The Australian Journal of
Agricultural and Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

The Australian Journal of Agricultural and Resource Economics, 55, pp. 239–256

Determinants of fruit and vegetable consumption in Malaysia: an ordinal system approach*

Steven T. Yen, Andrew K.G. Tan and Rodolfo M. Nayga Jr†

We examine the socio-demographic determinants of fruit and vegetable consumption using household survey data from Malaysia. A bivariate ordered probability model is developed by the copula approach. Results for a system of fruit and vegetable servings per week indicate that education, age, ethnicity, income, location of residence, smoking status and health conditions are significant predictors of fruit and vegetable consumption in Malaysia. Policy implications are suggested.

Key words: copula, fruits, ordered probability model, vegetables.

1. Introduction

A good diet is key to good health. The World Health Organization (WHO) estimated that low consumption of fruits and vegetables (FV) causes 19 per cent of gastrointestinal cancers, 31 per cent of ischaemic heart diseases and 11 per cent of strokes globally in 2002. Low FV intake is among the top-10 risk factors of global mortality, and up to 2.7 million lives could potentially be saved annually, given sufficient FV consumption (WHO 2003). Despite these health benefits, world statistics reflect under-consumption of FV in many parts of the world, developed and developing alike. This phenomenon can be attributed to existing economic, cultural and agricultural conditions in the respective countries (WHO 2003).

In Malaysia, statistics from the Food and Agriculture Organization (FAO) indicate that between 1980 and 2003, average consumption was about 150 g of fruits and 78 g of vegetables per capita per day (FAOSTAT 2009). However, the combined FV consumption of 228 g per day was far below the 400 g (or five servings) recommended by the WHO for the prevention of chronic diseases such as heart diseases, cancer, diabetes and obesity (WHO 2003).

* Research support from the Universiti Sains Malaysia Research University (RU) grant (Grant No. 1001/PSOSIAL/816072) is acknowledged. We wish to thank the Director General of the Ministry of Health Malaysia for sharing the data and permission to publish.

† Steven T. Yen (email: syen@utk.edu; http://web.utk.edu/~syen/) is at Department of Agricultural and Resource Economics, The University of Tennessee, Knoxville, TN, USA. Andrew K.G. Tan is at the School of Social Sciences, Universiti Sains Malaysia, Penang, Malaysia. Rodolfo M. Nayga Jr, is at Department of Agricultural Economics and Agribusiness, University of Arkansas, Fayetteville, AR, USA.

doi: 10.1111/j.1467-8489.2011.00536.x

Considering the importance of FV to health, it is important to know the socio-demographic determinants of FV consumption and the profile of people with insufficient intake. Studies have investigated this issue in western countries (Cox and Wohlgenant 1986; Huang and Lin 2000; Blisard *et al.* 2004; Stewart *et al.* 2004; Lin *et al.* 2010). However, the socio-demographic determinants of FV consumption in Malaysia have not been investigated. This study attempts to fill this void. Knowledge of the roles of socio-demographic characteristics in FV demand is important to policy makers interested in the nutritional well-being of the population and to FV marketers interested in identifying their target market. This study contributes to the empirical literature in three ways. First, to accommodate the discrete data feature and correlations among unobserved factors, we develop a bivariate ordered probability model using the copula approach. Second, we include several health-related variables, such as being diagnosed with hypercholesterolemia, hypertension, diabetes and smoking status, in a novel attempt to examine their effects on FV consumption patterns. Third, we focus on a developing country, Malaysia, for which few demand studies existed.

2. Insights from the literature

The relationship between FV consumption and health as well as socio-demographic factors has received widespread attention. Stewart *et al.* (2004) and Casagrande *et al.* (2007) found that Caucasians, African-Americans and non-Hispanics differ in their FV consumption and dietary patterns. Because ethnicity is found to play a role in FV consumption, the unique racial composition in Malaysia, consisting of Malay, Chinese, Indian, and a proportion of other races, allows a novel examination of ethnic factors in FV consumption decision making.

Huang and Lin (2000) and Stewart *et al.* (2004) found that college-educated individuals allocate more of their food budgets to FV than those without tertiary education. Blisard *et al.* (2004) reasoned that those who have invested time and effort in obtaining college education value the future more highly than those without tertiary education. Better educated individuals also view healthy diets as an avenue to achieve future economic and social goals. In contrast, individuals less certain about their future place more emphasis on the present and therefore are less concerned about the effect of current diet on future well-being.

Cox and Wohlgenant (1986) found age to be an important factor in FV consumption. Casagrande *et al.* (2007) and Krebs-Smith and Kantor (2001) found under-consumption of FV amongst low-income consumers. Blisard *et al.* (2004) and Stewart *et al.* (2004) highlighted the positive effect of income on vegetable purchases, suggesting that low-income households spend significantly less on FV than high-income households.

The effect of gender on FV consumption patterns is mixed. Cox and Wohlgenant (1986) concluded that men and women have a similar pattern in fresh vegetable consumption, but Bleich *et al*. (2007) found that the odds of women tracking their FV intake were 50 per cent higher than men. Lin and Huang (2008) showed that young singles purchased industrially processed vegetables less frequently, while married households were more likely than singles to buy organic fresh fruits.

3. Methodology

As the WHO-recommended consumption levels are five servings of either fruits or vegetables and FV are presumably substitutable, it is important from a policy perspective to estimate the FV equations jointly. Further, levels of FV consumption are measured in servings per week in our data. To accommodate the discrete nature of and correlation between these variables, we develop a bivariate ordered probability model. In what follows, observation subscripts are suppressed for brevity.

3.1. Empirical specification

To motivate the econometric specification, assume separability of FV from all other goods and consider an individual with personal characteristics c (a vector) facing a choice set $q = [q_1, q_2]'$ (containing FV) with prices $p = [p_1, p_2]'$. The individual maximises her utility subject to a fixed FV budget m:

$$\max_q \{U(q, c) \,|\, p'q = m\}. \tag{1}$$

Assuming the utility function $U(q, c)$ is continuous, increasing, and quasi-concave in q, optimal levels of quantities can be expressed as a function of prices, budget and personal characteristics, *viz*., $q = f(p, m, c)$. With a single cross section, prices are not available so regional variables are used as proxies. Income category variables are also used in lieu of FV budget (discussed below). Then, using a vector x to represent explanatory variables, a linear function with conformable parameter vector β_i to (first-order) approximate each deterministic demand function, and a random error u_i to capture the unobservable, the demand functions are expressed as

$$q_i = x'\beta_i + u_i, i = 1, 2 \tag{2}$$

3.2. A bivariate ordered probability model

Because FV servings are observed as counts and re-coded into ordinal variables, the use of a bivariate ordered probability model is appropriate. The approach is to begin with a univariate ordinal probability model for each equation and then link the two probabilities with a copula. Consider two ordered probability models for dependent variables q_1 and q_2

$$q_1 = j \text{ if } \mu_j \leq x'\beta_1 + u_1 < \mu_{j+1},\ j = 0, ..., J \tag{3}$$

$$q_2 = h \text{ if } \xi_h < x'\beta_2 + u_2 \leq \xi_{h+1},\ h = 0, ..., H \tag{4}$$

where the μ's and ξ's are threshold parameters, such that $\mu_0 = -\infty$, $\mu_1 = 0$, $\mu_{J+1} = \infty$, $\xi_0 = -\infty$, $\xi_1 = 0$, $\xi_{h+1} = \infty$, and $\mu_2, ..., \mu_J$ and $\xi_2, ..., \xi_h$ are estimable. The error terms u_1 and u_2 are bivariate distributed, not necessarily normal as specified in existing literature, with zero means, unitary variances, and a correlation structure specified below. The likelihood contribution for a sample observation is

$$L = \prod_{j=0}^{J} \prod_{h=0}^{H} \{\Pr(q_1 = j, q_2 = h)\}^{1(q_1=j, q_2=h)} \tag{5}$$

where $1(\bullet)$ is a dichotomous indicator function. Relative to existing single-equation ordered probability models, the bivariate system in Equations (3) and (4) allows dependent errors and yields more efficient estimates. Apart from the generalised error distribution described in the following paragraphs, the system is identical to the bivariate Gaussian ordered probit model (Calhoun 1989; Butler and Chatterjee 1997; Greene 2008, pp. 835–837), which includes the bivariate probit model (Greene 2008, pp. 817–822) as a special case. Because distributional assumptions are important in discrete response modelling (Horowitz 1993, p. 70), we extend this existing model to one with non-Gaussian error distributions.

Define the bivariate cumulative distribution function (CDF) $F(\tau_1, \tau_2) = \Pr(t_1 \leq \tau_1, t_2 \leq \tau_2)$ with marginal CDFs $F_1(\tau_1) = \Pr(t_1 \leq \tau_1)$ and $F_2(\tau_2) = \Pr(t_2 \leq \tau_2)$. Then, the component probabilities in Equation (5) are

$$\begin{aligned} \Pr(q_1 = j, q_2 = h) &= \int_{\mu_j - x'\beta_1}^{\mu_{j+1} - x'\beta_1} \int_{\xi_h - x'\beta_2}^{\xi_{h+1} - x'\beta_2} f(u_1, u_2)\, du_2 du_1 \\ &= F(\mu_{j+1} - x'\beta_1, \xi_{h+1} - x'\beta_2) - F(\mu_{j+1} - x'\beta_1, \xi_h - x'\beta_2) \\ &\quad - F(\mu_j - x'\beta_1, \xi_{h+1} - x'\beta_2) + F(\mu_j - x'\beta_1, \xi_h - x'\beta_2) \\ &\quad j = 0, ..., J,\ \ h = 0, ..., H. \end{aligned} \tag{6}$$

Each bivariate CDF on the right-hand side of Equation (6) is specified as a general CDF using the copula approach. The procedure involves specifying two marginal CDFs and linking these ''marginals'' with a copula function. For example, express

$$F_1 = F_1(\mu_{j+1} - x'\beta_1),\ F_2 = F_2(\xi_{h+1} - x'\beta_2). \tag{7}$$

Then, the first bivariate CDF on the right-hand side of (6) can be expressed as

$$F(\mu_{j+1} - x'\beta_1, \xi_{h+1} - x'\beta_2) = C[F_1(\mu_{j+1} - x'\beta_1),\ F_2(\xi_{h+1} - x'\beta_2); \theta] \tag{8}$$

where C is a copula function and θ is a concordance parameter which is a measure of association between the two random variables implied. Details on specific copulas used in this study (Gaussian, Frank, and Clayton copulas) are presented in the Appendix; also see Nelsen (2006).

Most copulas, except the Gaussian copula, admit skewness in the random errors even with symmetric marginals. Additional skewness can be accommodated by using skewed marginals. We consider two forms of marginals for F_1 and F_2. The first is the benchmark Gaussian CDF which corresponds to a symmetric probability density function, and the other is the generalised log-Burr CDF for standardised random variable u_i (Burr 1942):

$$F_i(u_i; \kappa_i) = 1 - (1 + \kappa_i e^{u_i})^{-1/\kappa_i},\ -\infty < u_i < \infty,\ i = 1, 2 \tag{9}$$

Including the logistic ($\kappa_i = 1$) and type-1 extreme-value ($\kappa_i \rightarrow 0$) distributions as special cases (Johnson *et al.* 1995, pp. 2–14, 116), the generalised log-Burr distribution can deliver very different probabilities even with a moderate range of skewness.

To demonstrate the copula approach in the present context, for a model with Gaussian copula (see Appendix) and generalised log-Burr marginals (henceforth, Gaussian–Burr model), the preferred specification in this study, the first probability on the right-hand side of Equation (6) is obtained by substituting $F_1(u_1;\ \kappa_1)$ and $F_2(u_2;\ \kappa_2)$ from Equation (9) into Equation (8), with a specific form in Equation (A.2):

$$\begin{aligned} &F(\mu_{j+1} - x'\beta_1, \xi_{h+1} - x'\beta_2; \kappa_1, \kappa_2) \\ &\quad = \Phi_2\{\Phi^{-1}[1 - (1 + \kappa_1 \exp(\mu_{j+1} - x'\beta_1))^{-1/\kappa_1},], \\ &\qquad \Phi^{-1}[1 - (1 + \kappa_2 \exp(\xi_{h+1} - x'\beta_2))^{-1/\kappa_2}]; \theta\} \end{aligned} \tag{10}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the univariate standard normal CDF and θ is Pearson's correlation coefficient between random variables u_1 and u_2. The remaining probabilities in Equation (6) are similar with different threshold parameters μ's and ξ's.

We calculated the marginal effects of positive and maximum numbers of servings for fruits and vegetables, respectively:[1]

$$\Pr(q_1 > 0) = 1 - F_1(0 - x'\beta_1; \kappa_1);\ \Pr(q_1 = J) = 1 - F_1(\mu_J - x'\beta_1; \kappa_1) \tag{11}$$

$$\Pr(q_2 > 0) = 1 - F_2(0 - x'\beta_2; \kappa_2);\ \Pr(q_2 = H) = 1 - F_2(\xi_H - x'\beta_2; \kappa_2). \tag{12}$$

as well as the means of the two dependent variables,

[1] It is possible to examine the marginal effects on all category probabilities but, owing to space consideration, we focus on the two extreme categories.

$$E(q_1) = \sum_{j=0}^{J} \Pr(q_1 = j)\, \bar{M}_{1j} \tag{13}$$

$$E(q_2) = \sum_{h=0}^{H} \Pr(q_2 = h)\, \bar{M}_{2h} \tag{14}$$

where $\bar{M}_{1j}$ ($j = 1,\ldots,J$) are the category means of q_1, $\bar{M}_{2h}$ ($h = 1,\ldots, H$) are the category means of q_2, and the marginal probabilities Pr ($q_1 = j$) and Pr ($q_2 = h$) follow from the joint probability in Equation (6). Marginal effects of explanatory variables are derived by differentiating Equations (11)–(14).

4. Data and variable definitions

4.1. The survey

Data for this study were obtained from the Malaysia Non-Communicable Disease Surveillance-1 (MyNCDS-1) collected by the Ministry of Health Malaysia (2006). The cross-sectional population-based survey covered the fourteen states in Malaysia. Data were collected from September 2005 to February 2006 according to a two-stage stratified random sampling procedure to ensure that the sample is representative of the Malaysian population. During the survey, field survey teams described the survey to household members to gather socio-demographic information, own medical history, family medical history and lifestyle behaviours. Inclusion criteria are those between 25 and 64 years of age and across all ethnic groups of both genders. From an initial sample of 3040, 2572 (84.6 per cent) respondents were retained in the survey. Our final sample contains 2447 observations after excluding observations with missing data.

4.2. Definition of variables

Data for our dependent variables, FV consumption, were collected as counts—numbers of servings per week.[2] Figure 1 shows the raw counts of weekly fruit servings. The co-existence of extremely high and low counts, multi-modal distribution and, more importantly, empty cells suggests that the variations in servings are not likely to be adequately explained by any count distribution. Thus, these counts are consolidated into fewer categories. See

[2] Our data also contain servings of FV per day, which are more likely to be subject to infrequency of purchase. Estimation of the servings per day system suggests the Gumbel-Gaussian model as the preferred specification. Marginal effects are generally similar to the per-week estimates in signs but sparser in statistical significance. The complete set of results for the per-day system are available upon request.

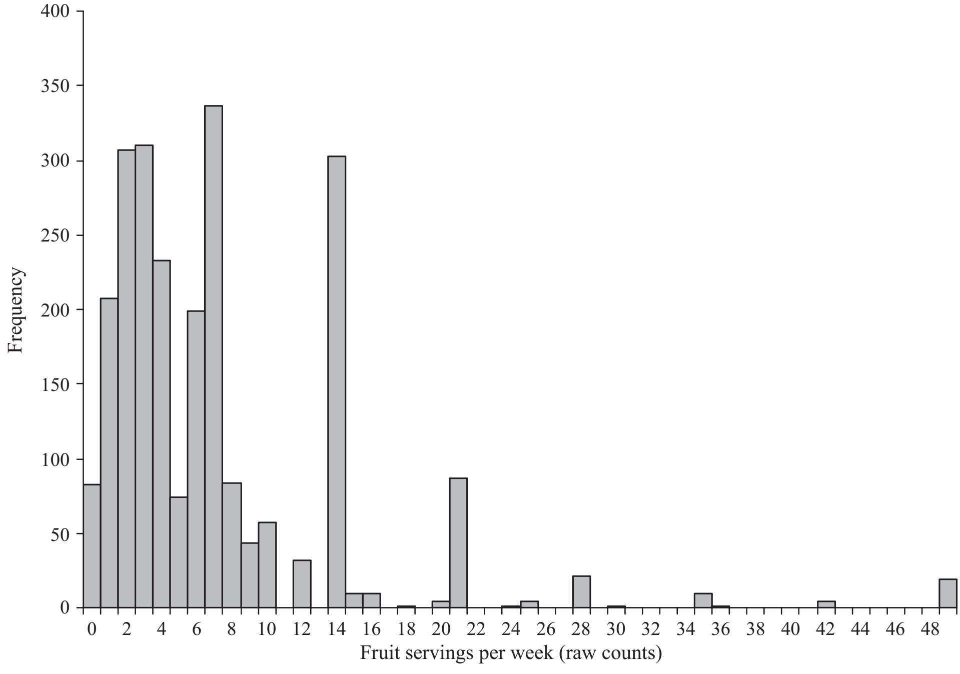

Figure 1 Frequency histograms of fruit servings per week (raw counts).

Table 1 for details on coding. Figure 2 illustrates the frequency histogram of the re-coded categories for fruit servings, suggesting these categories are likely to be explained better with an ordered probability model. Ordered probability models are also found to perform better than count data models with excessive zeros and sporadic high counts such as the number of cigarettes smoked (Kasteridis *et al*. 2010). Frequency histograms for vegetable counts show similar patterns and are omitted for brevity.

The following socio-demographic variables are hypothesised to influence FV demand: length of a typical work day (work hours), education levels, age brackets, ethnicity/race, income levels, gender, marital status, smoking status, health status, and location of residence (Table 2). Length of a typical work

Table 1 Sample statistics of dependent variables

Variable	Mean	SD	Minimum	Maximum
Fruits				
Servings per week	7.08	7.02	0	49
Re-coded counts	2.50	1.66	0	6
Vegetables				
Servings per week	13.08	7.58	0	49
Re-coded counts	4.29	1.41	0	6

Servings per week are recoded into seven categories: 0, 1 (1–3 servings), 2 (4–6), 3 (7–9), 4 (10–13), 5 (14–16), and 6 (> 16).

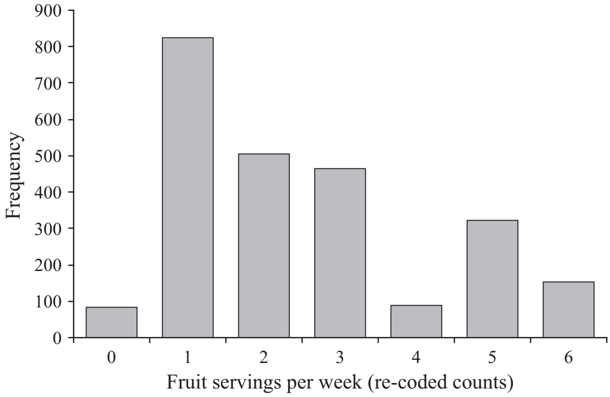

Figure 2 Frequency histograms of fruit servings per week (re-coded counts).

day is used to proxy the time available for healthy food consumption. The respondent's highest academic qualification is denoted by primary/grade-school (reference), junior-high, senior-high and tertiary. Age groups of the respondent are represented by younger (age ≤30) (reference), middle-age younger (age 31–40), middle-age older (age 41–58), and retiree (age ≥59). Ethnicity is represented by Malay, Chinese, Indian, and other races (reference). Monthly household income is defined by five income categories: poverty, low, middle-low, middle-high and high-income (reference) brackets. Two dummy variables are also included to indicate gender (male) and marital status (single).

Several unique socio-demographic variables are also considered. For instance, in order to capture the effect of unobserved time preferences of consumers on diet choice, cigarette smoking status (smoker) is included as a proxy for time preference, besides the inclusion of the respondent's highest academic qualification as previously mentioned. It is posited by Fuchs (1982) and Huston and Finke (2003) that individuals with higher discount rates, as measured by variables, including smoking and education, tend to have less healthy diets. In such cases, time preference variations affect individual smoking behaviour, given that smokers place less value on the future than non-smokers. In contrast, those with better education are hypothesised to have a lower time preference while placing more emphasis than average on their well-being in the further future, and thus are more likely to adopt healthy diet practices.

The respondent's current health status is hypothesised to affect FV demand as those diagnosed with a medical condition(s) are expected to take better care of their health and well-being. In this case, respondents diagnosed with hypercholesterolemia (hyperchol), hypertension (high BP) and diabetes are hypothesised to have a positive relationship with FV.

Finally, given that data from MyNCDS-1 do not contain FV price information, state of residence of the respondent is constructed as a proxy for

Table 2 Definitions and sample means of explanatory variables ($n = 2,440$)

Variable	Definition	Mean
Continuous variable		
Work hours	Length of typical work day (hours)	7.40 (SD = 3.17)
Binary variables (1 = yes; 0 = no)		
Education, highest level of		
Primary	Primary school (reference)	0.42
Junior-high	Junior-high	0.22
Senior-high	Senior-high	0.26
Tertiary	Tertiary education	0.10
Age categories (years)		
Age ≤30	≤30 (reference)	0.13
Age 31–40	31–40	0.27
Age 41–58	41–58	0.51
Age ≥59	≥59	0.09
Ethnicity		
Malay	Malay	0.55
Chinese	Chinese	0.18
Indian	Indian	0.09
Others	Others (reference)	0.18
Monthly income categories		
Poverty income	RM0–399	0.11
Low income	RM400–999	0.36
Middle-low income	RM1000–2999	0.38
Middle-high income	RM3000–5999	0.06
High income	≥RM6000 (reference)	0.03
Male	Male	0.41
Single	Single, divorced or widowed	0.13
Smoker	Currently smoking cigarettes	0.21
Diagnosed with health problems		
Hyperchol	Hypercholesterolemia	0.56
High BP	High-blood-pressure	0.32
Diabetes	Diabetes	0.13
Regions		
Region1	Penang, Selangor, Federal Territory	0.19
Region2	Perlis, Kedah, Perak, Melaka, Negeri Sembilan, Johor, Pahang, Kelantan, Terengganu (reference)	0.56
Region3	Sabah, Sarawak	0.25

Compiled from Ministry of Health Malaysia (2006). US\$1.00 = RM3.72 (February 2006).

price. While FV prices may vary between locations, the standard of living in metropolitan states with higher population densities would be invariably higher compared to that of less metropolitan states. As such, dummy variables denoted by Region 1 (consisting of respondents hailing from the metropolitan states of Penang, Selangor and the Federal Territory in West Peninsular Malaysia), Region 2 (representing the less metropolitan states of Perlis, Kedah, Perak, Melaka, Negeri Sembilan, Johor, Pahang, Kelantan and Terengganu in West Peninsular Malaysia) (reference), and Region 3 (representing the Borneo states of Sabah and Sarawak in East Malaysia) are included to examine if the standard of living (and hence prices) in various locations in the country may affect FV consumption patterns. Prices in a single cross section reflect mainly regional variations, and these regional variables are expected to capture such variations.

5. Maximum-likelihood estimation and model selection

The FV serving system is estimated with alternative copulas and marginals. The Gaussian and Frank copulas are first used, and after obtaining positive error correlation with both of these copulas, the Clayton copula, which admits only positive correlation, is also used. The latter two are members of the *Archimedean* family of copulas most commonly used in empirical applications with discrete dependent variables (Smith 2003).[3] The models with different copulas and marginals are non-nested, and choice among these models can be made using a non-nested specification test procedure. Specifically, let r_i and s_i be the maximum log likelihood contributions of sample observation i for two competing specifications and define differences $d_i = r_i - s_i$ for $i = 1,\ldots,n$ with sample mean $\bar{d}$ and standard deviation s_d. Then, under the null hypothesis of no difference between the two models, Vuong's (1989, equations (3.1), (4.2), (5.6)) standard normal statistic is $z = n^{\frac{1}{2}}\bar{d}/s_d \sim \mathcal{N}(0, 1)$.

Estimation was successful only with the generalised log-Burr marginals and with only three copulas. The Gaussian–Burr model is found to perform better than both the Clayton–Burr (z = 1.89, P-value = 0.06) and Frank–Burr (z = 3.61, P-value < 0.01) models. For brevity, we only present results for the Gaussian–Burr model (Table 3).

The Pearson's error correlation is significant at the 1 per cent significance level. In addition, all threshold parameters are positive and significant at the 1 per cent significance level for both the FV equations, suggesting that these threshold parameters are successful in delimiting the consumption categories. A negative threshold coefficient(s) would have implied misspecification of the model, and insignificant threshold coefficient(s) would have called for alternative categorisation of the consumption categories.

Estimates for the skewness parameters are significantly different from 0 and 1 for both FV equations, rejecting both the nested logistic and extreme-value marginals. Although the generalised log-Burr distribution does not nest the Gaussian distribution, it nests the symmetric logistic distribution. The skewness uncovered, with coefficients of 0.37 and 3.81 for the error terms of fruits and vegetables respectively, suggests that the Gaussian marginals would not have been acceptable. These skewed error distributions may be one reason why our attempt with the Gaussian marginals was unsuccessful.

[3] Another Archimedean copula, the Gumbel copula (Nelsen 2006, p. 116), is also considered but estimation was unsuccessful. The Gumbel copula was preferred for the per-day system (not reported).

Table 3 Maximum-likelihood estimates of ordinal equation systems for fruit and vegetable demands: Gaussian–Burr model

	Fruits		Vegetables	
Variable	Estimate	SE	Estimate	SE
Constant	2.398***	0.221	5.759***	0.526
Work hours	0.021**	0.009	0.044**	0.022
Junior-high	−0.038	0.077	0.132	0.183
Senior-high	0.349***	0.084	0.462**	0.193
Tertiary	0.327***	0.117	0.527*	0.277
Age 31–40	0.092	0.096	0.077	0.222
Age 41–58	0.456***	0.102	0.658***	0.231
Age ≥ 59	0.513***	0.148	0.631**	0.322
Malay	0.222**	0.105	−0.911***	0.256
Chinese	0.483***	0.117	−0.468*	0.275
Indian	0.295**	0.142	−0.095	0.336
Poverty income	−0.322**	0.135	−0.684**	0.310
Low income	−0.064	0.109	−0.596**	0.260
Middle-low income	−0.080	0.106	0.518**	0.251
Middle-high income	0.121	0.138	1.597***	0.385
Male	−0.120*	0.066	−0.217	0.169
Single	0.142*	0.085	0.205	0.203
Smoker	−0.219***	0.080	−0.317*	0.192
Hyperchol	0.262***	0.063	0.269**	0.138
High BP	0.036	0.065	−0.200	0.154
Diabetes	−0.144	0.088	−0.476**	0.203
Region 1	0.589***	0.087	−1.794***	0.229
Region 3	0.501***	0.102	0.890***	0.237
μ_2, ξ_2	2.710***	0.121	2.672***	0.270
μ_3, ξ_3	3.442***	0.139	3.457***	0.286
μ_4, ξ_4	4.135***	0.173	5.715***	0.356
μ_5, ξ_5	4.294***	0.185	5.969***	0.371
μ_6, ξ_6	5.131***	0.273	12.345***	1.151
Skewness (κ*)	0.365***	0.136	3.814***	0.602
θ	0.205***	0.021		
Log likelihood	−7135.428			

Asymptotic standard errors in parentheses. Asterisks indicate levels of significance: *** = 1%, ** = 5%, * = 10%.

Nearly two-thirds of the explanatory variables are significant at the 10 per cent level of significance or lower. As the effects of explanatory variables can be examined in greater depth by calculating marginal effects, we defer further discussions of the effects of explanatory variables to the next section.

6. Marginal effects of explanatory variables

Marginal effects of explanatory variables on the probabilities and conditional means of FV servings are presented in Table 4. Work hours play a positive, albeit small, role in FV consumption. All else equal, a one-hour increase in work hours increases the probability of consuming maximum (≥17) servings of fruits by 0.19 per cent and of vegetables by 0.15 per cent. The corresponding effects on levels are 0.09 serving of fruits and 0.07 serving of vegetables.

Compared to their less-educated cohort, those with senior-high school and tertiary education are 3.47 per cent and 3.19 per cent more likely, respectively, to eat the maximum servings of fruits. On levels, these individuals consume 1.58 and 1.47 more servings of fruits, respectively. For vegetables, individuals with senior-high school (1.58 per cent) and tertiary (1.82 per cent) education are more likely to eat the maximum servings and eat 0.75 and 0.85 more serving, respectively, compared to their less-educated cohorts. It is worth noting that education may be proxying the impacts of differing time preference rates on demand for healthy eating. In this case, a lower time preference amongst those who are higher educated is observed, given their inclination for healthy diet choices.

Individuals between 41 and 58 years and those with age ≥59 years consume more FV than their younger (age ≤31) cohort. These groups of consumers are also more likely to consume the maximum servings of FV. One possible reason for these positive effects of age is that individuals above 40 years may be more cautious about their health and diet and, hence, consume more FV than their younger counterparts.

Ethnicity is a statistically significant factor in FV consumption. Relative to individuals of other ethnicity, Malays and Chinese are both more likely to consume any fruits and the maximum servings of fruits whereas the effects on vegetables are the opposite. Those of Indian ethnicity consume more fruits compared with others. In terms of levels, the Malay, Chinese, and Indians consume 0.90, 2.12 and 1.23 more servings of fruits, and 1.46, 0.74, and 0.15 (insignificant) less servings of vegetables, respectively.

FV consumption is dependent on income levels. Relative to high-income individuals, those in the poverty-income bracket are 1.08 per cent less likely to consume any fruits, 2.58 per cent less likely to consume the maximum servings, and consume 1.35 less servings of fruits. Compared with high-income individuals, poverty and low-income earners are 2.12 per cent and 1.86 per cent less likely, respectively, to consume the maximum number of servings of vegetables, which translate to 1.13 and 0.99 fewer servings of vegetables. However, the effects are reverse amongst middle-low and middle-high income earners, who are 1.87 per cent and 6.7 per cent more likely to consume the maximum servings of vegetables compared with high-income individuals. Middle-low and middle-high income earners also consume 0.82 and 2.4 more servings of vegetables than high-income individuals. Overall, the results show that while lower income individuals may be eating less FV, consumption invariably increases with income, implying that FV may be viewed as normal goods.

The effects of gender suggest that men consume less fruits while the effects on vegetables are insignificant. The effects of marital status are seen only in fruit servings as singles are 0.42 per cent more likely to consume any fruits and eat 0.64 more serving of fruits compared with married persons.

Table 4 Marginal effects of explanatory variables on probabilities and conditional means of fruit and vegetable servings per week

Variable	Fruits		Mean number of servings	Vegetables		Mean number of servings
	Probability (× 100) of			Probability (× 100) of		
	> 0 serving	17–49 servings		> 0 serving	17–49 servings	
Continuous explanatory variable						
Work hours	0.066	0.193**	0.094**	0.012	0.149**	0.071*
Binary explanatory variables						
Junior-high	−0.135	−0.283	−0.156	0.041	0.432	0.216
Senior-high	1.024***	3.470***	1.580***	0.124*	1.579**	0.747**
Tertiary	0.967***	3.190**	1.469***	0.138*	1.816*	0.849*
Age 31–40	0.362	0.548	0.344	0.031	0.238	0.128
Age 41–58	1.528***	3.724***	1.913***	0.200*	2.183***	1.072***
Age ≥ 59	1.674***	4.381***	2.185***	0.194*	2.088*	1.030**
Malay	0.775**	1.659**	0.904**	−0.226***	−3.209***	−1.459***
Chinese	1.499***	4.455***	2.120***	−0.091	−1.743*	−0.737*
Indian	0.997**	2.344**	1.229**	−0.015	−0.371	−0.147
Poverty income	−1.081**	−2.583**	−1.345**	−0.271*	−2.115**	−1.133**
Low income	−0.189	−0.618	−0.287	−0.225**	−1.863**	−0.985**
Middle-low income	−0.238	−0.763	−0.357	0.113	1.872**	0.823**
Middle-high income	0.327	1.333	0.570	0.223***	6.696***	2.397***
Male	−0.375*	−1.066*	−0.521*	−0.061	−0.730	−0.352
Single	0.419*	1.385	0.639*	0.052	0.707	0.330
Smoker	−0.722**	−1.801***	−0.925***	−0.096	−1.049*	−0.517*
Hyperchol	0.824***	2.314***	1.134***	0.076*	0.909*	0.437**
High BP	0.111	0.330	0.159	−0.057	−0.671	−0.325
Diabetes	−0.468	−1.195*	−0.609*	−0.157*	−1.540**	−0.780**
Region 1	1.554***	6.940***	2.831***	−0.954***	−5.317***	−2.995***
Region 3	1.387***	5.477***	2.344***	0.202***	3.204***	1.416***

Asterisks indicate levels of significance: *** = 1%, ** = 5%, * = 10%.

Smokers are 0.72 per cent less likely to consume any fruits and 1.80 per cent less likely to eat the maximum servings of fruits. Hence, these individuals consume 0.93 less serving of fruits than non-smokers. In addition, smokers are 1.05 per cent less likely to consume the maximum servings and consume 0.52 less serving of vegetables than non-smokers. This outcome may not necessarily be viewed negatively as recent scientific studies have shown that while high intakes of FV may lower health risks among non-smokers, it may even have adverse effects on smokers. As noted by van Duijnhoven *et al*. (2009), substances within FV may increase the carcinogenic potential of tobacco smoke to cause colon cancer instead. It is worth noting that while this result does not diminish the health benefits of consuming FV, it should instead renew efforts to cease smoking. Further, as smoking may have important associations with dietary consumption, this suggests that time preference variations across individuals matter in healthy diet choices.

Hypercholesterolemic patients seem to be aware of the importance of eating healthy foods as their propensity to consume (any) fruits is 0.82 per cent higher. Their propensity to consume the maximum servings are 2.31 per cent higher, which corresponds to 1.13 more servings of fruits. Similarly, hypercholesterolemic patients are 0.08 per cent more likely to consume any vegetables, 0.91 per cent more likely to consume the maximum servings of vegetables, and consume 0.44 more servings compared with those without this ailment. Conversely, patients with diabetes are more complacent of their diets as they are less likely to consume the maximum servings of fruits (1.20 per cent). This results in negative effects on serving levels as diabetics eat 0.61 less servings of fruits compared with non-diabetics. Patients with diabetes are 1.54 per cent less likely to consume the maximum servings and also consume 0.78 less servings of vegetables than non-diabetics.

Although faced with the prospects of higher standards of living, individuals in the metropolitan areas of Region 1 are 1.55 per cent more likely to consume any fruits and 6.94 per cent more likely to consume the maximum servings compared with residents in the less metropolitan states of Region 2. These positive effects contribute to the additional consumption of 2.83 servings amongst residents of Region 1. In contrast, the propensities to consume any vegetables (0.95 per cent) and the maximum servings (5.32 per cent) are lower compared with their counterparts in states with relatively lower living costs. Subsequently, these negative effects are reflected in the lower consumption of 3.00 servings of vegetables for residents in Region 1 compared with those in Region 2. These results, which support the findings of Huang and Lin (2000), indicate that higher prices in the metropolitan cities in Malaysia clearly do not deter the consumption of fruits as consumers may even be inclined to pay for imported or "status" fruits such as apples, oranges, pears, grapes and so forth. Meanwhile, in contrast to the results of Krebs-Smith and Kantor (2001) and Gustavsen and Rickertsen (2006), individuals from the relatively rural states and with lower standard of living may be more inclined to consume vegetables, given that vegetable farming is considered as one of the primary cottage industries in Malaysia.

It is interesting to note that consumers in the Borneo states of Sabah and Sarawak (Region 3) have a 1.39 per cent higher propensity to consume any fruits and a 5.48 per cent higher propensity to eat the maximum servings of fruits, whilst consuming 2.34 more servings of fruits compared with residents in the less metropolitan states of Peninsular Malaysia. In addition, residents in Region 3 are 3.20 per cent more likely to consume the maximum servings and also consume 1.42 more servings of vegetables compared with those in Region 2. This is despite the fact that transportation costs, and hence consumer prices, are conceivably higher in these East Malaysian states compared with the less metropolitan states in Peninsular Malaysia.

7. Discussion and concluding remarks

Some of the world's most widespread and debilitating nutritional disorders, including birth defects, mental and physical retardation, weakened immune systems, blindness, and even death, are caused by diets lacking in FV. Therefore, promoting consumption of FV often tops the list of priorities for nutrition educators. Although a number of studies on FV consumption have focused on western countries, scant information is available on the determinants of FV consumption in developing countries, including Malaysia.

In this study, we analyse the determinants of FV consumption in Malaysia with a bivariate ordered probability model using the copula approach. Results of the study indicate that socio-demographic factors such as education, age, ethnicity, income, location of residence, as well as smoking status and health conditions significantly affect FV consumption. Specifically, education affects consumption patterns of FV as those with at least high school education consume significantly more servings compared with primary school–educated individuals. Individuals between 41 and 58 years of age consume more FV than those below 30 years of age. The Chinese consume more fruits than others while Malays, Chinese and Indians eat lesser vegetables than those of other ethnic backgrounds. Low-income individuals also consume less FV.

The responses of smokers towards fruit consumption are less encouraging compared with non-smokers, thus suggesting the possibility that such individuals may have a higher time preference and value immediate utility over delayed utility. Meanwhile, individual health conditions are significant contributors of FV consumption as evidenced by the higher consumption rates of fruits amongst those with hypercholesterolemia and lower consumption levels of fruits amongst diabetics compared to those without these ailments.

Our results call for measures to direct Malaysians to healthier dietary choices. Interventions that increase FV consumption by changing behaviours should be considered, as should those that increase public awareness of the benefits of FV in the diet. However, nutritional interventions should go beyond increasing awareness and targeting specific groups of individuals. For example, nutritionists dealing with lower educated groups might attempt to eliminate barriers to healthy eating, provide support for individuals making healthy changes, increase resources for populations with greater need and emphasise nutritional policies that have an impact on the society. Simply put, intervention programs should be targeted at and tailored towards those who have lower FV consumption. Based on our findings, these groups in Malaysia generally include the less educated, the young, the poor and smokers.

Although the virtues of consuming FV are often taught in schools, it may be worthwhile to continue educating younger individuals about the benefits of FV, especially given our findings that they are less likely to consume FV than older individuals. Hence, educational programs should target the younger and less-educated groups in order to sustain lifelong beneficial effects. Also, because poorer individuals consume less FV than middle- and high-income individuals, government policies towards providing food assistance to the poor could be geared towards increasing FV consumption. Price and income subsidies are found to be effective program tools in promoting FV consumption in the United States (Lin *et al.* 2010). Other programs might include subsidising the agricultural sector, particularly on FV.

While this study provides interesting new findings for a developing country like Malaysia, future research might also focus on identification of barriers to eating more FV and on evaluating environmental changes that could potentially increase FV consumption (e.g., increasing the proportion of FV in vending machines, promoting healthful food advertising and availability of healthful foods). Finally, while we are able to proxy the missing price information with regional dummy variables, further studies might consider the estimation of a utility-theoretic demand system when survey data containing prices become available.

References

Bleich, S., Blendon, R. and Adams, A. (2007). Trust in scientific experts on obesity: implications for awareness and behavior change, *Obesity* 15(8), 2145–2156.

Blisard, N., Stewart, H. and Jolliffe, D. (2004). *Low-income households' expenditures on fruits and vegetables*, Agricultural Economic Report No. 833, Economic Research Service, U.S. Department of Agriculture, Washington, DC.

Burr, I.W. (1942). Cumulative frequency functions, *The Annals of Mathematical Statistics* 13(2), 215–232.

Butler, J.S. and Chatterjee, P. (1997). Tests of the specification of univariate and bivariate ordered probit, *The Review of Economics and Statistics* 79(2), 343–347.

Calhoun, C. (1989). Estimating the distribution of desired family size and excess fertility, *Journal of Human Resources* 24(4), 709–724.

Casagrande, S.S., Wang, Y., Anderson, C. and Gary, T.L. (2007). Have Americans increased the fruit and vegetable intake? *American Journal of Preventive Medicine* 32(4), 257–263.

Cox, T.L. and Wohlgenant, M.K. (1986). Prices and quality effects in cross-sectional demand analysis, *American Journal of Agricultural Economics* 68(4), 908–919.

van Duijnhoven, F.J., Bueno-De-Mesquita, H.B., Ferrari, P., Jenab, M., Boshuizen, H.C., Ros, M.M., Casagrande, C., Tjønneland, A., Olsen, A., Overvad, K., Thorlacius-Ussing, O., Clavel-Chapelon, F., Boutron-Ruault, M.C., Morois, S., Kaaks, R., Linseisen, J., Boeing, H., Nöthlings, U., Trichopoulou, A., Trichopoulos, D., Misirli, G., Palli, D., Sieri, S., Panico, S., Tumino, R., Vineis, P., Peeters, P.H., van Gils, C.H., Ocké, M.C., Lund, E., Engeset, D., Skeie, G., Suárez, L.R., González, C.A., Sánchez, M.J., Dorronsoro, M., Navarro, C., Barricarte, A., Berglund, G., Manjer, J., Hallmans, G., Palmqvist, R., Bingham, S.A., Khaw, K.T., Key, T.J., Allen, N.E., Boffetta, P., Slimani, N., Rinaldi, S., Gallo, V., Norat, T. and Riboli, E. (2009). Fruit, vegetables, and colorectal cancer risk: the European prospective investigation into cancer and nutrition, *American Journal of Clinical Nutrition* 89(5), 1441–1452.

FAOSTAT, Food and Agriculture Organization Statistics Division (2009). Available from URL: http://faostat.fao.org [accessed 3 June 2009].

Fuchs, V.R. (1982). *Time Preference and health: an exploratory study*, in Fuchs, V.R. (ed), *Economic Aspects of Health*. University of Chicago Press, Chicago, Illinois, pp. 93–120.

Greene, W.H. (2008). *Econometric Analysis*, 6th edn. Prentice Hall, Upper Saddle River, NJ.

Gustavsen, G.W. and Rickertsen, K. (2006). A censored quantile regression analysis of vegetable demand: the effects of changes in prices and total expenditure, *Canadian Journal of Agricultural Economics* 54(4), 631–645.

Horowitz, J.L. (1993). Semiparametric and nonparametric estimation of quantal response models, in Maddala, G.S., Rao, C.R. and Vinod, H. (eds), *Handbook of Statistics*, Vol 11. Elsevier Science Publishing Co., North Holland, Amsterdam, pp. 45–72.

Huang, K. and Lin, B. (2000). *Estimation of food demand and nutrient elasticities from household survey data*, Technical Bulletin No. 1887, Economic Research Service, U.S. Department of Agriculture, Washington, DC.

Huston, S. and Finke, M.S. (2003). Diet choice and the role of time preference, *Journal of Consumer Affairs* 37(1), 143–160.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2. Wiley, New York.

Kasteridis, P.P., Munkin, M.K. and Yen, S.T. (2010). A binary-ordered probit model of cigarette demand, *Applied Economics* 42(4), 413–426.

Krebs-Smith, S. and Kantor, L.S. (2001). Choose a variety of fruits and vegetables daily: understanding the complexities, *Journal of Nutrition* 131(Suppl 2), 487S–501S.

Lin, B. and Huang, C.L. (2008). Demand for organic and conventional fresh fruits. Paper presented at the American Agricultural Economics Annual Meetings, 27–29 Jul 2008, Orlando, Florida, USA.

Lin, B., Yen, S.T., Dong, D. and Smallwood, D.M. (2010). Economic incentives in dietary improvement among food stamp recipients, *Contemporary Economic Policy* 28(4), 524–536.

Ministry of Health Malaysia (2006). *Malaysia NCD Surveillance 2006: NCD Risk Factors in Malaysia*. Ministry of Health, Malaysia: Disease Control Division, Putrajaya, Malaysia.

Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer, New York.

Smith, M.D. (2003). Modelling sample selection using Archimedean copulas, *Econometrics Journal* 6(1), 99–123.

Stewart, H., Harris, J.M. and Guthrie, J. (2004). *What determines the variety of a household's vegetable purchases?* Agriculture Information Bulletin Number 792–3, Economic Research Service, U.S. Department of Agriculture, Washington, DC.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses, *Econometrica* 57(2), 307–333.

World Health Organization (WHO) (2003). World Health Report 2002: Reducing risks, promoting healthy life. Geneva: World Health Organization. Available from URL: http://www.who.int/whr/2002/en/whr02_en.pdf [accessed 2 June 2009].

Appendix

Let H be a joint CDF with marginals F_1 and F_2. Then Sklar's theorem (Nelsen 2006, p. 18) states that there exists a copula C such that for all u_1, u_2 in the extended real line,

$$H(u_1, u_2) = C[F_1(u_2),\ F_2(u_2)]. \tag{A.1}$$

If F_1 and F_2 are continuous, then C is unique. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function defined by (A.1) is a joint CDF with marginals F_1 and F_2. Econometric specifications with the copula approach require choices of the marginals F_1 and F_2 and a copula. We consider the Gaussian copula and two members of the *Archimedean* copulas: the Frank and Clayton copulas (Nelsen 2006, Table 4.1, pp. 116–119).

The Gaussian copula has the form

$$\begin{aligned} C(u_1, u_2) &= \Phi_2[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta] \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left[\frac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)} \right] dsdt,\ -1 < \theta < 1 \end{aligned} \tag{A.2}$$

where Φ and Φ_2 are the univariate and bivariate standard normal CDFs, respectively, and θ is Pearson's correlation coefficient between random variables u_1 and u_2. We also use the Frank copula

$$C_\theta(u_1, u_2) = -\theta^{-1} \ln(1 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)/(e^{-\theta} - 1)),\ -\infty < \theta < \infty \tag{A.3}$$

and the Clayton copula

$$C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta},\ 0 \le \theta < \infty. \tag{A.4}$$

In Equations (A.3)–(A.4), θ is a measure of *concordance* (Nelsen 2006, pp. 157–158) between u_1 and u_2. A more useful measure of association, *the Kendall's* τ, can also be derived from the concordance parameter; see Nelsen (2006, p. 163).