Academic Education Publishing House

Consumption "Barometer": A Numerical "Tango" Between CCI, EEAI Indices and Household Expenditure

Shifeng Qian*, Haoran Zhao, Zongyang Wu
School of Liangxin, China Jiliang University, Hangzhou, China
**Corresponding Author*

Abstract: Against the backdrop of global economic integration and digital transformation, research on the behavioral patterns and influencing factors of household consumption – a core driver of economic growth – has become particularly significant. This study aims to delve into the impact of consumer confidence and e-commerce activity on household expenditure by constructing a standardized CCI and EEAI. It incorporates 19 indicators such as sub-item CPI data, urban and rural household income and expenditure from China's National Bureau of Statistics (2021-2025), along with 7 e-commerce metrics including online retail sales. Innovatively, this research proposes a hybrid modeling framework termed "PCA-ARIMAX-GBRT". By integrating PCA for dimensionality reduction, multiple lag regression models, ARIMAX for dynamic modeling, and the GBRT machine learning algorithm, it enables dynamic coupling analysis between macro-level psychological expectations and micro-level consumption behaviors.

Keywords: CCI; EEAI; PCA; ARIMAX; ML; Regression Model

1. Introduction

In today's digital and globalized economic landscape, statistical modeling, as a pivotal tool in economic research, is deeply embedded in all facets of economic analysis[1]. Leveraging advanced algorithms and data analytics technologies, it offers quantitative interpretations of complex economic phenomena, accurately captures the subtle interrelationships between economic variables, and holds profound significance for economic development forecasting and policy formulation evaluation[2]. As a core driver of economic growth, household consumption occupies a critical position in the national economic system. It not only reflects residents' quality of life and welfare levels but also is closely intertwined with the stability and prosperity of the macroeconomy[3]. Against the backdrop of the current slow global economic recovery and the accelerated adjustment of domestic economic structures, household consumption behaviors have exhibited notable changes, with the coexistence of consumption upgrading and consumption differentiation. Research into issues related to household consumption and analysis of the economic laws underlying consumption phenomena are crucial for formulating policies aimed at stabilizing economic growth, optimizing industrial structures, and improving social security systems. Such research helps the government implement targeted policies to achieve sustainable and healthy economic development. Meanwhile, for enterprises, it enables them to more precisely grasp market demands, optimize the supply of products and services, and enhance market competitiveness.

Existing consumer economy theories have limitations in explaining the dynamic relationships between residents' perceptions, confidence, and consumption behaviors. This study focuses on this domain, delves into the intrinsic connections among the three, and contributes to bridging theoretical gaps, improving the theoretical system of consumer economics, and providing new perspectives and theoretical foundations for subsequent research. Through empirical research utilizing the pioneering "PCA-ARIMAX-GBRT" hybrid modeling framework, this study uncovers the influencing factors and mechanisms of residents' consumption decisions under different economic environments, expands the boundaries of consumer behavior theory, and enhances its explanatory and predictive power for real-world economic phenomena. It provides robust data support for government departments in formulating macroeconomic policies. By accurately analyzing key indicators such as the Consumer Price Index and Consumer Confidence Index, the government can promptly discern changes in household consumption trends and formulate targeted fiscal, monetary, and industrial policies—such as precise price regulation, stimulating consumption growth, and optimizing industrial structures—to promote stable economic growth and harmonious social development. For enterprises, the research findings can assist them in gaining an in-depth understanding of changes in consumer demand, optimizing product research and development, pricing strategies, and marketing plans, increasing market share, and achieving sustainable corporate development. Additionally, it offers decision-making references for financial institutions, aiding them in the rational allocation of credit resources and the reduction of financial risks.

2. Literature Review and Theoretical Basis

Theory of Consumer Behavior: Drawing on classical consumption theories such as Keynes's Absolute Income Hypothesis[4], Duesenberry's Relative Income Hypothesis[5], and Friedman's Permanent Income Hypothesis[6], this paper elaborates on the relationship between household consumption behavior and income. These theories provide a basic framework for explaining household consumption decisions, emphasizing that income is a key factor affecting consumption, while also considering the impact of factors such as consumers' spending habits, relative income levels, and expectations of permanent income on consumption behavior.

Inflation Theory: Based on inflation theories such as the Phillips Curve and the Fisher Effect[7,8], this paper analyzes the interaction mechanism between inflation (measured by CPI) and household consumption behavior. Inflation affects residents' real income and consumption costs, thereby changing their consumption decisions and consumption structure; at the same time, residents' consumption behavior also feeds back on inflation, influencing fluctuations in price levels.

Theory of Consumer Confidence: With reference to Katona's theory of consumer confidence and expectation theory, this paper emphasizes the important role of consumer confidence in consumption decisions[9]. Consumer confidence reflects consumers' expectations of economic conditions, future income, and expenditure, and these expectations affect consumers' willingness and behavior to consume. When consumer confidence is high, people tend to increase consumption expenditure; conversely, they will reduce consumption.

This paper conducts an in-depth analysis of the deficiencies and gaps in current research in terms of theoretical frameworks, research methods, and empirical results. For example, some studies are insufficient in variable selection, failing to fully consider the impact of emerging consumption factors; the sample selection in some studies has limitations, leading to restricted generalizability of research results; most studies lack in-depth discussions on differences in consumption behavior among different regions and income groups. Based on this, the research direction of this paper is clarified: on the basis of making up for the deficiencies of existing studies, a research model will be constructed from new perspectives (such as introducing psychological expectation variables and considering regional heterogeneity), striving to achieve innovative breakthroughs in research methods and conclusions.

3. Data and Variable Description

3.1 Data Sources

This paper uses data from the official website of the National Bureau of Statistics of China, downloading and collating 60 months of data, including 19 confidence-related indicators such as sub-indices of the Consumer Price Index (CPI), urban and rural residents' income and expenditure, total retail sales of social consumer goods, and macroeconomic control variables (GDP and employment rate), as well as 7 e-commerce indicators such as cumulative online retail sales. The data acquisition paths, data release times, and data update frequencies are recorded in detail to ensure the accuracy and timeliness of the data.

3.2 Variable Definition

3.2.1 Consumer price index (CPI sub-indices)

This index is a relative figure reflecting the changing trends and degrees of price levels of goods and services consumed by residents over a certain period. It is subdivided into categories such as food, clothing, and housing. Its statistical scope follows the relevant regulations of the National Bureau of Statistics, and it is calculated by combining fixed-base indices and chain indices, which can accurately reflect price changes of sub-category goods and services.

3.2.2 Consumer expenditure

This indicator refers to the actual amount spent by individuals or households on purchasing goods and services within a certain period. As one of the important indicators for measuring economic activities, it can usually be obtained through direct observations. Consumer expenditure can be directly observed through various means, such as retail sales data, credit card transaction records, and transaction volume statistics on e-commerce platforms. These data sources can provide real-time and precise information on consumption behaviors.

3.2.3 Time variables

These variables are used to mark the time dimension in which data occur. They provide a temporal context for data analysis and help observe trends and cyclical changes. Time variables not only identify the time points of data but also facilitate the application of time series analysis methods (e.g., ARIMA models) to predict future trends.

3.2.4 E-commerce engagement activity index (EEAI)

The E-commerce Engagement Activity Index is a comprehensive indicator measuring user participation and business health of e-commerce platforms or stores, typically covering indicators such as cumulative online retail sales, cumulative online retail sales of physical goods, and current values of mail volume. As an important part of the current social consumption economy, its analysis helps make conclusions more comprehensive and reliable.

3.2.5 Other economic indicators

In addition to consumer expenditure and time variables, there are other macroeconomic indicators related to consumption. These indicators can help gain a more comprehensive understanding of consumption behaviors and their underlying driving factors, such as GDP, unemployment rate, inflation rate, and interest rate.

4. Model Construction and Data Analysis

4.1 Model Construction

4.1.1 Principal component analysis (PCA)

The Consumer Confidence Index (CCI) and E-commerce Engagement Activity Index (EEAI) are constructed by extracting key components from various economic data using the PCA method. As a dimensionality reduction technique, PCA transforms multiple correlated variables into a smaller set of uncorrelated principal components, thereby simplifying the data structure while retaining most of the information. Based on the variance explanation rate of the principal components, the top few components (typically those with a cumulative variance explanation rate reaching 80%-85%) are selected, and a composite score (i.e., CCI) is calculated according to their respective weights. This step ensures that the composite score can represent the main characteristics of the original data.

$$Var(PC_k) = \lambda_k \tag{1}$$

$$CR(PC_k) = x = \frac{\lambda_k}{\sum_{i-1}^{p} \lambda_i} \tag{2}$$

$$CCR(k) = \sum_{i=1}^{k} \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} \tag{3}$$

$$me(k) = \sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{p} \lambda_i \geq \alpha \tag{4}$$

$$Z_{\mathrm{i}}=(\mathrm{F}_{i1} * \lambda_1+\mathrm{F}_{i2} * \lambda_2+\cdots+\mathrm{F}_{ik} * \lambda_k)/(\lambda_1+\lambda_2+\cdots+\lambda_k) \tag{5}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0 \geq a_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{pmatrix}, a_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{pmatrix}, \cdots, a_p = \begin{pmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{pmatrix} \tag{6}$$

$$F_i = a_{1i}\mathrm{X}_1 + a_{2i}\mathrm{X}_2 + \cdots + a_{pi}\mathrm{X}_p, \mathrm{i}=1,2,\cdots p \tag{7}$$

$$Z_i=(\mathrm{F}_{i1} * \lambda_1+\mathrm{F}_{i2} * \lambda_2+\cdots+\mathrm{F}_{ik} * \lambda_k)/(\lambda_1+\lambda_2+\cdots+\lambda_k) \tag{8}$$

Through the Principal Component Analysis (PCA) of the Consumer Confidence Index (CCI) and Economic Environment Activity Index (EEAI) in Tables 1 and 2, the study reveals the following:: (1) The principal components of CCI are relatively scattered, reflecting that consumer confidence is subject to the comprehensive influence of multiple factors, particularly in terms of the service and employment markets, core inflation, and unemployment differentiation; (2) The principal components of EEAI are relatively concentrated, mainly driven by e-commerce retail and express logistics, followed by the impact of traditional postal services and international logistics; (3) The composite score of CCI fluctuates significantly, reflecting the uncertainty of consumer confidence, while the composite score of EEAI shows a steady upward trend, indicating the continuous vitality of the e-commerce industry. These analytical results provide an important foundation for subsequent regression analysis and model construction. In particular, when selecting features and explanatory variables, more targeted consideration can be given to the different principal components of CCI and EEAI.

Table 1. Contribution Rates of Principal Components of CCI and EEAI

	Individual contribution rate (%)	Cumulative contribution rate (%)	Naming of Principal Components
PC1	29.151	29.151	Fluctuations in the Service and Employment Market
PC2	19.752	48.904	Fluctuations in the Service and Employment Market
PC3	12.659	61.563	Core Inflation and Unemployment Differentiation
PC4	10.125	71.688	Producer Prices and Income Distribution
PC5	7.0827	78.77	Energy and Housing Costs
	Individual contribution rate (%)	Cumulative contribution rate (%)	Naming of Principal Components
PC1	62.057	62.057	Core Drivers of E-commerce Retail and Express Logistics
PC2	17.585	79.642	Traditional Postal Services and Intra-city Logistics
PC3	11.858	91.499	International Logistics and Reverse Fluctuations

Table 2. Principal Component Scores and Composite Scores (Confidence Index and E-commerce Engagement Activity Index)

	PC1	PC2	PC3	PC4	PC5	Composite score of CCI
2025.02	-1.8	3.2	1.0	0.8	2.2	0.601
2025.01	4.8	4.0	-0.4	-0.9	0.9	2.671
2024.12	-0.5	1.7	-1.3	-0.0	-0.1	0.030
2024.11	-3.8	1.7	-0.9	-0.0	0.1	-1.133
2024.10	-1.3	2.5	-0.9	-1.0	-0.4	-0.173

	PC1	PC2	PC3	Composite score of EEAI
2025.02	-0.368	-2.642	0.931	-0.637
2025.01	-0.369	-2.642	0.931	-0.637
2024.12	5.783	-0.188	0.689	3.975
2024.11	4.707	-0.292	0.541	3.207
2024.10	3.809	-0.700	0.298	2.488

4.1.2 Multiple regression model (including nonlinear terms)

(1) Extended Extended Features: In addition to CCI and EEAI themselves, squared terms of CCI and EEAI (CCI_sq), first-order lag terms (CCI_lag1), and time trends (Time) are introduced to capture the nonlinear relationships between CCI, EEAI, and consumer expenditure.

(2) Model Construction: A multiple regression model is employed, incorporating linear terms, quadratic terms of CCI and EEAI, and other extended features to construct a more complex regression equation. Through the estimation of regression coefficients, the impacts of different forms of CCI on consumer expenditure can be analyzed. The specific formula is constructed as follows:

$$\begin{aligned} y = \beta_0 + \beta_1 \cdot CCI + \beta_2 \cdot CCI^2 + \beta_3 \cdot CCI_{t-1} + \beta_4 \cdot \mu + \alpha_1 \cdot EEAI + \\ \alpha_2 \cdot EEAI^2 + \alpha_3 \cdot EEAI_{t-1} + \alpha_4 \cdot \mu + \varepsilon \end{aligned} \tag{9}$$

4.1.3 ARIMAX model

(1) Incorporating Time Series Characteristics: The ARIMAX model not only considers the direct impact of CCI and EEAI on consumer expenditure but also takes into account the autoregressive (AR) and moving average (MA) characteristics of the time series, as well as the influence of exogenous variables (such as CCI and its lagged terms).

(2) Dynamic Modeling: Through the ARIMAX model, the dynamic changes in the time series of consumer expenditure can be better captured, and future trends can be predicted[10]. The specific formula is constructed as follows:

$$\begin{aligned} y = \beta_0 + \beta_1 \cdot CCI + \beta_2 \cdot CCI_{t-1} + \beta_3 \cdot \mu + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ + \alpha_1 \cdot CCI + \alpha_2 \cdot CCI_{t-1} + \alpha_3 \cdot \mu + \sum_{i=1}^{p} \vartheta_i x_{t-i} + \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j} + \varepsilon_t \end{aligned} \tag{10}$$

4.1.4 Machine learning model (gradient boosting regression tree, GBRT)

(1) Feature Selection: Similarly, CCI, EEAI and their extended features (such as CCI_sq,

CCI_lag1, Time) are used as input features.
(2) Model Training: The Gradient Boosting Regression Tree (GBRT) algorithm is employed to gradually construct decision trees through iteration. Each tree is optimized based on the residuals of the previous tree, ultimately forming a powerful ensemble model.
(3) Performance Evaluation: The model performance is evaluated through cross-validation and test set, and indicators such as R-squared (R^2) and Mean Squared Error (MSE) are calculated to verify the predictive ability of the model. The specific formula is constructed as follows:

$$\hat{y} = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{11}$$

4.2 Data Analysis

Based on Figures 1, 2, and 3, the study shows that: (1) Relationship between CCI and consumer expenditure: CCI is positively correlated with consumer expenditure, with a quadratic term effect. (2) Relationship between EEAI and consumer expenditure: EEAI is also positively correlated with consumer expenditure, but its quadratic term effect is negative. (3) Comparison of model predictions: The prediction results of different models show differences in certain time periods, especially that GBRT performs less stably than multiple regression and ARIMAX on the test set. (4) Feature importance: Through the feature importance analysis of the GBRT model, it can be found that CCI, EEAI and their interaction terms are of high importance to model prediction.

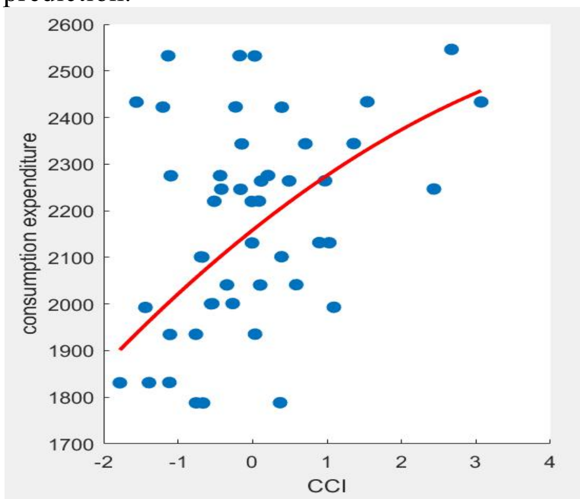

Figure 1. The Relationship between CCI and Consumer Expenditure

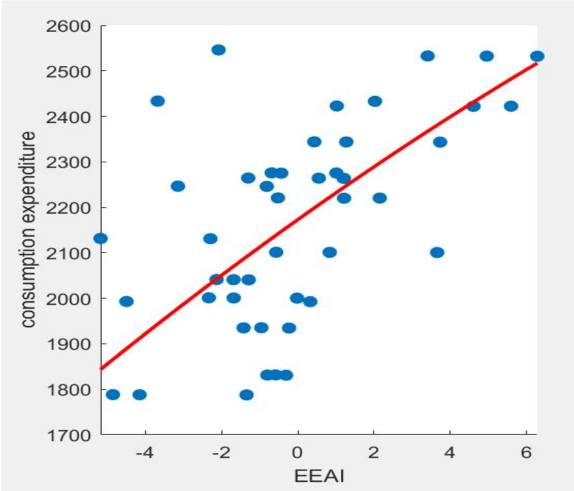

Figure 2. The Relationship between EEAI and Consumer Expenditure

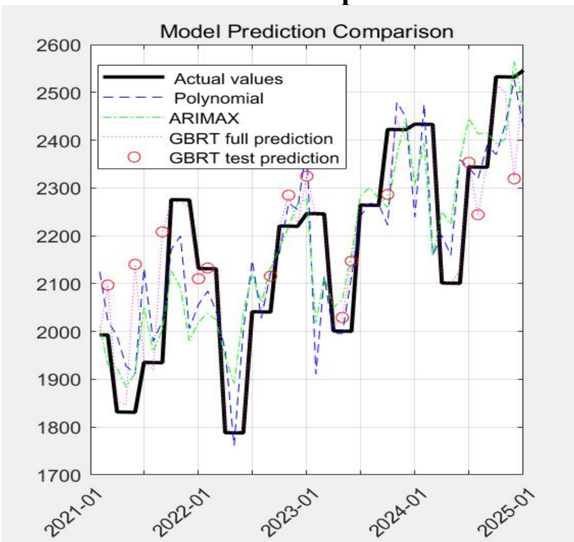

Figure 3. Comparison of Model Predictions

5. Regression Analysis

According to Table 3, the linear regression model is simple and easy to interpret, but its explanatory power is limited. The ARIMAX model takes into account the characteristics of time series, yet it does not significantly improve the prediction effect in this case. The GBRT model performs excellently, especially in capturing nonlinear relationships, and has a relatively high R-squared on the test set. To sum up, through comprehensive analysis that uses PCA to extract principal components, constructs CCI and EEAI indices, and combines linear regression, ARIMAX, and GBRT models, the study finds that there exists a certain relationship between CCI, EEAI, and consumer expenditure. In particular, the GBRT model performs outstandingly in predicting consumer expenditure and has high practical value. Future research can further explore the relationships between CCI, EEAI, and other economic indicators to improve economic forecasting models.

Table 3. Data of the Multiple Regression Model

CCI	0.43082 (p=1.2307e-09)
EEAI	0.48045 (p=1.2307e-09)
CCI	0.017201 (p=1.2307e-09)
EEAI	-0.0048246 (p=1.2307e-09)
CCI-EEAI	-0.17465 (p=1.2307e-09)
R^2	0.68638

6. Conclusion

6.1 Relationships between Consumer Confidence Index (CCI), E-commerce Engagement Activity Index (EEAI), and Household Expenditure

(1) Both CCI and EEAI exert a significant positive impact on consumer expenditure, with a certain interaction effect between them. (2) The coefficient of the quadratic term of CCI is positive, indicating that the impact of CCI on consumer expenditure strengthens as CCI increases; whereas the coefficient of the quadratic term of EEAI is negative, suggesting that the impact of EEAI on consumer expenditure weakens as EEAI rises. (3) The coefficient of the interaction term between CCI and EEAI is negative, indicating that their synergistic effect is weak and there may even be a certain offsetting effect.

6.2 Advantages of Machine Learning Models

(1) Excellent performance of the GBRT model: The R-squared of the GBRT (Gradient Boosting Regression Tree) model on the test set reaches 0.6684, significantly higher than that of the linear regression and ARIMAX models. This demonstrates that machine learning methods have obvious advantages in handling complex nonlinear relationships and time series data. (2) Feature importance analysis: The GBRT model can provide a ranking of feature importance, better reflecting the factors that have the most significant impact on household expenditure. For instance, CCI, the first-order lag of CCI, and other time trends may be key factors.

6.3 Policy Recommendations

(1) Application of threshold effect: Based on the significance of the coefficient of the quadratic term of CCI, specific CCI thresholds can be set. For example, when CCI is below a certain critical value (e.g., 105), the government can adopt measures to stimulate consumption and prevent economic downturns. (2) Timing of policy intervention: Due to the existence of nonlinear relationships between CCI, EEAI, and household expenditure, policymakers can take actions in advance according to the changing trend of CCI, rather than intervening only when CCI drops sharply.

6.4 Theoretical Contributions and Methodological Innovations

(1) Discovery of threshold effect: The study reveals a "threshold effect" between CCI and household expenditure, which helps to better understand consumer behavior patterns and provides theoretical support for macroeconomic policies. (2) Hybrid modeling framework: The proposed "PCA-ARIMAX-GBRT" hybrid modeling framework combines the advantages of multiple methods, which not only improves prediction accuracy but also enhances the interpretability of the model. This method can be promoted and applied in other similar studies.

6.5 Future Research Directions

(1) Incorporation of more factors: In addition to CCI and EEAI, other economic indicators (such as GDP growth rate and unemployment rate) can be considered to construct a more comprehensive model. (2) Long-term lag effect: Exploring the long-term lag effect of CCI on household expenditure may reveal more complex time-dependent relationships. (3) Cross-regional comparison: The relationship between CCI and household expenditure may vary across regions. Future studies can conduct cross-regional comparative research to uncover more regular patterns.

References

[1] Wang K Y. Research on the Application Problems and Countermeasures of Enterprise Management Accounting. Finance and Accounting for International Business, 2025, (07): 45 - 48.
[2] Yang J D, Gao Q N, Liu S Y. The Integrated Development of Digital Economy, Digital Product Space and Real - economy: From the Perspective of Coupling and Coordination of Technological Correlation and Comparative

Advantage. China Industrial Economics, 2025, (03): 78 - 96.
[3] Hou B W, Chen K T, Ji L, et al. Research Progress on Demand Limitation and Economic Growth. Journal of Yunnan University of Finance and Economics, 2025, 41(04): 38 - 51.
[4] Wen X X. The Impact of Urban Affordable Housing Acquisition on the Household Consumption of Agricultural Transfer Population. China Rural Survey, 2025, (02): 50 - 68.
[5] Zhou G L, Zhai X G, Zhang S Y. Research on the Boosting Effect of Digital Rural Development on Rural Residents' Consumption. Enterprise Economy, 2025, 44(02): 150 - 160.
[6] Ye X L, Sui L P, Sun R H. The Impact of Household Assets on Rural Residents' Household Tourism Consumption: Based on the Perspective of Income Mediation and Asset Change. Tourism Science, 2025, 1 - 17.
[7] Yao J. Empirical Research on China's Monetary Policy Rules from the Perspective of Non - linear Phillips Curve. Shenzhen University, 2019.
[8] Liu Z. Research on the Impact of Inflation on China's Stock Yield. Shandong University of Finance and Economics, 2012.
[9] Xing C. Research on the Marketing Strategy of Y Refrigeration Equipment Company. Dalian University of Technology, 2021.
[10] Gelper, S, & Croux, C. (2010). Robust forecasting with exponential and Holt-Winters smoothing. Journal of Forecasting, 29(3), 285–300.