

A Study of Middle Class Proportion in Saudi Arabia using Engel's Coefficient

Khaled Mohammed Alqahtani

School of Business, Shaqra University, 11662 Riyadh, P.O. Box 88058, Saudi Arabia

Abstract

Based on Engel coefficient, this paper makes a descriptive statistical analysis of the current situation of the middle class in Saudi Arabia. The proportion of the middle class is measured by Engel's coefficient. The research aims to identify the changing trend and development rules of the middle class. This is of great significance for achieving Saudi 2030 and building a moderately prosperous society in all aspects.

Keywords: middle class, Saudi Arabia, Engel coefficient, kernel density estimate, functional data

DOI: 10.7176/JRDM/55-06

Publication date: May 31st 2019

1. Introduction

In order to develop into a sustainable modern society, Saudi Arabia needs to cultivate a huge social intermediate force, namely, the middle class. However, there are a series of problems are worth studying. How large is the middle class? What is the current development situation? What is the changing trend? Scholars at home and abroad have made relatively comprehensive and in-depth researches on the middle class. Economists often start with income indicators, while the sociologists often use questionnaires or surveys to study the middle class from multiple indicators. On the basis of previous scholars' research, this paper will make a descriptive statistical analysis of the current situation of the middle class defined by the single index of Engel coefficient, measure its proportion, and then study the changing trend of the middle class. This research aims to find out the development rules of the middle class, which is of great significance for achieving Saudi 2030 and building a moderately prosperous society in all aspects.

2. Definition of Middle Class Concept

The middle class is a description given to individuals and households who fall between the working class and the upper class within a societal hierarchy. Through the study of relevant literature review, it is found that persons in the middle class tend to have a higher proportion of college degrees than those in the working class, have more income available for consumption and may own property. Those in the middle class often are employed as professionals, managers and civil servants. In general, most scholars believe that the middle class is closely related to occupation, which is defined by income, property or consumption indicators. To avoid the dispersion of multi-variables, this research equates the concept of middle class with that of middle-income group, all of which are generally called as Middle Class.

3. Data Description and Collection

In view of the fact that income may be affected by various factors such as the disunity of the income definition, the difficulty of investigation, the difference in regional income and consumption level, people's psychological effect of avoiding showing off their properties on purpose, the vagueness and taboo of respondents' answers, the existence of invisible income, etc., this research chooses Engel coefficient as analytical model. The Engel coefficient is easier to be measured accurately to define the middle class. It can also more systematically reflect a person's living habits as well as quality of life.

Engel coefficient is commonly used in consumption indicators. Considering that Saudi Arabia is still in the stage of economic transition, in addition to the residents' own living habits, this research modifies Engel coefficient to define the criteria of the middle class. This paper chooses to subtract 0.1 from the internationally standard Engel coefficient, that is, Engel coefficient is between 0.3 and 0.4 to define the middle class in Saudi Arabia.

The data in this investigation are considered from two aspects. One is the macro-level data from General Authority of Statistics and Ministry of Economy and Planning, including urban data and rural data. The other is to adopt the latest micro-level data released by the Saudi General Social Survey (SGSS), including the social comprehensive survey data in 2014 and 2016. Moreover, due to the uncertainty of students' income and consumption behavior, the data from students are difficultly used as a strong basis. Therefore, this research selects the sample data of the age-appropriate social population with statistical analysis value, which is the non-student group aged 25-55 years old in SGSS. Their number of valid questionnaires was 1051 in 2014 and 825 in 2016, respectively.

4. Analysis of the Current Situation of Middle Class

According to the General Authority of Statistics, a very important change in Saudi Arabia's consumption structure was the decrease of Engel's coefficient from 30.1% in 2017 to 29.3% in 2018. This showed that the living standard of the residents had greatly improved. The Engel coefficients in Saudi Arabia in recent years were between 30% and 40% across the Kingdom, including cities and villages. However, it is obviously not reasonable to assume that all the people in Saudi Arabia belong to the middle class defined by Engel coefficient. We just know that the proportion of the middle class is expanding from some aspects.

According to the data from the micro SGSS data (see Table 1.1 below), the Engel coefficient of the age-appropriate social population in 2016 is between 30% and 40%. This suggests that Saudi Arabia has reached a modest level of prosperity. We can see from the Table 1.1 that the standard deviation is lower than the average value, which shows that the data dispersion degree is low, Engel coefficient is relatively concentrated, and the consumption behavior is similar. According to the definition and standard of middle class, the proportion of the middle class in consumption in Saudi Arabia has reached 18.22%. Compared with 17.97% in 2014, the Engel coefficient and the proportion of middle class have both increased slightly. Due to regional differences in consumption level, Engel coefficients are further calculated based on the three regions of developed region, relatively developed region and underdeveloped region. It can be seen that the average value of developed region decreased and the proportion of middle class decreased significantly in 2016, which indicates that the quality of life is better and the gap between rich and poor is widening in developed region. However, the average value of Engel coefficient is significantly reduced, and the proportion of middle class is increasing in underdeveloped region, indicating that the living standard of residents in underdeveloped regions has improved.

Table 1.1 Statistics of Engel Coefficient (Unit: %)

Engel coefficient	Mean value	Standard deviation	Median	(30%-40%) middle class
Whole country in 2016	35.72	95.25	30.00	18.22
Developed region in 2016	33.45	19.95	30.71	15.38
Relatively developed region in 2016	34.33	20.42	31.93	19.97
Underdeveloped region in 2016	30.54	18.81	28.34	18.26
Whole country in 2014	32.42	19.67	29.94	17.97
Developed region in 2014	35.28	21.43	31.25	19.38
Relatively developed region in 2014	34.17	20.57	31.25	18.26
Underdeveloped region in 2014	37.83	153.69	29.7	17.18

Note: The data are from SGSS in 2014 and 2016

5. Measurement of the Proportion of Middle Class

Firstly, the density function is constructed to measure the Engel coefficient of the middle class. Secondly, the kernel density function is used to estimate it to obtain the kernel density function of Engel coefficient. Then, the upper and lower limits of Engel coefficient of the middle class are defined to be 0.3 - 0.4. Finally, the kernel density function is numerically integrated to obtain the proportion of the middle class.

5.1 Measurement Model of Kernel

Density estimation is a method for estimating unknown density function, which belongs to one of modern nonparametric test methods. If $f(x)$ is a density function of one-dimensional population, let $k(\cdot)$ be a given Borel measurable function on \mathbb{R} , $h_n > 0$ is a constant related to n , which satisfies $\lim_{n \rightarrow \infty} h_n = 0$, define:

$$\hat{f}(x, h_n) = \frac{1}{nh_n^2} \sum_{i=1}^n k\left(\frac{X_i - x}{h_n}\right) \quad (1)$$

$\hat{f}(x, h_n)$ is called the kernel density estimation of one-dimensional population density function $f(x)$. Among them, $k(\cdot)$ is called kernel function, because different choices of kernel function are not sensitive in kernel density estimation, when n is very large, it has little influence on the estimation result, so Gaussian kernel

$k(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ is used in this research. h_n is called bandwidth, i.e. smooth parameter. Generally speaking, after a sample is given, the performance of kernel estimation mainly depends on whether the selection of bandwidth h_n is appropriate. Therefore, bandwidth should be adjusted, i.e. the integral mean square error MISE $\hat{f}(x, h_n)$ proposed

by Rosenblat (1956) should be minimized to obtain the optimal bandwidth h_n^* :

$$h_n^* = \left\{ \frac{\int_{-\infty}^{\infty} k^2(t) dt}{d_k^2 \int_{-\infty}^{\infty} [f^{(k)}(x)]^2 dx} \right\}^{1/5} n^{-1/5} \quad (2)$$

However, there is an unknown quantity $f(x)$ in equation (12). In this research, the empirical law proposed by Sliverman 1986 is applied, that is, suppose $f(x)$ is a normal density function $N(0, \sigma^2)$, and Gaussian kernel is selected, the optimal bandwidth h_n^* is:

$$\hat{h}_n^* = 1.06 \hat{\sigma} n^{-1/5} \quad (3)$$

Finally, the one-dimensional kernel density function is:

$$\hat{f}(x, h_n) = \frac{1}{nh_n^*} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} K\left(\frac{X_i - x}{h_n^*}\right) \quad (4)$$

5.2 Proportion of Middle Class

Engel coefficients in 2014 and 2016 are calculated based on the SGSS data, and the figure of kernel density is drawn by R software as shown in Figure 1.1. As can be seen from Figure 1.1, Engel coefficient of most people in Saudi Arabia is about 0.3. The figure of kernel density also shows a slightly rightward shape, which shows that the proportion of people with large food consumption expenditure is relatively small, and the living standards of people vary greatly. Compared with 2014, the kernel density curve in 2016 shows the following characteristics: the curve shifted slightly to the right, indicating that the residents' food expenditure level increased and the family living standard decreased slightly; the peak value of the curve is slightly higher, the top rises and the width decreases slightly, which indicates that the income Engel coefficient gap narrowed and the proportion of the middle class is slightly increased. Furthermore, based on the kernel density figure of total household income in 2014 and 2016 for urban and rural areas, the kernel density figure of total household income in 2014 and 2016 for developed areas, relatively developed areas and underdeveloped areas are drawn by R software, which gains the conclusion that: the Engel coefficient of urban families in 2016 has not changed much from that in 2014, and the Engel's coefficient of rural families has obviously shifted to the right, indicating that the living standard of rural families has declined. The kernel density figure in developed areas widened in 2016 compared with 2014, that in relatively developed areas remained basically unchanged, and that in underdeveloped areas shifted to the right as a whole, which shows that the quality of living in developed areas widened, the quality of living in more developed areas remained basically unchanged, and the living standards of people in underdeveloped areas deteriorated.

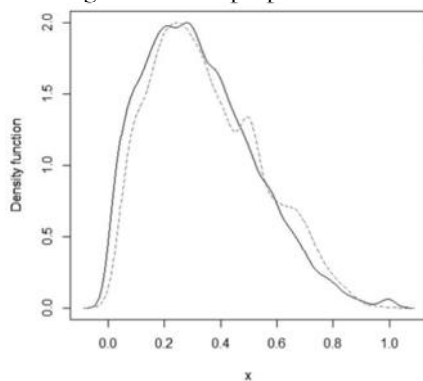


Figure 1.1 Engel Coefficient of Families in 2014 and 2016 (---2014/ —2016)

According to the upper and lower limits of the middle class defined by consumption given above, that is, Engel coefficient is between 0.3 and 0.4, the proportion of the middle class can be obtained as shown in Table 1.2 below. As can be seen, the proportion of the middle class of the whole country in 2016 slightly increased compared with that in 2014, the proportion of the urban middle class slightly decreased, and the proportion of the middle class in developed regions also decreased. The reason may be related to changes in economic growth and income changes, but also to greater pressure including housing and consumption on people in cities and developed areas.

Table 1.2 Proportion of Middle Class (Unit: %)

Time	Whole country	Cities	Rural regions	Developed regions	Relatively developed regions	Underdeveloped Regions
2016	23.22	22.57	24.24	20.38	24.97	23.26
2014	22.97	23.07	22.84	24.38	23.26	22.18

6 Analysis of the Changes of Middle Class

The analysis method of functional data is to consider each sample observation as a whole one. Moreover, in the analysis of economic functional data, scholars often need to find out the main variation modes of the variables that they are interested in that change with time, and also want to know how many such variation modes or forms can better fit the original curve samples, i.e. they need to discuss the main components of data variation by determining the typical functional characteristics of curve data. And principal component analysis can solve this kind of problem very well. This research will study the change rule of the middle class formed by Engel coefficient and find the change trend of the middle class from time to time, so the principal component analysis of functional data is introduced here. Moreover, because the microscopic data (only two years' data) are insufficient and they are not enough to represent the changes of the middle class in recent years, this research uses the urban seven-group data and the rural five-group data published by the General Authority of Statistics to expand the basis of functional data and introduces functional principal component analysis into the changes analysis of the middle class.

6.1 Basis Expansion of Functional Data

In this research, B-spline base is used to transform discrete data into functions. Assuming that the primary function $f(x_i)$ is a spline function, a cubic spline function with k nodes can be composed of linear combinations of $b_1(x_1)$, $b_2(x_2)$, $b_3(x_3)$, ..., $b_{k+3}(x_i)$, among which $b_i(x_i)$ has various selection methods. In this research, a cubic polynomial is selected as the basis, and then a truncation basis is added to each node, namely:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 h(x_i, \xi_i) + \dots + \beta_{k+3} h(x_i, \xi_k) + \varepsilon_i \quad (5)$$

Where, $h(x, \xi) = (x - \xi)^3, (x > \xi)$, $h(x, \xi) = 0, (x < \xi)$. ξ is a node. When estimating the model, the least square method is used to estimate $k+4$ coefficients.

6.2 The Analysis of Functional Principal Component

6.2.1 Mathematical model of functional principal component

In classical multivariate statistical analysis, eigenvectors are to be found. In the principal component analysis of functional data, this eigenvector is a variable function, called as principal component weight function and denoted as $\xi(s)$, where s changes within a T interval and $\xi(s)$ is quadratically integrable.

The weight function $\xi_j(s)$ of the principal component of the j -th function satisfies the following mathematical model conditions:

$$\begin{cases} \max N^{-1} \sum_{i=1}^N z_{ij}^2 \\ \int_T \xi_j(t)^2 dt = 1 \\ \int_T \xi_j(t) \xi_l(t) dt = \dots = \int_T \xi_j(t) \xi_{j-1}(t) dt = 0 \end{cases} \quad (6)$$

Where, $z_{ij} = \int_T \xi_j(t) x_i(t) dt$ is the j -th principal component score of i -th sample, $i=1, 2, \dots, N$. And many weight functions $\xi_j(s)$ obtained satisfy the standard orthogonal constraint conditions, namely:

$$\int_T \xi_j(t) \xi_l(t) dt = \dots = \int_T \xi_j(t) \xi_{j-1}(t) dt = 0$$

6.2.2 Method for solving functional principal components

In multivariate statistical analysis, the solution of principal component is to find the eigenvalues and eigenvectors of covariance matrix or correlation coefficient matrix. In functional principal component analysis, it is required to solve the characteristic equation $V \xi = \rho \xi$ and the corresponding characteristic function.

(1) Solving the characteristic equation

If the $p \times p$ order matrix $V = N^{-1} X' X$, i.e. V represents sample variance-covariance matrix, similar to multivariate statistical analysis, solving the functional principal component weight function $\xi_j(s)$ is transformed into solving the following characteristic equation:

$$\int_T v(s, t) \xi(t) dt = \rho \xi(s) \quad (7)$$

Where, let the covariance functions of $x_i(s)$ and $x_i(t)$ be $v(s, t)$, $s, t \in T$, i.e.:

$$v(s, t) = N^{-1} \sum_{i=1}^N x_i(s) x_i(t) \quad (8)$$

If an operator is defined:

$$V \xi = \int_T v(\bullet, t) \xi(t) dt \quad (9)$$

That is, V is an integral transformation of the weight function ξ and is called covariance operator. Therefore, the characteristic equation of equation (7) can be expressed as $V\xi = \rho\xi$ (note that ξ here is a characteristic function, not a eigenvector).

(2) Solving weight function by discretization method

Let the time points t_1, t_2, \dots, t_n of observation $x_i(t)$ be equally distributed in the interval T , i.e. values are taken at n equal diversion points of the interval T , thus obtaining the multivariate data set X :

$$X = \begin{pmatrix} x_1(t_1) & x_2(t_2) & \cdots & x_1(t_n) \\ x_2(t_1) & x_2(t_2) & \cdots & x_2(t_n) \\ \vdots & \vdots & \cdots & \vdots \\ x_N(t_1) & x_N(t_2) & \cdots & x_N(t_n) \end{pmatrix} \quad (10)$$

Similar to multivariate statistical analysis, eigenvalue and eigenvector satisfying equation (11) can be obtained:

$$Vu = \lambda u \quad (11)$$

Where u is an n -dimensional vector.

The elements of the sample variance-covariance matrix $V = N^{-1}X'X$ are $v(t_j, t_k)$. For the given function ξ , let ξ be an n -dimensional column vector composed of $\xi(t_j)$, and l is the length of interval T , $w = l/n$. Therefore, for any t_j , there are:

$$V\xi(t_j) = \int_T v(t_j, s)\xi(s)ds \approx w \sum_{k=1}^n v(t_j, t_k)\xi(t_k) \quad (12)$$

Therefore, the functional characteristic equation $V\xi = \rho\xi$ has an approximate discrete form:

The solution of this equation will correspond to the solution of the formula $V\xi = \rho\xi$, and the relation between eigenvalues is $\rho = w\lambda$. The approximate discrete form of the normalization constraint $\int_T \xi(t)^2 dt = 1$ is $w\|\xi\|^2 = 1$, so if the vector u is the normalized eigenvector of the matrix V , then $\xi = w^{-1/2}u$. After ξ is obtained, the approximate characteristic function ξ can be obtained from the discrete value ξ using any simple interpolation method.

6.3 Data Analysis of Middle Class Changes

In this research, B-spline basis is selected to expand by R software, and Engel coefficient of urban and rural seven-group data and their rate of change are plotted as smoothing curves. The conclusion is that: Apart from the relatively large fluctuation of Engel coefficient in the lowest income group, the Engel coefficient in the other six groups only has slight fluctuation, but has maintained a stable development on the whole. The fluctuation characteristics of Engel coefficient of the middle three curves (middle lower, and middle upper groups) measuring middle class groups are consistent, all of which firstly increase and then decrease periodically. Besides, the Engel coefficient reached the maximum increase in 2004 and 2008, and had a downward trend after 2012. Similarly, according to the five-group rural Engel coefficient and its rate of change, B-spline smoothing is also carried out, which gains the conclusion that: the five-group curve in rural areas changes more sharply than the seven-group curve in cities and towns, and its overall fluctuation trend drops. It reached to the top point in 2004, which may be due to soaring prices in 2004 without the income increase correspondingly. Besides, the middle three curves, which measure the middle class group, had a declining trend since 2012, which indicates that the living standard of the rural middle class group has been improving in recent years. However, from the five-group data of rural Engel coefficient change rate, the fluctuation rate of other groups is basically the same except that the change rate of low-income groups is extremely low near 2012.

Furthermore, the effect diagrams of the mean deviation of principal components of the change rate of seven-group urban and five-group rural Engel coefficients are drawn by R software, respectively. As can be seen from Figures 1.2 and 1.3, both urban and rural Engel coefficients deviate from the average value more and have large fluctuation, and explain most of the changes in functions of the function. Moreover, their first principal components caused the great changes of five-group Engel coefficient before 2006 and after 2010. In fact, the promulgation of a series of national policies benefiting the people in 2006 and 2010 increased people's income and correspondingly increased consumer spending. Therefore, under the control of various policies, the five-group Engel coefficient changed.

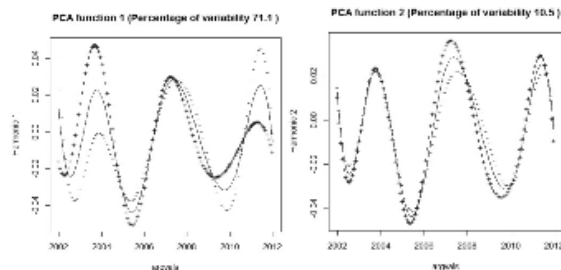


Figure 1.2 Effect Diagram of Mean Deviation of Principal Components of the Change Rate of Seven-group Urban Engel Coefficient

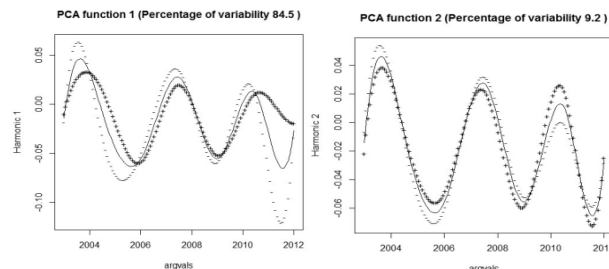


Figure 1.3 Effect Diagram of Mean Deviation of Principal Components of the Change Rate of Five-group Rural Engel Coefficient

7. Conclusion

The density function of the middle class group defined by Engel coefficient shows a slightly right-sided distribution, which shows that people's living standards are obviously different. Compared with 2014, the quality of life in 2016 declined, but the difference of the rich and the poor narrowed. From the perspective of urban and rural areas, the Engel coefficient in urban areas has not changed much, and that in the rural areas has moved to the right obviously, which indicates that the quality of life in rural areas has declined. From the regional point of view, it suggests that the quality of living in developed areas widened, the quality of living in more developed areas remained basically unchanged, and the living standards of people in underdeveloped areas deteriorated. Generally speaking, the proportion of Saudi Arabia's middle class is not large, but the overall trend is increasing.

And the results of functional data analysis show that: the fluctuation characteristics of Engel coefficient curve of middle-class groups are consistent, all of which firstly increase and then decrease periodically, and had a downward trend after 2012, indicating that the food expenditure of people in all income groups has decreased and the living standard has improved. The smoothing curve of the fluctuation rate of urban and rural Engel coefficient fluctuates roughly the same, but the curve of rural Engel coefficient fluctuates more violently than urban one. From the effect diagram of mean deviation of principal components of Engel coefficient, it can be seen that both urban cities and rural areas caused substantial changes of Engel coefficient before 2006 and after 2010.

In short, the proportion of the middle class in Saudi Arabia is generally increasing, whereas the scale is not large. There is still a big gap compared with the proportion of the middle class in developed countries such as Europe and the United States. Although the quality of life of people has indeed improved, the task of expanding middle class is still very arduous.

References

- Chanda, K.C., (1974). Strong mixing properties of liner stochastic processes. *Journal of Applied Probability*, 32, 43-54.
- Ioannides, G. A., (1987). *Stochastic Analysis and Applications*, New York: Macmillan.
- Parzen, E., (1962). On estimation of a probability density function and mode. *Annals of Mathematics*, 43, 15-19.
- Su, H, N., (2003). Strive to expand the proportion of middle-income people in developing countries. *Macroeconomic Research*, 13, 26-39.
- Wu, Q. Y., Zeng, X., (2013). Consistence r-th Mean of Density Kernel Estimators for a Mixing Sequences. *Journal of Huaqiao University*, 34, 143-152.
- Wei, S., (2013). A Literature Review of Studies on the Concept of the Middle Class. *Theoretical Research*, 62, 184-198.
- Yan, M, Y., (2010). *Analysis method and economic application of functional data*. Beijing Finance Publishing House.