Zero inflated ordered probit approach to modeling mushroom consumption in the US

Yuan Jiang (jyspring@ufl.edu), Graduate Student
Hyeyoung Kim(kim1978@ufl.edu), Assistant Professor

Lisa House (lahouse@ufl.edu), Professor

Food and Resource Economics Department, University of Florida

Susan S. Percival (percival@ufl.edu), Professor

Department of Food Science and Human Nutrition, University of Florida

Selected Poster prepared for presentation at the

2016 Agricultural & Applied Economics Association's Annual Meeting, Boston, MA, July 31-August 2

Zero inflated ordered probit approach to modeling mushroom consumption in the US

Introduction

The United States is the world's second-largest producer of mushrooms, with 16 percent of world output, following China. In terms of value of production, mushrooms are a leading U.S specialty crop, exceeded only by potatoes, tomatoes and lettuce. Consumption of mushrooms has been on the rise in the United States over the past several decades. Typically used as a vegetable, per capita consumption has quadrupled since 1965 (the first year for which reliable data are available). According to data compiled by the U.S. Department of Agriculture's (USDA) Economics Research Service (ERS), per capita use of all mushrooms totaled about 4.0 pounds in 2011, compared with about 0.69 pounds in 1965 (Figure 1).

The mushroom market can be divided into two main categories: fresh and processed. Fresh mushrooms accounted for three-fourths of domestic consumption in 2011. Since 1990, per capita consumption of fresh mushrooms has increased dramatically, while per capita consumption of processed mushrooms (mostly canned mushrooms) has gradually declined over the same time period.

There has been a lot of research investigating both the nutritional and medical benefits of mushrooms (Chang and Buswell, 1996; Sullivan and Smith, 2006), however, relatively little detailed information has been published concerning consumer behavior in the mushroom market. Aside from the basic USDA disappearance data and retail sales information, little research explores factors influencing mushroom consumption at the individual level. One exception is a study by Lucier et al. (2003) which demonstrated that Asian and non-Hispanic white consumers were the strongest consumers of mushrooms and per capita mushroom consumption was positively correlated with income. The study also found that men and women between 20 and 39 years old were the leading mushroom consumers, representing about 32% of the population, yet consuming 43% of all mushrooms (Lucier et al.2003).

To help fill this gap and investigate this growing market, the purpose of this research is to investigate and compare the determinants of the fresh and processed mushroom consumption. By employing the individual-level survey data to record consumer's consumption behavior for mushrooms, we use a model that allows us to examine the decision to consume and the frequency of consumption for consumers. In this format, there are many "non-consumption" observations in the sample, which causes the problem of "excess zeros". "Non-consumption" can mean two different things. It can represent those that "never" consume (non-consumers) mushrooms, as well as those that indicate that they did not consume in the specific time period indicated (we will identify these as "potential" consumers). Therefore, the two sources of zero observations may be from two distinct sources (Harris, Zhao; 2007).

In the case of mushroom consumption, we expect people choose to never consume mushroom due to some stable reasons, like tastes and preferences. However, given the lack of frequency of consumption, there are also many zero observations for consumers who did not consume in a specific time period, which create a corner solution of the standard consumer demand problem. We expect reasons for this to be more related to temporary reasons, such as price and income level. Thus, the two types of zero observations might be driven by different consumer behaviors. The potential consumers who choose "not consumed recently" might share similar characteristics with the current consumers, which will be influenced by the standard consumer demand factors such as price and income, while the non-consumers who choose never consumed mushroom before were driven by a separate process relating to their sociological and healthy factors. As a result, the zero-inflated ordered probit model is employed to analyze the mushroom data in this paper, by taking into account the two different types of zero observations, and explore the determinants of the fresh and processed mushroom consumption in the United States

Literature Review

There has been much research exploring the issue of "excess zeroes", and many approaches to account for this have been proposed in the past few decades. When the dependent variable is count data, the most popular method to account for the zeroes is to use either a zero-inflated count or hurdle count data model. Lambert proposed the zero-inflated Poisson (ZIP) model in 1995, and applied this model to the case of manufacturing defects. Following this, a number of related models have been used, including the Poisson-negative binomial and modified Poisson suggested to address inequality of the mean and variance (as equality is assumed for the Poisson distribution) (Consul and Jain, 1973; Famoye and Singh, 2006). In the presence of over-dispersion in the data, the negative binomial model (when Poisson mean has a gamma distribution) was preferred to Poisson. Mwalili, Lesaffre and Declerck (2008) proposed the zero-inflated negative binomial (ZINB) model as a natural extension of ZIP model with the presence of over-dispersion. The zero inflated count data model assumes that the zero count are from two distinct sources: "sampling zero" and "structured zero". The sampling zeros are due to the usual Poisson (Negative Binomial) distribution, which assumes that those zero observations happened by chance. The structured zero are observed due to the specific structure in the data. When applying the zero inflated model to the consumption case, the observed zero consumption will be recorded when the consumer is genuine non-participant (structure zero), or when the zero consumption is the corner solution of a standard consumer demand problem (sampling zero). According to this assumption, it is assumed that with probability p, the only possible observation is 0 (genuine non-participants), and with probability (1-p), a Poisson/Negative-binomial random variable is observed.

Another popular approach to model the excess zeros in count data is the hurdle count model. The hurdle model was first developed by Cragg (1971) as an example of truncated models, relaxing the Tobit model by allowing separate stochastic processes for the participation and consumption decisions (Yen and Huang, 1996). Mullahy (1986) modified the count data model by allowing excess zeroes, and specified the hurdle count data model. In Mullahy's model (Single hurdle count-data model[1]), zero

[1] The term borrowed from Shonkwiler and Shaw (1996)

observation was only allowed in the first stage (decide on whether to consume), and in the second stage, the consumption behavior was truncated at zero. Thus, in the single hurdle count-data model, it assumes that all the zeros are from a "structural" source. Shonkwiler and Shaw (1996) extended Mullahy's specification by allowing zero observations in both the first and second stage. In this research, Shonkwiler and Shaw applied the hurdle count data model to analyze recreation demand, and they classified people into three categories: "user", "potential user", and "non-user". Thus, in Shonkwiler and Shaw's model (Double hurdle count-data model[2]), there are two mechanisms generating zero observations: zero observations could either happen in the first stage by choosing not consume (non-consumer), or in the second stage by choosing consume zero frequency (potential-consumer).

When the outcome is an ordered discrete variable, Harris and Zhao (2007) proposed the zero-inflated ordered probit model (ZIOP) as a manner analogous to the zero-inflated count data model and double-hurdle count data model. The ZIOP model is an extension to the ordered probit model by allowing the excess zero observations to come from two different aspects of individual behaviors (Harris and Zhao, 2007). Similar to the zero-inflated count data model, the ZIOP model also allows zero observations coming from two different sources. But different from ZIP/ZINB, or double-hurdle count data model which have the passion or negative binomial regression framework, the ZIOP model consisted of a probit "splitting" model, which generated the non-participants and participants, and an ordered probit model, which generated zero-consumption participants and positive-consumption participants. Furthermore, Harris and Zhao also defined the ZIOPC model which allows the error terms of the binary probit model and ordered probit model being correlated.

When analyzing food consumption behavior, the hurdle count data model is widely employed (for example, with tobacco (Jones, 1989), alcohol (Yen), cheese (Yen and Jones, 1997), seafood (House, Hanson, and Sureshwaran, 2003), and fresh

[2] The term borrowed from Shonkwiler and Shaw (1996)

fruit (Shi and House, 2011)). There has also been much research using the zero-inflated count data model to analyze consumers' behaviors. Becker et.al. (2016) employed the ZIP model to analyze the influence of attentional bias and emotional dimensions on food consumption. Edmeades and Smale (2006) analyzed the potential demand for a genetically engineered food crop using the ZIP model. Lim et al. (2008) employed the ZIP model to investigated children's consumption of soft drinks, milk and fruit juice. However, very little research has been conducted using the ZIOP model. One use of the model was Downward, Lere-Lopez, and Rasciute (2011), and they used the ZIOP model to analyze sports participation. A second study employed the ZIOP model to analyze two types of peace in social science (Bagozze et al.2004). To our knowledge, this is the first study to use the zero-inflated ordered model to analyze consumers' food consumption behaviors.

Furthermore, most previous papers employed either zero inflated count data model or double hurdle count data model focusing on the discussion of the behaviors of non-consumers and consumers, and analyze the factors influencing their different behaviors. Different from the previous researches, this paper will not only explore the different factors driven consumers' participation and consumption, but also provide an understanding of the market construct of the three different types of consumers: non-consumers, potential-consumers and consumers.

Data

A consumer survey was designed to determine consumers' consumption behavior of fresh and processed mushrooms. In September, 2012, a random sample of 1,217 respondents in the United States was recruited through a national survey panel. The target sample included adults, aged 18 or older, living in the United States. A total of 1,217 respondent initiated the survey and 674 respondents completed the survey and met the screening questions, for a response rate of 55.4%. The survey included three parts. In the first part, respondents were asked to answer a series of questions asking attitudes and perceptions about health benefits of food in general, then mushrooms specifically.

In the second part, respondents were asked to indicate whether they had ever purchased fresh/processed mushrooms before, and if so, what was their consumption frequency in the past month. For those who have purchased mushrooms before, they were further been asked to indicate the reasons why they purchase fresh/processed mushrooms, and the most important factors influencing their purchase decision, while those that did not consume were asked to indicate the top reasons they did not consume.

In the final part, demographic information including socioeconomic status, age, gender, income, education, food budget, and any food restrictions was collected.

Due to the difficulty in collecting the price information across different purchase locations, purchase frequency information is used to represent the consumption amount for each household. In the survey, we first asked whether the respondent had ever purchased fresh or processed mushrooms and then asked whether they had purchased fresh or processed mushroom in the month before the survey was taken. For those respondents who had purchased in the prior month, a follow up question asked how often they purchased during that time period. One month was used to help with the accuracy of the data as it is usually difficult for people to recall purchases more than one month ago.

According to the dataset, 37.2% and 55.8% of participants did not purchase fresh or processed mushrooms in the past month, respectively. Of those that had not purchased, 53.7%and 37.1% reported that they have never purchased fresh or processed mushrooms, respectively (Figure 2). Of the total respondents, approximately 18.5% indicated they were aware of health benefits of mushroom, and approximately 21.6% reported that they believed that mushrooms would help with immunity.

Demographic information is summarized in Table 1. The majority of respondents were female (54.9%) and Caucasian (74.5%) with a college degree (77.0%). The average age of participants is 45 years old and the average household income is approximate $42,500 per year.

Method

Concerning the survey method employed in this research, specifically consumers' choice towards mushroom consumption, the category of "zero consumption" includes two different types of zeroes: genuine non-participants and zero consumption potential consumers. In the case of discrete levels of reported mushroom consumption, zero might reported for non-consumers who choose not to consume mushrooms because of some stable reasons, like eating habits, and personal concerns, and they will not change their mind easily in a short time At the same time, zero consumption might also be reported for potential mushroom consumers who did not consume during the prior month. The potential mushroom consumer made their decision of "non-consumption" as a corner solution to the standard demand problem. It can be expected that the underlying processing driving the behaviors of these two types of zero-consumption consumers will be quite different. If we use the simple discrete choice model (i.e. an ordered probit/logit model) the two types of zeroes will not be differentiated and the model may fail to capture the true reasons behind the zero observations

Harris and Zhao (2007) proposed the zero-inflated ordered probit model (ZIOP), which consists of a split probit model and an ordered probit model with potentially different sets of covariates. The system of the probit model and the ordered probit is generated by two latent equations which allows for the differentiation between the two separate processes generating the zero observations. Furthermore, the error terms of these two latent equations are allowed to be correlated.

In this study, we define X reflecting individuals' characteristics such as income, age, gender, awareness of mushroom benefits, etc. Z reflects the mushrooms characteristics such as taste, price, availability, etc. We also let the matrix W include both X and Z. The split probit equation can be expressed as

$$R_i^* = {X_i}'\alpha + u_i \quad R_i = \begin{cases} 1 & if \quad {X_i}'\alpha + u_i > 0 \\ 0 & otherwise \end{cases} \tag{1}$$

Where R_i^* is the latent variable measuring consumers' propensity for participation to purchase mushrooms, and R_i is a dichotomous variable of observation indicating whether or not consumers decided to participate. X is the vector of explanatory variables; u is the error term.

Conditioning on participation (R=1), consumers need to further decide how much to consume, and the consumption levels are represented by a discrete variable d, which is generated by an ordered probit model through the second latent variable d^*. The ordered probit equation is expressed as equation (2)

$$D_i^* = W_i{}'\beta + \varepsilon_i, \text{ if } R_i\text{=1}; \; D_i = \begin{cases} 0 \;\; if \;\; D_i^* \leq 0 \\ 1 \;\; if \;\; 0 < D_i^* \leq \alpha_1 \\ 2 \;\; if \;\; \alpha_1 < D_i^* \leq \alpha_2 \\ and \;\; so \;\; on \end{cases} \qquad (2)$$

In equation 2, W is the set of explanatory variables, including both X and Z, ε is the error term for the ordered probit equation, and α is the cutoff parameters. We only include three consumption levels in the survey (0 = no consumption in the past month, 1 = ??, and 2 = more than ?? times per month), thus we only have one cutoff parameter.

The error terms of equation (1) and equation (2) are allowed to be correlated, and the joint distribution function of (u_i, ε_i) is assumed to be Gaussian, with zero means, unit variances, and correlation coefficient defined as ρ.

Since, in the ZIOP model, the decision on whether or not to participate and how much to consume are not separately determined, the indicators d and r are also not individually observed. To observe the consumption level Y, it was given the following criteria: Y=R*D. According to this criteria, a positive Y was observed when R=1 and D>0; Y was observed as zero when R=0 or D=0.

Thus, with this model specification, the probability of non-participation is

$$\Pr \; (\mathrm{R} = 0) \;\; = \;\; 1 - \;\; \Phi(X_i{}'\alpha) \qquad (3)$$

The probability of observing zero-level consumption is

$$\Pr(\mathrm{Y} = 0) = \Pr(R = 0) + \Pr(R = 1, D = 0)$$
$$= [1 - \Phi(X_i'\alpha)] + \; \Phi_2(X_i'\alpha, -W_i'\beta; \; -\rho) \qquad (4)$$

The probability of observing a positive consumption level Y=j (j>0) is

Pr(Y = j)

$$= \begin{cases} \Pr(Y=1) = \Pr(\ R=1\ \&\ D=1) = \Phi_2(X_i'\alpha,\ \gamma_1 - W_i'\beta;\ -\rho) - \Phi_2(X_i'\alpha, -W_i'\beta;\ -\rho) \\ \Pr(Y=j) = \Pr(\ R=1\ \&\ D=2) = \Phi_2\left(X_i'\alpha,\ \gamma_j - W_i'\beta;\ -\rho\right) - \Phi_2\left(X_i'\alpha,\ \gamma_{j-1} - W_i'\beta;\ -\rho\right) \\ (j = 2,3 \ldots J-1) \\ \Pr(Y=J) = \Pr(\ R=1\ \&\ D=J) = \Phi_2\left(X_i'\alpha, W_i'\beta - \ \gamma_{j-1};\ -\rho\right) \end{cases} \quad (5)$$

From equation (4), we could indicate that the probability of observing zero level consumption includes two separate processes: the probability of non-participation (R=0) and the joint probability of the choice to participate, but choose to purchase zero. Equation (5) indicates that the probability of observing a positive consumption level is the joint probability of the choice to participate and to consume at j-level intensity.

For almost all the discrete choice models, the marginal effects are useful to indicate the effectiveness of covariates on the probability changes. For ZIOP model, there are different sets of marginal effects which would be of interest to analyze. At first, it would be interesting to analyze the effectiveness of variables on the probability of "participation". Then, comparing the effectiveness of the independent variables on the probability of different level of consumption intensity conditional on participation will be useful to interpret consumers' behaviors. What's more, based on the construct of ZIOP model, we need also pay attention to the marginal effects of the sets of explanatory variables on the overall probability for different level of observed consumption.

In ZIOP model, the marginal effect of variable x_k on the participation probability is

$$ME_{\Pr(R=0)} = \frac{\partial \Pr(R=0)}{\partial x_k} = -\phi(X'\alpha)\ \alpha_k \quad (6)$$

As mentioned before, the W matrix includes both X and Z matrixes (W= (X, Z)), where the variables in Z matrix are the distinct variables that only appear in the second latent equation---consumption equation. Thus, the set of associated coefficient vectors for W was $\alpha^* = (\alpha_x{}', o')'$ and $\beta^* = (\beta_x{}', \beta_z{}')'$. Followed by Harris and

Zhao (2007), the marginal effects of the explanatory variable vector on the overall probabilities of observing zero-level consumption in equation (4) is

$$ME_{\Pr(Y=0)} = -\phi(X_i'\alpha)\alpha^* + \Phi\left(\frac{-W_i'\beta+\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right)\phi(X_i'\alpha)\alpha^* - \Phi\left(\frac{X_i'\alpha+\rho(-W_i'\beta)}{\sqrt{1-\rho^2}}\right)\phi(W_i'\beta)\beta^* \quad (7)$$

The marginal effect on the overall probabilities of observing different positive level of consumption in equation 5 is

$$ME_{\Pr(Y=1)} = [\Phi\left(\frac{\gamma_1-W_i'\beta+\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{-W_i'\beta+\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right)]\ \phi(X_i'\alpha)\alpha^*$$

$$+[\ \phi(-W_i'\beta)\Phi\left(\frac{X_i'\alpha+\rho(-W_i'\beta)}{\sqrt{1-\rho^2}}\right) - \phi(\ \gamma_1 - W_i'\beta)\Phi\left(\frac{X_i'\alpha+\rho(\ \gamma_1-W_i'\beta)}{\sqrt{1-\rho^2}}\right)]\ \beta^* \quad (8)$$

$$ME_{\Pr(Y=j)} = [\Phi\left(\frac{\gamma_j-W_i'\beta+\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\gamma_{j-1}-W_i'\beta+\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right)]\ \phi(X_i'\alpha)\alpha^*$$

$$+[\ \phi(\ \gamma_{j-1}-W_i'\beta)\Phi\left(\frac{X_i'\alpha+\rho(\ \gamma_{j-1}-W_i'\beta)}{\sqrt{1-\rho^2}}\right)\phi(\ \gamma_j W_i'\beta)\Phi\left(\frac{X_i'\alpha+\rho(\ \gamma_j-W_i'\beta)}{\sqrt{1-\rho^2}}\right)]\ \beta^* \quad (9)$$

$$ME_{\Pr(Y=J)} = \Phi\left(\frac{W_i'\beta-\ \gamma_{j-1}-\rho X_i'\alpha}{\sqrt{1-\rho^2}}\right)\phi(X_i'\alpha)\alpha^* + \Phi\left(\frac{X_i'\alpha-\rho(W_i'\beta-\ \gamma_{j-1})}{\sqrt{1-\rho^2}}\right)\phi\big(W_i'\beta - \gamma_{j-1}\big)\alpha^* \quad (10)$$

where $\phi(.)$ is the p.d.f of the standard univariate normal distribution, and $\Phi(.)$ is the c.d.f of the standard univariate normal distribution. The standard errors of the marginal effects in this study could be either calculated using the Delta method (Greene, 2003) or using the simulated asymptotic sampling techniques. Considering the complexity of the marginal effect, in this case the sampling technique is used. To be more specific, we randomly draw θ (where θ is the parameters in the ZIOP model) from MVN($\hat{\theta}$, $\widehat{var}[\theta]$) 10,000 times, and for each draw we calculate the marginal effect based on equation (7) to equation (10), and then calculate the standard errors.

These empirical standard deviations of the simulated marginal effect are the valid asymptotic estimates of the true marginal effects' standard errors.

Furthermore, from the ZIOP model, we also calculate the expected probability of observing different levels of consumption: the probability of observing a non-consumer is expressed in equation (11); and the probability of observing different levels of consumption given R=1 is given in equation (12)

$$E(R_i=0) = \Pr(R_i=0|X) = 1 - \Phi(X_i'\alpha) \quad (11)$$

$$E(D_i=j|R_i=1)=\frac{\Pr(D_i=d, R_i=1|X,W)}{\Pr(R_i=1|X)} = \frac{\Phi_2(X_i'\alpha,\ \gamma_j - W_i'\beta;\ -\rho) - \Phi_2(X_i'\alpha,\ \gamma_{j-1} - W_i'\beta;\ -\rho)}{\Phi(X_i'\alpha)} \quad (12)$$

Results

Fresh mushroom consumption

Regression results for the model of fresh mushroom consumption are shown in Table 2. The ZIOP model results are displayed in the third column, and the binary probit and ordered probit models, which treat all zeros indifferently, are listed in the first and second column as a comparison. Table 3 displays the marginal effects on Pr(Y=0) using a ZIOP model, comparing with the results from the probit and ordered probit models. For the ZIOP model, the overall marginal effect on Pr(y=0) was divided into two parts: the effect on non-participation (Pr(r=0)), and the effect on the participation with zero consumption Pr(r=1, y=0). In table 4, we present marginal effects on the unconditional probabilities of positive levels of consumption (y=1, 2), using an ordered probit model versus the ZIOP model.

The estimated results illustrate that consumers who indicated that they were aware of the health benefits of mushrooms are significantly more likely to consume more fresh mushrooms in the binary probit and ordered probit models. However, when looking at the results from the ZIOP model, we found that consumers' awareness of health benefits only significantly influences the decision of participation, not consumption frequency. The same results for consumers' knowledge of mushroom (the effectiveness of mushroom towards certain symptom). The variable of consumers 'knowledge of the effectiveness of mushrooms on certain symptom is statistically significant on all of the three models, and within the ZIOP model, it indicates that it is only significant in the first stage (decision to participate).Thus it indicated that consumers' knowledge and awareness of mushroom benefits are significant factors influencing fresh mushroom consumption, but they only have significantly positive effect on consumers' participation decision, they do not actually influence frequency of consumption.

Of the socio-economic characteristics, the variable age is significantly negative in the ordered probit(OP) and zero-inflated ordered probit（ZIOP）model, while within the ZIOP model, age is significant only in the participation stage, indicating that younger people are more likely to consume fresh mushrooms. The variables for race/ethnicity is statistically significant in influencing consumers' consumption frequency of fresh mushrooms in both the ZIOP model and OP model. From the model results, compared to Caucasians, Hispanic, Black and other race will consume fresh mushrooms more often. Income is significant in the binary probit model and the participation stage of ZIOP model, which indicates that people with higher income are more willing to consume fresh mushrooms, but higher incomes don't translate to more frequent consumption. Budget is significantly positive in the probit model, and OP model. When looking into ZIOP model, we find that budget is significantly correlated with higher fresh mushroom consumption frequency, but it is not significantly influencing consumers' participation decision. Thus, comparing the variable income and budget, we could see that people with higher household income would more willing to try the fresh mushrooms, and people with higher food budget would buy the fresh mushroom more often.

When looking at mushroom characteristics, we find that taste, convenience and food safety are the top three factors influencing consumers' consumption decision. As revealed in the Table 3, the dominant influence on fresh mushroom consumption and consumption frequency is consumers' awareness of mushroom benefits, followed by consumers' ethnicity /race.

In terms of the potential unobserved effect, the ZIOP model for fresh mushroom suggests that there is no significant correlation existing between the participation decision and consumption decision. It indicates that for the promotion of fresh mushrooms, the marketing strategies of attracting new consumers might not work significantly on the increase of consumption frequency, and vice versa.

The estimated probabilities of different types of consumers are displayed in table 8. Overall, the probability of non-consumers of fresh mushroom is 18% (compared to the observed percent of 37.3%). The percentage of potential-consumers is 19.5% (compared to the observed percent of 17.3%).

Processed mushroom consumption

Regression results for processed mushroom consumption are shown in Table 5, and Tables 6 and 7 display the marginal effects on the choice probabilities for processed mushroom consumption. Results indicate that consumers' awareness of health benefits of mushrooms is not significant in all equations. As for the consumers' knowledge of the effectiveness of mushroom towards some certain symptoms, it is not a statistically significant factor except for the binary probit model.However, consumers' knowledge of the effectiveness of mushroom enhancing immunity is significant in both the binary probit model and the ZIOP model. The variable of consumers' knowledge of the effectiveness of the mushroom enhancing immunity is significant in both of the participation stage and the consumption stage, but with opposite directions. Consumers' knowledge of the effectiveness of mushroom enhancing immunity will significantly decrease the probability of non-participation (Pr(r=0)), while significantly increase the probability of zero-consumption consumers (Pr(r=1, y=0)). It indicates that although people's knowledge of mushroom health might attract consumers to try processed mushrooms, it does not impact frequency of consumption and might make consumers more likely to consume processed mushrooms infrequently.

Focusing on the socio-economic variables, age is not significant except for the binary probit model and income is not significant in the binary probit model nor the OP model, but it does being significant in the ZIOP model. To be more specific, income is negatively related to the consumption participation decision, yet, positively related to the consumption frequency. To interpret the meaning, it helps to examine the marginal effects. Although income does not significantly influencing the probability of zero-consumption, when the ZIOP model is used, we assume that the observed consumption categories are the result of two distinct decisions of "participation" and "levels of consumption conditional on participation", on which income can have opposite effects. As shown in Table 6, an increase in income significantly increases the probability of non-participation (Pr(r=0)), while significantly decreasing the probability of zero-consumption consumers (Pr(r=1, y=0)). Thus, the effect of income on the overall zero consumption is approximately zero as the two impacts counter effect each other in the binary and OP models. However, using the ZIOP model, we can see that consumers with higher income are more likely to be non-consumers, but if they do consume, they are more likely to consume recently (less likely to be potential consumers). The variable of budget is significantly positive in the probit model, OP model, and the participation-stage of ZIOP model, but is not correlated with consumers' consumption frequency. It suggests that people with a higher food budget will be more willing to consume processed mushrooms.

Considering mushroom characteristics, we find that for the processed mushrooms, taste and price are the two factors that consumers care about the most. Processed mushrooms with better taste and lower price will significantly attract more consumption. Different from the fresh mushrooms, the factor of convenience and health are not significant factors influencing consumers' choice.

Based on the marginal effects, ethnicity is the dominant factor influencing consumers' participation and consumption frequency of processed mushrooms. Compared to Caucasians, people of other races are more willing to consume more processed mushrooms.

In terms of the potential unobserved effect the ZIOP model of processed mushrooms, it suggests that there exists a significant negative correlation between the two-stage decision, meaning the factors which could increase the probability of consumption might cause consumers consume less frequently. This relationship suggests that for the promotion of processed mushrooms, attracting new consumers and then increasing their consumption frequency are entirely different challenges.

The estimated probabilities of different types of consumers of processed mushrooms are displayed in table 9. Overall, the probability of non-consumers of processed mushroom is 32.8%, which is much higher than the percentage of non-consumers of fresh mushrooms (18%). The predicted percentage of potential-consumers of processed mushrooms is 24%, which slightly higher than that for fresh mushrooms (19.5%).

Conclusion

In this paper, a ZIOP model was employed to analyze the significant factors influencing consumers' behavior in the fresh and processed mushroom market accounting simultaneously for the probability of consumption participation and frequency. The ZIOP model allows us to take into account the excessive zero observations in the survey data which might come from two different sources: genuine non-consumers and zero-consumption participants (potential consumers). The latter were considered as potential consumers because these zero-consumption participants choose not to consume at current time, but would do so if the circumstance were different.

Comparing the fresh and processed mushroom market, the mushroom market is dominated by fresh mushrooms. The portions of non-consumers of processed mushrooms is much larger than that of fresh mushroom. However, the percentage of potential-consumers of processed mushrooms is larger than that of fresh mushroom, which indicates that the processed mushroom has potential in the future if appropriate marketing strategies are applied.

For fresh mushrooms, results indicate that the decision to consume or not and the decision on how much to consume are driven by different factors, and there is no significant relationship between the unobserved factors. Consumers' awareness and knowledge about mushroom benefits are significantly positively related to consumers' participation. Age affects consumers participation, where younger consumers are more likely to consume fresh mushrooms. Income also significantly increases consumers' probability of participation. Consumption frequency is significantly related to the consumers' food budget, where higher food expenditures drives higher consumption frequency. Ethnicity also significantly influence fresh mushroom consumption frequency with Caucasians consuming fresh mushroom less often than others. What is more, we also find that once consumers make their purchase decision, convenience, health and taste are the most important factors to increase frequency of fresh mushroom consumption.

As for the processed mushrooms, we found that the decision to consume or not and the decision on how much to consume were also driven by different observed factors, as well as being negatively related based on the unobserved factors. Consumers' knowledge of mushroom health was positively related towards consumers' participation, but negatively related with the consumption frequency. Income was also negatively related with the probability of participation, and positively related with the consumption frequency, which indicates that people with higher income level will be less likely consume processed mushrooms, however if people have already are processed-mushroom consumers, they would consume more often with higher income, which shows that the processed mushrooms is a normal good. The probability of participating also increases with higher food budgets. Compared with the Caucasians, other races seems to be consume processed mushrooms less often. What is more, we also find that once consumers make their purchase decisions on processed mushrooms, price and taste are the only two significant factors that influence frequency of consumption.

Mushroom consumption participation seems to be significantly related with consumers' awareness and knowledge of mushroom benefits, which is especially true for fresh mushrooms. This suggests that a policy of advertising the mushroom benefits would be a good method to encourage mushroom consumption. Considering the products' characteristics, taste, convenience and health should be the key points promoting fresh mushrooms, while lower prices and better taste would help processed mushroom attracting more consumption.

Considering the unobserved factors, there is no significant correlation between the decisions to participate and the consumption frequency for fresh mushrooms, which suggest that marketing strategies which succeed in attracting new consumers might not work well on the increase of consumption frequency, and vice versa. For processed mushrooms, there is negative unobserved correlation between the two-stage decisions of consumption, thus, marketing strategies could harness different aspects of the consumption behavior: those for whom the primary issue is to being attracted to the market, and those, more committed, who make a choice about how often to consume the processed mushrooms.

References

Bagozzi, B. E., Hill, D. W., Moore, W. H., & Mukherjee, B. (2015). Modeling Two Types of Peace The Zero-inflated Ordered Probit (ZiOP) Model in Conflict Research. *Journal of Conflict Resolution*, 59(4), 728-752.

Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4), 791-799.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829-844.

Chang, S. T., & Buswell, J. A. (1996). Mushroom nutriceuticals. *World Journal of Microbiology and Biotechnology*, 12(5), 473-476.

Downward, P., Lera-Lopez, F., & Rasciute, S. (2011). The zero-inflated ordered probit approach to modelling sports participation. *Economic Modelling*, 28(6), 2469-2477.

Famoye, F., & Singh, K. P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4(1), 117-130.

Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of applied econometrics*, 4(1), 23-39.

Harris, M. N., & Zhao, X. (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics*,141(2), 1073-1099.

House, L., Hanson, T. R., & Sureshwaran, S. (2003). U. S. consumers: Examining the decision to consume oysters and the decision of how frequently to consume oysters. *Journal of Shellfish Research*, 22(1), 51-59.

Lucier, G., Allshouse, J. E., & Lin, B. H. (2003). Factors affecting US mushroom consumption. US Department of Agriculture, *Economic Research Service*

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3), 341-365.

Mwalili, S. M., Lesaffre, E., & Declerck, D. (2008). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical Methods in Medical Research*, 17(2), 123-139.

Shonkwiler, J. S., & Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. *Journal of Agricultural & Resource Economics*, 21(2).

Shi, L., House, L., & Gao, Z. (2011, May). Consumer Structure of the Blueberry Market: A Double Hurdle Model Approach. *In 2011 Annual Meeting, July 24-26, 2011, Pittsburgh, Pennsylvania (No. 103630).* Agricultural and Applied Economics Association.

Sullivan, R., Rowan, N. J., & Smith, J. E. (2006). Medicinal mushrooms and cancer therapy: translating a traditional practice into Western medicine. *Perspectives in biology and medicine,* 49(2), 159-170.

Yen, S. T., Dellenbarger, L. E., & Schupp, A. R. (1995). Determinants Of Participation And Consumption: The Case Of Crawfish In South Louisiana. *Journal of Agricultural and Applied Economics*, 27, 253-253.

Yen, S. T., & Jones, A. M. (1997). Household consumption of cheese: an inverse hyperbolic sine double-hurdle model with dependent errors. *American journal of agricultural economics*, 79(1), 246-251.

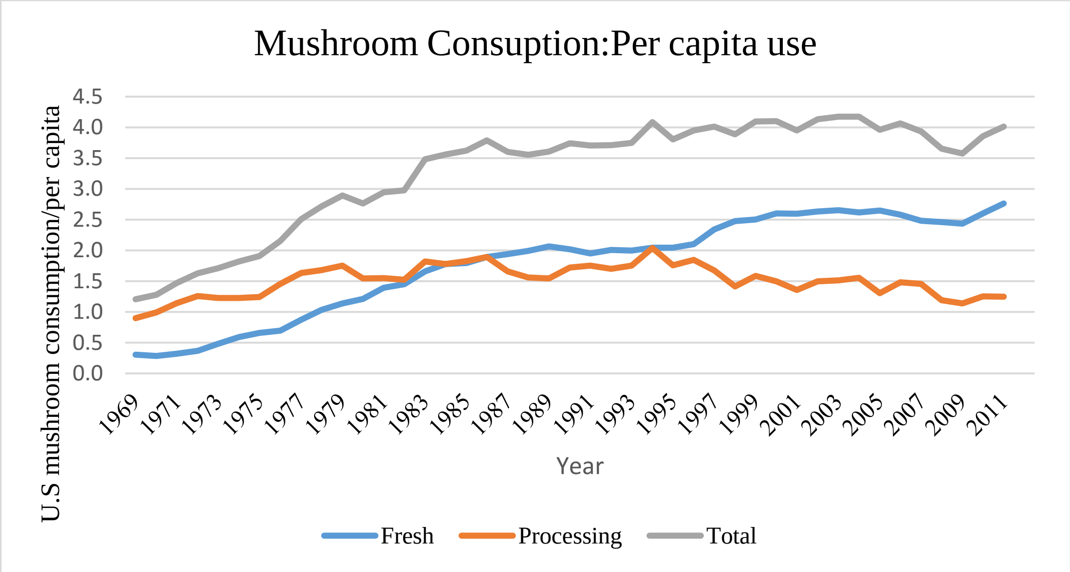

Figure 1. Mushroom Consumption Changes

Table 1. Demographics of survey respondents

Category	Percent	Category	Percent
Percent Female	54.91	Employment Status	
		Full-time	36.13
Age		Part-time	15.46
18-29	22.54	Student	5.78
30-39	18.93	Retired	16.76
40-49	21.39	Not employed	19.22
50-69	31.22	Other	6.65
69+	5.32		
Diet		Income level	
vegetarian	6.54	$24,999 or less	27.02
Non-vegetarian	93.46	$25,000-$34,999	10.69
		$35,000-$49,999	15.32
Ethnicity/Race		$50,000-$74,999	21.53
White/Caucasian	75.53	$75,000-$99,999	10.55
African American	11.13	$100,000 or more	14.88
Hispanic	7.40		
Asia	7.68		
Other	2.70	Education level	
		High school graduate or less	23.04
Awareness of mushroom benefits		Some College or higher	76.96
Yes	18.5	Food budget per week	
No	81.5	Less $49	12.2
		$59-$99	36.3
Knowledge about mushroom towards immunity		$100-$149	32.7
Yes	11.0	$150-$199	10.8
No	89.0	$200-$249	3.5
		$241 and above	4.5

Table 2 Fresh Mushoom consumption: regression results

	Binary Probit Model		Ordered Probit model		ZIOP model			
					Participation		Frequency	
	coeffs	t-stats	coeffs	t-stats	coeffs	t-stats	coeffs	t-stats
male	0.234**	2.11	0.072	0.72	0.260	1.40	-0.032	-0.22
college	0.116	0.90	-0.046	-0.37	-0.059	-0.27	-0.052	-0.30
age	0.006	0.17	-0.096**	-2.90	-0.127**	-1.99	-0.066	-1.15
income	0.062**	2.09	0.038	1.39	0.097*	1.73	-0.008	-0.18
hispanic	0.055	0.21	0.706***	2.71	0.741	0.89	0.693**	2.23
black	-0.314*	-1.80	-0.062	-0.35	-0.178	-0.55	0.032	0.13
asian	0.686***	2.66	0.374**	2.00	0.296	0.76	0.423*	1.77
otherrace	0.548	1.54	0.614**	2.07	-0.058	-0.13	0.941**	2.15
immunity	0.057	0.33	-0.187	-1.08	-0.372	-0.72	-0.008	-0.03
symptom	0.478**	3.88	0.293**	2.45	0.392*	1.86	0.165	0.83
awr	0.606**	3.79	0.299**	2.19	0.558**	2.21	0.211	1.07
budget	0.145**	2.82	0.155***	3.34	0.081	0.89	0.187***	3.06
vegan	0.085	0.35	-0.011	-0.05	0.389	0.83	-0.132	-0.52
taste	-		0.202***	4.70	-		0.244***	4.28
price	-		0.046	0.94	-		-0.004	-0.07
convenience	-		0.064	1.22	-		0.162**	2.09
mushhealth	-		0.122***	2.95	-		0.156***	2.82
diversity	-		0.022	0.45	-		0.046	0.74
Constant	-1.03**	-3.67			0.740	0.97	-2.110*	4.37
Threshold			3.48***	6.70			2.099**	8.42
Rho(u,e)					0.150	0.20		
# of obs	648		648		648			
Log-likelihood	-382.174		-523.314		-513.891			
Wald TestX_2	76.83(df=13)		233.22(df=18)		161.13(df=18)			

Table 3: Fresh mushroom consump;iton: marginal effect for non-participation and zero

			ZIOP model		
	Binary Probit	Ordered Probit	Non-Participation	Zero consumption	Full
	Pr(y=0)	Pr(y=0)	Pr(r=0)	Pr(r=1, y=0)	Pr(y=0)
male	-.087**	-.027	-0.061	0.021	-0.040
	(.041)	(.037)	(0.044)	(0.024)	(0.031)
college	-.043	.017	0.014	0.002	0.016
	(.049)	(.046)	(0.052)	(0.030)	(0.037)
age	- .002	.036**	0.029**	-0.001	0.028***
	(.013)	(.012)	(0.014)	(0.009)	(0.010)
income	-.023**	-.014	-0.022*	0.008	-0.015**
	(.011)	(.010)	(0.012)	(0.007)	(0.008)
hispanic	-.020	-.221***	-0.175	-0.027	-0.20
	(.096)	(.064)	(0.198)	(0.076)	(0.142)
black	.121*	.023	0.041	-0.015	0.025
	(.069)	(.067)	(0.076)	(0.043)	(0.053)
asian	-.216***	-.130**	-0.071*	-0.028	-0.099*
	(.063)	(.059)	(0.042)	(0.047)	(0.065)
otherrace	-.176*	-.197	0.010	-0.111	-0.100
	(.094)	(.076)	(0.104)	(0.071)	(0.082)
immunity	-.021	.068	0.081	-0.022	0.059
	(.066)	(.061)	(0.117)	(0.065)	(0.068)
symptom	-.182***	-.112**	-0.092**	0.007	-0.085***
	(.047)	(.046)	(0.050)	(0.031)	(0.038)
awr	-.202***	-.108**	-0.132***	0.012	-0.120***
	(.046)	(.047)	(0.064)	(0.034)	(0.046)
budget	-.053***	-.058***	-0.019	-0.016*	-0.036***
	(.019)	(.017)	(0.022)	(0.010)	(0.015)
vegan	.031	.004	-0.092	0.042	-0.050
	(.088)	(.078)	(0.112)	(0.049)	(0.080)
taste	-	-.076***	-	-0.028***	-0.028***
		(.016)		(0.009)	(0.009)
price	-	-.017	-	- 0.000	- 0.000
		(.018)		(0.008)	(0.008)
convenience	-	-.024	-	-0.018**	-0.018**
		(.020)		(0.009)	(0.009)
mushhealth	-	-.046**	-	-0.018***	-0.018***
		(.015)		(0.007)	(0.007)
diversity	-	-.008	-	-0.005	-0.005
		(.018)		(0.008)	(0.008)

Table 4: Fresh mushroom consumption: marginal effect for non-zero consumption levels

	OP	ZIOP	OP	ZIOP
	Pr(y=1)	Pr(y=1)	Pr(y=2)	Pr(y=2)
male	.011	0.034	.016	0.005
	(.016)	(0.026)	(.022)	(0.025)
college	-.007	-0.004	-.010	-0.013
	(.018)	(0.032)	(.028)	(0.031)
age	-.015***	-0.011	-.021***	-0.018*
	(.005)	(0.010)	(.007)	(0.008)
income	.006	0.012*	.008	0.003
	(.004)	(0.008)	(.006)	(0.007)
hispanic	.016	0.040	.206**	0.162**
	(.033)	(0.112)	(.093)	(0.068)
black	-.010	-0.024	-.013	-0.001
	(.031)	(0.047)	(.036)	(0.044)
asian	.035***	0.005	.096*	0.094*
	(.010)	(0.055)	(.055)	(0.046)
otherrace	.021	-0.074	.176*	0.174*
	(.030)	(0.085)	(.103)	(0.085)
immunity	-.024	-0.042	-.044	-0.016
	(.018)	(0.074)	(.044)	(0.046)
symptom	.052**	0.037	.059***	0.048
	(.025)	(0.032)	(.023)	(0.033)
awr	.035***	0.054*	.072*	0.066*
	(.012)	(0.034)	(.036)	(0.034)
budget	.024***	-0.003	.034***	0.039***
	(.008)	(0.014)	(.009)	(0.012)
vegan	-.002	0.056	-.002	-0.006
	(.033)	(0.061)	(.011)	(0.049)
taste	.032***	-0.017	.044***	0.045***
	(.009)	(0.013)	(.009)	(0.012)
price	.007	0.001	.010	-0.001
	(.008)	(0.006)	(.011)	(0.012)
convenience	.010	-0.012	.014	0.030*
	(.008)	(0.011)	(.011)	(0.015)
mushhealth	.019***	-0.011	.027***	0.029**
	(.007)	(0.009)	(.009)	(0.011)
diversity	.003	-0.003	.005	0.009
	(.008)	(0.006)	(.011)	(0.011)

Table 5 Processed Mushoom consumption: regression results

	Binary Probit Model		Ordered Probit model		ZIOP model			
					Participation		Frequency	
	coeffs	t-stats	coeffs	t-stats	coeffs	t-stats	coeffs	t-stats
male	0.027	0.26	-0.060	-0.59	0.124	0.64	-0.313	-1.67
college	0051	-0.40	-0.169	-1.37	-0.019	-0.09	-0.155	-0.76
age	0.073**	2.20	0.0108	0.33	0.087	1.29	-0.072	-1.11
income	-0.004	-0.14	-0.012	-0.44	-0.088*	-1.71	0.077*	1.69
hispanic	-0.162	-0.65	0.178	0.71	1.645	0.88	-0.843	-1.62
Black	-0.322*	-1.85	-0.186	-1.04	0.104	0.19	-0.279	-0.56
asian	-0.179	-0.88	-0.222	-1.17	-0.196	-0.65	-0.070	-0.21
otherrace	-0.035	-0.11	-0.341	-1.11	0.956	1.32	-1.65**	-2.84
immunity	0.291*	1.65	0.094	0.52	0.810**	3.17	-.976**	-3.56
symptom	0.348**	2.82	.197	1.61	0.089	0.41	-0.010	-0.03
awr	0.152	1.11	-0.0800	-0.59	0.137	0.52	-0.278	-1.13
budget	0.176***	3.62	0.153***	3.24	0.200***	2.48	-0.027	-0.30
vegan	0.336	1.54	0.339*	1.68	5.961	0.01	-0.750**	-2.57
taste	-	-	0.086**	1.96	-	-	0.065*	1.74
price	-	-	0.081*	1.65	-	-	0.112**	2.11
convenience	-	-	0.113**	2.10	-	-	0.091	1.31
mushhealth	-	-	0.050	1.17	-	-	0.059	1.28
diversity	-	-	-0.005	-0.11	-	-	0.064	1.19
Constant	-1.39**	-5.03			-1.041	-2.140	0.579	1.08
Threshold			3.118**	6.70	1.254			
Rho(u,e)					-0.907***	-7.945		
# of obs	652		652			652		
Log-likelihood	-422.82		-542.588		-524.368			
Wald testX_2	44.25(df=13)		130.68(df=18)		87.14(df=18)			

Table 6: Processed mushroom consumption: marginal effect for non-participation and zero

			ZIOP model		
	Binary Probit	Ordered Probit	Non-Participation	Zero consumption	Full
	Pr(y=0)	Pr(y=0)	Pr(r=0)	Pr(r=1, y=0)	Pr(y=0)
male	-.011	.023	-0.037	0.051	0.013
	(.041)	(.039)	(0.058)	(0.039)	(0.031)
college	.020	.066	0.006	0.021	0.027
	(.049)	(.049)	(0.067)	(0.043)	(0.036)
age	-.028**	-.004	0.025	-0.016	-0.009
	(.013)	(.013)	(0.019)	(0.014)	(0.009)
income	.002	.004	0.026*	-0.017*	-0.009*
	(.011)	(.011)	(0.015)	(0.010)	(0.005)
hispanic	.063	-.070	-0.488	0.231	-0.257
	(.094)	(.101)	(0.552)	(0.198)	(0.368)
black	.122*	.070	-0.027	0.051	0.024
	(.063)	(.066)	(0.157)	(0.106)	(0.067)
asian	.069	.083	0.059	-0.003	0.056
	(.077)	(.069)	(0.091)	(0.063)	(0.048)
otherrace	.014	.124	-1.747***	0.637***	-1.111***
	(.122)	(.103)	(0.205)	(0.110)	(0.197)
immunity	-.111*	-.036	-0.241***	0.192***	-0.048
	(.064)	(.068)	(0.070)	(0.067)	(0.050)
symptom	-.134**	-.075	-0.027	0.005	-0.022
	(.046)	(.049)	(0.066)	(0.046)	(0.032)
awr	-.060	.031	-0.040	0.049	0.008
	(.054)	(.052)	(0.078)	(0.052)	(0.040)
budget	-.069***	-.059***	-0.060***	0.017	-0.042***
	(.019)	(.018)	(0.024)	(0.016)	(0.014)
vegan	-.133	-.134*	-1.699	0.48	-1.218
	(.086)	(.080)	(4.482)	(4.918)	(3.55)
taste	-	-.033**	-	-0.009*	-0.009*
		(.017)		(0.005)	(0.005)
price	-	-.031*	-	- 0.015**	-0.015**
		(.020)		(0.008)	(0.008)
convenience	-	-.044**	-	-0.012	-0.012
		(.021)		(0.10)	(0.10)
mushhealth	-	-.019	-	-0.008	-0.008
		(.016)		(0.007)	(0.007)
diversity	-	.002	-	-0.008	-0.008
		(.020)		(0.007)	(0.007)

Table 7: Processed mushroom consumption: marginal effect for non-zero consumption levels

	OP	ZIOPC	OP	ZIOP
	Pr(y=1)	Pr(y=1)	Pr(y=2)	Pr(y=2)
male	-.012	0.018	-.010	-0.032
	(.021)	(0.034)	(.017)	(0.016)
college	-.035	-0.006	-.031	-0.021
	(.024)	(0.037)	(.024)	(0.018)
age	.002	0.013	.002	-0.004
	(.007)	(0.010)	(.006)	(0.005)
income	-.003	-0.014	-.002	0.004
	(.006)	(0.010)	(.005)	(0.005)
hispanic	.036	0.268	.034	-0.011
	(.048)	(0.308)	(.053)	(0.071)
black	-.041	0.007	-.029	-0.031
	(.041)	(0.084)	(.025)	(0.035)
asian	-.050	-0.035	-.034	-0.021
	(.044)	(0.050)	(.025)	(0.031)
otherrace	-.077	0.977*	-.047	0.133*
	(.071)	(0.136)	(.033)	(0.081)
immunity	.021	0.125*	.015	-0.077*
	(.04)	(0.045)	(.028)	(0.028)
symptom	.043	0.016	.032*	0.006
	(.027)	(0.036)	(.019)	(0.020)
awr	-.017	0.020	-.013	-0.028
	(.030)	(0.043)	(.022)	(0.021)
budget	.032***	0.034**	.026***	0.008
	(.011)	(0.013)	(.008)	(0.009)
vegan	.063**	0.934	.070	0.283
	(.032)	(5.696)	(.049)	(8.864)
taste	.018*	0.000	.015**	0.008*
	(.010)	(0.003)	(.008)	(0.005)
price	.017*	0.001	.014*	0.014**
	(.011)	(0.004)	(.009)	(0.006)
convenience	.024**	0.001	.019**	0.012
	(.012)	(0.004)	(.009)	(0.009)
mushhealth	.011	0.000	.009	0.008
	(.009)	(0.003)	(.007)	(0.006)
diversity	-.001	0.000	-.001	0.008
	(.011)	(0.003)	(.009)	(0.007)

Table 8 Estimated probabilities for fresh mushroom consumption

Binary Probit		Ordered Probit		ZIOP model	
Pr(y=0)	0.351	Pr(y=0)	0.361	Pr(r=0)	0.180
				Pr(r=1, y=0)	0.195
Pr(y>0)	0.649	Pr(y=1)	0.504	Pr(r=1, y=1)	0.392
		Pr(y=2)	0.135	Pr(r=1, y=2)	0.233

Table 9 Estimated probabilities for Processed mushroom consumption

Binary Probit		Ordered Probit		ZIOP model	
Pr(y=0)	0.571	Pr(y=0)	0.601	Pr(r=0)	0.328
				Pr(r=1, y=0)	0.240
Pr(y>0)	0.429	Pr(y=1)	0.302	Pr(r=0, y=1)	0.404
		Pr(y=2)	0.097	Pr(r=0, y=2)	0.028