Sains Malaysiana 53(4)(2024): 969-982
http://doi.org/10.17576/jsm-2024-5304-18

Latent Hidden Factor Model for Malaysian Consumer Price Index

(Model Faktor Pendam bagi Indeks Harga Pengguna Malaysia)

NURULKAMAL MASSERAN[1,2,*], SITI NORSALSABILA AHMAD FAROUK[1], R. NUR-FIRYAL[1,2] & MAHAYAUDIN M. MANSOR[3]

[1]Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
[2]Center for Modelling and Data Analysis (DELTA), Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
[3]School of Mathematical Sciences, College of Computing, Informatics & Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Received: 10 October 2023/Accepted: 1 March 2024

ABSTRACT

The consumer price index (CPI) is one of the significant indicators that can be used to track the inflation rate of a country and assess changes in the cost of living. Generally, the CPI measures the change in the average price of goods and services used by households. The increase in inflation can have a negative socioeconomic effects and the changes in the CPI value needs to be monitored to ensure that the country does not experience a serious inflation rate. In this work, exploratory factor analysis (EFA) was employed to interpret the importance of each variable in the CPI and determine how the latent factors influence the CPI structure in Malaysia for the period of 2003–2022. The findings showed that there are two main latent factors that can be formed. Factor 1 can be classified as 'Household Expenditures and Lifestyle Choices' and consists of six variables, namely, 'alcoholic beverages and tobacco', 'housing, water, electricity, gas, and other fuels', 'decoration, hardware, and household maintenance', 'health', 'recreational and cultural services', and 'education'. In contrast, Factor 2 can be classified as 'Daily Necessities and Lifestyle Convenience' which consists of six variables, namely, 'food and non-alcoholic beverages', 'clothing and shoes', 'transportation', 'communication', 'restaurants and hotels', and 'various goods and services'. In addition, the results of our analysis showed that Factor 1 is more dominant in influencing the CPI structure in Malaysia.

Keywords: Factor analysis; latent structure; statistical model description

ABSTRAK

Indeks harga pengguna (IHP) adalah salah satu petunjuk penting yang boleh digunakan untuk mengesan kadar inflasi sesebuah negara dan juga untuk menilai perubahan dalam kos sara hidup. Secara amnya, IHP mengukur perubahan dalam harga purata barangan dan perkhidmatan yang digunakan oleh isi rumah. Kenaikan inflasi boleh mendatangkan kesan sosioekonomi yang negatif dan perubahan nilai IHP perlu dipantau bagi memastikan negara tidak mengalami kadar inflasi yang serius. Dalam kajian ini, analisis faktor terokaan (AFT) digunakan untuk mentafsir kepentingan setiap pemboleh ubah dalam IHP dan menentukan bagaimana faktor pendam mempengaruhi struktur IHP di Malaysia untuk tempoh 2003–2022. Keputusan kajian menunjukkan terdapat dua faktor pendam utama yang boleh dibentuk. Faktor 1 boleh dikelaskan sebagai 'Perbelanjaan Isi Rumah dan Pilihan Gaya Hidup' dan terdiri daripada enam pemboleh ubah, iaitu, 'minuman beralkohol dan tembakau', 'perumahan, air, elektrik, gas dan bahan api lain', 'perkakasan, hiasan dan penyelenggaraan isi rumah', 'kesihatan', 'perkhidmatan rekreasi dan kebudayaan' dan 'pendidikan'. Sebaliknya, Faktor 2 boleh dikelaskan sebagai 'Keperluan Harian dan Kemudahan Gaya Hidup' yang terdiri daripada enam pemboleh ubah, iaitu 'makanan dan minuman bukan alkohol', 'pakaian dan kasut', 'pengangkutan', 'komunikasi', 'restoran dan hotel' serta 'barangan dan perkhidmatan lain'. Di samping itu, hasil analisis kami menunjukkan Faktor 1 lebih dominan dalam mempengaruhi struktur IHP di Malaysia.

Kata kunci: Analisis faktor; huraian model berstatistik; struktur pendam

Introduction

Generally, price changes and inflation rate are activities that gives a great impact to almost any issue related to a country economy (Hobijn & Lagakos 2005; Jaravel & O'Connell 2020; Raza et al. 2023). In fact, inflation can be categorized as endemic in the field of economics since the stability of price levels or inflation is a precondition for long-term economic growth, and price level target is a popular approach for central banks as the ultimate target to enact monetary policy (Abdul Karim, Mod Asri & Wajdi 2006; Cúrdia & Woodford 2011). Inflation is the increase in the prices of goods and services over a period of time and it is commonly measured by the consumer price index (CPI), which is a numerical value of the estimated average price of goods and services paid by consumers over a period based on fixed basket of goods and services. Furthermore, changes in the CPI depict the price movement of goods and services paid by consumers and reflect the purchasing power of currency (Hassan 2022) where an increase in the price of goods can be interpreted as a decrease in the value of a country's currency and vice versa. Thus, the CPI is a very important criterion to determine the country's economic performance and is a cost-of-living indicator that is useful for households in budgeting their expenses (Shaari 2019).

CPI were calculated based on certain types of goods and services has been set based on certain classifications. Based on Department of Statistics Malaysia (2020), the goods and services included in the fixed basket are classified following the Classification of Individual Consumption According to Purpose (COICOP), namely, 'food and non-alcoholic beverages', 'alcoholic beverages and tobacco', 'clothing and footwear', 'housing, water, electricity, gas, and other fuels', 'decorations, appliances, and household maintenance', 'health', 'transportation', 'communication', 'recreational and cultural services', 'education', 'restaurants and hotels', and 'various goods and services'. The CPI is measured by comparing the average price of these goods and services in the current year based on a certain base year, as well as considering the 'basket' of goods and services in total, which refers to goods and services that have unchanged quality and quantity. The price used in the CPI calculation is the retail price through the 'Survey on Consumer Price', which is conducted weekly and monthly regardless of whether it is stated as a normal or special price, including all sales tax and excise tax imposed on certain commodities.

The CPI is associated with a country's economic policy; however, it can also contribute to an individual's socioeconomic life (Allen et al. 2011; Chu et al. 2022). For example, Malaysia currently shows an increasing poverty rate in the last 2–3 years because of the COVID-19 pandemic (Daud 2021). Based on the 2020 Household Income and Poverty Estimates reported by Department of Statistics Malaysia (2021), the number of households included in the lower income groups of B40 and M40 in Malaysia is about 2.91 million individues, while the higher income group known as T20 consists of 1.46 million individues. However, because of COVID-19, the poverty level is found to increase: with 639,800 households that fell into the poverty group in 2020 compared to 405,400 households in 2019. Because of the rising cost of living and the challenging economic situation have become, individual might need to find side job to as the basic salary is not sufficient to meet the most basic needs because of the additional costs of goods and services.

Since CPI is one of the significant indicator of economic performance and as a proxy to inflation rate, it is important to understand the impact of increase in CPI rate. Although it is difficult to control inflation and rising commodity prices because both are closely related to the current global economy, we need to understand the impact of these can have on individual socioeconomic and country's economy. The Chief Economist of the International Monetary Fund, Gourinchas (2022) stated that the global economy is facing a worrying economy slowdown but a critical priority for national policymakers, especially leaders, is to control the rise in inflation by enacting a well-thought-out policy to reduce the burden on the people, particularly regarding the cost of living, which has recently become increasingly worrying. To ensure the stability of inflation in the country, a study on CPI can be used to understand the change over a period of time to avoid the increase in inflation. Hence, the responsible governing body needs to monitor the items in the fixed basket of goods and services that often contribute to the increase in CPI in addition to making it a reference for future preparation. Therefore, the objective of this study was to determine the latent factors that can explain the relationship among the variables in the Malaysian CPI.

Malaysia CPI Data

In this work, the Malaysia CPI data for the period of 2003–2022 was utilized, which was obtained from the Department of Statistics Malaysia (2021) and consists of 12 variables total up to 240 observations related to the CPI. Malaysia CPI consist of composition related to basket

of goods and services that represent the consumption patterns of average Malaysian households. In general, these goods and services are important information that is used by the government for tracking a change in the cost of living over time. In fact, it also provides an important insights related to inflation and economic conditions over time. The composition of goods and services in the CPI basket can change depending on current consumption patterns. However, 12 sub-indexes as shows in Table 1 commonly referred as the main categories of goods and services used in CPI data. In the category of 'food and non-alcoholic beverages', among the items included in this basket are food, grains, bread, fish, milk, fruits, cheese, meat, eggs, vegetables, and non-alcoholic beverages. While, for 'alcoholic beverages and tobacco' category, the items included are beer, wine, and cigarettes. For 'Clothing and Shoes', this category considers various items involving clothing as well as shoes for men and women. For 'Housing, Water, Electricity, Gas and Other Fuels', this category consider the expenditure of Malaysians covering related items such as; housing (including rent), water, electricity, gas, and other fuels. While for 'Decoration, Hardware and Household Maintenance', this category includes expenses related to furniture, household equipment, and also routine maintenance of the household. For 'Health' category, its consider expenses related to medical products, appliances, health equipment, as well as the cost outpatient services in Malaysia. For 'Transportation' category, the items considered are includes expenses related to transport services as well as cost of vehicles purchasing. For 'Communication' category, the items considered includes to be include in CPI are expenses related to communication services such as internet services and telephone. For 'Recreational and Cultural Services', a considered expense is related to recreational and cultural activities, such as; audio-visual equipment, services for cultural and recreational activities, and also any recreational equipment. For 'Education' category, among the items included are books, tuition fees, and other educational related education equipment and services. For 'Restaurant and Hotel' category, a considered expense is related to hotel accommodation to general cost for eating out. For the last category which is 'Various Goods and Services', among the items included in this basket are financial services, personal care products, insurance, various goods and services not included in other CPI sub-categories, and miscellaneous items. All of these CPI subcategories are designed by the government for the purpose to cover various types of goods and services. In parallel, it reflects the consumption pattern of the average household in Malaysia. Therefore, it allows the CPI to function as an important measure of inflation and changes in the cost of living in Malaysia. A details description about CPI categories can be referred to a report by Department of Statistics Malaysia (2023) and Graf (2020).

Each observation in the data is an index value calculated using the CPI formula considering the weight for each category of goods and services. The weights used were based on a base year 2010 (Table 1). Thus, the CPI is also known as a weighted average of price changes for a group of goods and services which reflects the relative importance of goods and services. The purpose of weighing is to determine the effect of price changes on the entire index. The formula for CPI computation is given as follows:

$$CPI = \frac{market\ basket\ cost\ in\ a\ given\ year}{market\ basket\ cost\ in\ the\ base\ year}, \quad (1)$$

THE IMPORTANCE OF FACTOR ANALYSIS ON CONSUMER PRICE DATA

As mentioned earlier, CPI data is important economic indicator that commonly used to measure inflation, changes in the cost of living, and also economic condition in general. Thus, there is a lot of analysis available in the literature which has been done for the sake of extracting important information from CPI data. For instance, the assessment on the impact of oil price changes on domestic price inflation (Husaini, Puah & Lean 2019; Sek 2023, 2017), CPI as one of important measurement for inflation (Murdipi & Law 2016; Yusof et al. 2021), impact of monetary policy on CPI (Kassim & Manap 2008), time series behaviors of CPI data (Mhd Ruslan & Mokhtar 2020), and linkage between CPI with purchasing power parity (Chen & Hu 2017).

However, to the best of our knowledge, there no specific study has been done to investigate the existence of latent factors on Malaysia CPI data. Thus, this study tries to look at a different perfective by investigating the inherent structure of Malaysia CPI data through the technique of factor analysis. Factor analysis is important technique that can be used to describe about the existence of latent structure in consumer price dataset. Commonly, most of the consumer price dataset having a quite large number of variables which represent several categories

TABLE 1. Weighted CPI (2010 = 100)

Basket of goods and services	Weighted (%)
Food and Non-Alcoholic Beverages (A)	29.5
Alcoholic Beverages and Tobacco (B)	2.4
Clothing and Shoes (C)	3.2
Housing, Water, Electricity, Gas and Other Fuels (D)	23.8
Decoration, Hardware and Household Maintenance (E)	4.1
Health (F)	1.9
Transportation (G)	14.6
Communication (H)	4.8
Recreational and Cultural Services (I)	4.8
Education (J)	1.3
Restaurant and Hotel (K)	2.9
Various Goods and Services (L)	6.7
Total	100

of goods and services, thus factor analysis is a suitable approach to be used in reducing this dimensionality, and to make the consumer price data more manageable and easier to interpret (Pérez-Barea et al. 2018). In fact, the determining latent factors provide information regarding the patterns of correlations among the CPI variables. Thus, unobserved underlying structure of the consumer price dataset can be shown. This underlying structure represent a valuable insights regarding the relationships among the different categories of the CPI index. Thus, it could be a useful information that can help policymakers and economists to better understand the drivers of inflation and how changes in specific sectors or categories contribute to overall price movements.

Factor analysis also can be used as a tool for improving CPI index construction. Particularly, based on the results of factor analysis, each sub-indices of consumer price variables that are most closely related to each factor can be identified (Victor et al. 2018). Thus, this information can be used in refining the CPI components to better reflect changes in consumer spending patterns and preferences.

STATISTICAL METHODOLOGIES

Factor analysis is a statistical method used to analyze the relationship among several variables and explain those variables in the form of certain latent factors (Hair et al. 2010) to summarize the data with the aim of interpreting and understanding the relationships and patterns of the observed variables. In other words, observed variables are regrouped in a lower dimension of unobserved latent variables based on the concept of shared variance to the variables or describe and provide an intrinsic meaning for each category (Watkins 2021). However, before performing a factor analysis, the data must be verified for appropriateness criteria by checking the correlation to determine the relationship between these variables. When calculating the correlation factor, it is assumed that there is a linear relationship between the latent factor and the variable. A high correlation allows a factor to be formed by at least two variables, while a low correlation indicates a weak relationship between the variables and no formation of a factor. Generally, the correlation coefficient should be at least 0.32, and variables with a lower correlation should be excluded from factor analysis. In addition, a suitable sample size needs to be considered despite the number of suitable sample size is still under debate based on prior literature, but the recommended sample size is about 150 samples (Beavers et al. 2013).

DATA RELIABILITY TEST

Based on the data obtained, reliability testing must be performed to ensure that each CPI value is appropriate and then factor analysis can be used to remove possible latent factors. The two tests that can be used, namely, Bartlett's sphericity test and the Kaiser Meyer Olkin (KMO) test to determine the relationship strength and assess the factorability of the variables (Siswanah 2021). A significant result in Bartlett's test statistic indicates that at least some of the variables among the 12 groups in CPI are significant which concludes that the correlation matrix is statistically different from the identity matrix and confirms the existence of a linear combination. The χ^2 statistic for Bartlett's sphericity test is given as follow:

$$\chi^2 = -\left[(N-1)-\left(\frac{2p-5}{6}\right)\right]\ln|R|, \tag{2}$$

where N is the number of observations; p is the number of variables and; $|R|$ is the determinant for the matrix correlation (Bartlett 1951). The statistic follows a distribution with degree of freedom. In contrast, the KMO test for sampling adequacy was used to determine the suitability of the correlation matrix of this data set to undergo factor analysis (Amit, Sapiri & Yusof 2022). Basically, it is a measure of shared variance in variables. The KMO test formula is given as follows:

$$KMO = \frac{\sum r_{ij}^2}{\sum r_{ij}^2 + \sum a_{ij}^2} \tag{3}$$

where r_{ij} is the correlation coefficient between the variables X_i and X_j, and a_{ij} is a partial correlation coefficient between the variables X_i and X_j. The value for the KMO test is between 0 and 1 which KMO value close to 1 indicates that the data is significant and sufficient to be used in factor analysis. However, if the KMO value is ≤0.59, the data cannot be used because there is no proper correlation structure that can be analyzed by factor analysis. The guidelines for KMO values are presented in Table 2, where it classifies the criteria for the acceptance level in the KMO test for sampling adequacy (Howard 2023; Kaiser 1970).

MODEL FOR THE LATENT FACTOR

As mentioned earlier, factor analysis is a statistical data reduction technique used to explain the covariance among observed variables in the form of unobserved (latent) variables with smaller dimensions. It aims to find hidden factors in the original data, and it is assumed that there exists a set of latent factors F_j, $j = 1, ..., k$ that can be derived from the original data. Herein, factor analysis characterizes the nature of the dependence among the attributes of the original data through factors with smaller dimensions (Backhaus et al. 2021). For example, Figure 1 presents the mechanism of exploratory factor analysis (EFA), where Y_1, Y_2, Y_3, Y_4, Y_5, and Y_6 are the observed characteristics of some events formulated in Equations (4) to (9), respectively. A latent factor with a lower dimension, such as two-dimension, could be used to describe these six observed characteristics. The two-dimension latent factors (F_1 and F_2) are hidden because they are not being observed directly but are determined from the hidden structures that lead to the generation of an observed data of Y_1, Y_2, Y_3, Y_4, Y_5, and Y_6.

TABLE 2. Criteria for the KMO test

KMO value	Level of Acceptance
≥ 0.9	Great
0.8 – 0.89	Very Good
0.7 – 0.79	Good
0.6 – 0.69	Moderate
0.5 – 0.59	Bad
≤ 0.5	Unacceptable

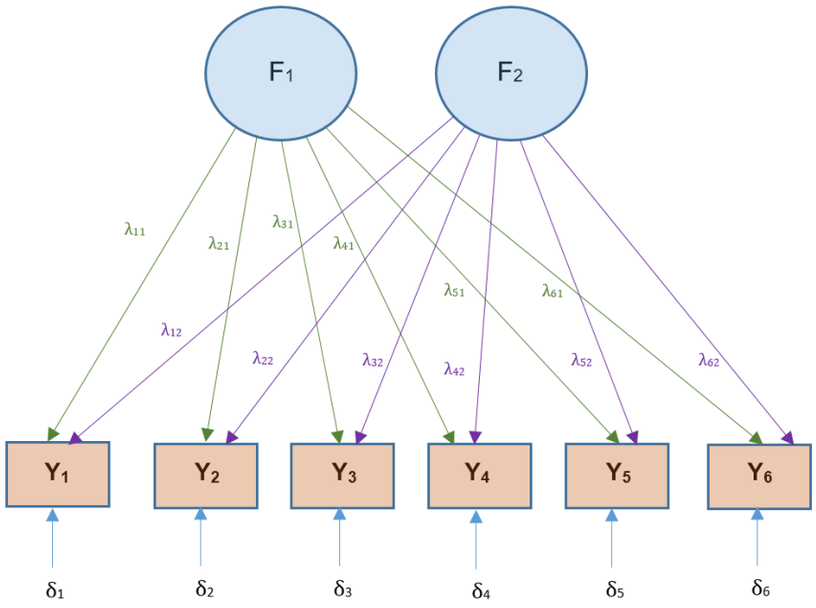

FIGURE 1. Example of a model with two latent factors

From Figure 1, the factor model can be described as a system of linear equations:

$$Y_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \delta_1, \tag{4}$$

$$Y_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \delta_2, \tag{5}$$

$$Y_3 = \lambda_{31}F_1 + \lambda_{32}F_2 + \delta_3, \tag{6}$$

$$Y_4 = \lambda_{41}F_1 + \lambda_{42}F_2 + \delta_4, \tag{7}$$

$$Y_5 = \lambda_{51}F_1 + \lambda_{52}F_2 + \delta_5, \tag{8}$$

$$Y_6 = \lambda_{61}F_1 + \lambda_{62}F_2 + \delta_6, \tag{9}$$

where F_1 and F_2 are the causal factors for Y_1, Y_2, Y_3, Y_4, Y_5, and Y_6. F_1 and F_2 are independent of the term δ_j, such that $cov(F_1, \delta_j) = cov(F_2, \delta_j) = 0$. Moreover, δ_i and δ_j are independent of each other for all $i \neq j$, such that $cov(\delta_i, \delta_j) = 0$. This model has the properties of conditional independence, such that the variables Y_i and Y_j are independent of each other given the factors F_1 and F_2, such that $cov(Y_i, Y_j \mid F_1, F_2) = 0$ for all $i \neq j$. Furthermore, the factors F_1 and F_2 are orthogonal with each other, that is, $cov(F_1, F_2) = 0$. Thus, based on the factor model, factor loading can be calculated as follows:

$$\lambda_{ij} = corr\left(Y_i, F_j\right), \tag{10}$$

where the factor loading λ_{ij} is the magnitude of the relationship between the observed variable Y_i and latent factor F_j. Next, based on the factor loadings, a communality measure for variable Y_i can determined as follows:

$$h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2, \tag{11}$$

where the communality is the percentage of variance for variable Y_i that can be explained by the latent factors F_1 and F_2. Likewise, a uniqueness for variable Y_i corresponding to a latent factor F_i can be calculated as follows:

$$unique\left(Y_i, F_i\right) = 1 - h_i^2, \tag{12}$$

where the uniqueness measure is the variance of residual for variable Y_i that cannot be captured by latent factor F_i (Gorsuch 2014).

DETERMINING THE NUMBER OF LATENT FACTORS

Generally, if we retain a large latent factor dimension, it does not contribute significantly to the analytical solution, or the factors can be difficult to analyze. Thus, it is important to correctly determine the appropriate number of factors. To obtain the correct number of latent factors, total variance is observed to determine the number of factors that is sufficient to represent the data. The first factor in the list of latent factors formed in the analysis results is the one that explains the most variance and the amount of explained variance decreases continuously with the increase of subsequent factor. The amount of this variance will determine the number of factors that should be kept.

In this work, the number of latent factors was determined by employing Cattell's scree plot, proposed by Cattell (1966) as a graphical method to show the magnitude of factor eigenvalues against the factor ordinal number. Cattell's scree plot test is a graph of the factors in the x-axis and their corresponding eigenvalues on the y-axis, as illustrated in Figure 2. There are 12 variables available in the CPI data thus the maximum number of possible factors that can be achieved is 12 latent variables. Each factor will have its own eigenvalue, where it will show the amount of variation in the variable. If there are factors that contribute to part of the variance, then the remaining factors will have smaller eigenvalues. The term 'scree' is derived from the word meaning 'rubble at the base of a mountain' (Newsom 2005). Factors that are at the corner point are considered latent factors, and the remaining factors are error variances.

Following the data reliability test, the maximum likelihood method, a type of factor extraction, was performed on the significant data to extract latent factors that can explain the relationship among the measured variables. Here, the 12 variables of goods and services in the CPI are summarized by removing latent factors to explain the relationship among the 12 variables. Assuming the data follows a multivariate normal distribution, factor weights can be estimated. The priority for the results obtained in factor extraction depends on the weighing of the resulting factors (loadings). A factor weight (represented by λ) is essentially a correlation coefficient for each variable in the latent factor. It shows the variance explained by a variable on a given factor. A factor weight of 0.7 and above indicates that the factor extracts sufficient variance from the variable.

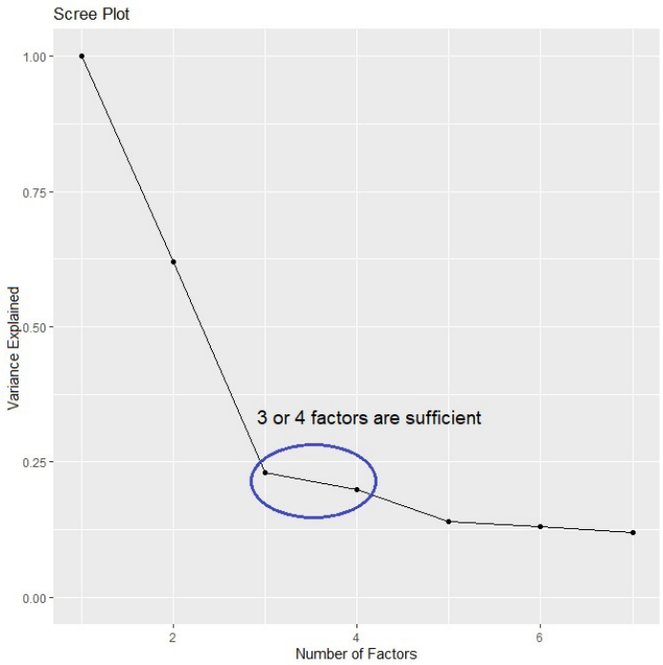

FIGURE 2. Scree plot to determine the number of appropriate factors

FACTOR ROTATION AND INTERPRETATION

Since each observed variable is separated according to specific factors that are based on the correlation value on each variable, these will form several latent factors. However, the latent factor obtained before rotation can be ambiguous and difficult to interpret. Therefore, the factor rotation process should be included to eliminate uncertainty and provide a better interpretation.

Tabachnick and Fidell (2019) showed that none of the subtraction techniques provide interpretable solutions without rotation. Rotation method clarifies and strengthens the factor weights in each factor by maximizing the high factor weights on each variable and provide a clearer and simpler factor structure for easy interpretation. Figure 3 exhibits the formation of factors before and after application of factor rotation.

In this work, it is assumed that the factors formed have orthogonal properties where each factor is independent, that is, no correlation exists among them. The varimax rotation method was employed for this study, which can further explain variables with too high or too low factor weights for each factor to obtain the simplest factor structure. It can maximize the amount of variance for the square of the weighing across the 12 variables in the CPI data where the results of the analysis provide the components of the factors with each clear variable in addition to being able to include and interpret all the information for all the variables in the CPI. Latent factors that were formed from the complete factor analysis and interpreted were given appropriate names. The interpretation of the names of the factors was based on the criteria of the 12 main groups of the basket of goods and services that make up each factor. The naming of these factors is intended to make it easier for researchers to explain and provide an overview of the influence of these factors on the CPI in Malaysia. Further, communality was discussed to assess how well the variables were explained by the extracted factors.

RESULTS AND DISCUSSION

As mentioned earlier, before the application of factor analysis, the observation dataset needed to be verified for its suitability criteria by checking the correlation value of the dataset and sample size to ensure that the factor analysis method is suitable for analyzing the data. Figure 4 shows a very high correlation value among the 12 variables. The letters A, B, …, K, and L represent the variables basket of goods and services in the order shown in Table 1, where A represents food and non-alcoholic beverages, and B represents alcoholic beverages and tobacco. The obtained correlation value $|r|$ was observed to be between 0.8 and 1, exhibiting a very strong relationship in the data and thus confirming that the data was sufficient and satisfactory for factor analysis. Further, the CPI data contains as many as 204 observations. Based on prior literature, a sample size of 200 was deemed sufficient and eligible for analysis use.

The data was further analyzed by data reliability tests, namely, Bartlett's sphericity test and KMO test, and found that they provided significant results to proceed

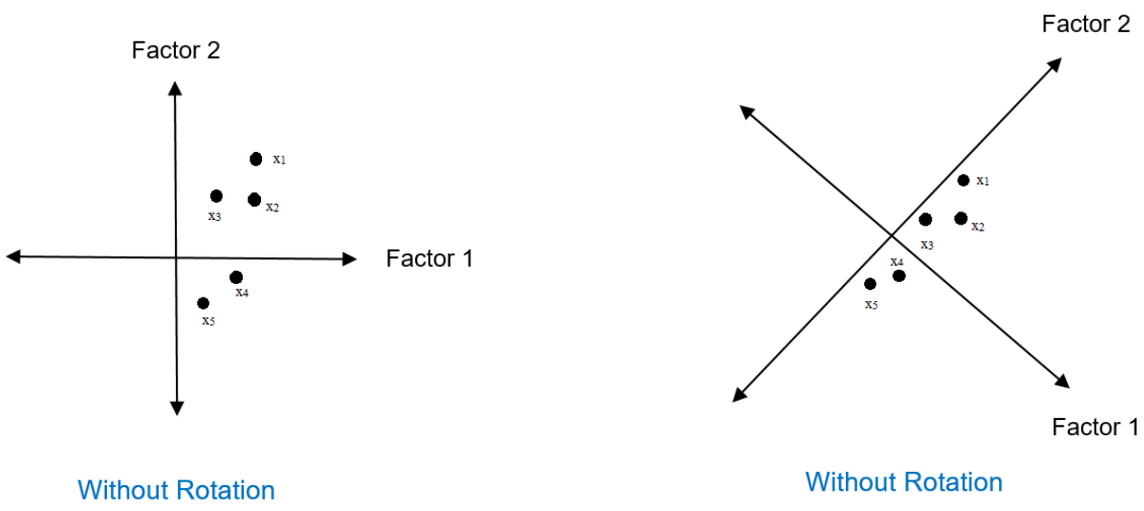

FIGURE 3. Factor rotation

with factor analysis. In Bartlett's sphericity test, the null hypothesis assumes that the matrix for the data is an identity matrix, while the alternative hypothesis assumes that the matrix is correlated. From the analysis results, it was found that the chi-square value and probability were significant, and these concluded that all the 12 variables of CPI data are correlated and have a strong relationship with one another (Figure 4). Next, the KMO sampling adequacy test determined an overall sampling adequacy measure of 0.85, which is in the 'Good' category in the KMO test's goodness criteria, and more than 0.7 for each variable of goods and services, as shown in Table 3. The KMO test concludes that each variable in this data can be factored and decides that the data correlation matrix is appropriate to be analyzed using the factor analysis method as set out in this study. Based on the two tests, the CPI data is appropriate for use in EFA.

FIGURE 4. Correlation among CPI variables

TABLE 3. KMO test results

Basket of goods and services	Sampling Adequacy Measure (MSA)	Overall KMO
Food and Non-Alcoholic Beverages (A)	0.81	
Alcoholic Beverages and Tobacco (B)	0.92	
Clothing and Shoes (C)	0.82	
Housing, Water, Electricity, Gas and Other Fuels (D)	0.81	
Decoration, Hardware and Household Maintenance (E)	0.82	
Health (F)	0.87	
Transportation (G)	0.89	
Communication (H)	0.96	
Recreational and Cultural Services (I)	0.89	0.85
Education (J)	0.87	
Restaurant and Hotel (K)	0.82	
Various Goods and Services (L)	0.80	
Total	100	

The first step in EFA is to determine the number of latent factors and the latent factors that exist should represent the entire data. The coordinates of the resulting points on the scree plot are based on the variance of each factor explained, where this value defines the amount of variance in the variable associated with each factor with the first factor has the highest variance and it decreases for the subsequent factors, as presented in Table 4. Kaiser (1970) rules that all factors with an eigenvalue greater than 1 should remain and for this dataset only the first factor with the eigenvalue of 1.1413 meets Kaiser criteria. However, if there is only one latent factor existing in the dataset, then there is no use for factor analysis, so the second factor were also considered, though it is moderate (0.34402). However, when compared in term of the % of variance, there is quite a large amount of variance that is being explained by the first two latent factor.

To investigate further, Figure 5 exhibits a scree plot to view the position of the eigenvalues for determining the number of latent factors. Figure 5 illustrates that the bend point or 'collapse' point on the second factor and clearly shows that the amount of variance in the third factor up to the 12th factor is considered as error variance due to the very small values. Although the position of the point on Factor 2 is a little vague, the difference can still be observed with the point on the third factor. Therefore, it is concluded that 85.91% of the total variance in the CPI data that can be explained by the two-factor model is significant (Factors 1 and 2). This high percentage also means that the obtained model is very efficient because it can explain almost the entire variation in the CPI data. Based on the proportion of variance, Factor 1 is more dominant in influencing the CPI in Malaysia compared to Factor 2.

After determining the appropriate number of factors in this study, the analysis was continued by extracting factors based on factor loadings (weights). Table 5 shows the resulting factor loadings for each 12 variables of goods and services that are the degree of strength or correlation with each factor. The variables to be included in a certain factor are selected by choosing the largest factor weighing value for each factor. In other words, variables that contribute to certain factors will have the highest correlation coefficient values. Based on Table 5, Factor 1 shows that the relationship between all the variables are high with clothing and communication have negative relationship. These shows that spending more on the other ten variables will spend less on clothing and shoes and communication services and vice versa.

TABLE 4. Results of the factor analysis on the CPI data

Number of latent factors	Eigen value	% of variance	% of cumulative variance
1	1.1413	66.0759	66.0759
2	0. 3426	19.8386	85.9145
3	0. 1084	6.2753	92.1899
4	0.0837	4.8495	97.0394
5	0.0290	1.6840	98.7235
6	0.0128	0.7421	99.4656
7	0.0053	0.3085	99.7742
8	0.0015	0.0914	99.8656
9	0.0013	0.0793	99.9450
10	0.0005	0.0324	99.9774
11	0.0002	0.0156	99.9930
12	0.0001	0.0069	100.000

However, almost all the variables in Factor 1 have high loadings which make it harder to interpret as we cannot disclose the main variables in this factor. Interpretation for Factor 2 is much harder since the values are close to zero means that there might be no correlation between all the variables. Therefore, these results are not informative and varimax rotation were implemented to fix this issue.

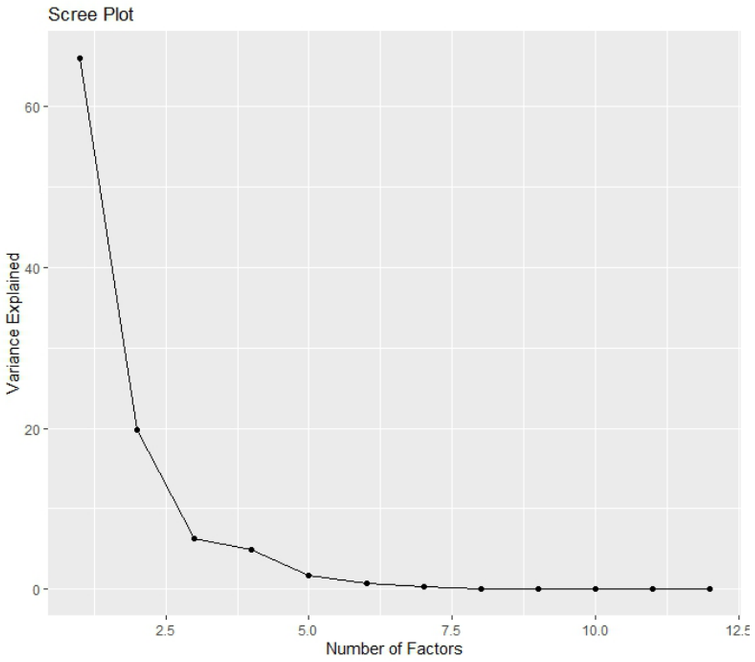

FIGURE 5. Scree plot on cumulative variance explained by latent factors

TABLE 5. Latent factor extraction without rotation

Basket of goods and services	Latent Factor 1	Latent Factor 2
Food and Non-Alcoholic Beverages (A)	0.99839	-0.00040
Alcoholic Beverages and Tobacco (B)	0.98921	0.00110
Clothing and Shoes (C)	-0.94990	0.15037
Housing, Water, Electricity, Gas and Other Fuels (D)	0.99387	0.10283
Decoration, Hardware and Household Maintenance (E)	0.99875	0.00333
Health (F)	0.99891	0.06224
Transportation (G)	0.83112	-0.16172
Communication (H)	-0.91377	0.00249
Recreational and Cultural Services (I)	0.99751	0.00037
Education (J)	0.99884	0.00006
Restaurant and Hotel (K)	0.99821	-0.00030
Various Goods and Services (L)	0.99259	-0.10009

Table 6 shows the varimax factor rotation results and the correlation of the variables on each latent factor becomes clearer to interpret. The weighing range of the factor is −1 to 1 determines the positions of the 12 variables and indicates the strong influence on the variable of the factor. Thus, variables with high factor weighing values were observed, and usually weighing values ≥0.7 were considered high and significant to be in the factor group. The selected variable on each factor was determined based on the maximum loading value on each variable corresponding to the latent factor. For instance, the loading values of the 'food and non-alcoholic beverages' variable are 0.77307 and 0.63304 for Factors 1 and 2, respectively. Since the loadings on Factor 1 are higher, the variable is placed under Factor 1. It was found that nine variables of the CPI could be considered under Factors 1 and 2 consist of the remaining three variables. To assess the extent to which these variables are explained well by each factor formed, the analysis continued by obtaining the communality value of each variable of the CPI basket of goods and services. Communality is calculated by taking the sum of the squares of the factor weights according to the number of factors extracted. A communality value close to 1 indicates that the variable is well explained by the factors.

Table 6 lists the communality values. All the variables were observed to approach 1, and it is interpreted that each variable is well explained by two factors and once again proves the efficiency of the two-factor model in this analysis. In a similar vein, a uniqueness measure shows the unique or specific variance in each observed variable that cannot be explained by the latent factors. Based on Table 6, it is found that all the CPI variables having a low values of uniqueness correspond to model of two-latent factors. Meaning that, for each observed CPI variable, two-latent factors model able to describe a large proportion of the variance in each variable. This also implies that the observed variables are well-represented by the underlying latent factors with only little unique or specific variance in each variable that is not explained by the factors. On other interpretation, this result suggests that the obtained two latent factors model provide a meaningful construct that explain a large part of the relationships between the CPI variables.

Based on the obtained latent factor, an interpretation needed to be provided on the factor itself, that is, by naming latent factors obtained based on the type of variable that makes up the factor, as it is impractical to label them as Factors 1 and 2. In this study, there are two latent factors that are expected to have an impact on the pattern of the CPI in Malaysia. Based on the above analysis results, Factor 1 consists of six variables, while Factor 2 also consists of six variables. The variables in Factor 1 are 'alcoholic beverages and tobacco', 'housing, water, electricity, gas, and other fuels', 'decoration, hardware, and household maintenance', 'health', 'recreational and cultural services', and 'education'.

TABLE 6. Latent factor extraction with rotation

Basket of goods and services	Latent Factor 1	Latent Factor 2	Communality	Uniqueness
Food and Non-Alcoholic Beverages (A)	0.701	**0.711**	0.9976	0.0023
Alcoholic Beverages and Tobacco (B)	**0.778**	0.616	0.9851	0.0148
Clothing and Shoes (C)	-0.589	**-0.760**	0.9237	0.0762
Housing, Water, Electricity, Gas and Other Fuels (D)	**0.793**	0.692	0.9966	0.0033
Decoration, Hardware and Household Maintenance (E)	**0.720**	0.598	0.9970	0.0029
Health (F)	**0.762**	0.646	0.9980	0.0019
Transportation (G)	0.495	**0.686**	0.7161	0.2838
Communication (H)	-0.602	**-0.692**	0.8411	0.1588
Recreational and Cultural Services (I)	**0.723**	0.686	0.9930	0.0069
Education (J)	**0.730**	0.682	0.9978	0.0021
Restaurant and Hotel (K)	0.705	**0.707**	0.9966	0.0033
Various Goods and Services (L)	0.654	**0.753**	0.9941	0.0058

From the study, it can show that all these are highly related to each other where one price increase can lead to increase in price in other categories since all these variables have positive loadings. To represent all these variables, Factor 1 can be labeled as a 'Household Expenditures and Lifestyle Choices'. Meanwhile, Factor 2 includes, 'food and non-alcoholic beverages' 'clothing and shoes', 'transportation', 'communication', 'restaurants and hotels', and 'various goods and services'. Factor 2 can be labeled as a 'Daily Necessities and Lifestyle Convenience'. However, Factor 2 shows slightly different results where 'clothing and shoes' and 'communication' have negative loadings while 'food and non-alcoholic beverages', 'transportation', 'restaurants and hotels', and 'various goods and services' have positive loadings. This shows that 'clothing and shoes' and 'communication' are moving in different direction compared to 'food and non-alcoholic beverages', 'transportation', 'restaurants and hotels', and 'various goods and services'. This can be seen based on increase in communication technology and fabric industries, these two baskets of goods the prices might decrease over time but it has a different impact on 'food and non-alcoholic beverages', 'transportation', 'restaurants and hotels', and 'various goods and services' variables.

Conclusion

This study aims to investigate the latent factors in Malaysia CPI data between 2003 and 2022. By performing factor analysis, a two-factor model was successfully produced and found to be significant in representing a large proportion of variance in the Malaysia CPI data. In addition, two-factor model also provide a meaningful construct that explain a large part of the relationships between the Malaysia CPI variables. A total of 12 variables in the CPI were successfully summarized to form two factors with the first factor labeled as 'Household Expenditures and Lifestyle Choices', which consists of six variables, and the second factor labeled as 'Daily Necessities and Lifestyle Convenience', which consists of other six variables in the basket of goods and services. Based on the amount of variance obtained in factor extraction, it was found that these two factors have a total variance of 66.07% and 19.83%, respectively. Thus, the two factors formed in this analysis can explain as much as 85.91% of the 12 variables of the basket of goods and services in the CPI data in Malaysia. Although the influencing factors of the CPI pattern have been identified, there is no guarantee that the index value for the coming years will show better changes in ensuring the stability of inflation in the country. However, when viewed from a different perspective, exposure to these factors allows the parties involved to identify and provide better economic measures to ensure that the cost of the goods and services remains under control, thus helping in upgrading the socioeconomic level of society in Malaysia.

Although the findings of this study meet the desired objectives, we suggest a more comprehensive factor analysis of the CPI based on the Malaysian states should be conducted in the future study. In particular, a separate CPI study on each state in Malaysia may produce better results, which can show latent factors that further describe the life patterns of household expenses and the lifestyle of the community, considering the cost of living in each state in Malaysia.

Acknowledgements

The authors are indebted to the Department of Statistics Malaysia for providing the CPI data. This research would not be possible without the sponsorship from the Universiti Kebangsaan Malaysia (Grant number GP-K020446).

References

Abdul Karim, Z., Mod Asri, N. & Wajdi, M.F. 2006. Hubungan antara Indeks Harga Pengguna (CPI) mengikut wilayah di Malaysia. *Jurnal Ekonomi Pembangunan* 7(2): 184-201.

Amit, N., Sapiri, H. & Yusof, Z.M. 2022. Factors affecting housing price in Malaysia using structural equation modeling approach. *Sains Malaysiana* 51(12): 4161-4173.

Allen, R.C. Bassino, J-P., Ma, D., Moll-Murata, C. & Van Zanden, J.L. 2011. Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. *Econ. Hist. Rev*. 64(s1): 8-38.

Backhaus, K., Erichson, B., Gensler, S., Weiber, R. & Weiber, T. 2021. *Multivariate Analysis: An Application-Oriented Introduction.* Wiesbaden: Springer.

Bartlett, M.S. 1951. A further note on tests of significance in factor analysis. *Br. J. Res*. 4: 1-2.

Beavers, A.S., Lounsbury, J.W., Richards, J.K., Huck, S.W., Skolits, G.J. & Esquivel, S.L. 2013. Practical considerations for using exploratory factor analysis in educational research. *Pract. Assess. Res. Evaluation*. 18(1): 6.

Cattell, R.B. 1996. The scree test for the number of factors. *Multivariate Behavioral Research* 1(2): 245-276. Newsom 2005

Chen, M. & Hu, X. 2017. Linkage between consumer price index and purchasing power parity: Theoretic and empirical study. *J. Int. Trade Econ. Dev.* 27(7): 729-760.

Chu, M.Y., Law, T.H., Hamid, H., Law, S.H. & Lee, J.C. 2022. Examining the effects of urbanization and purchasing power on the relationship between motorcycle ownership and economic development: A panel data. *Int. J. Transp. Sci. Technol.* 11(1): 72-82.

Cúrdia, V. & Woodford, M. 2011. The central-bank balance sheet as an instrument of monetary policy. *J. Monet. Econ.* 58(1): 54-79.

Daud, S. 2021. The COVID-19 pandemic crisis in Malaysia and the social protection program. *J. Dev. Soc.* 37(4): 480-501.

Department of Statistics Malaysia. 2020. *Indeks Harga Pengguna*. https://www.dosm.gov.my/v1/uploads/files/6_Newsletter/Newsletter%202020/DOSM_BPHPP_5-2020_Siri-44.pdf Accessed on 16 May 2023.

Department of Statistics Malaysia. 2021. Pendapatan purata isi rumah merosot 10.3 peratus kesan pandemik COVID-19 pada tahun 2020. *Laporan Anggaran Pendapatan Isi Rumah dan Insiden Kemiskinan, Malaysia*. https://www.dosm.gov.my/v1/uploads/files/5_Gallery/2_Media/4_Stats%40media/4-Press_Statement/2021/8%20Ogos/Kenyataan%20Media%20Anggaran%20kemiskinan%202020%206%20Ogos%202021_COMBINE.pdf Accessed on 3 July 2023.

Department of Statistics Malaysia. 2023. Consumer price index. https://www.dosm.gov.my/admin/images/icon/contentdoc_pdf_20220907235910.pdf Accessed on 29 January 2024.

Gorsuch, R.L. 2014. *Factor Analysis.* New York: Routledge.

Gourinchas, P.O. 2022. *Policymakers Need Steady Hand as Storm Clouds Gather over Global Economy*. IMF Blog. https://www.imf.org/en/Blogs/Articles/2022/10/11/policymakers-need-steady-hand-as-storm-clouds-gather-over-global-economy

Graf, B. 2020. *Consumer Price Index Manual: Concepts and Methods*: Washington: International Monetary Fund. https://www.elibrary.imf.org/display/book/9781484354841/9781484354841.xml Accessed on 29 January 2024.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. 2010. *Multivariate Data Analysis.* 7th ed. New York: Prentice Hall International Inc.

Hassan, M.S. 2022. Kuasa beli menurun, harga barang naik. *Harian Metro.* https://www.hmetro.com.my/utama/2022/07/864528/kuasa-beli-menurun-harga-barang-naik

Hobijn, B. & Lagakos, D. 2005. Inflation inequality in the United States. *Rev. Income Wealth* 51(4): 581-606.

Howard, M.C. 2023. A systematic literature review of exploratory factor analyses in management. *J. Bus. Res.* 164: 113969.

Husaini, D.H., Puah, C-H. & Lean, H.H. 2019. Energy subsidy and oil price fluctuation, and price behavior in Malaysia: A time series analysis. *Energy* 171: 1000-1008.

Jaravel, X. & O'Connell, M. 2020. Real-time price indices: Inflation spike and falling product variety during the Great Lockdown. *J. Public Econ.* 191: 104270.

Kaiser, H.F. 1970. A second-generation little jiffy. *Psychometrika* 35(4): 401-415.

Kassim, S.H. & Manap, T.A.A. 2008. Consumer credit and monetary policy in Malaysia. *Int. J. Consum. Stud.* 32(3): 188-193.

Mhd Ruslan, S.M. & Mokhtar, K. 2020. Structural break and consumer prices: The case of Malaysia. *Cogent Bus. Manag.* 7(1): 1767328.

Murdipi, R. & Law, S.H. 2016. Dynamic linkages between price indices and inflation in Malaysia. *Jurnal Ekonomi Malaysia* 50(1): 41-52.

Newsom, J.T. 2005. A quick primer on exploratory factor analysis. Available at: https://web.pdx.edu/~newsomj/semclass/ho_efa.pdf. (Accessed 5 July 2023).

Pérez-Barea, J.J., Fernández-Navarro, F., Montero-Simó, M.J. & Araque-Padilla, R. 2018. A socially responsible consumption index based on non-linear dimensionality reduction and global sensitivity analysis. *Appl. Soft Comput.* 69: 599-609.

Raza, H., Laurentjoye, T., Byrialsen, M.R. & Valdecantos, S. 2023. Inflation and the role of macroeconomic policies: A model for the case of Denmark. *Struct. Change Econ. Dyn.* 67: 32-43.

Sek, S.K. 2023. A new look at asymmetric effect of oil price changes on inflation: Evidence from Malaysia. *Energy Environ.* 34(5): 1524-1547.

Sek, S.K. 2017. Impact of oil price changes on domestic price inflation at disaggregated levels: Evidence from linear and nonlinear ARDL modeling. *Energy* 130: 204-217.

Shaari, A.F. 2019. Peningkatan kos hidup: Apa kata data? *Malaysia Kini.* https://www.malaysiakini.com/news/497984

Siswanah, E. 2021. Mathematical Study: The CPI in the expenditure sub group using factor analysis. *J. Phys.: Conf. Ser.* 1796: 012118.

Tabachnick, B.G. & Fidell, L.S. 2019. *Using Multivariate Statistics*. Boston: Pearson Education.

Victor, V., Joy Thoppan, J., Jeyakumar Nathan, R. & Farkas Maria, F. 2018. Factors influencing consumer behavior and prospective purchase decisions in a dynamic pricing environment - An exploratory factor analysis approach. *Soc. Sci*. 7: 153.

Watkins, M. 2021. *A Step-by-Step Guide to Exploratory Factor Analysis with Stata*. New York: Routledge.

Yusof, N., Nin, L.F., Kamal, H.K.M., Taslim, J.R.A. & Zainoddin, A.I. 2021. Factors that influence the inflation rate in Malaysia. *Int. J. Acad. Res. Bus. Soc. Sci.* 11(9): 626-637.

*Corresponding author; email: kamalmsn@ukm.edu.my