ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 5 No. 1 (2025) 346-355
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Analysis of Household Income and Expenditure in Malaysia Using Principal Component Analysis and Maximum Likelihood Estimation

Alyaa Afza Atan[1], Mohd Saifullah Rusiman[1]*, Norziha Che Him[1], Muhammad Luqman Hakim Mislan[1]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: saifulah@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2025.05.01.041

Article Info

Received: 30 December 2024
Accepted: 21 January 2025
Available online: 30 July 2025

Keywords

Household Expenditure, Principal Component Analysis, Maximum Likelihood Estimation, Malaysia, Economic Analysis

Abstract

The relationship between household income, expenditure, and economic stability has been the focus of extensive research due to its implications for policy and economic planning. This study examines the underlying factors influencing household income and expenditure in Malaysia using Principal Component Analysis (PCA) and Maximum Likelihood Estimation (MLE). The objectives of the research are to identify the primary factors influencing spending behaviour, examine the relationships among expenditure categories, and compare the performance of PCA and MLE. Key findings highlight the significant role of income variables and socio-demographic factors in shaping expenditure behaviour. PCA was employed to reduce dimensionality and highlight significant components, while MLE focused on parameter estimation. Results indicate distinct differences in cumulative variance explained by both methods, showcasing the strengths and limitations of each approach. These findings contribute to the development of targeted economic policies that address the needs of diverse households.

1. Introduction

Household expenditure plays a pivotal role in shaping the economic landscape, as it directly reflects consumer behaviour and the allocation of resources within a country. It is not only a crucial indicator of economic stability but also an essential input for policymakers in formulating strategies to address economic inequalities and promote sustainable growth. According to [1], understanding expenditure patterns provides valuable insights into economic development and policy planning. This sentiment is echoed by [2], who emphasized that analysing household behaviours, especially among low-income groups, is vital for preparing for economic uncertainties.

The evolution of Malaysia's economic structure has brought significant changes to household expenditure components. A study by [3][4] highlighted that rising food and utility prices disproportionately affect low-income households, as they spend a larger percentage of their income on essential goods. Furthermore, [5] reported an increase in Malaysian household consumption expenditure per capita in 2022, underlining the importance of understanding how economic instability and price fluctuations impact discretionary income.

Empirical research has demonstrated that household income and demographic factors significantly influence spending behaviour. For instance, [6] revealed a strong correlation between household income and demographic attributes such as gender, ethnicity, and education level. These findings align with the work of [7], who explored the positive impact of government spending on education and healthcare in driving economic growth. Similarly, [8] confirmed the long-term correlation between public expenditure and economic advancement in Malaysia, underscoring the need for targeted policy interventions.

This study applied two advanced statistical methods, Principal Component Analysis (PCA) and Maximum Likelihood Estimation (MLE) are increasingly utilized to unravel the complex relationships within multidimensional data. It found MLE to be particularly effective in identifying latent constructs, providing a robust alternative to exploratory factor analysis [9]. PCA simplifies data by reducing its dimensionality, enabling researchers to identify primary components influencing spending behaviour while still retain most of the original information [10]. This is especially useful when there are many correlated variables. Other methods might require labels for classification tasks, whereas PCA is unsupervised and can be used for feature extraction in general. PCA, on the other hand, focuses on explaining the variance directly without assuming any hidden structure. Meanwhile, MLE offers a probabilistic approach to estimating latent factors, providing a nuanced understanding of the underlying structures in household expenditure data [11]. MLE can handle more general distributions, tends to be simpler than other methods and often provides similar results when no strong prior is available.

The economic landscape in Malaysia is characterized by rising living costs, regional disparities, and demographic inequalities that significantly affect household expenditure [12]. While past research [13] has explored the relationship between income and expenditure, gaps remain in identifying and understanding the latent factors influencing spending behaviour. Additionally, there is limited use of advanced statistical methods like PCA [14] and MLE to comprehensively analyse these patterns in the Malaysian context.

This study aims to identify the primary components influencing household expenditure patterns in Malaysia using advanced statistical techniques. It seeks to explore the relationships among various expenditure categories and demographic variables, offering a clearer understanding of the key drivers of spending behaviours. Furthermore, the research evaluates and compares the effectiveness of PCA and MLE in explaining variance and interpreting factors, providing insights into the suitability of these methods for complex economic analyses. By achieving these objectives, the study aspires to guide policymakers in addressing income disparities and improving household economic resilience.

2. Materials and Methods

2.1 Data Source and Variables

Data for this study shown in Table 1 were extracted from Malaysia's Household Expenditure Survey (HES) 2022 obtain from Department of Statistics Malaysia DOSM website (https://shorturl.at/Gpemm). The sample covers 8.4 million households, providing a comprehensive dataset categorized by income, expenditure types, and demographic attributes. Expenditure data were classified using the Classification of Individual Consumption According to Purpose (COICOP) framework.

Table 1 *List of classifications*

Group	Classification	Variable Type
01	Gender	Qualitative (Binary)
02	Certification	Qualitative (Ordinal)
03	Ethnic	Qualitative (Nominal)
04	Strata	Qualitative (Binary)
05	State	Qualitative (Nominal)
06	Marital status	Qualitative (Nominal)
07	Activity status	Qualitative (Nominal)
08	Industry of head of household	Qualitative (Nominal)
09	Occupation of head of household	Qualitative (Nominal)
10	Income Earn	Quantitative (Continuous)
11	Current Transfers	Quantitative (Continuous)
12	Property income	Quantitative (Continuous)
13	Net total	Quantitative (Continuous)

2.2 Statistical Techniques

2.2.1 Pre-Tests for Suitability

In this study, missing values were handled using mean substitution to ensure the completeness of the dataset. However, this approach may reduce the variance of the data and fail to capture any underlying patterns in the missing values. Outlier detection was performed using the z-score approach, which identifies extreme values by measuring how far they are from the mean. This method is effective for normally distributed data, but it may not work well for data with non-normal distributions and might overlook important variations in extreme values. Next, data normalization was applied to ensure that all variables had zero mean and unit variance, making them comparable across different scales. This is crucial for many machine learning algorithms, but it can be affected by outliers, which might distort the scaling.

After these preprocessing techniques were applied, this study used Kaiser-Meyer-Olkin (KMO) Test and Bartlett's Test of Sphericity before conducting factor analysis, such as Principal Component Analysis (PCA) or Maximum Likelihood Estimation (MLE), to ensure that the data is suitable for such analyses. KMO statistic quantifies the proportion of variance in the variables that might be caused by underlying factors. A KMO value greater than 0.6 indicates that the data are suitable for factor analysis [12][13]. Bartlett's is a test of sphericity. This test checks whether the correlation matrix is significantly different from an identity matrix. A significant Bartlett's test ($p < 0.05$) confirms that the variables are interrelated and appropriate for PCA and MLE [9].

2.2.2 Principal Component Analysis

PCA is a multivariate statistical technique widely applied for dimensionality reduction while preserving most of the dataset's variance [11]. It transforms correlated variables into uncorrelated principal components, which represent clusters of related variables and highlight the most significant patterns in household expenditure [12]. PCA involves assessing data suitability, extracting components based on eigenvalues observe from scree plot and the greater Kaiser Criterion, and enhancing interpretability using Varimax rotation to achieve a simpler and more interpretable structure [8][13]. The threshold values for eigenvalues were chosen based on the Kaiser criterion, where components with eigenvalues greater than 1 were selected because they explain more variance than a single original variable, and the significance of the results was tested using Chi-Square Test to ensure that the findings were not due to random chance.

PCA has been proven effective in uncovering patterns in expenditure data by summarizing key variables like income and household demographics [10][14]. Through the identification and combination of related variables into fewer components, PCA is an effective technique for reducing the complexity of data while preserving the greatest degree of variation present in the original dataset [15]. It's critical to understand that PCA works by converting the original variables into principal components, which are linear combinations of the original variables. The principal components analysis model is given as in equation (1) [16]:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}a_{3j} + \cdots + a_{im}x_{mj} \tag{1}$$

where, *z* is the component score, *a* is the component loading, *x* is the measured value, *i* is the number of components, *j* is the sample number and *m* is the total number of variables. For variance formula, in PCA each eigenvalue λ_i represents the variance explained by the *i-th* principal component. The total variance in the data is the sum of the eigenvalues as in equation (2) [17].

$$\textit{Total Variance} = \sum_{i=1}^{p} \lambda_i \tag{2}$$

2.2.3 Maximum Likelihood Estimation

MLE is another advanced statistical technique employed in this study, focusing on estimating model parameters by maximizing the likelihood function, ensuring that the observed data are most probable under the specified model [17]. In factor analysis, MLE identifies latent factors that explain relationships among observed economic variables, such as expenditure categories [18]. The threshold for significance is typically determined by likelihood ratio tests, which compare the fit of two models to assess if the additional parameters in the full model significantly improve the fit. The significance level (0.05) is used to determine if the likelihood ratio statistic is large enough to reject the null hypothesis, ensuring the results are not due to random chance [19][20]. Rotation, particularly Varimax, improves the interpretability of factor loadings by redistributing variance across factors. Unlike PCA, MLE adopts a probabilistic framework, offering a deeper understanding of latent structures influencing expenditure behaviours [21].

The method relies on a probabilistic framework, assuming that the observed variables are linear combinations of latent factors and unique variances. Key assumptions include multivariate normality and linearity. Initial eigenvalues are then examined using a scree plot and the Kaiser Criterion to decide the number of factors to retain. The factors are rotated using the Varimax method for better interpretability [22]. The maximum likelihood model as in equation (3) [22];

$$X = \Lambda F + \epsilon \tag{3}$$

where, X is the $p \times 1$ vector of measurements, Λ is the $p \times m$ matrix of factor loadings, F is the $m \times 1$ vector of common factors and ϵ the $p \times 1$ vector of residuals/error. For variance in MLE, the total variance of each observed variable X_i can be decomposed into common and unique variances as in equation (4) [17];

$$\Sigma = AA^T + \psi \tag{4}$$

where, AA^T is the communality (common variance explained by the factors) and ψ is the unique variance.

2.2.4 Pearson Correlation Coefficient

The Pearson correlation coefficient is use as it quantifies the strength and direction of the linear relationship between pairs of variables. By calculating these coefficients, variables that are strongly correlated and how they move together can be identify, which is crucial for understanding data patterns and relationships. This information can guide the selection of variables for further analysis, help validate assumptions and support the interpretation of results from methods like PCA and MLE, ensuring that the analysis shows meaningful associations within the data. The Pearson correlation coefficient quantifies the strength and direction of a linear relationship between two variables. It ranges from -1 to 1:

r = 1 : Perfect positive correlation (variables increase together).
r =−1: Perfect negative correlation (one variable increases as the other decreases).
r = 0 : No linear relationship.

The formula for the Pearson correlation coefficient, often denoted as *r*, that measures the strength and direction of the linear relationship between two variables *X* and *Y*. as in equation (5) [23];

$$r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 (Y_i - \bar{Y})^2}} \tag{5}$$

where, r is the Pearson correlation coefficient, *X* is independent variable, *Y* is dependent variable, $\bar{X}$ is the mean of the *X* values and $\bar{Y}$ is the mean of the *Y* values.

3. Results and Discussion

3.1 Principal Component Analysis and Maximum Likelihood Estimation

Bartlett's test must have a *p*-value less than 0.05 and the Kaiser-Meyer-Olkin (KMO) value more than 0.5 for the PCA and MLE test to be considered significant. The KMO value for this dataset is 0.630 and the Bartlett's test result of 0.000, and the test is considered acceptable

Table 2 Total Variance Explained of all components

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.972	22.858	22.858
2	1.654	12.721	35.579
3	1.309	10.070	45.649
4	1.203	9.257	54.907
5	1.024	7.877	62.783
6	0.978	7.527	70.310
7	0.868	6.676	76.986
8	0.802	6.171	83.157
9	0.659	5.070	88.227

10	0.647	4.975	93.203
11	0.608	4.677	97.879
12	0.177	1.364	99.243
13	0.098	0.757	100.000

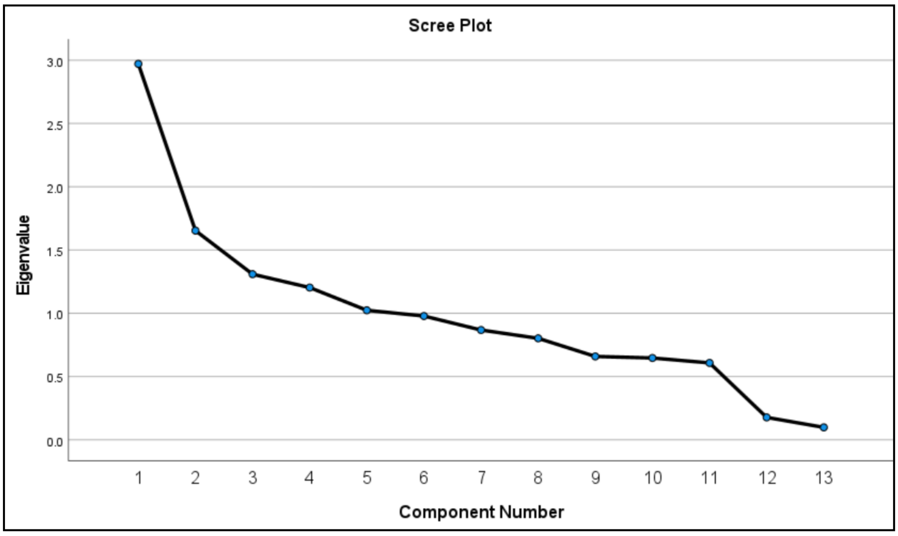

Fig. 1 Scree Plot

From Table 2, it is evident that the first five principal components have eigenvalues greater than 1, indicating they are significant based on the Kaiser criteria. Together, these components explain 62.78% of the total variance, with the highest fraction of the explained variance being 22.86% for the first component and the lowest being 7.53% for the fifth. As the data spread increases, so does the variance, emphasizing the role of these principal components in representing the dataset effectively. Fig. 1 shows the scree plot associated with this analysis likely shows a clear "elbow" after the fifth component, where eigenvalues begin to decline and form a straight line. If the explained cumulative variance of 62.78% is deemed sufficient, then the first five components can be considered adequate for the analysis. This approach aligns with the goal of identifying meaningful patterns while reducing dimensionality, offering a systematic method for selecting principal components.

3.1.1 Principal Component Analysis

Table 3 Rotated component matrix (PCA)

Variable	Component				
	1	2	3	4	5
Gender	.044	-.014	.167	.787	-.063
Certification	-.291	.275	.490	.002	-.013
Ethnic	-.197	.657	-.313	.155	.036
Strata	-.151	.736	.074	.035	-.039
State	.255	.053	.073	-.093	-.079
Marital status	-.047	-.052	.104	-.763	.010
Activity status	.043	-.061	.839	.015	-.015
Industry of head of household	-.217	-.601	-.306	.166	-.034
Occupation of head of household	.403	.012	-.419	-.285	.164
Income Earn	.896	-.153	-.189	.142	-.033
Current Transfers	-.042	-.018	-.046	-.060	.966
Property income	-.759	.159	-.009	-.123	.519
Net total	.858	-.144	-.143	.137	.212

From Table 3, Component 1 now distinctly represents economic factors, with strong loadings for Income Earn (0.896), and Net Total (0.868). Component 2 focuses on socio-demographic factors, including Ethnic

(0.657), Activity Status (-0.736), and Industry of Head of Household (-0.601). Component 3 captures variables related to employment and education, such as Certification (0.490) and the Occupation of Head of Household (-0.419). Component 4 highlights personal and demographic attributes, dominated by Marital Status (-0.763) and Gender (0.787). Finally, Component 5 includes variables related to residual income measures, such as Current transfer (0.966) and Property Income (0.519). Rotation redistributes the variance and enhances the clarity of relationships, as each component now represents specific themes such as economic factors, socio-demographics, employment, and personal attributes. This clearer structure simplifies interpretation, revealing the key dimensions as shown in Table 4.

Table 4 Category of PCA factors

Component	Factor
Component 1	Economic factors
Component 2	Socio-demographic factors
Component 3	Employment and qualification factors
Component 4	Personal and demographic attributes
Component 5	Residual income-related measures

Table 5 Total variance explained of extracted components by PCA

Component	Total Variance Explained					
	Extraction Sum of Squared Loadings			Rotated Sum of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.972	22.858	22.858	2.560	19.696	19.696
2	1.654	12.721	35.579	1.490	11.465	31.161
3	1.309	10.070	45.649	1.419	10.918	42.079
4	1.203	9.257	54.907	1.401	10.776	52.855
5	1.024	7.877	62.783	1.291	9.928	62.783

From Table 5, the Extraction Sum of Squared Loadings indicates that the first five components collectively explain 62.78% of the total variance in the dataset. This implies that the five components capture most of the information contained in the variables. The first component alone explains 22.86% of the variance, highlighting a dominant factor likely related to income and economic variables, as evidenced by their high communalities. Subsequent components explain decreasing proportions of variance while the second component accounts for 12.72%, and the third, 10.07%. This suggests a hierarchical structure where the first few components represent the most influential aspects of spending behaviour.

From Table 5, after Varimax rotation, the variance becomes more evenly distributed across the five components. The first component now explains 19.70%, and the second accounts for 12.72%, indicating a clearer distinction between the factors. Rotation aligns variables with specific factors, enhancing interpretability. For instance, income-related variables cluster under one component, while demographic or household variables group under others. This separation reduces overlap among components and improves clarity in identifying the factors influencing spending.

3.1.2 Maximum Likelihood Estimation

Table 6 Rotated component matrix (MLE)

Variables	Component				
	1	2	3	4	5
Gender	.004	-.063	.099	-.003	.741
Certification	-.276	-.029	.212	.258	.015
Ethnic	-.159	.065	-.167	.352	.064
Strata	-.137	-.029	-.002	.777	.026
State	.129	-.040	.034	-.029	-.016
Marital status	-.042	.042	.045	.003	-.430
Activity status	-.022	-.044	.854	.035	.004
Industry of head of household	.005	-.011	-.105	-.233	.043
Occupation of head of household	.299	.080	-.244	-.078	-.172

Income Earn	.915	-.060	-.131	-.133	.125
Current Transfers	.009	.739	-.046	.009	-.110
Property income	-.756	.629	-.073	.121	-.106
Net total	.893	.198	-.086	-.112	.122

From Table 6, the Rotated Factor Matrix provides a clearer interpretation of the underlying dimensions among the variables. Factor 1 captures economic factors, with strong positive loadings for Income Earn (0.915) and Net Total (0.893), while Property Income (-0.756) negatively contributes, highlighting an inverse relationship. Factor 2 focuses on income from Current Transfers (0.739) and Property Income (0.629), distinguishing these financial sources from direct earnings. Factor 3 represents Socio-demographic factors, driven by Activity Status (-0.854) and Strata (0.777), reflecting the interaction between employment and residential stratification. Factor 4 emphasizes education and occupational roles, with contributions from Certification (0.258) and Occupation of Head of Household (-0.244). Lastly, Factor 5 captures demographic attributes, dominated by Gender (0.741) and Marital Status (-0.430). This structure simplifies the complex relationships in the dataset, consolidating variables into distinct themes such as income, socioeconomic factors, and demographics, which can guide targeted analysis or interventions as Table 7.

Table 7 Category of MLE factors

Component	Factor
Component 1	Economic factors
Component 2	Residual income-related measures
Component 3	Socio-demographic factors
Component 4	Employment and qualification factors
Component 5	Personal and demographic attributes

Table 8 Total variance explained of extracted components by MLE

Component	Total Variance Explained					
	Extraction Sum of Squared Loadings			Rotated Sum of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.312	17.784	17.784	2.435	18.733	18.733
2	1.497	11.512	29.296	1.007	7.745	26.477
3	.764	5.877	35.173	.918	7.065	33.542
4	.813	6.252	41.425	.902	6.939	40.481
5	.702	5.397	46.822	.824	6.342	46.822

From Table 8, the Extraction Sum of Squared Loadings for MLE reveal that the first five lantent factors account for 46.82% of total variance, with the first factor contributing 17.78%. Compared to PCA, the variance explained is lower due to the stricter assumptions of MLE, which aims to optimize the likelihood of the observed correlations rather than maximize variance explained. Like PCA, rotation redistributes the variance more evenly among factors. The first factor now explains 18.73%, and the second contributes 7.75%, reflecting a more balanced representation of variables. The rotated solution better isolates distinct factors, such as economic and household demographics, providing clearer insights into the underlying dimensions of spending behaviour.

3.2 Pearson Correlation Coefficient

Fig. 2 reveals significant relationships among expenditure categories, with "Income Earn" showing the strongest positive correlation with "Net Total" (r=0.85). This indicates that income is a critical factor in determining household financial stability, as higher earnings directly contribute to an increased financial balance. Additionally, a weak positive relationship between "Income Earn" and "Property Income" (r=0.10) suggests that higher earners might generate some additional income from property investments, though the effect is marginal.

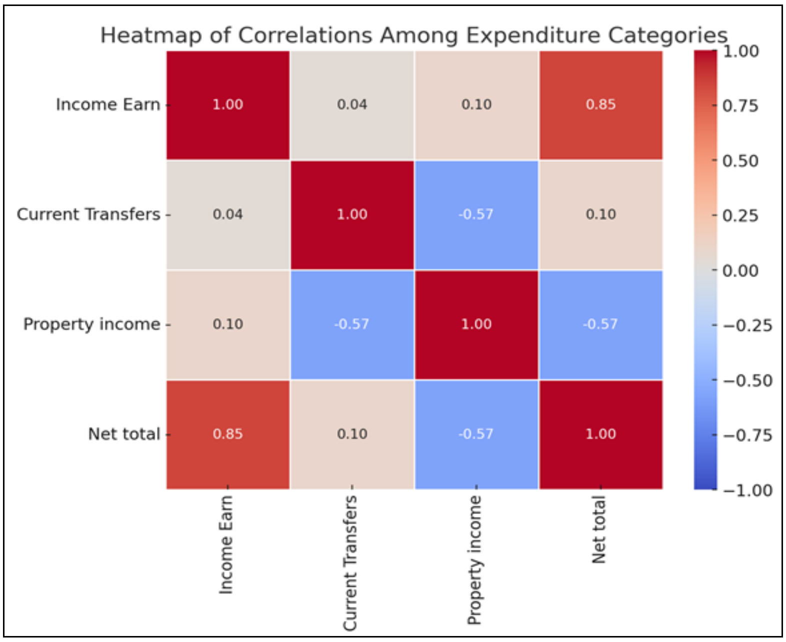

Fig. 2 Heatmap of Correlations

On the other hand, there are notable negative correlations. "Property Income" and "Current Transfers" exhibit a strong negative relationship (r=−0.57), suggesting a trade-off where households receiving higher external financial assistance (such as pensions or remittances) are less reliant on property income. Similarly, the negative correlation between "Property Income" and "Net Total" (r=−0.57) implies that households with higher property income might reinvest their earnings or incur related expenses, reducing their overall net balance. A weak positive relationship between "Current Transfers" and "Net Total" (r =0.10) indicates that financial aid or transfers contribute slightly to improving household financial standing. These findings highlight "Income Earn" as the primary driver of financial well-being, while "Current Transfers" and "Property Income" represent distinct financial strategies. Households' dependent on current transfers may have less access to property income, while investment-oriented households may prioritize property-related expenditures over maintaining a high net total.

3.3 Comparisons (PCA vs MLE)

Table 9 Cumulative Variance

Method	Cumulative Variance Explained (%)	Rotated Cumulative Variance Explained (%)
PCA	62.78	62.78
MLE	46.82	46.82

Based on Table 9, MLE explained a lower cumulative variance compared to PCA, reflecting its focus on parameter estimation rather than variance decomposition, PCA explains 62.78% of the total variance, whereas MLE explains only 46.82%. This indicates that PCA is more efficient in capturing the overall variance in the data compared to MLE. Interestingly, the rotation does not affect the cumulative variance for either method, as the values remain the same after rotation. This suggests that while rotation redistributes the variance across the factors to improve interpretability, it does not increase the total variance explained.

Table 10 Sum of Squared Loadings for each Factors Extracted

		Sum of squared loadings			
		PCA	PCA (Rotated)	MLE	MLE (Rotated)
Factor	1	2.972	2.560	2.312	2.435
	2	1.654	1.490	1.497	1.007
	3	1.309	1.419	0.764	0.918
	4	1.203	1.401	0.813	0.902
	5	1.024	1.291	0.702	0.824

From Table 10, it shows comparisons of the sum of squared loadings for factors extracted using Principal Component Analysis (PCA) and Maximum Likelihood Estimation (MLE), both before and after rotation. Factor 1 accounts for the highest variance across methods, with PCA explaining the most (2.972) and MLE slightly less (2.312), though rotation redistributes this variance (PCA to 2.560 and MLE to 2.435). For subsequent factors, the variance explained decreases, with PCA generally capturing more variance than MLE for each factor. Rotation further balances the variance across factors, improving interpretability by redistributing the explained variance such as Factors 3–5 show slight increases post-rotation for both methods). Overall, PCA prioritizes maximizing variance, while MLE focuses on fitting a probabilistic model, with rotation enhancing factor clarity in both cases.

4. Conclusion and Recommendations

4.1 Conclusion

This study provides in depth exploration of household expenditure patterns in Malaysia, utilizing advanced statistical methods to uncover the key factors influencing household expenditure. The findings highlight economic measures as the most significant determinant, with income variables such as net income and property income playing a central role. Policymakers can leverage these findings to create more targeted interventions, particularly in addressing income disparities and promoting policies that foster equitable economic growth.

Through a comparative analysis of PCA and MLE, the study established PCA's superiority in explaining variance and simplifying component interpretation. The cumulative variance of 62.78% achieved by PCA underscores its efficiency in capturing the underlying structure of the data. In contrast, MLE's shows a lower cumulative variance of 46.82%, though it provided nuanced insights into latent factors. The rotation process in both methods further enhanced interpretability, offering clearer groupings of variables and revealing distinct dimensions of expenditure behaviour. In conclusion, based on the findings, Principal Component Analysis (PCA) is considered the better method as it explains a higher percentage of variance (62.78% compared to 46.82% by MLE), effectively simplifies complex data, and provides clearer and more interpretable results for identifying key patterns in household expenditure. This research contributes to the growing body of knowledge on household expenditure by demonstrating the utility of PCA and MLE in economic analysis. The findings not only enhance our understanding of spending behaviour in Malaysia but also provide a foundation for developing effective strategies to address economic challenges and foster sustainable growth.

4.2 Recommendation

By integrating these findings into future policy frameworks, the government could design more effective programs that focus on income redistribution, education accessibility, and gender equality. This includes providing subsidies and tax incentives for low-income households to alleviate financial burdens on essential goods and services. Additionally, enhancing financial literacy through educational campaigns can empower individuals to manage their resources effectively, fostering long-term financial resilience. Furthermore, leveraging advanced statistical tools such as PCA in regular analyses can help identify emerging expenditure trends, enabling data-driven policy formulation. Finally, the study also highlights the need for ongoing research in this domain. Future studies could incorporate longitudinal data to track changes in expenditure patterns over time and explore additional variables, such as cultural and psychological factors, to provide a more comprehensive understanding of household financial behaviour. By leveraging advanced analytical techniques and expanding the scope of research, valuable insights can be generated to guide policy decisions and improve household well-being.

Acknowledgement

The research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Multidisciplinary Research Grant (MDR) Vot Q702 and GPPS Grant Vot Q668.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Alyaa Afza Atan, Mohd Saifullah Rusiman;* ***data collection:*** *Alyaa Afza Atan* ***analysis and interpretation of results:*** *Alyaa Afza Atan, Mohd Saifullah Rusiman, Muhammad Luqman Hakim Mislan;* ***draft manuscript preparation:*** *Alyaa Afza Atan, Mohd Saifullah Rusiman, Norziha Che Him. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Deaton, A. S., Ruiz-Castillo, J., & Thomas, D. (1989). The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence. *Journal of Political Economy*, *97*(1). https://doi.org/10.1086/261597

[2] Hashim, N., Mustapha, N., Kulub, N., Rashid, A., & Nasir, A. (2011). Household Income and Expenditure Relationships: A Simultaneous Equation Approach. In Chinese Business Review (Vol. 10, Issue 6).

[3] Ismail, R., & Tendot, N. A. B. (2012). The relationship between income, expenditure and household savings in Peninsular Malaysia. *Malaysian Journal of Consumer and Family Economics*, *15*(1).

[4] Aguiar, M., & Hurst, E. (2005). Consumption versus expenditure. Journal of Political Economy, 113(5). https://doi.org/10.1086/491590

[5] Statista Research Department. (2024). Malaysia: Household expenditure per capita 2022. Statista.

[6] May, T. K., Saharan, S., & Rusiman, M. S. (2023). Analysis of Income and Expenditure of Households in Peninsular Malaysia. https://doi.org/10.1007/978-981-99-2850-7_17

[7] Lee, J. C., Won, Y. J., & Jei, S. Y. (2019). Study of the relationship between government expenditures and economic growth for China and Korea. Sustainability (Switzerland), 11(22). https://doi.org/10.3390/su11226344

[8] SING YUN, W., & YUSOFF, R. (2019). REVISITING LONG-RUN DYNAMICS OF THE DETERMINANTS OF PUBLIC EDUCATION EXPENDITURE IN MALAYSIA. Journal of the Asian Academy of Applied Business (JAAAB), 5. https://doi.org/10.51200/jaaab.v0i0.1781

[9] Greenacre,Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1).https://doi.org/10.1038/s43586-022-00184-w

[10] Youngblut, J. M., & Bolton, F. P. (2013). Comparison of Factor Analysis Options Using the Home/ Employment Orientation Scale.

[11] Norhisham, M. A. N., Rusiman, M. S., & Che Him, N. (2023). A Multivariate Analysis on the response of crop to fertilizer and soil type. Enhanced Knowledge in Sciences and Technology, 3(1), 216–225.

[12] A. A. Saad, S., Adam, A., & H. Abdelateef, A. (2016). The relationship between household income and educational level. (south Darfur rural areas-Sudan) statistical study. International Journal of Advanced Statistics and Probability, 4(1).https://doi.org/10.14419/ijasp.v4i1.5655

[13] Wu, L., & Zhang, X. (2022). Household expenditure patterns in urban China: An exploratory factor analysis. Journal of Economic Studies, 49(3), 456–470.

[14] Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. Tutorials in Quantitative Methods for Psychology, 9(2).https://doi.org/10.20982/tqmp.09.2.p079

[15] Sedgwick, P. (2012). Pearson's correlation coefficient. BMJ, 345(jul04 1), e4483–e4483. https://doi.org/10.1136/bmj.e4483

[16] Grant-Jacob, J. A., Zervas, M. N., & Mills, B. (2024). Spatial analysis of femtosecond laser generated plasma using principal component analysis. Scientific Reports, 14(1), 30301. https://doi.org/10.1038/s41598-024-81389-9

[17] Marukatat, S. (2023). Tutorial on PCA and approximate PCA and approximate kernel PCA. Artificial Intelligence Review, 56(6). https://doi.org/10.1007/s10462-022-10297-z

[18] Shrestha, S., Kazama, F., & Nakamura, T. (2008). Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. Journal of Hydroinformatics, 10(1). https://doi.org/10.2166/hydro.2008.008

[19] Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis.: Pearson Prentice Hall. In Pearson Prentice Hall.

[20] Bekalo, D. B., Wanjoya, A. K., & Mwalili, S. M. (2024). On Bayesian estimation of a latent trait model defined by a rank-based likelihood. Scientific Reports, 14(1), 28965. https://doi.org/10.1038/s41598-024-80145-3

[21] Rahman, S., Chowdhury, A. H., & Akhter, T. (2021). Comparative analysis of expenditure patterns in high- and low-income countries. Global Economic Review, 50(1), 76–98.

[22] Adachi, K. (2015). A New Algorithm for Generalized Least Squares Factor Analysis with a Majorization Technique. Open Journal of Statistics, 05(03), 165–172. https://doi.org/10.4236/ojs.2015.53020

[23] Obilor, E. I., & Amadi, E. C. (2018). Test for Significance of Pearson's Correlation Coefficient ( ). International Journal of Innovative Mathematics, Statistics & Energy Policies, 6(1).