*, JunJun Wu
College of Education, Capital Normal University, Beijing 100048, China.
Corresponding Author: XiaoYu Liu, Email: 1598593819@qq.com

Abstract: China's recent curriculum reform for primary education has placed a strong emphasis on teaching driven by core literacy. To this end, "large-unit teaching" has emerged as an innovative approach to teaching language in secondary school classrooms. Nonetheless, it is challenging to move at a snail's speed when teaching a substantial language unit. Large-unit instruction must be implemented creatively in the classroom, single-unit instruction must be used logically to create a ladder that will help students develop reading comprehension skills, and single-unit instruction must be viewed dialectically. It is blending everyday life, distilling essential ideas, constructing authentic scenarios, and structuring work teams. In order to accomplish the integration of teaching and research, we simultaneously fortify the framework of the school language large-unit teaching discipline, amass outstanding large-unit teaching experience, enhance teachers' control over large-unit teaching, and actively collaborate with colleges and universities.
Keywords: Core literacy; Large-unit instruction; Contextual tasks; College collaboration

1 INTRODUCTION

Large-unit Teaching is a teaching design idea based on the current educational goal of nurturing core literacy in countries worldwide through the continual development of unit teaching and classroom teaching changes. Large-unit teaching differs from the usual single-unit-oriented teaching model in enhancing students' overall abilities. It is dedicated to developing students' core academic literacy through global and systematic thinking, organizing and designing relevant contextual tasks, integrating learning resources, logically linking unit learning content, acquiring knowledge and skills, and developing conceptual understanding through experience and task completion.

2 THE RISE OF LARGE-UNIT TEACHING IN LANGUAGES

To completely understand large-unit teaching, we must first define unit teaching. The "New Education Movement" and the "Progressive Education" movements evolved in Western countries around the close of the nineteenth and beginning of the twentieth centuries, and the unit teaching approach gained popularity. In this context, Kobrecht introduced the design teaching technique, which emphasizes the creation of transdisciplinary units while focusing on children's mental representations and lifestyles. This pedagogy creates appropriate situations and conditions, stimulates interest, respects personality and the child's autonomy, and encourages active engagement. After introducing this teaching method in China, Chen Heqin and others experimented with it in early childhood. They provided a summary of the "Experiment-Reference-Publication-Review" sequence of actions for implementation. Xia Mianzun and Ye Shaojun edited the Hundred Eight Lessons of Chinese Literature in the 1930s. Each lesson includes four fundamental components: "Literary Words," "Selected Writings," "Grammar," and "Rhetoric." This allowed teachers to become fluent in Chinese while giving them a foundation for unit teaching. Following the establishment of New China, unit teaching was implemented, mainly using textbook units. Throughout more than a half-century, basic education research has reaped the benefits of unit writing and organizing in many textbooks and gained extensive experience in unit teaching, thereby establishing the foundation for large-unit instruction. Currently, China's ministry-edited language teaching materials are still structured in a unit setting, with the unit structure organized in a dual path combining content themes and language literacy and dispersing language knowledge and ability development, as well as thinking and habit formation. In the design of textual guides or practice questions for each unit with a logical structure that ranges from the simple to the complex. Such an architecture and configuration also provide chances and conditions for the growth of large-unit Teaching.

3 PROBLEMS IN THE IMPLEMENTATION OF LARGE-UNIT TEACHING IN LANGUAGES

Large-unit teaching has emerged as a new model that essential education areas are vying to explore in the current historical stage of growing curricular reform. However, when it is put into practice, how things stand and whether or not the desired outcome can be achieved make it a situation that needs to be monitored and understood constantly. For this reason, the author conducts a reflective inquiry on language large-unit teaching and conducts interviews on large-unit Teaching with several language teachers of different school grades in a specific city to understand the problems of large-unit Teaching in the actual teaching situation from the perspective of frontline teachers and to make relevant suggestions to improve language large-unit teaching and promote the realization of core literacy.

3.1 Deviation in Teachers' Concepts and Solidification of Contextual Settings

With the continuous evolution and significant transformations in education, many intricate new terms and concepts have surfaced in the language education sphere. Phrases like "core literacy," "whole book reading," "big concept teaching," "thematic reading," and "interdisciplinary teaching" have become focal points in contemporary educational research and application. These educational principles offer educators valuable insights and innovative avenues for Teaching, expand their pedagogical perspectives, and encourage diverse explorations of teaching methodologies and strategies. However, these new educational ideologies have also contributed to educational practices becoming overly internalized. Educators grapple with a constant influx of fresh terminology and theoretical frameworks, necessitating swift comprehension and application within their constrained teaching schedules, heightening their cognitive load and professional pressures.

Moreover, the overlapping and interconnected nature of these teaching concepts, each with distinctive focal points and practical contexts, often confounds educators, leading to misconceptions and partiality in their interpretation and implementation. Consequently, striking a harmonious balance between the swift evolution of educational theory and practice, facilitating educators' precise comprehension and seamless integration of these emerging teaching paradigms, and steering clear of unquestioningly adhering to trends and superficial concepts has emerged as a pivotal challenge in contemporary education reform and advancement. This scenario profoundly impacts teacher A, who is employed at a high school in a specific urban locale.

Nowadays, we educators are concerned that we may be falling behind, not so much in terms of performance but teaching methodologies and principles. While other educational institutions may have progressed significantly with school-based initiatives and other areas, we are still at a basic level regarding comprehensive Teaching and task-oriented learning. However, it is worth mentioning that teaching approaches are constantly evolving. You may only be somewhat familiar with one before it changes, and then you quickly adapt to the next one. These approaches seem similar, but there is still a slight gap. I often cannot be confident that my comprehension is flawless as uncertainties and ambiguities persist (Teacher A of a high school, 15 years of teaching experience).

The continuous introduction of novel teaching theories and the resemblances among these theories have placed stress on the comprehension and implementation of educators at the forefront, resulting in partiality in their interpretation of the evolving teaching theories. For instance, regarding the approach of teaching language in larger units, some educators may misinterpret it as focusing on the amalgamation of various texts, using an article as the focal point to guide the reading of multiple texts to enhance students' skills, which is mistaken for the concept of group text reading. Such misconceptions will likely lead to discrepancies in the subsequent planning and execution of teaching methods.

Large-unit teaching, which is task-oriented and emphasizes contextualization, relies on the learning context to facilitate large-scale learning. Language and text are central, originating from a specific historical context and intertwined with our mother tongue, conveying a tapestry of emotions, thoughts, and meanings that shape the work's framework. Students engaging with such material may need help to grasp the profound messages conveyed by the authors solely through textual cues, often missing the broader context created by the works. To address this challenge, it is crucial to pique students' interest by establishing a familiar context that immerses them in the study, deepening their comprehension and integrating newfound knowledge into their cognitive framework. This approach, as explored by Dai Xiaoe in "Situational Task Activity: An Exploration of Large Unit Teaching Towards Chinese Literacy,"[1] It is critical to fostering genuine learning experiences.

The setting and selection of context play a crucial and challenging role. Some educators establish a simplistic and rigid context that not only diminishes the allure of the language subject but also diminishes students' anticipation and engagement in the language course. Within such a rigid context, students memorize standardized response formats, which are scrutinized as "response templates" in secondary school language examinations. This criticized using "response templates" in secondary school language exams, which hinders the development of students' relevant skills and competencies within a broad teaching framework.

3.2 Difficulty in Implementing Classroom Teaching Due to the Varying Levels of Student Proficiency

As an innovative teaching mode, the core of large-unit Teaching lies in breaking the boundaries and framework of traditional unit teaching, which is no longer confined to the established chapters or units of the textbook but reorganizes and reorganizes the teaching content based on specific teaching themes or core concepts. This change in teaching mode requires educators to have a high degree of curriculum awareness and integration capabilities, to be able to analyze the content of the textbook in depth, to extract the knowledge and skills that have an intrinsic logical connection, and then build a more systematic, coherent and in-depth teaching system around the selected teaching theme. When implementing large-unit Teaching, teachers need to clarify the wholeness and coherence of the teaching objectives and ensure that the chosen theme can run through the whole teaching unit and become a link between different knowledge points and activities. In this regard, Lu Zhiping believes that "language large-unit teaching through the refinement of the relatively appropriate unit theme, and strive to explore the unit humanities theme and the organic links between various elements of language, the unity of several factors, so that the humanities is no longer separated from the language

learning process of a label."[2] For this reason, in the design of the large unit, for the sake of the thematic needs, the selection of the texts in a thematic unit may be a fusion of multiple topics; for example, there are texts selected in prose, argumentative essays, and literature, which, in the In practice, this is a great challenge to teachers' Teaching and students' learning. Ms. B, who has been teaching in a junior high school in a city for eight years, also talked about this point in the interview:

To design a large unit of Teaching, sometimes with the unit theme related to a variety of (genre) articles added, we are designed to enrich the content and even feel a great sense of accomplishment; I think that I set the theme of the particular good, so the Teaching is exciting, but you come to the classroom, you are dumbfounded. Some students think, wow, this class is so interesting. The teacher said the task was challenging, and he quickly entered the state. However, some kids get a big headache when they look at it. They cannot even read a text, and now they have different genres, tasks, difficulty levels, and activities added in. He feels that he cannot keep up, and often, he can only follow the "chorus" of his classmates, but he doesn't gain anything or very little. This significant unit teaching is difficult for children with poor fundamentals (Teacher B of a junior high school, eight years of teaching experience).

Under the guise of a class-teaching system, there exists variability in the foundational language proficiency levels among students within a class, coupled with varying learning aptitudes and reception. Disparities are observable in student performance within the encompassing classroom setting of large-group instruction and in the knowledge and skills they acquire. The theoretical concepts held by educators when formulating large-group instruction curricula need more effective implementation in practical teaching scenarios. Throughout instruction, the objectives of large-group Teaching gradually veer from their original intent, leading to a misalignment between students' classroom responses and performances, creating an overall skewed classroom dynamic. Furthermore, the comprehensive nature of large-group Teaching, entailing the integration of multiple chapters and the incorporation of diverse tasks throughout the learning process, places heightened demands on students' learning capabilities, presenting significant challenges to their linguistic proficiency. This scenario exacerbates the divide in students' language acquisition, as those with a solid foundation in literacy may achieve a "higher level" with integrated large-group instruction. In contrast, students lacking in the language knowledge base may exhibit low engagement levels, struggle to keep pace with classroom advancements, possess a shaky grasp of acquired knowledge, and ultimately experience a decline in learning efficacy over time. Consequently, prolonged exposure to such conditions could intensify student frustration toward language studies and dampen their enthusiasm for learning.

3.3 Lack of Support for School Teaching and Research and Lack of Teacher Experience

Large-unit teaching is not only a brand-new teaching concept and practice mode for teachers but also poses a significant challenge to the school's philosophy and school-based Teaching and research system. Under the traditional teaching mode, schools often arrange and manage the curriculum according to the established teaching material units. At the same time, large-unit Teaching requires schools to break this routine, reorganize the teaching resources more flexibly and innovatively, and design cross-disciplinary and cross-chapter teaching themes, which undoubtedly puts higher requirements on the schools' teaching organization and coordination ability. More importantly, most schools need more experience when they first try to implement large-unit Teaching in their schools. Due to the lack of precedent cases and mature experiences, schools often need help promoting large-unit teaching effectively, providing teachers with the necessary support and guidance, and assessing the effectiveness of their Teaching. This lack of experience increases the risk and uncertainty of school reform. It leads to confusion and frustration among teachers in practicing, thus affecting the smooth implementation and in-depth promotion of large-unit Teaching. Teacher C, who teaches in a junior high school in a city, talked about this:

When the new form of large-unit Teaching was introduced, the teachers in the school needed to learn more about it; after all, it was a new form, and we needed to engage in it more. Whenever we have a relevant teaching and research meeting, we are still very enthusiastic and active in giving our opinions on large-unit Teaching. Moreover, suppose the group wants to do a demonstration lesson on large-unit Teaching. In that case, the teachers will come together to revise the teaching design and give advice to the lecturer during the Teaching and research meeting. The main reason is that they need to gain experience. Some schools in the district are doing well, but they are adapting to their situation, so we may need help to use them. So, the older teachers don't have much experience, and the newer teachers have a lot of ideas, but it is different when they go to the class, and they need to gain experience, too (Teacher C of a junior high school, 17 years of teaching experience).

The advancement of comprehensive teaching methods and the evolution of our comprehensive teaching approach into a distinctive educational attribute necessitate extensive long-term knowledge accumulation and synthesis. Academic institutions need more pertinent teaching expertise. Consequently, subject-specific collaborative teaching and research remain confined to individual experiential boundaries. Many educators need more structured training in executing comprehensive teaching practices. Both educators and institutions are presently navigating through unfamiliar territory and can only enhance their understanding of comprehensive teaching by progressively accumulating relevant pedagogical insights through ongoing training.

4 PATHS TO OPTIMIZING THE REALIZATION OF LARGE-UNIT TEACHING OF LANGUAGES

According to the problems of large-unit Teaching of languages mentioned above, to optimize the teaching effectiveness of large-unit Teaching of languages, the following three aspects can be taken into account:

4.1 Understanding the Learning Situation, Using Single-article Teaching as a Ladder to Progressively Develop Large-unit Teaching

Large-unit teaching is considered a prevailing trend and is proposed as a holistic approach to conventional unit teaching methods, such as standalone lectures. The stance presented by the author is viewed as biased. It is contended that while large-unit teaching may be appropriate for certain educational institutions and all students, it poses a significant challenge for many students with inadequate language fundamentals. While real-life situations and social engagement are crucial, the concept of "indirect learning" holds equal, if not more significant, importance.[3]The cognitive development of students varies across age groups, with classroom lectures and practical exercises serving as fundamental means of knowledge dissemination. Through detailed explanations provided by educators, students passively absorb information, fostering their expertise and skills—an indispensable aspect of the learning process. Hence, in the realm of language education, deeply ingrained aspects of Chinese language, writing, and cultural literary knowledge highlight the language domain's capacity to exemplify tacit education's effectiveness. While extensive multi-chapter unit teaching can enhance students' overall competencies, it may compromise the in-depth analysis of texts and the appreciation of literary aesthetics. Consequently, the adoption of single-chapter teaching should be advocated as a foundational step, supporting subsequent learning progression.

In the field of education, the integration of individualized instruction and group instruction can be mutually beneficial. Educators can utilize individualized instruction to enhance students' understanding of language and text by engaging in thorough textual analysis to uncover the historical context of the works, the author's background, life experiences, national sentiments, ideals, and beliefs. As Zhu Xi once stated, "Instruction should be meticulous in its details and refined in its execution. Merely indulging in surface-level learning, akin to hastily consuming a lavish feast, is not conducive to learning."[4] Through individualized instruction, educators nurture students' reading comprehension and problem-solving skills, laying a solid groundwork for them to progress to more comprehensive group instruction. Additionally, teachers provide targeted support to students who may be struggling in language studies, assisting them in mastering effective reading strategies and fostering their ability to engage with and learn from group instruction sessions.

4.2 Extracting Core Concepts and Creating Contextualized Tasks

At the commencement of a significant unit in instructional design, educators must initially clarify that the term "large unit" does not denote "ample capacity" akin to "group reading." It does not pertain to the pace at which students read, nor is it solely about implementing the unit and structure of the textbook in a sequential manner. To enhance students' fundamental language literacy, large-unit teaching aims to entirely eliminate the adverse effects caused by exclusively "double-basic" teaching and the predominance of scientism rooted in unit teaching and to effectively eradicate the teaching inertia arising from the linear arrangement of linguistic knowledge points and the fragmented analysis of linguistic competence points.[5] Teachers ought to meticulously consider the standards' requisites, use the textbook as a resource but not entirely rely on it, comprehend students' actual needs in the language classroom, fuse objectives into the context where the content embodies the theme, activities reflect the methodology and practical application fosters the development of skills.

Addressing authentic problems, real-life scenarios, interests, and active engagement realizes core language literacy. Teachers should concentrate on extracting core concepts when structuring language instruction in extensive units. For instance, during a novel unit's learning process, educators may define "The novel as a literary genre centered on character portrayal, reflecting societal life through a comprehensive storyline and environmental depiction, encapsulating vast human history, amalgamating literary and ideological value" as the focal point of study. This core concept encompasses fundamental knowledge related to novel comprehension in the language domain, the primary route for students to grasp novels, and the essential significance of mastering novel reading. In former teaching practices, students mainly assumed a "passive receiver" role; in extensive unit teaching, students are situated as "active learners in exploration." Thus, when crafting contextual assignments, educators should align them with students' oral proficiency, cater to their cognitive capacities, consider language learning traits, and closely adhere to the language subject's demands.

4.3 Seek Cooperation from Universities and Realize the Integration of Teaching and Research

As pioneers in advancing educational theory and innovation, teacher-training institutions such as colleges and universities have amassed a wealth of high-quality educational resources by leveraging their deep expertise in educational disciplines and vast academic networks. These resources encompass state-of-the-art research findings in educational theory and encompass the latest developments and advancements in research, as well as cutting-edge

teaching methodologies and technologies. By engaging in continual scientific research and academic collaborations, teacher-training colleges and universities can facilitate the comprehensive integration of educational theory and practical application, ensuring that their research findings are both scientifically sound and hold practical value.

In this particular case, the research findings generated by educational institutions dedicated to teacher training serve not only to enhance comprehension of educational phenomena and principles but also to furnish substantial theoretical backing and methodological direction for educational implementation. Of particular significance to on-the-ground educational endeavors, these findings can be skillfully translated into specific instructional approaches, curriculum formation, or educational evaluation tools to enhance the instructional process, elevate teaching standards, and bolster holistic student advancement. Aside from addressing and sharing the challenges language educators face in executing comprehensive instructional units, there is also an opportunity to solicit support and guidance from a consortium of university professors to enhance frontline teachers' grasp of extensive instructional units. Furthermore, schools have the option to enlist the expertise of professionals and academics to periodically train their educators in comprehensive unit instruction, thereby perpetually enhancing their capacity to conduct extensive unit instruction.

Large-unit instruction in three key areas is crucial: firstly, identifying student needs at the outset, comprehending both the school environment and the student's proficiency level in the target language; secondly, avoiding hasty trends during implementation; and finally, conducting large-unit teaching and research within a task-based framework using contextual backgrounds. Emphasis is placed on task series within contextual settings to enhance students' cognitive skills. Following each teaching stage, reflections on challenges and successes should drive the creation of integrated college-supported teaching scenarios. This involves offering training to educators engaging in large-unit instruction, leveraging university partnerships to address practical issues and share experiences, thus building a repository of best practices and pedagogical insights for future large-unit teaching endeavors.

Currently, the primary language curriculum greatly emphasizes holistic learning, aiming to enhance students' language proficiency within authentic contexts and unifying educational materials to tackle the issue of fragmented instruction in classrooms. It also promotes the professional development of educators and steers language lessons towards a more coherent and efficient path of progress. This methodology represents a progressive transition towards a unified and dynamic language education framework that caters comprehensively to students' diverse needs. Seamlessly blending practical application with theoretical foundations enriches the learning process and fosters a more profound comprehension of language nuances. This curriculum will enhance language teaching to unprecedented efficacy and productivity by providing teachers with the essential tools and strategies to navigate this evolving educational landscape.

Large-unit teaching is a product of the era of core literacy, and this kind of creative teaching points to the necessary character, key abilities, and values that students should possess after learning day by day. It integrates the concepts of curriculum reform with the demands of talent cultivation in the new era. Based on the integration of curriculum content, it takes the learning of big tasks in real situations as the organization of the curriculum, to make students' learning a comprehensive, contextual, and experiential language practice activity. The design of large-unit teaching reflects its own systematic and internal structure of hierarchical order, which is in line with the teaching law of language as well as the learning characteristics of secondary school students. Also, it reflects the status of the learner as the main body of the implementation and evaluation of the curriculum. Only in this way can the drawbacks of fragmentation of subject knowledge points and traditional teaching be changed, and the learning objectives of core literacy be implemented into teaching through teaching design.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Dai Xiaoe. Situational Task Activity: Exploring extensive unit teaching towards Chinese Literacy. Basic Education Curriculum, 2019, 250(10): 7-11.

[2] Lu Zhiping. The pursuit of extensive Chinese unit teaching. Chinese Building, No. 2019, 419(11): 4-7. DOI: 10.16412/j.carol carroll nki.1001-8476.2019.11.003.

[3] Ren Haixia, Guan Ranrong. The text still needs "article reading" -- Cognition and reflection on "Big concept and big unit teaching". Chinese Teaching in Middle Schools, 2021, 502(04): 8-12.

[4] Zhu Xi. Zhu Zi's Reading Method. Zhang Hong, Qi Xi, Ed. Li Xiaoguo, Dong Liping. Tianjin; Tianjin Academy of Social Sciences Press, 2016, 209.

[5] Xu Peng. Reflection on Large Unit Teaching in the Context of Core Literacy. Chinese Teaching in Middle Schools, 2021, 502(04): 4-8.

Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3038

INDIVIDUAL ENJOYED TEACHING: THE DEMANDS OF TEACHING UNDER THE BACKGROUND OF THE GLOBAL COMMON GOOD

JunJun Wu *, XiaoYu Liu
College of Education, Capital Normal University, Beijing 100048, China.
Corresponding Author: JunJun Wu, Email: 2216556374@qq.com

Abstract: In 2015 UNESCO released a new report, Rethinking Education: Towards a Conceptual Shift towards the 'Global Public Good', which redefines knowledge, learning and education, emphasizes the humanist spirit of education, and suggests that education and knowledge should be defined as a 'common good' that requires the collective efforts of society. The "common good". However, in the existing teaching process, the lack of humanism and the proliferation of instrumental theories are still the curse of teaching, which makes individual enjoyment obscured, and the teaching falls into the situation of "mechanical thinking" and "man-made tools", so that the educational value of the global common good cannot be realized. The value of education for the global common good cannot be realized. Adhering to the spirit of the report, it is not difficult to find that individual enjoyment is the same as the report. Only by paying more attention to the individual enjoyment function of teaching and constructing an evaluation system for individual enjoyment can we break through the fence of "mechanical thinking" and get rid of the "man-made tools". The shackles of "man-made apparatus" will be broken through the fence of "mechanical thinking" and get rid of the shackles of "man-made apparatus", and it will also become a practical way to promote the common value of education for human beings.
Keywords: Individual enjoyed teaching; The global common good; Pedagogical claims

1 INTRODUCTION

In 2015, UNESCO released a new study, Rethinking Education: a Conceptual Shift Toward the Global Common Good? which reaffirms the core concept of a humanist view of education as a guide to education. Looking at teaching and learning in the light of the report's core concepts, we find that existing teaching and learning have become mechanical tools for inculcating knowledge. As Thomas Berry describes it, "To let children live only in connection with concrete, steel, wires, wheels, machines, computers, and plastics, and to let them experience hardly any original reality, not even to teach them to look up at the stars at night, is a deprivation of the soul that deprives them of the deepest experience of life.[1]" In other words, if our teaching imparts only one-sided, detached, fragmented knowledge that is outside the life and practice of society, such knowledge will not only fail to interest students, let alone enjoy them, and certainly will not promote the spirit of humanism, but it will also embark on the path of no return to atrophy. Thus, this paper takes the concept of the global common good as a perspective, diagnoses what is wrong with the existing teaching, and proposes the teaching of individual enjoyment on the basis of it, so as to answer the question: how should the teaching in the context of the era of globalization practice the spirit of humanism?

2 INTERPRETATION OF INDIVIDUAL ACCESS TO INSTRUCTION

Individual enjoyment refers to the use of something by an individual for material or spiritual fulfillment. When our teaching is infused with individual enjoyment, it also means that it must provide spiritual fulfillment to our students. Whether the individual enjoyment function of teaching is innate or acquired, we need to re-examine the concept of teaching in order to better understand the essence of individual enjoyment.

2.1 Review of the Concept of Teaching and Learning

The concept of teaching has evolved from the day it appeared to present a variety of interpretations. In order to understand the true meaning of teaching, we need to clear up the definition from these various interpretations.In Chinese vocabulary grammar, teaching consists of two words: teaching and learning,the meaning is that teaching focuses on imparting and receiving behavior, and learning favors inner feelings and gains. Taken together, we can grasp the meaning of teaching from the two dimensions of "imparting imitation" and "gaining in the heart".

2.1.1 The dimension of instructional technology

Education was created to satisfy the needs of life, so it is natural to think of education as a way of life or as a technique. It is for this reason that the nature of pedagogy can never be completely separated from the scope of the "normative disciplines". Therefore, "teaching" in pedagogy is undoubtedly regarded as a technology, a tool for acquiring knowledge through technology. As stated in the Chinese"Record of Learning":Preventing students from making mistakes before they occur is called prevention; Providing education at the appropriate time is called timely; Education that does not go beyond the talent and age characteristics of the educated is called conformity with order; Mutual learning and complementing each other's strengths and weaknesses is called mutual discussion. These four points are experiences of

successful teaching. In his teaching activities, Socrates in the West also used the question-and-answer method, thinking that the course of teaching was similar to the maternity technique, so he called his teaching method the maternity method. Plato, in his book "The Ideal State", while treating education as a tool of politics, also considered teaching as a technique of acquiring knowledge. In Herbart's book "General Pedagogy", although teaching is played out in more detail, it still focuses only on the method by which students receive the material.

These facts are clear examples of how teaching is viewed as a technology. It can be said that at this point in time, teaching is a "thing outside the body" independent of the individual, and the individual enjoyment of teaching has not yet really begun.

2.1.2 Teaching the dimension of individual enjoyment

Another view is that teaching involves people, their feelings and their values. For example, Confucius said, "It is not a pleasure to learn and to be learned. In the nineteenth century, under the flourishing of natural sciences and humanities, teaching and learning continued to be divided into different types, and people began to look at "teaching and learning" from a psychological point of view, reflecting the important influence of "enjoyment" on the individual within the individual. As a result, the individual enjoyment of teaching and learning has gradually become more valuable. For example, Rousseau's "Emile" proposes to use love to make children grow naturally, and advocates the implementation of the education of love rather than the application of coercion, which should be human-centered. After Rousseau, Johann Heinrich Pestalozzi brought Rousseau's education of love into full play. With the opening of modernization, we are faced with a complex world, both material and spiritual. A variety of philosophical trends are colorful, a variety of pedagogical pluralism coexist, but also on the "human" this proposition has never stopped discussing. Such as Whitehead on the "human" has a profound understanding, in his eyes, each student is "a living human being, is a creative and aesthetic interest in the concrete existence"[2]. Here, Whitehead will be "people" individual understanding of further deepening, and individual enjoyment function to a new level of understanding. Dewey after the twentieth century in the teaching practice, also fully interpreted the connotation of the individual enjoyment function and should achieve the goal. He believed that "education must begin with a psychological exploration of the energies, interests and habits of the child". In this way the child's learning becomes part of the child's life; it is no longer passive reception or listening, but active absorption or experience. "Going to school is a joy; child management is no longer a burden and learning is easier.[3]" It is thus clear that the individual enjoyment function of teaching exists objectively and has a historical tradition; it is not a far-fetched statement, let alone a subjective and artificial idea.

To sum up, only by giving full play to the enjoyment inherent in teaching itself can we satisfy the spiritual needs of the individual student.

2.2 Definition of Individual Access to Instruction

To summarize, for each individual, in addition to the acquisition of appropriate knowledge, teaching should also have a function of enjoyment, which points to the spiritual world within the "individual". Some scholars have summarized the essence of this function as follows: "In the process of receiving education, one obtains a sense of satisfaction and fulfillment of self-improvement, and experiences the freedom and happiness in education.[4]" Therefore, we call teaching with the function of individual enjoyment "individual enjoyment teaching".

The so-called individual enjoyment of teaching refers to the fact that teaching enables each individual to experience joy, happiness and a kind of spiritual enjoyment while acquiring the corresponding knowledge. This is in fact the kind of teaching that Confucius envisioned as being "enjoyable". Therefore, only by giving full play to the inherent enjoyment of teaching can we satisfy the spiritual needs of the individual student and enable him or her to have a happy experience while receiving knowledge.

3 DIAGNOSIS OF INDIVIDUAL ACCESS TO INSTRUCTION

In summary, the individual enjoyment function of teaching is innate. At some point, the individual enjoyment function of teaching has been obscured, and further restoration of this function requires a re-diagnosis of the current state of teaching.

3.1 Emotional Empathy of Indifference

As the process of globalization continues to advance, the multiplication of knowledge, information explosion makes people more and more in favor of technology, in favor of knowledge, resulting in the teaching of individual enjoyment of the function of obscuring, which leads to the blind pursuit of technological advances in educational activities, so the emphasis on the feelings of the pursuit of the spirit of the spirit of the spirit is ignored, so that the students have become emotionally unavailable devices. Cai Yuanpei once said: education is to help the educated person to give him the ability to develop his own, complete his personality, and human culture can do a part of the responsibility, not to be educated people into a special apparatus. However, when we are teaching knowledge, we still can't get out of the mechanical mode of thinking, and we eliminate emotions from our learning.

Narrow value orientation, so that the teaching of the lack of humanistic spiritual significance, so that the human emotions more and more indifferent, prompting our teaching has become a standardized "input - output" scientific assembly line, as if the teacher as long as to complete the process, will have completed the task of teaching. Such steps consistent teaching process, so that teachers with preset goals to "kidnap" the student's mind, to complete its mission of

educating people, with a thousand teaching routines, to the student's mind filled with a variety of "standard answers", become a rewrite knowledge "machine". This "machine" has no emotions, does not know how to question, and is afraid of making mistakes. When their spiritual needs are not satisfied from the teaching, feel the loneliness and emptiness brought by knowledge, they lose the enthusiasm for creativity and the joy of inquiry, the campus is full of indifferent faces exactly the same as their own, and what a sad picture!

3.2 Lack of Humanism

Whereas the humanistic spirit originally endowed the individual with subjectivity and was the cornerstone of the individual's enjoyment function, today we are facing the challenge of a lack of faith in the humanistic spirit. This challenge extends to teaching and learning, obscuring its function of individual enjoyment. Under the major mission of learning "knowledge", we have almost forgotten that the most fundamental connotation of education is to educate people.

As far as teachers are concerned, society lacks the necessary humanistic care for teachers, and still regards them as mere "pedagogues", believing that the task of teachers is to do a good job of teaching so that children can achieve high grades. Teachers feel the pressure, and thus in the teaching of "step by step", diligently, they can not feel from the teaching process to create the joy and happiness of interaction. As far as students are concerned, the loss of the humanistic spirit of certain courses of knowledge more and more alienated from the students' lives, easy to cause students to resist knowledge, naturally, no enjoyment can be said. Teachers and students in the field of knowledge in the diligent labor, but such labor is only superficial, ignoring the soil under the field, where the roots of plants reach out, but also ignored the wonderful scenery around. This stays at a shallow level of learning is undoubtedly to quench the thirst of plums, lack of emotion and value of the depth of learning makes the teaching process boring and tedious. It can be seen that our students in the most passionate golden years passively received too much, difficult to understand the knowledge instilled, poor to cope with a variety of duckling teaching and examination, this inhumane teaching, is the double strangulation of the students' souls and bodies.

3.3 The Ills of Instrumental Rationality

Since the Industrial Revolution, modern science has brought unprecedented changes to people's lives. People try every possible way to extract rich returns on knowledge, in this competitive game of interests, instrumental rationality is favored by the people, so in the field of education, the emergence of a "technology-centered" one-sided instrumental rationality orientation, showing excessive enthusiasm for new technologies and the blind use of the scene, people have to improve the technology to the efficiency of the classroom, the pursuit of knowledge teaching returns, "tainted" teaching is more like a norm, a technology. To the classroom for efficiency, the pursuit of knowledge teaching returns, "tasteless" teaching is more like a norm, a kind of technology. Under such a trend, such characteristic words as "exam-oriented education", "load-shedding" and "college entrance examination pledge" have come into being in the Chinese context, which in itself is paradoxical: the logical starting point of teaching itself is not purely for the purpose of learning and teaching. The logical starting point of teaching itself is not purely for the examination, when knowledge becomes the first, the status and value of knowledge overrides the faith, how can we talk about letting people to pursue poetry? Teaching, as a tool, has no sense of beauty, and when weighing material gain and spiritual enjoyment, people do not hesitate to choose the former and abandon the latter.

There is nothing inherently wrong with techno-instrumental rationality, because human beings have the need to pursue certain utilitarian values to create material civilization. "But this utilitarian value must not obscure the significance of knowledge for the spiritual world of man, nor threaten his pursuit of a meaningful life, or else knowledge will most likely be alienated into an antagonistic force for man's happy existence.[5]" We can't ignore the demand for examination, but within the framework of "examination oriented", teachers use traditional teaching methods and modern Internet means to process and package knowledge, so that students have an interest in active learning.When a person becomes a slave to the rules and regulations, he or she is not a creative and free being, trampling on the individual's right to enjoy teaching and learning, and extinguishing the individual's sense of innovation.

4 INDIVIDUAL ACCESS TO TEACHING AS A CLAIM TO THE CONCEPT OF THE COMMON GOOD

Lack of enjoyment of teaching has been in trouble, it is as a scholar in our country described, "education is to lead students to the 'living' road, rather than with inanimate knowledge of the accumulation of their original creative space, forcing them into a dead end.[6]" It seems that the return of individual enjoyment of teaching has been imperative, and we can no longer afford to be half-hearted. The return of individual enjoyment of teaching involves the spiritual world of each individual, and will inevitably put forward higher requirements for us.

4.1 Breaking Through the Fence of "Mechanical Thinking", Revitalizes Teaching and Learning

Society should be the forerunner in breaking down the fence of mechanical thinking, i.e., breaking down the utilitarian view of teaching and advocating an enjoyable value of knowledge teaching. In the conceptual transformation of the whole society, individuals will naturally be influenced by this general environment and regard teaching as an object that can be enjoyed.

Secondly, society should reshape the humanistic spirit. Today, material civilization has reached a certain level, in contrast to the lack of spiritual civilization. We should realize that material poverty may not mean spiritual poverty, but material wealth never means spiritual wealth. Therefore, the humanistic spirit advocated by society should be a spirit that makes people extremely rich in spirit and realize the meaning of "human" existence. In the teaching of knowledge, the individual who is imbued with the humanistic spirit is "a subject of knowledge who has awakened to his or her inner nature.[7]" With this humanistic atmosphere and culture as a foundation, the individual's function of enjoyment can be better realized.

Therefore, teachers should shift from the single teaching value of imparting knowledge to a teaching value that emphasizes both "teaching" and "enjoyment". The core concept should be: teaching is not "teaching" for the sake of accomplishing tasks, but "teaching" for the sake of spiritual fulfillment. Sometimes, looking at things from a different perspective may have a different effect. For example, when dealing with advanced students, teachers should hold the idea that they still have unlimited possibilities for improvement, and that turning such possibilities into reality is the embodiment of self-worth, the improvement of quality, and the ability to bring self-spiritual enjoyment. In this way, teachers will not only focus on the amount of knowledge taught to students, but also find the possibility of enjoyment from students, then teachers will go smoothly on the road of knowledge teaching.

Accordingly, students should change from a passive view of knowledge learning to a view of knowledge learning that emphasizes both "learning" and "enjoyment". The core concept should be: knowledge learning is the process of improving one's own quality and enriching one's own spiritual world. For example, when receiving new knowledge, students do not think too much about "what can I do with this knowledge", but should think more about "what aspects of this knowledge can bring me spiritual enjoyment". Only in this way can individual enjoyment of teaching return.

4.2 Freeing Teaching and Learning from the Shackles of the "Human Instrument"

Classroom as the most important form of teaching organization, related to individual enjoyment of the function can be effectively highlighted, so we need to pay attention to the art of classroom teaching, which requires every teacher in the state of mind, the momentum of the "heavy as light" style, and dare to pursue the highest realm of the art of teaching, and dare to challenge the authority, and dare to transcend the spirit of self! The spirit of teaching. In the preparation of each teaching link, they are like walking on thin ice, like being in an abyss, and carving carefully.

First of all, to enhance the art of classroom teaching, this art does not mean that the teacher to the lesson how fancy, ups and downs, but really through the teacher's thinking, rooted in the students' lives, according to the art of the person. Knowledge away from the students' life experience is difficult to arouse the interest of the students, and not let the students from scratch to construct knowledge, which is too cumbersome and time-consuming, not to directly instill the results to the students, which becomes the examination-oriented education. Rather, it is about carefully selecting the key, focal points in the process of knowledge generation and formation, and helping students to make connections between knowledge and their own experiences through this node. The art of teaching usually requires careful carving and polishing, and teachers need to exert great initiative to study and reflect. Teachers need to be able to master the classroom so that people can feel the satisfaction and happiness of creation through this artful activity. At the same time, when the teacher painstakingly to operate a kind of teaching art, which is itself a creative enjoyment. From the choice of teaching content to the selection of teaching methods, from the formation of teaching strategies to the growth of teaching wisdom, in the practice and reflection, the teacher's "artistic means" constantly renovated, "artistic approach" more clever, "artistic style". "Artistic style" is more distinctive, teachers and students can enter the "artistic realm" that brings infinite enjoyment.

Secondly, an equal teacher-student relationship should be established to rectify the "dominant" status of teachers in the past, and the relationship between teachers and students in teaching should be developed into one of equality, understanding and inter-subjectivity. This relationship is dynamic, it leaves room for students to play creatively, leaving teachers and students in the "you and me" resonance to expand the space for individual development, so as to enjoy the joy of growth. At the same time, this relationship is also democratic and open, the teaching process of the evaluation of students is no longer a perspective of a voice, under the pluralism to meet the needs of students in all aspects of development, "open to avoid monism, in order to avoid unidirectional linear, in order to be full of poetry.[7]"

The third should reflect the emotional care in teaching , knowledge teaching must be changed from single development of students' cognition to "knowledge" and "emotion". There is no profession like teaching, the spirit of interaction between life so often, so that people can get the joy of communication, so from the teaching design to teacher-student dialogue should be full of emotional color and rhythm of the heart. Mechanical explanation will only make the knowledge into a dry "chicken ribs", teachers want to get students' recognition, emotional resonance, not only to grasp the content of the teaching and the spirit of the organic integration of students, taking into account the interests and characteristics of students and pay attention to their own charisma to care for students. After all, do a "no love" teachers than to do a "no knowledge" of the teachers of the students more harm.

4.3 Creating "Individually Accessible" Contexts for Enjoying Teaching and Learning

In order for teaching to reflect enjoyment, it is necessary to create "individual enjoyment" scenarios in teaching, abide by the principles of induction, authenticity, proximity, cooperation, harmony and unity of conflict, hierarchy, act on students to awaken their emotional responses, create a learning atmosphere, and enjoy the process of exploring

knowledge. Johann Amos Comenius once said, "All knowledge begins with the senses", and famous soviet educator Zankovalso proposed that "knowledge that is not reinforced and warmed by one's positive emotions will make one apathetic". Teachers warm up knowledge with emotion, add love to make knowledge softer, and such a teaching situation can satisfy the psychological needs of students. The atmosphere of the classroom is no longer dead, serious and rigid, teachers and students between the emotional stimulation and collision set off a thousand waves, resonance of the soul, after the class can also be in the hearts of the students ripples, recalling the memory. Such an atmosphere can greatly stimulate students' interest in learning, to achieve the purpose of letting students enjoy the learning process, increase their emotional experience.

In order not to let the individual enjoyment function of teaching fail or even degenerate, firstly, we should create a situation close to the students' life and understand what they think and need. Secondly, we should pay attention to the image, effectively stimulate students' association and imagination, combine the characteristics of the subject, explore the charm of the subject, closely follow the teaching content, help students understand the abstract content of the books, stimulate students' interest, and master the true knowledge. Once again, it should contain valuable questions to guide students in a goal-oriented manner and adapt to the current cognitive level of students, as well as novel and vivid to stimulate students' desire for knowledge. Finally, and consistently, it should include emotions that motivate, inspire, and promote "meaningful learning" for students. We should also create the conditions, when we teach into the life to cherish, to enjoy, we will not deviate from our meaning as "human", in order to let the freedom, joy, happiness and the light of beauty shine on the road of education.

As the scholar Li Zhaocun said: "Teaching is no longer just a matter of knowledge acquisition and skill cultivation, but also a matter of children's happiness and freedom, the legitimacy and reasonableness of curriculum knowledge, the morality and ethics of the teaching process, and knowledge acquisition and spiritual education in knowledge acquisition, and so on, which will be presented in front of our eyes[8]. "In the face of these problems, the individual enjoyment function of teaching can no longer remain "behind the scenes". As the core of education, if the teaching of knowledge cannot effectively fulfill this function of enjoyment, how can we expect education to cultivate individuals who are truly self-reliant? Imagine if we feel more in front of the classroom teachers and students to the classroom knowledge teaching aspirations, in the classroom to hear the students of the classroom expression of happiness and happiness, which in itself is what a happy thing.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Thomas Berry. The Great Work: Our Way into the Future. Life - Reading - Xinzhi Sanlian Bookstore, 2005: 96.
[2] Robert S. Brumbaugh. Whitehead's process philosophy of education. New York: State University of New York Press, 1982: 124.
[3] John Dewey.Democracy and Education. Translated by Wang Chengxu.Beijiing:People’s Education Press, 2019: 50-211.
[4] Feng Jianjun. The individual enjoyment function of education. Shanghai Education Research, 2002, (01): 28.
[5] Li Zhaocun, Curriculum knowledge theory. Shanghai: East China Normal University Press, 2009: 7.
[6] L. Wang. Process Philosophy and the Way of University. The future of higher education: proceedings of the international symposium on process thinking and higher education reform. Changchun: Jilin People's Publishing House, 2007: 152.
[7] Wang Shuai. Poetic Knowledge and the Poetic Construction of Knowledge Teaching. Global Education Perspectives, 2010, (03): 21-22.
[8] Li Zhaocun. Epistemological foundations of reflective knowledge teaching. Global Education Perspectives, 2006, (11): 21.