Malaysian Household Consumption Expenditure: Rural vs Urban

Wan Zawiah Wan Zin, Siti Fatin Nabilah
School of Mathematical Sciences, Faculty of Science & Technology,
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor

E-mail: w_zawiah@ukm.edu.my

Abstract

The objective of this study is to identify the determinants of household consumption expenditure in both the urban and rural areas in Malaysia. Data from the Household Expenditure Surveys (HES) conducted by Department of Statistics Malaysia; collected in 1998/1999, 2004/2005 and 2009/2010 time-frame is used in this analysis. Seven factors which may affect the household expenditure patterns are considered, which are age, gender, ethnicity, marital status, education levels, work status & household size were considered. Ordinary Least Squares and Quantile regression methods are used to analyze the difference between the urban and rural distributions of household consumption expenditure at various levels of expenditure. The findings show that education level increases the consumption expenditures in the urban area significantly whereas in the rural area, the household size and the work status of family head are the more important determinants of the household spending in 2004/5 and 2009/10 time frame, respectively.

Keywords: *Household consumption expenditure; HES; Ordinary Least Squares; Quantile Regression*

1. Introduction

Malaysia is a country that is experiencing rapid urbanization process in accordance with the status of a developing country. This has stimulated the process of urbanization as many rural communities are trying to seize the opportunities that exist in urban areas to improve their economic status. Although various efforts have been made, there are still socio-economic gap between rural and urban areas. This, in turn, affects the way people in rural and urban spending. Thus, factors affecting the expenditure pattern between residents in rural and urban areas may be investigated.

Household expenditure consists of expenditures incurred by a family to meet daily needs, regardless of the needs of the individual and the needs of the whole family. Examples of household expenditure are on meals, clothing, utility bills, health, transportation, education etc. Factors affecting the household expenditure are age, gender, ethnicity, marital status, strata, state, occupation, household size and religion.

Several studies have been carried out in several parts of the world to identify what could possibly be the factors influencing household expenditure pattern. For example, in Turkey, a study by Ebru Caglayan (2012) found that age increases the consumption

expenditures in general and urban estimations, while it decreases the consumption expenditures in the rural estimations. In rural estimates, only age, income, marital status, insurance and the size of the household are significant. Another study by Nguyen et al (2006) in Taiwan based on 1993 and 1998 data, they found that the difference between the urban and rural distributions due to factors such as education, ethnicity and age. However, in the latter survey analysis, they showed that the factors identified earlier are only true for lowest quantiles. A study on the behaviour of household's food expenditures by Massimo et al (2009) showed that households headed by elder, not employed and endowed with low education level individuals spent a relatively higher level of food expenditure based on 2000 and 2006 Consumption Expenditure Surveys at household level, implemented by the Italian National Statistical Institute (ISTAT). In most cases, it can be found that factors influencing expenditures for both urban and rural areas vary between countries, depending on covariates of interest.

Ordinary least squares method (OLS) and Quantile regression (QR) method are used to model the relationship between the expenditure and factors affecting it. The OLS method assumes that the total expenditure is a linear function of a set of socioeconomic households characteristics and proceeds to minimize the sum of squared residuals from the mean. On the other hand, QR, introduced by Koenker and Basset (1978) is a method to estimate the conditional of a variable. This regression has the potential of generating different responses in the dependent variable at different quantiles. According to Koenker and Bassett (1978), the QR can be considered as an extension of the conditional mean model when compared with the OLS model. In other words, the QR substitutes to the mean the different quantiles values and proceeds to minimize the weighted sum of the absolute residuals. In that, the median regression estimator can be considered as a central special case (Koenker and Hallock, 2001).

This study aims to investigate the main factors affecting household consumption expenditure in urban and rural areas in Malaysia using the Household Expenditure Surveys (HES) data collected in 1998/1999, 2004/2005 and 2009/2010. The aim of this study is to identify the main factors affecting household expenditure in rural and urban areas based on the method Ordinary Least Squares Regression (OLS) and Quantile Regression (QR) methods across the 3 census periods. The usage of QR enables the identification of factors which may be different across several levels of quantiles considered. From this study, the differences in the factors affecting the expenditures at both urban and rural areas may be identified. Seven factors which may affect the household expenditures, namely factors concerning the head of the household which are age, gender, ethnicity, marital status, academic status and employment status as well as household size were considered.

2. Methodology

Two statistical modelling methods is used in this study, which are the Ordinary Least Squares (OLS) and Quantile Regression (QR) models.

The Ordinary Least Squares (OLS) regression is a statistical technique used to find the closest estimate to the actual value of the data or usually termed as the best fit line. It assumes that the relationship between the dependent variable Y and the independent variable X is linear with the following equation:

$$Y = \alpha + \beta_i X_i + \varepsilon_i \qquad Y = \alpha + \beta i\ xi + ei \qquad (1)$$

where

Y = dependent variable
α = intercept on the y-axis
β_i = the ith regression coefficient
x_i = the i th independent variable
e_i = the ith error

Based on equation (1), α is the value of the dependent variable Y when x value is zero. The β coefficient describes the change in value of Y with every unit change of x. The coefficient β provides information on the average estimate of the model.

To get the best fit line, the fitted line must be the one that minimizes the total distance between the actual values of the data with the expected values based on the fitted line. The difference between these two values is called residual or error of the model. The best fitted line is the line that minimizes the total sum of squares between the fitted line and actual data, that is

$$\sum_i e_i^2 = \sum \left(Y_i - \hat{Y}_i \right)^2$$

with ei = error at the i-the observation,
Y_i = actual value, Y at the i-th observation
$\hat{Y}_i$ = estimated Y value at the i-th observation

The OLS method assumes that the total household expenditure is a linear function of a set of other household characteristics and proceeds to minimize the sum of squared residuals from the mean.

Quantile regression, introduced by Koenker and Basset (1978), is a method to estimate and draw inferences about conditional quantile functions. As opposed to OLS which estimates the conditional mean of the response variable given certain values of the predictor variables, QR aims at estimating either the conditional median or other quantiles of the response variable. In other words, this regression has the potential of generating different responses in the dependent variable at different quantiles. These different responses may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable.

Quantile regression models assume that the conditional quantile of a random variable Y is a linear in the regressors X with

$$Y_i = \alpha + \beta_\theta X_i + \varepsilon_{\theta i} \text{ with } Quant_\theta \left(Y_i \setminus X_i \right) = \beta_\theta X_i$$

where X_i is the vector of independent variables and β_θ is the vector of parameters. $Quant_\theta(Y\backslash X)$ is the θth conditional quantile of Y given X. Estimation of the quantile parameters is performed as the solution to

$$\min_{\beta \in R^{\to k}} \left(\sum_{i:Y_i > X_i\beta} \theta |Y_i - X_i\beta_\theta| + \sum_{i:Y_i < X_i\beta} (1-\theta)|Y_i - X_i\beta_\theta| \right)$$

The standard errors for the vector of parameters are obtainable by using the bootstrap method described in Buchinsky (1998). As QR looks into different quantiles of the model, it can provide a more complete description of the underlying conditional distribution compared to other mean-based estimators such as OLS.

To determine whether a factor is significant or not, the confidence interval for QR is computed and if the coefficient of one factor calculated based on OLS does not fall inside the confidence interval for the QR, this implies that the factor is significant at that quantile. A factor is considered as significant based on OLS using the ANOVA test that is by comparing the sum of squared error for that factor versus the overall model total sum of squares.

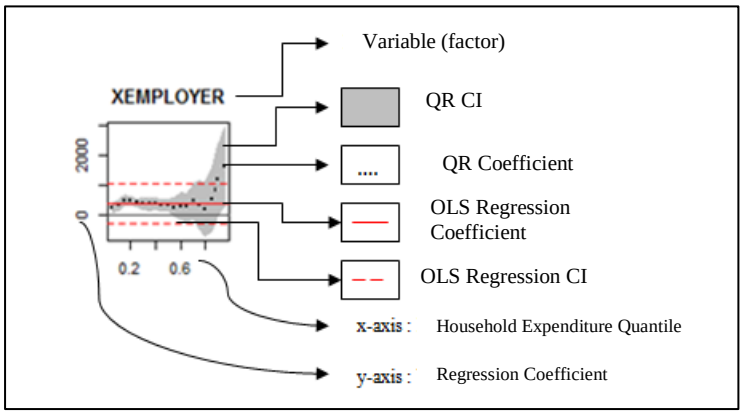

Figure 1: Example of OLS- QR graph with indicators

3. Data

This study aims to determine the factors affecting the household expenditures for rural and urban households in Malaysia. The data used in this study is part of the household expenditure data collected during the Household Expenditure Survey (HES) in 1998/1999, 2004/2005 and 2009/2010 covering both the rural and urban areas. The data constitutes 30% of total data collected by Department of Statistics Malaysia (DOSM) and the number of data, n used in this study is as in Table 1 below:

Table 1: Number of households, n involved in the HES for the 3 periods considered

Year	Rural	Urban
1998/1999	1191	1570
2004/2005	1385	2838
2009/2010	1997	4498

The dependent variables in this study the total household expenditure per month with seven independent variables recorded. Details of these independent variables can be found in Table 2.

Table 2: Description of independent variables in this study

Independent Variable	Term	Description	Dummy Variable
Age	AGE	Age of Head of household	Nil
Gender	MALE	Gender of Head of Household	MALE = 1 for male, 0 otherwise
Ethnic	BUMIPUTERA	Ethnic of Head of Household	BUMIPUTERA = 1 for bumiputera, 0 otherwise
Marital Status	MS	Marital Status of Head of Household	SINGLE = 1 MARRIED = 1 SEPARATED=1
Education Level	EDU	Highest education level of head of household.	EDU1= 1 if low level (PMR) EDU2= 1 medium level (SPM/SPMV/diploma/certificate) EDU3= 1 if high (university) EDU4= 1 if no formal education
Work Status	WS	Work status of the head of household	EMPLOYER = 1 WORKER = 1 OTHERS = 1 if not working
Household Size	SIZE	Number of household supported by head of household	SIZE1 = 1-3 people SIZE2 = 4-7 people SIZE3 $\geq$ 8 people

4. Results

Both the OLS and QR were fitted to the data and the best fitted line for each type of model was derived. For the QR model, three quantiles namely the 25^{th}, 50^{th} and 75^{th} quantiles were considered; which imply the low expenditure level, medium level and high expenditure level.

Table 3 shows the results of analysis based on data from urban areas for the three periods considered. Based on the analysis results, in urban areas, the highest level of education (EDU3) is the factor that contributes most to the expenditure regardless whether the analysis was based on mean or on low, middle and income range. For OLS, large household size is found to be highly significant for expenditure; however for the QR, for all quantiles considered, EDU2 is a very important factor. For all the regression

models considered, BUMIPUTERA gives a negative relationship with household expenditure. In 2004/2005 data, the household size is found not to have any relationship with expenditure, however in 2010/11 survey data, this is not true. In 2010/2011 time frame, the status of the head of household as employer has a significant positive relationship with expenditure for all models considered. The age of the head of household and work status as 'Worker' play an important role for the low and medium household expenditure in 1998/99 time frame.

Table 3: Analysis results for Urban Area for the three periods considered.

TOTAL HOUSEHOLD EXPENDITURE		1998/99				2004/5				2010/11			
		OLS	QR - 25th	QR - 50th	QR - 75th	OLS	QR - 25th	QR - 50th	QR - 75th	OLS	QR - 25th	QR - 50th	QR - 75th
AGE							5						
MALE								5	4				
BUMIPUTERA		5	3	4	4	4	3	3	3	3	4	4	4
MS	SINGLE					2		4	5	5			
	MARRIED												
EDU	EDU1						4						
	EDU2	2	2	2	2	3	2	2	2	4	3	2	3
	EDU3	1	1	1	1	1	1	1	1	1	1	1	1
WS	EMPLOYER					5				2	2	3	2
	WORKER												
HH SIZE	SIZE2	4	4	3	3								
	SIZE3	3	5	5	5						5	5	5

*The blue box indicates significant positive relationship, the pink box indicates significant negative relationship, numbers inside the box (1-5) indicates ranking based on significance with 1 indicate very significant. The green box indicates the factor which is most significant in all models considered with direction same as the effect of that particular factor in the same year.

The results on factors that significantly affect household expenditure are slightly different for rural areas. Based on OLS, the highest education level (EDU3) is the most significant factor which influences household expenditure in positive direction. However, detailed analysis showed that this factor does not have the most significant effect for all quantiles considered apart from at the low and middle expenditure in 1998/99 time frame. For the high expenditure in the same time frame, BUMIPUTERA is the most significant factor with negative relationship with expenditure. For the rural areas, being married is a factor that is significant in influencing expenditure in a positive relationship. Similarly, size of the family affects household expenditure significantly whereby the larger household size will translate to higher expenditure especially in all quantiles considered in 2004/5 as opposed to other factors It can also be seen that in 2009/10 time frame, age, BUMIPUTERA, and being single has significant effects compared to the low and middle expenditure. Table 4 shows the summary of results for rural area.

Table 4: Analysis results for Rural Area for the three periods considered.

TOTAL HOUSEHOLD EXPENDITURE		1998/99				2004/5				2009/10			
		OLS	QR - 25th	QR - 50th	QR - 75th	OLS	QR - 25th	QR - 50th	QR - 75th	OLS	QR - 25th	QR - 50th	QR - 75th
AGE													
MALE								5		4			
BUMIPUTERA		5			1								
MS	SINGLE							4	3				
	MARRIED	4			3	3	4	3	4		4	4	4
EDU	EDU1		5	5									
	EDU2	2	3	2	2	5			2	5	5		
	EDU3	1	1	1		1	2			1		3	3
WS	EMPLOYER						5		5	2	1	1	1
	WORKER												5
HH SIZE	SIZE2		4	4	5	4	3	2			3	5	
	SIZE3	3	2	3	4	2	1	1	1	3	2	2	2

* The blue box indicates significant positive relationship, the pink box indicates significant negative relationship, numbers inside the box (1-5) indicates ranking based on significance with 1 indicate very significant. The green box indicates the factor which is most significant in all models considered with direction same as the effect of that particular factor in the same year.

4. Conclusion & Discussion

This study manages to identify the factors that affected the household expenditure across the three time frames the HES were conducted using the OLS and QR techniques. It can be seen that in the urban areas, level of education of the household head plays a very important role in determining the household expenditure. However, this factor, although significant is not the main factor that determines household expenditure in the rural areas whereby it can be seen that when data is analysed using quantile regression method, although in 1998/99 for the low and middle expenditure, highest level of education plays a very significant role, nevertheless a person who falls in BUMIPUTERA category will results in decrease in household expenditure at the high expenditure category. In 2004/2005 time frame, large family size affects expenditure most in all categories of expenditure whereas when the head of the household is an employer, this result in higher expenditure for all quantiles considered. It can also be seen from the results that the latest time frame (2009/10) showed that for the low expenditure group in the urban area, being single and married are two significant factors that affect expenditure as opposed to earlier years. Similarly, in the rural areas, these two factors are also significant but at the high expenditure level for the same year. Age also affects expenditure but at low expenditure level in the urban areas during 1998/99 and 2004/5 time frames; which implies that the older the head of household is, the more expenditure is spent.

The results in this study may be used as a guideline to differentiate factors that influence household expenditure between urban and rural areas. However, there may be other factors which may influence the expenditure that were not available in this set of data such as income of the head of household or total household income. Similarly future study can also be done on the effect of independent variables on different types of household consumption such as medical, insurance, food, entertainment etc. It is hoped that the availability of more data with other possible influencing variables in the future will ensure a more valid study to be conducted.

Acknowledgements

The authors are indebted to the Department of Statistics Malaysia for providing the valuable data to make this paper possible.

References

Binh T. Nguyen, James W. Albrecht, Susan B. Vroman & M. Daniel Westbrook. 2006. *A Quantile Regression Decomposition Of Urban-Rural Inequality In Vietnam.* 31

Buchinsky, M. (1998): *Recent advances in quantile regression models: A practical guidline for empirical research*," The Journal of Human Resources, 33, 88–126

Ebru Caglayan (2012). *A Microeconometric Analysis Of Household Consumption Expenditure Determinants For Both Rural And Urban Areas In Turkey.* Februari. No. 2:8

Koenker, R. and Bassett, G. (1978). *Regressions quantiles*, Econometrica, 46, No. 1, 33 - 50.

Koenker, R. and Hallock, K. F. (2001), *Quantile regression*, Journal of Economic Perspectives, 15(4), 143-156.

Massimo Bagarani, Maria Forleo & Simona Zampino. (2009). *Households Food Expenditures Behaviours And Socioeconomic Welfare In Italy: A Microeconometric Analysis.* 1-16