MODELING FRESH ORGANIC PRODUCE CONSUMPTION: A GENERALIZED DOUBLE-HURDLE MODEL APPROACH

Feng Zhang, Chung L. Huang, and Biing-Hwan Lin

Authors' Affiliations:

Feng Zhang is a Ph.D. student and graduate research assistant, Department of Agricultural and Applied Economics, the University of Georgia. Chung L. Huang is a professor, Department of Agricultural and Applied Economics, the University of Georgia. Biing-Hwan Lin is an agricultural economist, Economic Research Service, U.S. Department of Agriculture.

Contact information:

Feng Zhang
207 Conner Hall,
University of Georgia,
Athens, GA 30602-7509
Email: fzhang@agecon.uga.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Orlando, FL, Feb 5-8, 2006

Abstract

Using actual retail data, this study is intended to provide an objective view of the consumers' social economic characteristics which contribute to the growth of the fresh organic produce market with a generalized double-hurdle model. The nested test shows that the generalized double-hurdle model performs significantly better than the Cragg's independent double-hurdle model and the commonly used Tobit model. The estimated results indicate that marketing strategies targeted at higher income, and higher educated consumers can be effective in both attracting new consumers and eliciting more purchases from the current consumers. Household size is not likely to be a factor affecting fresh organic produce consumption. Older people is found to be more likely to consume organic produce, but the age groups of the current consumers may not be a distinguishing factor for further promotions to aim at to elicit more purchases.

Keywords: double-hurdle model, scanner data, organic, fresh produce, demand

Modeling Fresh Organic Produce Consumption: a Generalized Double-hurdle Model Approach

Introduction

Concerns over health and environmental degradation have motivated US consumers to consume more organic produce over the recent years. Sales of organic commodities in natural foods stores approached $3.3 billion in 1998, compared with $2.08 billion in 1995, according to industry sources. Among various organic foods, fresh fruits and vegetables have much higher market penetration rates than others. For example, in 2002, organic fresh fruit and vegetable sales accounted for 4.5 percent of total fresh fruit and vegetable sales (NBJ, 2003). *Natural Foods Merchandiser* reported that sales of packaged fresh produce had the highest growth rate among sales of all organic products during 2002-2003, expanding 26 percent to $364 million.

Despite the projected high growth in consumption of fresh organic produce, consumer characteristics contributing its growth are not well understood. Most previous studies of organic produce have measured attitudes regarding the purchase of organic produce rather than actual purchase choices or behaviors. As an indication of such attitudes, these studies typically elicit willingness to pay for organic produce and the likelihood of consumption of organic produce relative to its conventional counterpart. And results from previous studies using surveys are often fragmentary and sometimes inconsistent. Thompson (1998) summarized studies prior to 1997 on the impact of demographic characteristics on the likelihood of consumption of the organic foods. His study revealed some contradictory findings about the effect of income, age, and educational attainment on likelihood of consuming organic foods. More recent survey studies also had different conclusions on the impact of income on consumption of organic food. A survey conducted by Hartman Group in 2002 showed that over half of those who frequently buy organic foods in the United States have incomes below $30,000, and African-Americans, Asian-Americans, and Hispanics use more organic products than Caucasians. The results of Hartman Group survey are interesting, given that a USDA ERS study that found that low-income households eat less fresh fruits and vegetables than higher-income households (Blisard et al., 2004). Thus, additional research on who buys organic foods is needed (Oberholtzer et al., 2005).

Objectives

Our objective in this study is to identify important consumer characteristics that are associated with fresh organic produce consumption and investigate their effects on consumption. To achieve this purpose, we utilized a generalized double-hurdle model which allows for different parameterizations of the participation and consumption processes, and the possible correlation between those two processes. The statistical performance of this model and its results will be compared with those from Cragg's independent double-hurdle model and Tobit model.

Econometric Model

For most of cross-sectional consumption data, zero consumption is one problem for any modeling effort to address. Tobit model developed by Tobin (1958) has been widely used to deal with censored observations. It attributes the censoring to a standard corner solution. However, this model is very restrictive. For one thing, Tobit model has been shown to be inadequate to characterize the two processes in consumption: the participation process and consumption process. Any variable which increases the probability of non-zero consumption must also increase the mean of the positive consumption, which is not always reasonable (Lin and Schmidt, 1984). As an example, consider a hypothetical sample of buildings, and suppose that we wish to analyze the dependent variable, “loss due to fire”, during some time period. Since this is often zero but otherwise positive, the Tobit model might be an obvious choice. However, it is not hard to imagine that newer (and more valuable) buildings might be less likely to have fires, but might have greater average losses when a fire did occur. The Tobit model can not accommodate this possibility.

The double-hurdle model, originally proposed by Cragg (1971), assumes that households make two decisions with respect to purchasing an item, each of which is determined by a different set of explanatory variables. In order to observe a positive level of expenditure, two separate hurdles must be passed. First, based on impediments to acquisition, the household decides whether or not to purchase the good, and second, according to the intensity of the desire for the good, the household decides on how much to purchase. A different latent variable is used to model each decision process, with a Probit part determining participation and a Tobit part determining the expenditure level (Blundell and Meghir, 1987). The double-hurdle model has been used widely since its introduction. Newman, et al. (2003) applied double-hurdle model to study Irish households’ expenditure on prepared meals for home consumption. Yen and Johns (1997) used the procedure for analysis of U.S. household consumption of cheese. Other studies have also applied the double-hurdle model to examine U.S. food expenditure away from home (Jensen and Yen, 1996) and household demand for finfish (Yen and Huang, 1996).

Most applications rejected Tobit model in favor of Cragg's independent double-hurdle model.

Though Cragg's independent double-hurdle model is an improvement of Tobit model, it is still limited in that it assumes that the shocks to the participation process and consumption process are independent. For consumers' demand for a particular commodity, this seems to be unrealistic. Drawing on the thought of correlated processes from the sample selection model of Heckman (1979), the generalized double-hurdle model extended Cragg's independent double-hurdle model to deal with correlated residuals from the participation process and the consumption process. Jones (1989 and 1992) first used this generalized double-hurdle in analyzing tobacco consumption in UK. Yen (2005) applied this approach to study the cigarette consumption in the United States and computed the elasticities using the MLE estimators. The nice feature of this model is that the common Tobit model and the Cragg's independent double-hurdle can be incorporated as special cases and tested against the generalized double-hurdle model. The specification of the generalized double-hurdle model is as follows:

(1) $$y = \begin{cases} x'\beta + v & \text{if } x'\beta + v > 0 \text{ and } z'\alpha + u > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = N\left\{0, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right\} \quad \text{where } v > -x'\beta,$$

where y is the expenditure; x and z are variables determining the participation process and the consumption process respectively; u and v are residual terms from those two processes, with a correlation coefficient ρ; α, β, ρ, and σ are parameters for estimation

Then the likelihood function can be written as

$$L = \prod_{y=0} \{1 - \psi(z'\alpha, x'\beta/\sigma; \rho)/\Phi(x'\beta/\sigma)\}$$

(2)
$$\times \prod_{y>0} \{\frac{1}{\sigma}\phi[(y - x'\beta/\sigma]\Phi[\frac{z'\alpha + \rho(y - x'\beta)/\sigma}{(1-\rho^2)^{1/2}}],$$

$$\times [\Phi(x'\beta/\sigma)]^{-1}\},$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are univariate standard normal CDF and PDF respectively; $\psi(\cdot)$ is the bivariate standard normal CDF with three arguments, bivariate means and the error term correlation. When $\rho = 0$, the above model reduces to Cragg's independent hurdle model. When $\rho = 0$, $x = z$, and $\alpha = \beta/\sigma$, it leads to the Tobit model. In this analysis, we used one set of explanatory variables for both processes, $x = z$, so that we can test our generalized hurdle model against two special cases: Cragg's independent hurdle model and Tobit model.

Data and Variables

AC Neilson Homescan data of 2003 is the data source of this study. AC Nielsen Homescan data is unique in that each panelist was supplied with a scanner device that he/she used at home to record grocery items purchased at any grocery store, or other type of store throughout a given time period. Each panelist represents a unique household, with each household having eighteen known demographic characteristics. By investigating the relationship between consumption of fresh organic produce and consumer characteristics, we can identify those potential consumers of fresh organic produce.

In 2003, there are 8,833 households included in the AC Neilson consumer panel. The date, expenditure, and quantity of each purchase are recorded with the supplied scanner. To avoid the data problem of inadvertent recording by some households, we include only those households who made purchases of fresh produce for at least 10 months in 2003, which reduces our sample to 7,052 households.

The organic expenditure is specified as the following equation:

$$(3) \quad \begin{aligned} ORGCOST = f\,(&HHSIZE, INCOME, AGE2, AGE3, EDUC2, EDUC3, CHILD6, \\ &EAST, CENTRAL, WEST, URBAN, BLACK, HISPANIC, ORIENTAL). \end{aligned}$$

This functional form is used in both participation and consumption processes expressed in the equation 1. The response variable of our model, the expenditure of fresh organic produce ($ORGCOST$), is modeled as a function of various consumers' social economic variables, which are listed and described in Table 1. The problem with the fresh organic produce expenditure is that its distribution is highly skewed. If used directly as response variable, it may cause inconsistency and nonnormality of error terms (Newman et al., 2003). In this study, we used natural logarithm of positive fresh organic produce consumptions since the transformed variable is more likely to be normally distributed. Figure 1 shows the histograms of both original and transformed positive expenditures. In addition, natural logarithmic transformation of the response variable is more amenable in computing elasticities of organic consumption with respect to demographic variables. For example, for dummy variables, the estimated parameters ($\partial \ln y / \partial x$) are elasticities per se. For continuous variables like income, the elascititities are calculated as the estimated parameters ($\partial \ln y / \partial x$) times mean level of explanatory variable, $\bar{x}$.

Empirical Results

Estimation results are presented in Table 2. Since the generalized double-hurdle model nests Cragg's independent model that in turn nests Tobit model as a special case, the standard log likelihood ratio test between the restricted and unrestricted models applies in this case. The log likelihood values of the generalized double-hurdle model, Cragg's independent double-hurdle model, and, Tobit model are -6761, -9589, and -12192 respectively. All the likelihood ratio tests show that the generalized double-hurdle model is the best one among the three models. The P-values of the likelihood ratio tests among three models are highly significant in favor of the generalized double-hurdle model used in this study.

The elasticities of consumption probability and level (both conditional and unconditional for the latter) are computed by referencing to Yen's (2005) formula. The probabilities of consumption (i.e., a positive observation) is,

(4) $\Pr(y > 0) = \psi(z'\alpha, x'\beta/\sigma; \rho)/\Phi(x'\beta/\sigma)$,

which depends on both participation and consumption process parameters. The conditional and unconditional means of the dependent variable are listed as follows:

$$\begin{aligned} E(y \mid y > 0) &= x'\beta + E(v \mid u > -z'\alpha, v > -x'\beta) \\ &= x'\beta + [\psi(z'\alpha, x'\beta/\sigma; \rho)]^{-1}\sigma \times \{\phi(x'\beta/\sigma)\Phi[(z'\alpha - \rho x'\beta/\sigma)/(1-\rho^2)^{1/2}] \\ &\quad + \rho\phi(z'\alpha)\Phi[(x'\beta/\sigma - \rho z'\alpha)/(1-\rho^2)^{1/2}]\}, \end{aligned} \tag{5}$$

(6) $E(y) = \Pr(y > 0) \times E(y \mid y > 0)$.

The elasticities of the probability, conditional level, and unconditional level are calculated at the sample mean of continuous variables (household size and household income) for the baseline group -- young white people with no more than high school education, without child under 6 years old, dwelling in the rural area in the south. For dummy variables, the elasticities are computed as the percentage change in probability or level of consumption with respect to discrete change in the status of the dummy variable concerned. Since our dependent variable is in natural logarithmic form, the conditional and unconditional elasticities of the consumption level are computed as $(\partial E(y \mid y > 0) / \partial x) * \bar{x}$ and $(\partial E(y) / \partial x) * \bar{x}$ respectively for continuous variables, $(\Delta E(y \mid y > 0) / \Delta x)$ and $(\Delta E(y) / \Delta x)$ respectively for dummy variables. Since the impacts of all significant variables on the market participation and consumption decisions are in the same direction in this study, the unconditional elasticities are higher in magnitude than the conditional elasticities.

As indicated in the maximum likelihood estimates in Table 2, household size is not a significant determinant in household decision on whether to buy or how much to buy fresh organic produce. In contrast, the economic factor, the household income, has a positive and significant effect on household expenditure decisions. The unconditional income elasticity of the consumption is about 27%, and income elasticity for the probability of entering into the organic fresh produce market is 16%.

The effect of age of the head of household on household expenditure on fresh organic produce is mixed in the market participation and consumption decisions. As shown in the results, among the three age groups, the older age group is the only one found to be significantly more likely to buy fresh organic produce. Of households that do participate in the market, there is no significant difference among these three age groups in the level of consumption. Educational levels are highly significant in explaining both the market participation and consumption of the fresh organic produce. The result implies that the higher educational level the household head is, the more likely the household will buy fresh organic produce. Of the households that are already in the market, higher educational level of the household head is also associated with higher level of consumption. Among all dummy variables, the post-college degree (EDUC3) is associated with highest elasticity for consumption probability and consumption level.

The binary effects also show that, in 2003, *ceteris paribus*, households living in urban areas spend about 30% more on fresh organic produce than those living in rural areas, are 5% more likely to participate in this market than rural households, and of all households that purchase the fresh organic produce, spend 5% more than the rural households.

Results for the geographic dummy variables indicate that the area associated with highest probability and level of fresh organic produce consumption is the west, followed by the east area, the south, and the central area at the last position. This result echoes the facts that the west area in the U.S. has the highest organic produce production and that the east area has the highest percentage of certified organic acreage. California is the biggest organic vegetable producer in 2001, accounting for 41 percent of U.S. certified organic vegetable acreage, while the certified organic acreage accounted for over 10 percent of the vegetable acreage in Vermont, New Hampshire, Maine, and Colorado in 2001 (Oberholtzer et al., 2005). Therefore, people in those areas have a broader access to or are more aware of the fresh organic produce than people in other areas. Among people with difference races, Hispanics, as a group, are more likely to consume and consume significantly higher level of fresh organic produce than any other group on average. This may reflect the increasing role of the Hispanics in conventional and organic produce farming in the United States.

Conclusions

Previous studies of consumer surveys based on contingent valuations gave inconsistent or even contradictory results on the impact of some consumer characteristics on organic foods consumption. Using the actual retail data, this study is intended to provide a more objective view of the consumers' characteristics which contribute to the growth of the fresh organic produce market.

By modeling the market participation and consumption levels at the same time with a maximum likelihood function, our generalized double-hurdle model utilizes more information from fresh organic produce consumption behaviors than a single Probit model on consumption probability or a Tobit model on consumption levels. The nested test shows that the generalized double-hurdle model is significantly better than the Cragg's independent hurdle model and the commonly used Tobit model.

The estimated results indicate that marketing strategies targeted at higher income, and higher educated consumers can be effective in both attracting new consumers and eliciting more sales from the current consumers. Household size is not likely to be a factor affecting fresh organic produce consumption. Even though older people may be more likely to consume organic produce, the age of the current consumers may not be a distinguishing factor for further promotions to aim at to increase sales.

References

Blisard, N., H. Stewart, and D. Jolliffe. "Low Income Households' Expenditures on Fruits and Vegetables." USDA, ERS, AER 833, May 2004, available at http://www.ers.usda.gov/publications/aer833/

Blundell, R. and Meghir, C., "Bivariate Alternatives to the Univariate Tobit Model." *Journal of Econometrics* (34) 1987: 179-200.

Cragg, J. G., "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica 39* (1971): 829-844.

Hartman Group. Hartman Organic Research Review: A Compilation of National Organic Research Conducted by the Hartman Group. Bellevue, WA. 2002.

Heckman, J., "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and Social Measurements 5(1976):475-495.

Jensen, H. and S. Yen. "Food Expenditures away from Home by Type of Meal." *Canadian Journal of Agricultural Economics* 44(1996):67-80.

Jones, A. M. "A Double-Hurdle Model of Cigarette Consumption." *Journal of Applied Econometrics* 4(1989):23-29.

Jones, A. M. "A Note on Consumption of the Double-Hurdle Model with Dependence with an Application to Tobacco Expenditure." *Bulletin of Economic Research* 44(1992):67-74.

Lin, T. and P. Schmidt. "A Test of the Tobit Specification against an Alternative Suggested by Cragg." *The Review of Economics and Statistics* 66(1984):174-177.

Newman, C., M. Henchion, and A. Matthews, "A Double-Hurdle Model of Irish Household Expenditure on Prepared Meals." *Applied Economics* 35(2003):1051-1061.

Nutrition Business Journal (NBJ). The NBJ/SPINS Organic Foods Report *2003*, Penton Media, Inc.

Oberholtzer L., C. Dimitri, and C. Greene, "Price Premiums Hold on as U.S. Organic Produce market Expands.", U.S. Department of Agriculture, Economic Research Service (USDA, ERS). Outlook Report No. (VGS30801) 22 pp, May 2005 http://www.ers.usda.gov/Publications/vgs/may05/VGS30801/

Tobin, J. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26(1958):24-36.

Thompson, G. D. "Consumption Demand for Organic Foods: What We Know and What We Need to Know." *American Journal of Agricultural Economics* 80(1998):1113-1118.

Yen, S. and C., Huang. "Household Demand for Finfish: a Generalized Double-Hurdle Approach." *Journal of Agricultural and Resource Economics* 21(1996):202-234.

Yen, S. and A. Jones. "Household Consumption of Cheese: an Inverse Hyperbolic Sine Double-hurdle Model with Dependent Errors." *American Journal of Agricultural Economics* 79(1997):246-251.

Yen, S. "Zero Observations and Gender Differences in Cigarette Consumption." *Applied Economics* 37(2005):1-11.

Table 1. Definition of Variables and Sample Statistics

Variables	Definition	Mean (SE)
Orgcost	Per household expenditure on organic fresh produce in 2003 (in cents)	594 (2758)
Hhsize	Household size - number of people in a household	2.45 (1.33)
Income	Income ($1000), calculated as the median of the selected income interval	52.34 (27.34)
Dummy variables (Yes = 1, no =0)		
Dumorg	Households buying organic fresh produce	0.42
Age1	The higher age of the male and female household heads is less than 40	0.13
Age2	The higher age of the male and female household heads is between 40 and 64	0.62
Age3	The higher age of the male and female household heads is 65 and above	0.25
Educ1	The higher education of the male and female household heads is high school	0.19
Educ2	The higher education of the male and female household heads is college	0.64
Educ3	The higher education of the male and female household heads is post college	0.16
Child6	Households with children under 6 years old	0.08
East	Residents in east region	0.21
Central	Residents in central region	0.19
South	Residents in south region	0.39
West	Residents in west region	0.21
Urban	Residents in urban areas	0.87
Rural	Residents in rural areas	0.13
White	White households	0.76
Black	Black households	0.12
Hispanic	Hispanic households	0.08
Oriental	Oriental households	0.02
Sample size		7,052

Source: Compiled from AC Neilson Homescan data 2003.

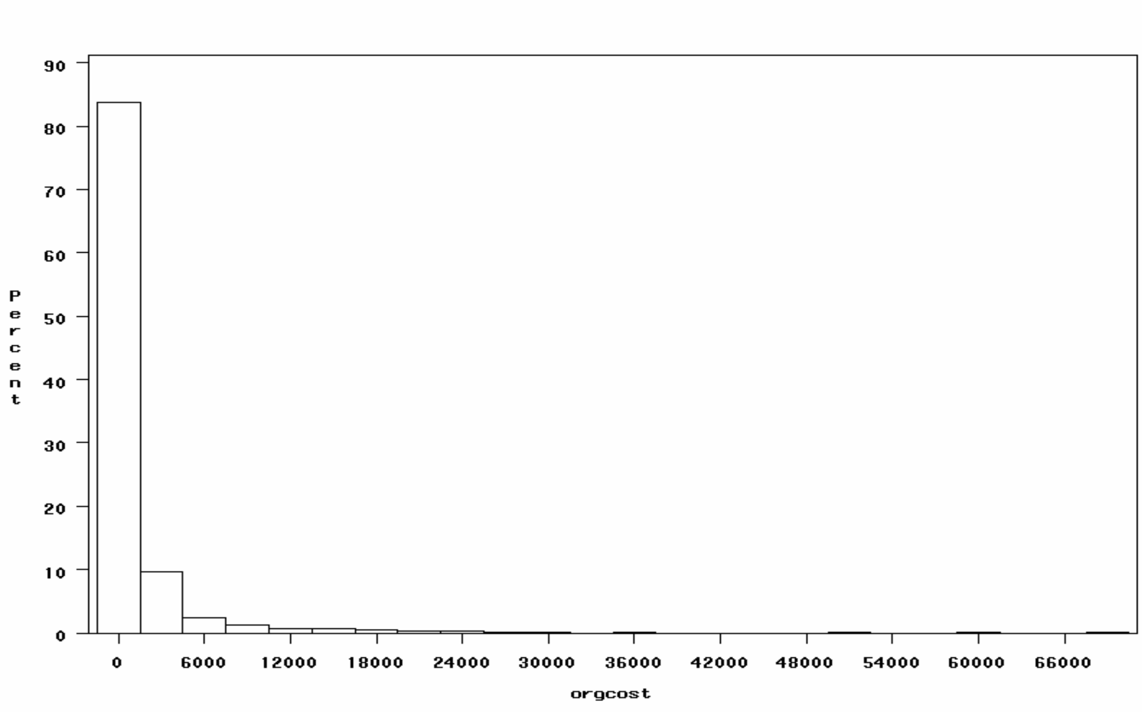
Table 2. Maximum Likelihood Estimates of the Generalized Double-hurdle Model

	Participation Process		Consumption Process		
	Parameters (S.E.)	Elasticities	Parameters (S.E.)	Elasticities (Conditional)	Unconditional Elasticities
Constant	-0.8227** (0.0816)	----	3.0681** (0.1938)	----	----
Hhsize	-0.0053 (0.0134)	----	-0.0083 (0.0288)	----	----
Income	0.0024** (0.0006)	0.1632	0.0065** (0.0014)	0.1486	0.2694
Age2	0.0554 (0.0490)	----	-0.1012 (0.1048)	----	----
Age3	0.1817** (0.0562)	0.0598	0.0468 (0.1200)	----	0.2908
Educ2	0.1689** (0.0422)	0.0554	0.3274** (0.0928)	0.0725	0.3535
Educ3	0.3440** (0.0556)	0.1183	0.7838** (0.1198)	0.2743	0.8076
Child6	0.0049 (0.0643)	----	-0.0004 (0.1367)	----	----
East	0.1808** (0.0405)	0.0595	0.4356** (0.0863)	0.1630	0.4055
Central	-0.0778* (0.0446)	-0.0234	-0.1056 (0.0975)	----	-0.1375
West	0.1754** (0.0412)	0.0576	0.4679** (0.0875)	0.2034	0.4058
Urban	0.1452** (0.0477)	0.0472	0.2661** (0.1051)	0.0464	0.2967
Black	-0.0009 (0.0476)	----	0.0477 (0.1027)	----	----
Histpanic	0.2280** (0.0562)	0.0761	0.3629** (0.1175)	0.0208	0.4628
Oriental	0.1188 (0.0978)	----	0.2382 (0.2024)	----	----
σ			2.1045** (0.0412)		
ρ			0.9550** (0.0050)		
Log Likelihood	-6761				

Note: 1. Double asterisks and single asterisk demote significance at 5% and 10% respectively.
2. Elasticities for the dummy variables are interpreted as the percentage change in organic consumption in response to 0/1 change in dummy variables

Figure 1. Distribution of fresh organic produce expenditure at the original scale and natural logarithm transformed scale (for positive consumptions):

Expenditure in original scale (in cents):



Expenditure in natural logarithm transformed scale