Hatice Ozer Balli¹ / Mohammad Amin Kouhbor² / Rosmy Jean Louis³

Towards Understanding Vegetables Consumption Behaviour in Iran: A Full Box-Cox Double-Hurdle Application

¹ Associate Professor in Economics, School of Economics and Finance, College of Business, Massey University, Room QB 3.44, Quadrangle Building B, Albany, New Zealand, E-mail: H.Ozer-Balli@massey.ac.nz

² Assistant Professor in Economics, Department of Economics, College of Marine Managements and Economic, Khorramshahr University of Marine Science and technology, Iran (the Islamic Republic of), E-mail: aminkuhbor@yahoo.com

³ Professor and Chair, Department of Economics, Vancouver Island University, 900 Fifth St, Nanaimo, BC V9R 5S5, Canada, E-mail: Rosmy.JeanLouis@viu.ca

Abstract:

Using Iran's 2010–2011 household survey data on income and expenditure, this paper estimates the demand for vegetable consumption. Based on the Vuong's (1989) Likelihood Ratio Test for Model Selection and Non-Nested Hypothesis, a full Box-Cox double-hurdle model adjusted for heteroskedasticity, dependency, and normality was estimated to uncover factors underlying Iranian households' decisions to purchase and consume vegetables. Results show that all demographic, socioeconomic, and geographical variables significantly explain vegetable consumption behaviour in Iran. A positive relationship exists between educational attainment and the decision to purchase and consume vegetables. As well, households' size and average age exert a statistically significant positive effect on vegetable consumption.

Keywords: consumer behaviour, food policy, limited dependent variables model, vegetables demand

DOI: 10.1515/rmeef-2016-0017

1 Introduction

According to the central bank of Iran, household monthly vegetable expenditures in real terms fell by about 80,000 Iran's Rial from 1965 to 2011 while consumption shrank by about 300 g over the period 1985–2011. These changes happened at the same time that the share of the vegetable food group in the food budget decreased to 8 % from 11 %. As shown in Figure 1, vegetable consumption displays a sharp decrease by approximately 20 % especially after 2003 along with a 35 % decline in per capita vegetable expenditures over a 25-year period.

Figure 1: (a) Amount of per capita daily vegetables consumption (in grams) (b) Per capita expenditure of vegetables in Iran (in Rials).

A number of factors are likely responsible for the reduction in vegetable consumption patterns observed over the years. First, due to the sheer size of the Iranian population and the competing activities that force consumers to juggle for time, a transition to the readily available convenient fast food might have taken place, as is often the case for younger households in many developing and developed nations. Second, while rising oil prices has provided the government with more revenues to fund the food aid program in which vegetables had no place, population growth has exacerbated these subsidies thereby leaving many with less financial

Mohammad Amin Kouhbor is the corresponding author.

© 2017 Walter de Gruyter GmbH, Berlin/Boston.

resources to pay for vegetables, which are essential for a good food diet but not crucial for the main course of a meal. Third, with the political shift in Iran and the imposition of sanctions by the international community based on the fear that Iran was close to producing a nuclear bomb, individuals' livelihood had geared towards a subsistence or survival mode where the consumption of vegetables was relegated to a secondary even to a tertiary level as income dwindled.

A study on the consumption pattern of vegetables is quite important. According to the World Health Organization (World Health Organization 2015), 1.7 million (2.8 %) of deaths worldwide are attributable to low fruit and vegetable consumption.

As is understood, from a health perspective, vegetables, like fruits are low in fat and contain good amounts of vitamins and minerals. All the green/yellow/orange vegetables are rich in calcium, magnesium, potassium, iron, beta-carotene, vitamin B-complex, vitamin-C, vitamin A, and vitamin K. They are home to many antioxidants that help protect the human body from oxidant stress, diseases and cancers, and also help the body develop the capacity to fight against diseases by boosting immunity.

Additionally, vegetables are packed with soluble as well as insoluble dietary fibre known as non-starch polysaccharides (NSP) such as cellulose, mucilage, hemi-cellulose, gums, pectin, and so forth. These substances absorb excess water in the colon, retain a good amount of moisture, and help its smooth passage out of the body. As per the Global Strategy Diet Information Sheet of World Health Organization (World Health Organization 2015), individuals should aim for at least nine servings (4½ cups) of vegetables and fruits per day.¹

Labour market outcomes also tend to improve with the consumption of vegetables. Since a healthy diet with vegetable content improves wellbeing, we cannot negate the resulting effects in terms of less sick days away from work, visits to doctors, hospitalization, complication of related diseases, and pharma care costs. The overall impact on government budget is a positive one if money is spent either as food subsidy or in promotion campaign to induce people to consume more vegetables. Medical research has from time to time come up with new findings that assert the health benefits of vegetable consumption while nutritionists continue to innovate on our lifestyle and eating habits. These initiatives though quite useful can only increase the incidence and the level of vegetable consumption if underlying demand factors are identified. Unfortunately, studies of this sort have not flourished in the literature, let alone in the case of Iran. There have been some attempts by Dettmann and Dimitri (2007) for the United States; Tiffin and Arnoult (2008) for the United Kingdom, and more recently Jabarin and Al-Karablieh (2011) for Jordan. These studies however are either descriptive or quite restrictive in their parameterization.

The expenditure variable is discrete in that agents who do not participate are assigned a zero value while those who do have a positive numerical value that is continuous since they report how much they consume. A number of estimators can be used to explore this type of censored data. The issue however is to find the most suitable one. To this end, we use the Vuong (1989) modified likelihood ratio test to choose among the various models. The results show that the double hurdle model is the best fit of all estimators. In this vein, we estimate a generalized double hurdle model that accounts for non-normality, heteroskedasticity, and dependence among the two hurdles to arrive at parameter estimates that are consistent. We test whether demographics affect participation decision and whether socio-economic variables can explain the level of vegetables consumption in Iran. Our choice of explanatory variables is by and large motivated by the subsidy reforms that started in 2008 with the main component of the subsidy phasing out.² It is a timely undertaking to assess the effects of these reforms on the vegetable food group and nutrients and on food security in Iran. Results support the hypothesis that both participation and consumption decisions are made simultaneously in the vegetables market. Demographic characteristics, economic factors, and geographical differences exert a statistically significant impact on both the probability of vegetables consumption and the level of consumption in Iran. Estimated elasticities suggest that vegetables are normal good while education and household size are key factors in determining consumption patterns of vegetables in Iran.

The rest of the paper is organized as follows. Section 2 discusses the theoretical underpinnings and the empirical models. Section 3 describes and analyses the data. Section 4 reports and assesses the empirical results. Section 5 concludes the paper.

2 Theoretical Models

2.1 Microeconomic Foundation

Following Huang (1996) and Kiani and Kouhbor (2008) in assuming weak separability between food demand and other non-durable goods, we postulate the objective of a representative consumer is to maximize the following utility function $u^f = u(x_1, x_2, \dots, x_n; w)$ subject to a budget constraint $m = P'X$ where the vector X represents

the n food main groups, w , the non-economic variables that influence food habit, P , the respective vector of food prices, and m , the household income. It is truism that for some food groups such as vegetables, households that are not potential consumers report zero consumption, as have been the case for tobacco and alcohol consumption documented in Aristei and Pieroni (2008), Garcia and Labeaga (1996), Yen and Jensen (1996), Bleylock and Blizard (1992), Jones (1989), and Labeaga (1999)) and food away from home documented in Keelan, Henchion, and Newman (2008). To account for this feature of the income-expenditure data, we split the set of goods consumed into two subgroups; vegetables or x_1 and other products $x_2, \dots, x_n$. Hence, the sub-utility function of the food group can be rewritten as:

$$u^f = du^{pv}(x_1, x_2, \dots, x_n) + (1 - d)u^{npv}(x_2, \dots, x_n) \quad [1]$$

where d is a discrete choice variable that takes 1 for those who derive satisfaction from participation (p) and consumption of vegetables (v), u^{pv} , for which x_1 is included in their basket, and 0 for those who do not. It is clear from eq. 1 that when $d = 0$, the first term vanishes and utility is only derived from the non-vegetable group, u^{npv} . The associated Marshallian demand function for vegetables derived from the constrained maximization problem is given by

$$x_1^* = f(p, m; w) \quad [2]$$

It is worth noting that quantity variation in cross section data is not only due to price differentiation but also to factors such as quality differentiation that is quite important for the vegetables market. To overcome this issue, authors like Newman, Henchion, and Mathews (2003), Aristei and Pieroni (2008), and Garcia and Labeaga (1996) excluded the price variable from their models and estimated Engle curves instead. An alternative approach, however, is to assume a constant price for vegetables to estimate vegetables expenditure as a function of economics variables such as food habit and consumption patterns.

$$e_1^* = f(m; w) \quad [3]$$

The remaining issue with this approach is to find suitable variables that influence food habit to incorporate in vector w . Unfortunately, the literature is mute on this matter and on the appropriate functional form for Eq. (3). One way is to include demographic variables to capture the participation hurdle and economic factors to explain the level of consumption (Jones 1992; Yen and Huang 1996). Another alternative is to include all variables in both equations and exclude non-suitable ones gradually (Aristei and Pieroni 2008).

2.2 Econometric Model

Cross-sectional data have been thus far the most suitable to analyse demand in the literature. However, the main issue is how to deal with zero consumption and missing observations. To overcome this problem, researchers have tried several methods. These regroup the use of limited dependent and zero inflated models such as Censored Tobit model (Tobin 1958),³ the Heckman's Selection model or Heckit (Heckman 1979) and Cragg's Double Hurdle (Cragg 1971) among others such as Truncated tobit and two part model. Results obtained from the Tobit model are often not robust across distributional assumptions (Arabmazar and Schmidt 1982; Amemiya 1984). The caveat with the Tobit model is its presumption that all zero expenditures occur because of unaffordability (high prices or low income or both) which forms the basis for observed standard corner solution. As is well-known zero consumption might occur due to reasons such as infrequent purchases (Garcia and Labeaga 1996; Sue and Yen 1996) or non-participation in market because of differences in consumption habits and culture, especially when it comes to products such as vegetables, cigarettes and derivatives, and beverages. Hence, there is a certain level of bias involved in coefficient estimates and also it is impossible from an econometric estimation standpoint to include different sets of variables for the probability and level of consumption using the Tobit methodology. To overcome this issue, Jabarin and Al-Karablieh (2011) used Heckman's two step procedure and the inverse Mills ratio to correct for the sample selectivity bias in their study of Jordanian fresh vegetable consumption.

The preference for the Heckman over the Tobit model is that in the Heckman model the factors affecting the probability and the level of consumption can differ from each other, whereas in the Tobit model coefficient estimates are efficient when factors affecting both the probability and the level of consumption are the same. It is in this vein that Humphreys (2013) suggests the use of Heckit as appropriate estimation model when zeros are generated because of non-observable responses. Accordingly, Probit is to be used for the selection equation (the first stage) and OLS with an inverse Mills ratio for the observation equation or second stage. A shortcoming of the Heckit model is that there is no possibility for assuming the two decisions are made simultaneously.

A relaxation of the Heckman model is the original Cragg (1971) double hurdle model that deals with zeros in the data.⁴ According to this approach, the demand for vegetables must pass two separate hurdles to observe a positive value; the decision to participate or purchase vegetables and the decision to consume certain amount. These two must take place concurrently. In other words, all nonparticipants must have zero vegetable consumption. As discussed in Garcia and Labeaga (1996), there are three reasons for observing these zeros in the data. The first relates to mismatch between purchase frequency and timing of survey. The second has to do with abstention whereby a conscious decision is made not to consume the good. The third one is low purchasing power; in that current income and prices do not allow the agents to consume the good during the survey's recording period. We conjecture that the Iranian survey data inherited are subjected to these same intricacies.

To overcome the conundrum of the Tobit and Heckit models, we focus on the Generalized Double Hurdle model originally proposed by Cragg (1971). This model postulates that the demand or expenditures have to pass two separate hurdles to observe a positive value. The basic idea is consumers do not make decision about how much to consume at the same time they are making the decision as to whether they want to purchase or not. In the context of our study, the first or participation hurdle is whether to purchase vegetables or not. The second hurdle, contingent on the first one, is if the decision is to purchase then how much, ergo the consumption hurdle. A feature of the double hurdle model is that it is linear in the parameters and additive in the disturbances. The original Cragg's Gaussian likelihood function can be written as:

$$L = \prod_0 \left[1 - \Phi(Z\beta_1, \frac{x\beta_2}{\sigma}) \right] * \prod_+ \left[\Phi(Z\beta_1) \left(\frac{1}{\sigma} \right) \phi(y^* - \beta x_i) / \sigma \right] \quad [4]$$

where ϕ is the standard normal probability distribution function, Φ is cumulative distribution function (CDF) for normal distribution, y^* , the latent variable that takes a positive value if a consumer participates in vegetables consumption, and σ , the standard deviation of participation hurdle. The first part of Eq. (4) is the probability of non-participation that is a function of socio-economic variables (Z_i), and the second part is the probability of positive consumption (X_i).

Although the Cragg's model assumes that the disturbance terms of the two hurdles are normally distributed, uncorrelated and homoscedastic, there is no guarantee that these problems are not present in the data. To deal with these limitations, the generalised Gaussian double hurdle model or "Full Double Hurdle Model" was developed by Jones (1992). This model nests the Cragg's model and accounts for error terms that are heteroskedasticity and interdependent (see Smith 2003). The likelihood function for the Jones model is given by:

$$L = \prod_0 [1 - \phi(\alpha'z, \beta x_i, \rho)] * \prod_+ \left[\Phi \left(\alpha'z + \frac{\frac{\rho}{\sigma_i}(y - \beta x)}{\sqrt{1 - \rho^2}} \right) \frac{1}{\sigma_i} \phi(y - \beta x) / \sigma_i \right] \quad [5]$$

where ρ is the correlation between the two parts of eq. 4 with all other parameters already defined therein. This model assumes that the error term of the two hurdles follow a bivariate normal distribution, which gives rise to inconsistent parameter estimates even if the normality assumption is violated. To correct for this problem, some researchers make use of the Inverse Hyperbolic Sine (IHS) (Newman, Henschon, and Mathews 2003), while others (Aristei and Pieroni 2008) employ the Box-Cox transformation:

$$y^{\text{Box-Cox}} = \frac{y^\gamma - 1}{\gamma}, \quad \text{where } 0 < \gamma < 1 \quad [6]$$

The γ parameter can be selected freely and has two extreme limitations. The first is to assume $\gamma = 0$ which gives the Box-Cox logarithmic transformation, and the second is to assume that $\gamma = 1$, which gives a linear transformation of y . It is however important to note that using the Box-Cox double hurdle model requires two important adjustments. The first is to adopt a selection criterion such as:

$$y^{\text{Box-Cox}} = \begin{cases} y^* & \text{if } y^* > -\frac{1}{\gamma} \\ 0 & \text{Otherwise} \end{cases} \quad [7]$$

The second is to tailor the likelihood function accordingly. In this vein the Box-Cox full double hurdle model can be written as:

$$L = \prod_0 \left[1 - \phi \left(\alpha'z, \frac{\beta x_i + 1/\gamma}{\sigma_i}, \rho \right) \right] * \prod_+ \left[\phi \left(\alpha'z + \frac{\frac{\rho}{\sigma_i}(y_i^{\gamma-1})/\gamma - \beta'x}{\sqrt{1 - \rho^2}} \right) y_i^{(\gamma-1)} \frac{1}{\sigma_i} \phi \left[(y_i^\gamma - 1)/\gamma - \beta x \right] / \sigma_i \right] \quad [8]$$

This is a nested model of all of the previous models discussed above.

2.3 Specification Tests

As mentioned earlier, the Tobit model nested in the double hurdle model can be derived by restricting the parameters of Probit participation hurdle to be equal to the parameters of the consumption coefficients.⁵ So, the double hurdle model can be easily tested against the Tobit using the standard likelihood ratio test:

$$LR = -2(LL_{\text{Double Hurdle}} - LL_{\text{Tobit}})$$

where LL is the log-likelihood value of each model shown in Table 3. The LR test follows a χ^2 distribution with degrees of freedom equal to the number of restricted parameters or slope coefficient estimates for the first hurdle. As evident from Table 3, the large computed value for this statistic strongly rejects the Tobit specification in favour of the Box-Cox double hurdle model. However, as Vong (1989) pointed out, the standard LR test is not suitable for non-nested models unless some adjustments are made. Accordingly, the modified LR test is given by:

$$LR_1 = -2(LL_{\text{Double Hurdle}} - LL_{\text{Heckit}})$$

and the Vuong statistic is computed as follows:

$$v_n = \left(\frac{1}{n}\right) LR_1^2 - \left[\left(\frac{1}{n}\right) LR_1\right]^2 \quad [9]$$

where n is the number of observation and LR_1 the difference between the log-likelihood of two non-nested models. As Vuong also shows $\sqrt{n} \frac{LR_1}{v_n}$ follows a normal standard distribution and larger difference between the maximum likelihood of the two alternative non-nested models leads to a greater Vuong statistic. The null hypothesis is that the Heckit model is more suitable than the double hurdle model. In other words, if the Vuong statistics lies in the critical region, the double hurdle model is a better fit.

3 Elasticity and Marginal Effects

We follow a two-step procedure to calculate the elasticity and marginal effects. Following Newman, Henschon, and Mathews (2003), Yen and Jensen (1996), and Jones (1992), the first step consists in finding the conditional and unconditional expected values for the dependent variables. That is, the probability of participation and the expected level of consumption as given by: $p(y_i > 0) = \phi(\alpha'z)\phi\left(\frac{\beta'x_i}{\sigma}\right)$

$$E(y_i|y_i > 0) = \Phi\left(\frac{\beta'x_i + 1/\gamma}{\sigma_i}\right)^{-1} \int_0^\infty \left(\frac{y_i^\gamma}{\sigma_i} \phi\left(\frac{y_i^{\text{Box-Cox}} - \beta'x_i}{\sigma_i}\right)\right) dy_i \quad [10]$$

With the unconditional expectation represented by:

$$E(y_i) = E(y_i|y_i > 0) \cdot p(y_i > 0)$$

Accordingly, we compute the

$$\text{Unconditional marginal effects for consumption} = \frac{\partial E(y_i)}{\partial (\text{Explanatory variable}_i)}$$

$$\text{Conditional marginal effects for consumption} = \frac{\partial E(y_i|y_i > 0)}{\partial (\text{Explanatory variable}_i)}$$

$$\text{Marginal effect of the probability} = \frac{\partial p(y_i)}{\partial (\text{Explanatory variable}_i)}$$

Having these marginal effects, we can proceed to the next step to calculate the elasticities at mean or at margin. However, since food subsidy policies usually aim at low income households, it is more interesting to calculate the elasticities at the marginal households.

Conditional elasticity of consumption with respect to

$$x_{ij} : e_j^u = \frac{\partial E(y_i)}{\partial x_{ji}} \cdot \frac{x_{ji}}{E(y_i)} \quad [11]$$

Unconditional marginal effects for consumption:

$$e_j^c = \frac{\partial E(y_i|y_i > 0)}{\partial x_{ji}} \cdot \frac{x_{ji}}{E(y_i|y_i > 0)} \quad [12]$$

Marginal effect of the probability:

$$e_j^p = \frac{\partial p(y_i > 0)}{\partial x_{ji}} \cdot \frac{x_{ji}}{p(y_i > 0)} \quad [13]$$

And of course:

$$e_j^u = e_j^c + e_j^p \quad [14]$$

4 Data and Descriptive Statistics

The required data for this study are the 2010–2011 Iran's income-expenditure survey data obtained from the Iranian Bureau of Statistics. We made use of the first three parts of the questionnaire that were about household income, vegetable consumption, and socio-demographic variables such as household size, the number of women in each household, the average age of each household, the education characteristics, among others. Also, we exploit features of the data related to household postal codes to differentiate between those living in rural, urban, West, East, North, South and the Centre of Iran.

For each code of vegetables and for each record (household), information about the quantity (in grams) consumed and expenditure (in Rial) during the last month (at the point of data gathering) was listed. The aggregate vegetable consumption and its corresponding expenditure were computed as the summation of the vegetables sub-categories. As discussed earlier, some households did report zero consumption during the data gathering period, which gave rise to the non-participation issue, discussed earlier, hence the appropriateness of the limited dependent models.

The original sample size was 31,075 household, of which 14,703 household were urban. We truncated the records for households that did not report any information on incomes and arrive at a sample size of 29,123 households with, 1,238, reporting zero consumption for vegetable during the month.

In our proposed empirical model to be estimated, the expenditure that each household allocates to vegetable divided by the household size (in a month) is the dependent variable for the second hurdle. In the first hurdle, the dependent variable takes the value of 1 if a household participate in the vegetables consumption and 0 otherwise. The explanatory variables are partitioned into three categories including economic, demographic, and geographical variables. The economic impact is measured with the natural logarithm of income. For the demographic variables, we use the size of households, the number of educated individuals per household, the average age of the household, and the educational attainment of the head of each household as to whether they have no diploma, a diploma, or more than one diploma. The geographical variables are categorical variables representing food habit formation that varies across North, Centre, South, East, and West. To differentiate between consumption habits across households living in urban and rural areas, we create a dummy variable that takes the value of 1 if the household lives in a city and 0 otherwise.

We present in Table 1 the list of variables along with their abbreviation and related descriptive statistics. The table shows that over 47 % of households in our sample live in the city, of which 17.8 % resides in the south, 53.6 % in the north, and the rest in the centre of Iran. More than 41 % of households reside in Western Iran, while less than 21 % live in Eastern Iran.

Table 1: Descriptive statistics.

Variable Name	Definition	Whole sample Mean	SD	Censored sample Mean	SD
Expenditure	Vegetables expenditure	204,417	148,645	212,866	146,583
Log Income	Natural logarithm of income for households in the last month	16.45	1.64	16.48	1.64
HH Size	Size of household	4.29	1.68	4.33	1.67
HH Age	Average age of household	30.22	12.21	30.06	11.86

Educated	The number of educated person in the household	3.17	1.59	3.21	1.58
Diploma	Number of diploma person in a household	0.11	0.35	0.11	0.36
BA	Number of person in a household with the education level of Bachelor of Arts (BA) or more	0.28	0.63	0.29	0.64
South	Dummy variable that take 1 if the house hold lives in the south of Iran	0.17	0.38	0.18	0.38
North	Dummy variable that take 1 if the house hold lives in the north of Iran	0.54	0.50	0.53	0.50
West	Dummy variable that take 1 if the house hold lives in the west of Iran	0.41	0.49	0.42	0.49
East	Dummy variable that take 1 if household lives in the east	0.22	0.41	0.21	0.41
Urban	Dummy variable that take 1 if the household live in the city	0.47	0.50	0.47	0.49

Note: SD denotes the standard deviation. HH stands for household.

We use a general to specific approach to arrive at the appropriate set of variables for each hurdle. Two discrete variables are used to capture the educational attainment of each household. The first is a count of educated individuals in each household. As Table 1 shows, on average about three quarters of individuals in each household are classified as educated. This is quite high given the large number of infants generally present in families. The second is a categorical variable related to the education level of the head of each household as a decision maker. In partitioning the households into three categories, we find that more than 10 % of the households head had a diploma (12 years of education but a college degree), less than 28 % had either a Bachelor of Arts or Science among other academic and religious education levels, while the remainder, about 61 % had training below the diploma level.

Table 2: Diagnostic tests.

Model	Homoscedasticity	P-values	Normality	p-values
Tobit	153.32(3)	0.00	142.31(2)	0.00
Heckit, 1 st	128.87(1)	0.00	89.4(2)	0.00
Heckit 2 nd	113.64(4)	0.00	145.35(2)	0.00
Double	134.6(1)	0.00	97.44(2)	0.00
Hurdle, 1 st				
Double	189.32(4)	0.00	125.50(2)	0.00
Hurdle, 2 nd				

Note: In each test statistic, the degree of freedom is written in the parentheses.

Finally, for the whole sample, the average age of a typical Iranian household was about 30 years old with a standard deviation of 12 and spending between 204,000 and 213,000 Rials monthly on vegetables as per the censored sample. This expenditure represents less than 11 % of their food budget of around 2.2 million Rials.

5 Empirical Results and Parameter Estimation

We used the Tobit, Heckit and the double hurdle models to arrive at the coefficient estimates of the determinants of vegetables consumption in Iran. Following Reynolds (1990), we used a general-to-specific modelling strategy in testing the explanatory power of a wide range of variables to arrive at a more parsimonious model capable of capturing the empirical evidence within our theoretical framework. As Arabmazar and Schmidt (1982) and Maddala and Nelson (1974) noted, the distributional assumption of the variables is crucial in testing the validity of limited dependent models. For example, testing for normality and homoscedasticity is paramount when it comes to choose a set of consistent parameters. To that end, we used the Pagan and Vella (1989) 's LM test. Table 2 shows that both non-normality and heteroskedasticity were present in the disturbance terms thereby justifying our choice for a heteroskedasticity double hurdle model with bivariate Box-Cox adjustment as per the likelihood function represented in eq. 8) (see Jones and Yen (1996, 2000) , Yen (1995) and Aristei and Pieroni (2008)).

Table 3 presents the results for the three different specification tests. Based on the LR test and the Vuong statistic, the Box-Cox double hurdle model with dependent error terms is the most suitable model to explain vegetable consumption in Iran. We find a rho stat of 0.34 for with a chi-square of 38.18 to support the assumption of a double hurdle decision making process. That is, households indeed first determine whether to consume or not, then make the decision as to how much if they must consume.

Table 3: Specification test.

Model	Test type	Statistics	P-values
Box-Cox dependent double hurdle vs Box-Cox independent double hurdle	LR	542,413.9	0.0000
Tobit vs Box-Cox independent double hurdle	LR	79,888.63	0.0000
Box-Cox Heckit vs Box-Cox independent double hurdle	Vuong	4,723,932	0.0000

Note: In each test, null hypothesis is that the nested model fits better against alternative.

The estimated parameters of a full double hurdle along with the Heckit and the Tobit models are presented in Table 4. The results show that all coefficient estimates but age, and dummy variables for North and East regions are statistically significant at the 5 % level, and for the most part at the 1 % level for the participation hurdle model. Based on the marginal effects estimated, the elasticities for vegetable expenditures with respect to each explanatory variable are computed at mean for variables that are continuous variables and at point for variables that are categorical.

Table 4: Regression results.

Variable name	Double Hurdle		Heckit		Tobit
	Participation	Expenditure	Participation	Expenditure	
Log Income	0.101***	0.048***	0.103***	0.047***	0.151***
HH Size	0.084***	0.061***	0.080***	0.061***	0.108***
HH Age	0.001	0.003***	0.001***	0.003	-0.002
Educated	0.143***	0.044***	0.139***	0.044***	0.175***
Diploma	0.116**	0.074***	0.116***	0.074***	0.141***
BA	0.154***	0.075***	0.156***	0.075***	0.149***
South	0.391***	0.226***	0.363***	0.227***	0.513***
North	0.031	0.039***	0.011***	0.040***	0.098***

West	0.291***	0.152***	0.299***	0.152***	0.437***
East	-0.062	-0.165***	-0.055***	-0.165	-0.222***
Urban	0.139***	0.026***	0.137***	0.026***	0.204***
Logarithm of Likelihood	-27,857.93		-27,859.03		-67,802.245

Note: ***, **, * denote significance at the 1 %, 5 % and 10 % levels respectively. Constant is included to all the regressions but not reported. HH stands for household. For variable definitions, please see Table 1. The dependent variable is vegetables consumption. The double hurdle model refers to Box-Cox Independent Double Hurdle model.

Original estimation of the models with income as an explanatory variable did not reveal any significant effect on vegetable expenditure. To test the veracity of this finding, we experimented with nonlinear specification such as the natural logarithmic transformed and squared income. For our data, we found a statistically significant impact of only log of income. That is, we find a positive effect of income on participation and expenditure decision for all models estimated; a one percentage point increase in income gives rise to 0.1 % increase in the probability of participation in vegetables consumption, 0.05% increase in conditional expenditure, and 0.15 % increase in unconditional vegetables expenditure.

In line with eqs (9), (10), and (11), we report in Table 5 the elasticity estimates of the preferred double hurdle model for participation, conditional and unconditional consumption with respect to each explanatory variable. The results show that the size of household strongly affects participation and expenditure decision when it comes to vegetables consumption in Iran. More populated households have a higher incidence of participation, hence, as expected, higher consumption and expenditures. For every one person extra in household size, the probability of vegetables consumption increases by 0.08 while expenditures rise by more than 0.06 Rial. Another important demographic factor is the average age of the household. The results are mixed; a one-percentage point increase in household's average age only leads to an increase of 0.01 % in vegetable expenditures. This low impact may be due to different food diet of younger households who are inclined to consume fast food due to time constraint as opposed to older households who adhere to traditional food diets with high content of vegetables.

Table 5: Elasticity estimates.

Variable name	Participation elasticity	Conditional elasticity of expenditure	Unconditional elasticity of expenditure
Log Income	0.101	0.048	0.148
HH Size	2.526	1.840	4.366
HH Age	0.003	0.008	0.010
Educated	0.451	0.141	0.593
Diploma	0.012	0.008	0.020
BA	0.043	0.021	0.064
South	0.069	0.041	0.111
North	0.017	0.021	0.038
West	0.119	0.063	0.182
East	-0.013	-0.035	-0.049
Urban	0.066	0.013	0.078

Note: For variable definitions, please see Table 1. HH stands for household.

We also find that households' education levels have a strong effect on both the participation and the consumption of vegetables consumption. The presence of one more educated person in a family increases the probability of participation in vegetables consumption by 0.14. As displayed in Table 4, the probability of participation in vegetables consumption increases with more years of education. This is a sensible finding since people with education tend to make better decisions for their health in general. The data show that this is also the case for Iran.

Being a relatively large country, food culture in Iran differs across geographical regions. We test the significance of such difference for vegetables consumption in both participation probability and expenditure decision. The results are quite enlightening; households from South of Iran spend 11 % more than those from the Centre (the reference group), while those from the North spend just 4 % more than that group. The probability of participation and the conditional expenditure in the South (0.0391 and 0.226 respectively) are more than those of the North (0.031 and 0.039), which in turn exceed those of the Centre. Although we could not find evidence of meaningful differences between participation probability and vegetables consumption in Central and Eastern regions, Eastern Iranian consume less vegetables than other groups while Western households are more likely

to participate, consume, and spend more money on vegetables. For example, Western Iranians spend about 0.06 % more than central Iranian households on vegetables.

6 Conclusions and Policy Implementations

Since vegetables are a natural source of many vitamins and micronutrients, increasing the amount of vegetables intake in food basket is clearly one solution for poor diet quality in Iran. The objective of this paper was to document the underlying determinants of vegetables consumption in Iran. Using cross-sectional data for approximately 29,000 rural and urban households in 2011–2012 based on Iran's income-expenditure survey of 2012, a full Box-Cox double hurdle model was selected to estimate and examine the effects of socio-economics factors. We postulated that both participation and consumption decisions were functions of economic, demographic and geographical characteristics of the households while accounting for heteroskedasticity and normality following Anderson and Reynolds (1990). Results of the study support the hypothesis that participation and consumption decisions are made simultaneously in the vegetables market. Also, demographic characteristics, economic factors, and geographical differences exert a statistically significant impact on both the probability of vegetables consumption and the level of consumption in Iran as in many other foods and beverage consumption. These results are in line with the Pilgrim's theory of food consumption determinants, in which, except of economic factors, three most important types of determinants are: (1) perception physical and chemical properties of the food, (2) factors related to the person engaged in food consumption, that is highly affected by demographical characters of households, and (3) environmental factors that is correlated with geographical factors (Pilgrim 1957). Estimated elasticities suggest that vegetables are normal goods (as is stated before by Deaton and Muellbauer 1980 among others), while education and household size are key factors in determining consumption patterns of vegetables in Iran. These results are robust with Pilgrim's theory of food consumption (Pilgrim 1957).

From an economic policy standpoint, the computed elasticities shed quite some lights on the effects of the latest food reforms in Iran, future food strategies, and food program implementations. For example, the elasticities of participation and consumption of vegetables with respect to income suggest that granting pecuniary aids induces new and existing consumers to increase both their participation and consumption of vegetables (As proposed by Dawoud 2005). A corollary of this finding is that emphasis must be placed on increasing participation since there is a constraint on daily nutrition as to the absorption capacity of the human body. A salient feature of our findings is that education plays a crucial role in both participation and consumption of vegetables. Households with higher education levels participate more and consume larger amount of vegetables than other groups. This is not surprising and approves Pilgrim's theory of food consumption since individuals with higher education tend to have more knowledge about nutrition facts and as such their revealed preferences towards vegetables are quite well noted. Also, we cannot negate the fact that advertisement and other programs do contribute to people's knowledge and benefits of a balanced diet with vegetables content. These social benefits may compensate for the price increments of vegetables that arise as a result of increased Iranian participation and consumption of vegetables. Our results also shed light on the cross-region differences when it comes to participation and vegetables consumption per capita and confirm the results of Anderson and Reynolds (1990). Households from the central region spend less than those from the North, whereas participation and expenditures are higher in the South. Overall, this paper has informed policy makers that more attention needs to be placed on households from central Iran when it comes to guaranteeing food security in the vegetable group.

One limitation of our research is to control the effect of house garden area and cost of own-produced vegetables on the both hurdle, especially on the second one. Also, one can examine the effect of unemployed ratio of household members on the participation decision.

Notes

¹See promoting fruit and vegetable consumption around the world, WHO, Global strategy Diet Information sheet.

²Ghanbari and Sadeghi (2010), Developments of the Iranian Economy, SAMT publication.

³Tobit (1958) model was the first attempt to deal with zero consumption problems.

⁴See Madden (2008) for a discussion on the relative merits of this model.

⁵See Humphreys (2013) for details.

References

- Amemiya, T. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 84: 3–61.
- Arabmazar, A., and P. Schmidt. 1982. "An Investigation of the Robustness of the Tobit Estimator to Non-Normality." *Econometrica* 50: 1055–1063.
- Aristei, D., and L. Pieroni. 2008. "A Double Hurdle Approach to Modelling Tobacco Consumption in Italy." *Applied Economics* 40 (19): 2463–2476.
- Bleylock, J., and W.N. Blizard. 1992. "Cigarette Consumption: The Case of Low Income Women." *American Journal of Agricultural Economics* 74: 698–705.
- Cragg, J. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for the Durable Goods." *Econometrica* 23: 59–80.
- Dawoud, S. 2005. M.A Thesis, Macau University, Denmark. Analysis of Food Consumption Patterns in Egypt.
- Deaton, A., and J. Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.
- Dettman, R.L., and C. Dimitri. (2007). Organic Consumers: A Demographic Portrayal of Organic Vegetables Consumption within the United States. the 105th EAAE Seminar of international marketing and international trade of quality food products. Bologna, Italy.
- Garcia, J., and J.M. Labeaga. 1996. "Alternative Approaches to Modelling Zero Expenditures: An Application to Spanish Demand for Tobacco." *Oxford Bulletin of Econometrics and Statistics* 58: 589–590.
- Ghanbari, A., and H. Sadeghi. 2010. *Developments of the Iranian Economy*. Iran: SAMT publication.
- Heckman, J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153–161.
- Huang, K.S. 1996. "Nutrient Elasticities in a Complete Food Demand System." *American Journal of Agricultural Economics* 72: 145–168.
- Humphreys, B.R. 2013. "working paper." Dealing with Zeros in Economic Data www.ualberta.ca/~bhumphre/class/zeros_v1.pdf.
- Jabarin, A.S., and E.K. Al-Karablieh. 2011. "Estimating the Fresh Vegetables Demand System in Jordan: A Linear Approximate Almost Ideal Demand System." *Journal of Agricultural Science and Technology* 5 (3): 322–331.
- Jones, A.M. 1989. "The UK Demand for Cigarettes 1954–86, A Double-Hurdle Approach." *Journal of Health Economics* 8: 133–141.
- Jones, A.M. 1992. "A Note on Computation of the Double Hurdle Model with Dependence with an Application of Tobacco Expenditures." *Bulletin of Economic Research* 44: 67–74.
- Jones, A.M., and S.T. Yen. 1996. "Individual Cigarette Consumption and Addiction: A Flexible Limited Dependent Variable Approach." *Health Economics* 5: 105–117.
- Jones, A.M., and S.T. Yen. 2000. "A Box-Cox Double Hurdle Model." *Manchester School* 68: 203–221.
- Keelan, C., M. Henschion, and C. Newman. 2008. "A Double-Hurdle Model of Irish Households' Foodservice Expenditure Patterns." *Journal of International Food & Agribusiness Marketing* 21 (4): 269–285.
- Kiani, K.H., and M.A. Kouhbor. 2008. "Estimating Nutrients Elasticities: An Application of the Tobit Model." *Journal of Quantitative Economics* 2: 42–60.
- Labeaga, J.M. 1999. "A Double Hurdle Rational Addiction Model with Heterogeneity: Estimating the Demand for Tobacco." *Journal of Econometrics* 93: 49–72.
- Maddala, G.S., and F.D. Nelson. 1974. "Maximum Likelihood Method for Models of Markets in Disequilibrium." *Econometrica* 42 (6): 1013–1030.
- Madden, D. 2008. "Sample Selection versus Two-Part Models Revisited: The Case of Female Smoking and Drinking." *Journal of Health Economics* 27 (2): 300–307.
- Newman, C., M. Henschion, and A. Mathews. 2003. "A Double Hurdle Model of Irish Household Expenditure on Prepared Meals." *Applied Economics* 35 (9): 1053–1061.
- Pagan, V., and P. Vella. 1989. "Diagnostic Tests for Models Based on Individual Data: A Survey." *Journal of Applied Econometrics* 4: 529–559.
- Pilgrim, F.J. 1957. "The Components of Food Acceptance and Their Measurement." *American Journal of Clinical Nutrition* 5: 171–175.
- Reynolds, A. 1990. "Analysing Fresh Vegetables Consumption from Household Survey Data." *Southern Journal of Agricultural Economics* 22: 31–38.
- Smith, M.D. 2003. "On Dependency in Double Hurdle Models." *Statistical Papers* 44: 581–595.
- Sue, S.J., and S.T. Yen. 1996. "Microeconomic Models of Infrequently Purchased Goods: An Application to Household Pork Consumption." *Empirical Economics* 21: 513–533.
- Tiffin, R., and M. Arnoult. (2008). Bayesian Estimation of the Infrequency of Purchase Model with an Application to Food Demand in the UK Working paper.
- Tobin, J. 1958. "Estimation of Relationship for Limited Dependent Variables." *Econometrica* 26: 24–36.
- Vuong, Q.H. 1989. "Likelihood Ratio Test for Model Selection and Non-Nested Hypothesis." *Econometrica* 57: 303–333.
- World Health Organization. 2015. *Promoting Fruit and Vegetable Consumption Around the World, Global Strategy on Diet, Physical Activity and Health Information Sheet*.

- Yen, S.T. 1995. "Alternative Transformation in a Class of Limited Dependent Variable Models: Alcohol Consumption by US Women." *Applied Economics Letters* 2: 258–262.
- Yen, S.T., and C.L. Huang. 1996. "Household Demand for Finfish: A Generalized Double Hurdle Model." *Journal of Agricultural and Resource Economics* 21 (2): 220–234.
- Yen, S.T., and H.H. Jensen. 1996. "Determinants of Household Expenditures on Alcohol." *The Journal of Consumers Affairs* 30: 48–67.