A Triple-Hurdle Count Data Model of Market Participation and Consumption

Yuan Jiang (jyspring@ufl.edu), Graduate Student
Lisa House (lahouse@ufl.edu), Professor
Food and Resource Economics Department
University of Florida

Selected Paper prepared for presentation at the
2017 Agricultural & Applied Economics Association's Annual Meeting, Chicago, IL, July 31-August 2

A Triple-Hurdle Count Data Model of Market Participation and Consumption

Introduction

In empirical economics, there has long been interest in modeling consumers' behaviors, in particular consumers' preferences and purchases, and analyzing and predicting market structure. When analyzing consumption behavior on the individual level, researchers frequently find themselves working with count data, especially when collecting primary data using survey instruments. Count data can be found when measuring consumers' consumption frequency or intensity during a certain time period. Data on the frequency of purchase of food products in a given period may provide unique challenges, as one might find many observations recorded as zero-consumption. For example, if consumers are asked "How often did you consume blueberries last month?", there may be many respondents who answer that they did not consume in the past month (hence a zero observation). Data with excess zero observations is referred to as zero-inflation.

Though there are models to deal with zero-inflation, for consumption data, there is an interesting quality as the zero observations may be caused by different reasons. One reason is that some individuals have a non-positive desire for this product. For some permanent reason, these individuals will not be consumers of the product (i.e. they might be allergic). However, a considerably different reason for a zero observation is that some individuals have a positive desire for the product, but they do not consume for some temporary reason (i.e. they may not be able to afford the good at the given price). In this case, zero consumption is the corner solution for the individual's utility-maximizing decision. Similarly, some individuals might have a positive desire to consume the product, but not during the recorded period (i.e. past month) due to infrequent or seasonal consumption. Individuals who show zero consumption in the first case are considered as non-consumers, and these zero observations are structural zeros. The individuals who have positive desire to consume, but were observed consuming zero units because of the second and third reasons are potential consumers, and the corresponding zero observations are considered as the sampling zeros. The particular interpretations given to these zero consumption observations can have a crucial bearing on the estimation techniques, and the interpretation of market segmentation.

In survey research, both non-consumers (those who do not have positive participation desire) and potential consumers (those who have positive market participation but choose to consume zero units) are observed to have zero-consumption and are often treated as one in modeling. However, the decision of market participation would be driven by a structurally different process than the subsequent consumption decision and consumption intensity decision for potential consumers compared to non-consumers. Thus, analyzing the different factors influencing consumers' participation and consumption decisions will provide researchers, retailers, and producers with a better understanding of consumers' behaviors if these two types of participants are modeled separately.

Existing analyses of market participation and consumption are mostly based on a "double-hurdle" modeling approach. The double-hurdle approach assumes that the process of generating zero-consumption is handled separately from the process of generating positive consumption. However, it fails to distinguish between potential consumers and non-consumers, which are all observed at zero-consumption. It is possible that marketing strategies developed on these models to target non-consumers might exert different influences on non-consumers and potential consumers. To address these limitations, this paper presents a "triple-hurdle" count data model which allows us to observe participation intention in the first hurdle, and conditional on the participation decision, consumers would further make the subsequent consumption and consumption intensity decisions (different from the double hurdle model by adding a step that allows a positive participation but zero consumption decision). This model is used to classify three types of consumers in the market: non-consumers, potential consumers, and consumers, and explores the appropriate structurally different reasons explaining the three groups on market participation, consumption, and consumption intensity in sequence.

The econometric modeling of count data for consumption behavior

When dealing with the problem of "excessive zeros", a variety of statistical techniques has been proposed and applied in economic literature. One of the most widely used is the Tobit model (Tobin, 1958). It was developed to account for the limited capacity of simple linear regression in the presence of a preponderance of zero observations. However, the Tobit model assumes zeros represent censored values of an underlying normally distributed latent variable that theoretically includes negative values. This results in a restrictive model that assumes all zero observations are structural zeros resulting from the same generating process (there is no allowance for the possibility of sampling zeros). The model is also restrictive by assuming that it is the same set of factors that influence both consumers' desire and acquisition. To solve the shortcoming of the Tobit model, a number of generalizations to the Tobit model have been developed.

The most popular generalization of Tobit model is the Heckman's sample selection model and double-hurdle model. When modelling consumption behavior, given the different reasons caused the zero-consumption, it assumes that individuals must pass two stages before being observed with a positive level of consumption: a participation decision and a consumption decision. he difference between the double-hurdle and sample selection models is in the assumption of dominance: whether the participation decision dominates the consumption decision

First considering the double-hurdle model, it assumes that positive consumption is observed only when consumers have overcome both of the stages. In other words, the observed consumption variable is given by $y=dy^*$ where d is the indicator for a consumers' desire to participate, and y^* is the indicator for the consumers' determination on the consumption level. In order to observe a positive consumption level, both d and y* must be positive. However, if we assume the participation decision dominants the consumption decision, it implies that all the zero observations are structural zeros, thus zero consumption does not arise from a standard corner solution. To express the dominance using the equation, it implies that $p(y^*>0|d=1)=1$

One significant problem with the Tobit model and its generalizations is that it assumes the latent variable is normally distributed, and it is very sensitive to violations of the assumption of normality (Arabmazer and Schmidt, 1982), thus Tobit model and its generalized models have significant restrictions when applying to the analysis of consumption behavior.

When the dependent variable is in the format of count data, the most popular regression technique is the Poisson regression Poisson regression is commonly used in economics to model the number of events, for example, the frequency of consumption. However, the Poisson model fails to provide an adequate fit when there exists the problem of "excessive zeros". The Poisson model has a basic assumption of mean-variance equality, which is violated when "excessive zeros" pull the mean towards zero. A number of modified Poisson regression models has been developed to account for excess zeroes, the most popular of which are zero-inflated/modified Poisson models and Hurdle models.

The zero-inflated Poisson (ZIP) model was proposed by Lambert in 1995. This zero-inflated count data model assumes that zero observation come from two distinct sources: "sampling zeros" and "structured zeros." When applying to the consumption analysis, it assumes that zero-consumption could either be recorded when the consumer is genuine non-participant (structural zero), or when the zero consumption is the corner solution of a standard consumer demand problem (sampling zero).

Different from the zero-inflated count data model, the hurdle model proposed by Mullahy (1986) assumes that all zeros are sampling zeros. When applying to the consumption analysis, it assumes that individuals need to pass two stages before being observed with a positive level of consumption: a participation decision and a consumption decision. Furthermore, the hurdle models assume participation dominant. Thus, all the zero observations are assumed generated in the first stage (decide on whether to consume), and in the second stage, the consumption behavior was truncated at zero.

Shonkwiler and Shaw (1996) extended the Hurdle model by allowing zero observations in both the first and second stages. Thus, in Shonkwiler and Shaw's model "Double hurdle count-data model, there are two mechanisms generating zero observations: zero observations could either happen in the first stage by choosing not consume, or in the second stage by choosing consume zero frequency. In their research, Shonkwiler and Shaw applied the double hurdle count data model to analyze recreation demand, and they classified people into three categories: "user", "potential user", and "non-user". They define a "user" as a person who is currently consuming the product, a non-user as a person who has never consumed the product before, and likely will not consume the product in the future, and a "potential-user" as a person who has ever consumed the product before, but is not consuming the product in the given period. In Shonkwiler and Shaw's research, they also made connections between the zero-inflated count data model to the Double-hurdle count data model by laying out the probability mass function for both models, and they concluded that the zero-inflated count data model and the double-hurdle count data model are essentially the same. They both allow the zero observations generated from two separate processes, allowing the zero observations to be either structural or sampling zeros.

However, although those previous studies assumed that the zero observations might relate to two distinct sources (non-participants and potential consumers), they fail to differentiate between the two types of consumer segments in their research. Harris and Zhao (2007) specified an acquisition variable that multiplies participation intention and consumption frequency. In this manner, the acquisition variable will be zero for either non-participation or zero consumption. Participation intention is not observable from the acquisition variable. If we could observe participation intention, it would be a triple hurdle model. That is, only consumers who have positive intention would be observed to participate in the market. Conditional on the participation decision, consumers would further make the subsequent consumption and consumption intensity decisions.

This study contributes to the literature by proposing a triple hurdle count data model, which allows us to observe the consumers' participation intention, and explore the appropriate structurally different reasons explaining consumers' participation and consumption decisions. The results of the triple hurdle count data model will provide more detailed information that can be useful to better classify markets in three segments: non-participants, potential consumers and consumers.

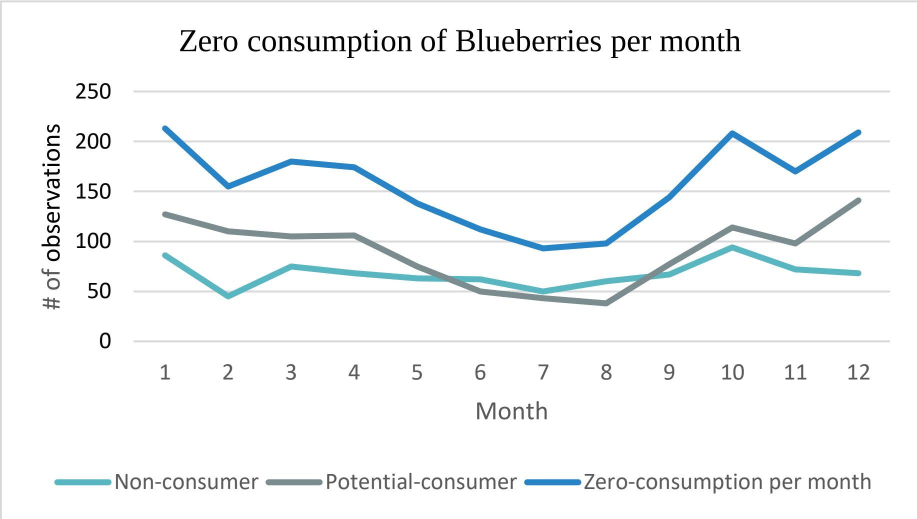
Background

The consumption of fresh produce is influenced by many different factors, which can be stable or unstable. For example, consumers might choose not to consume certain fruits or vegetables because of allergies, taste preferences, or diet constraints. These factors are considered stable factors, which cause consumers to virtually ignore that certain type of fresh produce in their decision making. These consumers would be expected to have non-positive market participation intention for specific item of produce, thus are considered non-consumers.

However, other consumers might be influenced by unstable reasons. One significant unstable factor for the consumption of fresh produce is seasonality. The consumption of fresh produce can change significantly in different seasons. This can be a result of decreased supply and availability leading to changes in prices and/or origin of producers at different times of the year. Even though some consumers might have positive participation intention for some fruits or vegetables, they may still choose not to consume during the off-season because of the high price. This is referred to as the corner solution of a standard consumption problem. These consumers influenced by seasonality (price) would easily change their consumption behaviors when the circumstances differ, thus are considered as potential consumers.

Taking the consumption of fresh blueberry as an example, the total observed zero-consumption per month is significantly higher in winter and lower in summer.

However, when we differentiate the observed zero-consumption into non-consumers (never purchase) and potential consumers (purchase before but not purchase last month), the number of non-consumers appears to be comparatively stable over the year, while the number of potential consumers changes significantly over the year.



Although both non-consumers and potential consumers report zero-consumption, it would be impropriate to treat all the zero observations the same for fresh produce consumption. The decision of market participation appears to be driven by a structurally different process than the subsequent consumption decision and consumption intensity decision. Analyzing the different factors influencing consumers' participation, consumption, and consumption intensity decisions can provide researchers, retailers, and producers better understandings of consumers' behaviors, thus help them develop effective and separate promotions strategies targeting non-participants, potential consumers, and current consumers.

Conceptual Framework

To develop the triple hurdle count data model, we first began by outlining the existing double-hurdle approach. Previous studies have theorized that the observed zero-consumption could be driven by two different mechanisms: non-consumers (who have non-positive market participation desire) and potential consumers (who have non-positive consumption intention given positive participation desire). Although these studies allow for the idea that factors influencing market participation could be different from the factors influencing consumers' consumption decision, they fail to observe the consumers' actual participation desire. They are also restrictive by assuming that it is the same mechanism that determines consumption intention and consumption frequency decisions, which might not always be true.

In the triple hurdle count data model, we relax the restrictions of double-hurdle approach, and extend the framework to differentiate three types of consumers, and allow three different mechanisms to generate the consumers' decisions on market participation, consumption intention and consumption intensity.

The full triple hurdle data model specification can be represented as:

R = R (consumers' characteristics, products' characteristics, seasonal effect)
D = D (consumers' characteristics, products' characteristics, seasonal effect) for participants
Y = Y (consumers' characteristics, products' characteristics, seasonal effect) for consumers

Where R is a binary indicator of whether the consumer has a positive desire to participate in the market, D is also a binary indicator of whether the consumer would have a positive consumption intention in the given time period given positive desire to participate, and Y is positive integers indicating consumption frequency/intensity.

Econometric Framework

In this section, we start by proposing the triple hurdle count data model with the three stages independent of each other, then we further allow the three stages to be correlated. Next, we outline the estimated strategy and discuss inference and interpretation of the results. We also report performance of the triple hurdle count data model compared to alternative model specifications of the double-hurdle count data model.

Triple hurdle count data model with independent stages

The triple hurdle count data model, a mixture of Poisson regression models, is an extension of the hurdle count data model proposed by Mullahy (1984). Mullahy's model included a market participation stage before the consumption intention and consumption intensity stages. Thus, the triple hurdle count data model involves three latent equations to indicate the three stages in succession, with the first two equations having binary outcomes indicating participation and consumption, and the third equation having positive count outcome indicating consumption intensity. This splits the observations into three regimes (non-participants, potential consumers and consumers) that relate to potentially three different sets of explanatory variables.

The model specification for the triple hurdle count data model is as follows:

Market participation stage

$$\Pr(R^*=r) = \frac{\exp(-\theta_1)*\theta_1{}^r}{r!} \quad r=0,1,2,3\ldots.$$

$$R = \begin{cases} 0 \;\; if \;\; R^* \leq 0 \\ 1 \;\; if \;\; R^* > 0 \end{cases}$$

Where R denotes the binary indicator of whether to participate or not (with R=0 for non-participants, and R=1 for participants). R is related to a latent variable R* via the mapping: R=1 for R*>0 and R=0 for R*≤0. The latent variable R* represents the propensity for market participation, specifically, we adopt the Poisson distribution[1]

[1] It is possible that R* could be a continuous variable and generated by other approaches, for example, R* could be possibly distributed with Normal distribution, then Pr (R*≤0) = Φ(-$x'\beta$) where Φ is the cumulative distribution function for the normal distribution. However, in order to derive the sample-

for R*. θ_1 is the parameter for the Poisson distribution, which can be parameterized as $\theta_1 = \exp(x'\beta)$, where x is a vector of covariates and β is a vector of unknown coefficients.

Consumption intention stage

Pr $(D^*$=d) = $\frac{\exp(-\theta_2)*\theta_2{}^d}{d!}$ d=0,1,2,3……

$$D = \begin{cases} 0 \ \ if \ \ D^* \leq 0 \\ 1 \ \ if \ \ D^* > 0 \end{cases}$$

Conditional on participation (R=1), consumers make the second decision on whether to consume during a specific time period. Let D denote a second binary indicator of whether to consume or not in the given period (with D=0 for non-consumption, and D=1 for positive consumption), where D is also related to a latent variable D* via the mapping: D=1 for D*>0 and D=0 for D*≤0. We also adopt the Poisson distribution for D*. θ_2 is the parameter for the Poisson distribution of D*, which can be parameterized as $\theta_2 = \exp(z'\alpha)$, where z is a vector of covariates that determine consumers' second choice and α is the corresponding unknown vector of parameters. Furthermore, there is no requirement that x=z.

Consumption intensity stage

Pr(Y*=y) = $\frac{\exp(-\theta_{3)}*\theta_3{}^y}{1-\exp(-\theta_3)}$ y=1,2,3,4…..

Conditional on consumption in the given period (D=1 and R=1), positive consumption frequency is observed, and the consumption intensity is represented by a latent variable Y* (Y*=1,2,3,…J) which is generated by a Poisson regression truncated at 0. θ_3 is the parameter for the Poisson distribution of Y*. θ_3 could be parametrized as $\theta_3 = \exp(w'\gamma)$, where w is a vector of covariates that determine consumers' consumption intensity, and γ is the corresponding unknown vector of parameters. In this stage, there is no requirement that w=z=x.

Accordingly, in order to observe a non-participant, we require that R=0; to observe a potential consumer, we require jointly that the individual is a participant (R=1) that chooses not consume in the given period (D=0); and to observed positive consumption, we require jointly that the individual is a participant (R=1), and that they choose to consume a positive intensity (D=1, and y*>0).

Under the assumption that the three stages are independent, the probability of an individual being a non-participant is:

Pr (R=0| x) = Pr(R*≤0|x) = exp(-θ_1)

The probability of an individual being a potential-consumer is:

Pr (D=0|x,z) = Pr(R*>0) * Pr(D*≤0) = (1-exp (-θ_1))*exp $(-\theta_2)$

And the probability of observing positive consumption intensity, y, is:

Pr(Y=y|x,z,w) = Pr(R*>0)*Pr(D*≤0) *Pr(Y*=y) = (1-exp (-θ_1))*(1-exp $(-\theta_2)$)

---

selected hurdle count data model with interdependence, we employ the Poisson regression for the latent variable R*.

$$\times \frac{\exp(-\theta_{3})*\theta_3{}^{y}}{1-\exp(-\theta_3)} \quad y=1,2,3,\ldots.$$

In this way, given the independence of the three stages, the probability of observing a non-participant is exp(-θ_1), the probability of observing a potential-consumer is (1-exp (-θ_1))*exp $(-\theta_2)$, and the probability of observing a positive consumption intensity is a combination of the three separate processes. Note that this specification differentiates zero observations into two different regimes coming from two different generating processes. The first process selects the individuals who have positive desire and the second process generates individuals who determines zero-consumption given positive participation desire

Once the full set of probabilities have been specified, for any given observation, i, the sample-selected hurdle count data model has the following likelihood function:

$$f(\mathrm{R}_i, \mathrm{D}_i, \mathrm{Y}_i, |\ \theta_{1,}\theta_{2,}\theta_3) =$$

$$[\exp(-\theta_1)]^{1[\mathrm{R}_i=0]} * \left((1-\exp(-\theta_1)) * \left(\begin{array}{c} [\exp(-\theta_2)]^{1[\mathrm{D}_i=0]} \\ \left[(1-\exp(-\theta_2)) * \frac{\exp(-\theta_3)*\theta_3{}^{y}}{1-\exp(-\theta_3)}\right]^{1[\mathrm{D}_i=1]} \end{array} \right) \right)^{1[\mathrm{R}_i=1]}$$

Where $\theta_1 = \exp(x'\beta)$, and β are the parameters on x in the first stage, $\theta_2 = \exp(z'\alpha)$, and α are the parameters on z in the second stage, and $\theta_3 = \exp(w'\gamma)$ with γ being the set of parameters on w in the third stage.

Triple- hurdle count data model with interdependence

The assumption that the three stages are not related is restrictive, it is quite plausible that the three stages are related. To accommodate that we now extend the model to have the three stages correlated, which requires the latent variables (D*, R*, Y*) follow a trivariate Poisson distribution. The full observability criteria of observing the three types of consumers are as follows:

A consumer is a non-participant if R=0, is a potential consumer if (R=1 and D=0) and is a positive consumer with a positive consumption level y if (R=1, D=1, and y*=y), which translates into the following expressions for the probabilities:

Non-participants: Pr(R=0|x) = Pr(R*=0|x)

Potential-consumers: Pr(D=0, R=1|x, z) = $\sum_{r=1}^{\infty} Pr(R^* = r,\ D^* = 0|x, z)$

and Positive consumption: Pr(Y=y|x,z,w)= $\sum_{d=1}^{\infty}\sum_{r=1}^{\infty} Pr(R^* = r, D^* = d, Y^* = \mathrm{y}|\mathrm{x, z, w})$
where r=0,1,2….; d=0,1,2,….; y=1,2,3….

Considering the trivariate Poisson distribution with two-way covariance structure (R^*, D^* Y^*) ~TP ($\theta_1, \theta_2, \theta_3, \theta_{12}, \theta_{13}, \theta_{23}$), which takes the form:

$R^* = Z_1 + Z_{12} + Z_{13}$

$D^* = Z_2 + Z_{12} + Z_{23}$

$Y^* = Z_3 + Z_{13} + Z_{23}$

Where Z_i ~Po(θ_i), i ∈ {1,2,3}, and Z_{ij} ~Po(θ_{ij}), i,j ∈ {1,2,3}, i<j. Then

R^* follows marginally a Poisson distribution with parameter $(\theta_1 + \theta_{12} + \theta_{13})$, D^* follows marginally a Poisson distribution with parameter $(\theta_2 + \theta_{12} + \theta_{23})$, and Y^* follows marginally a Poisson distribution with parameter $(\theta_3 + \theta_{13} + \theta_{23})$. $(R^*,\ D^*)$, $(R^*,\ Y^*)$, and $(D^*,\ Y^*)$ marginally follow the bivariate Poisson distributions as follows:

$(R^*,\ D^*) \sim$ BPoisson $(\theta_1 + \theta_{13}, \theta_2 + \theta_{23}, \theta_{12})$ with Cov$(R^*,\ D^*)=\ \theta_{12}$
$(R^*,\ Y^*) \sim$ BPoisson $(\theta_1 + \theta_{12}, \theta_3 + \theta_{23}, \theta_{13})$ with Cov$(R^*,\ Y^*) =\ \theta_{13}$
$(D^*,\ Y^*) \sim$ BPoisson $(\theta_2 + \theta_{12}, \theta_3 + \theta_{13}, \theta_{23})$ with Cov$(D^*,\ Y^*) =\ \theta_{23}$

Thus, given the general joint probability function of bivariate distribution for (X, Y) ~BP$(\theta_1, \theta_2, \theta_0)$, where θ_0 is the covariance parameter between X and Y.

$$P(X=x, Y=y) = \exp(-\theta_1 - \theta_2 - \theta_0)\ \frac{{\theta_1}^x}{x!}\ \frac{{\theta_2}^y}{y!}\ \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i!\ (\frac{\theta_0}{\theta_1 \theta_2})^i$$

(Johnson and Kotz, 1997)

And the trivariate Poisson distribution with two-way covariance structure $(R^*,\ D^*\ Y^*)$ ~TP $(\theta_1, \theta_2, \theta_3, \theta_{12}, \theta_{13}, \theta_{23}\ )$

$$\Pr(R^* = \mathrm{r}, D^* = \mathrm{d},\ Y^* = \mathrm{y}) = \exp(-\theta_1 - \theta_2 - \theta_3 - \theta_{12} - \theta_{13} - \theta_{23})$$
$$\times \sum_{(z_{12}, z_{13}, z_{23}) \in C} \{(r - z_{12} - z_{13})!\,(d - z_{12} - z_{23})!$$
$$\times (y - z_{13} - z_{23})!\ \ z_{12}!\, z_{13}!\, z_{23}!\}^{-1}$$
$$\times {\theta_1}^{r - z_{12} - z_{13}} {\theta_2}^{d - z_{12} - z_{23}} {\theta_3}^{y - z_{13} - z_{23}} {\theta_{12}}^{z_{12}} {\theta_{13}}^{z_{13}} {\theta_{23}}^{z_{23}}$$

Where the summation is over the set C$\in N^3$ defined as
C=$[(y_{12}, y_{13}, y_{23}) \in N^3\colon \{y_{12} + y_{13} \le x\} \cup\ \{y_{12} + y_{23} \le y\} \cup \{y_{13} + y_{23} \le z\} \neq \emptyset]$
(Karlis and Meligkotsidou , 2005)[2]

Under the assumption that the three stages are interdependent, the probability of an individual being a non-participant is

Pr(R=0|x) = Pr(R*=0|x) = exp (- $(\theta_1 + \theta_{12} + \theta_{13})$)

The probability of an individual being a potential-consumer is:

[2] In the case of the trivariate Poisson distribution with two-way covariance structure, the variance-covariance matrix of $(R^*,\ D^*\ Y^*)$ is as follows

$$\begin{pmatrix} \theta_1 + \theta_{12} + \theta_{13} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_2 + \theta_{12} + \theta_{23} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_3 + \theta_{13} + \theta_{23} \end{pmatrix}$$

Then the parameters of θ_{ij} , i,j=1,2,3, i≠j, have the straightforward interpretation of being the covariance between the each pair of the variables.

$\Pr(D=0, R=1|x, z) = \sum_{j=1}^{\infty} Pr\left(R^* = j,\ D^* = 0\right)$

$= \exp(-(\theta_2 + \theta_{12} + \theta_{23})) - \exp(-(\theta_1 + \theta_{13}) - (\theta_2 + \theta_{23}) - \theta_{12})$[3]

And the probability of an individual being observed with positive consumption intensity, y, is

$$\Pr(Y = y|x, z, w) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} Pr(R^* = j, D^* = k, Y^* = y)$$

$$= \exp(-(\theta_3 + \theta_{13} + \theta_{23}))$$
$$- \frac{\exp(-(\theta_1+\theta_{12})-(\theta_3+\theta_{23})-\theta_{13}) * (\theta_3+\theta_{23})^y}{y!}$$
$$- \frac{\exp(-(\theta_2+\theta_{12})-(\theta_3+\theta_{13})-\theta_{23}) * (\theta_3+\theta_{13})^y}{y!}$$
$$+ \frac{\exp(-\theta_1-\theta_2-\theta_3-\theta_{12}-\theta_{13}-\theta_{23}) * \theta_3{}^y}{y!}$$

Where y=1,2,3,…. [4]

Thus, under the assumption of interdependence of the three stages, the probability of observing a non-participant is exp (- $(\theta_1 + \theta_{12} + \theta_{13})$) and the probability of observing a potential-consumer is $\exp(-(\theta_2 + \theta_{12} + \theta_{23})) - \exp(-(\theta_1 + \theta_{13}) - (\theta_2 + \theta_{23}) - \theta_{12})$.

Considering the likelihood function, the parameters will be redefined as $(\theta_1, \theta_2, \theta_3, \theta_{12}, \theta_{13}, \theta_{23})$ in the case of interdependence, where $\theta_{12}, \theta_{13}, \theta_{23}$ are the correlation parameters between each pair of stages. A wald test of $\theta_{ij} = 0$ i,j∈{1,2,3}, i<j will be employed to test for the independence between each pair of stages. The likelihood function under interdependence is as follows

$f(R_i, D_i, Y_i, |\ \theta_1, \theta_2, \theta_3)$ =

---

[3] $\Pr(D=0, R=1) = \sum_{j=1}^{\infty} Pr\left(R^* = j,\ D^* = 0\right)$

$= \sum_{j=0}^{\infty} Pr(R^* = j,\ D^* = 0) - Pr(R^* = 0,\ D^* = 0)$

$= Pr(D^* = 0) - Pr(R^* = 0,\ D^* = 0)$

$= \exp(-(\theta_2 + \theta_{12} + \theta_{23})) - \exp(-(\theta_1 + \theta_{13}) - (\theta_2 + \theta_{23}) - \theta_{12})$

[4] In order to derive Pr(Y=y), we employ the marginal distribution of $Y^* \sim Po((\theta_3 + \theta_{13} + \theta_{23})$, and the marginal distribution of (R^*, Y^*) ~ BPoisson $(\theta_1 + \theta_{12}, \theta_3 + \theta_{23}, \theta_{13})$ and (D^*, Y^*) ~ BPoisson $(\theta_2 + \theta_{12}, \theta_3 + \theta_{13}, \theta_{23})$

$\Pr(Y = y) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} Pr(R^* = j, D^* = k, Y^* = y)$

$= \Pr(Y^* = y) - \Pr(Y^* = y, R^* = 0) - \Pr(Y^* = y, D^* = 0) + \Pr(Y^* = y, D^* = 0, R^* = 0)$

$= \exp(-(\theta_3 + \theta_{13} + \theta_{23})) - \frac{\exp(-(\theta_1+\theta_{12})-(\theta_3+\theta_{23})-\theta_{13}) * (\theta_3+\theta_{23})^y}{y!}$

$- \frac{\exp(-(\theta_2+\theta_{12})-(\theta_3+\theta_{13})-\theta_{23}) * (\theta_3+\theta_{13})^y}{y!}$

$+ \frac{\exp(-\theta_1-\theta_2-\theta_3-\theta_{12}-\theta_{13}-\theta_{23}) * \theta_3{}^y}{y!}$

$$[\exp\ (-\ (\theta_1 + \theta_{12} + \theta_{13}))]^{1[\mathrm{R}_i=0]}\ *$$

$$\left(\left(\begin{matrix}[\exp(-(\theta_2 + \theta_{12} + \theta_{23}))\ -\ \exp(-\ (\theta_1 + \theta_{13}) - (\theta_2 + \theta_{23}) - \theta_{12})]^{1[\mathrm{D}_i=0]} \\ \left[\exp(-(\theta_3 + \theta_{13} + \theta_{23})) - \frac{\exp(-(\theta_1 + \theta_{12}) - (\theta_3 + \theta_{23}) - \theta_{13}) * (\theta_3 + \theta_{23})^y}{y!}\ \frac{\exp(-(\theta_2 + \theta_{12}) - (\theta_3 + \theta_{13}) - \theta_{23}) * (\theta_3 + \theta_{13})^y}{y!} + \frac{\exp(-\theta_1 - \theta_2 - \theta_3 - \theta_{12} - \theta_{13} - \theta_{23}) * {\theta_3}^y}{y!}\right]^{1[\mathrm{D}_i=1]}\end{matrix}\right)\right)^{1[\mathrm{R}_i=1]}$$

Where $\theta_1 = \exp(x'\beta)$, and β are the parameters on x in the first stage, $\theta_2 = \exp(z'\alpha)$, and α are the parameters on z in the second stage, and $\theta_3 = \exp(w'\gamma)$ with γ being the set of parameters on w in the third stage.

Furthermore, from the likelihood 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 a potential consumer is expressed in Equation

Pr (Non-consumer) = Pr(R_i=0|x_i)= exp (- ($\theta_{1\mathrm{i}} + \theta_{12} + \theta_{13}$))

Pr(Potential-consumer) =Pr(D_i=0, R_i=1|x_i, z_i) =exp(-($\theta_{2\mathrm{i}} + \theta_{12} + \theta_{23}$)) − exp(− ($\theta_{1\mathrm{i}} + \theta_{13}$) −($\theta_{2\mathrm{i}} + \theta_{23}$)−$\theta_{12}$)

Marginal Effects and Interpreting Results

The overall effect of a given explanatory variable is determined by several different sets of marginal effects. For example, marginal effects of an explanatory variable can be determined on the probability of being "non-consumers" Pr(R=0), the probability of being a potential consumer, and on the probabilities for different levels of consumption Pr(Y=j). Calculating marginal effects for each stage of decisions allow for comparisons between non-consumers and potential-consumers (which has been lacking from previous models).

The marginal effect of a dummy variable is calculated as the difference between the probabilities given the dummy variable equals to 1 or 0. As for continuous variables, the probability expressions provided for each consumer category can be found from the numerical derivatives. Note that the explanatory variables in the three different stages might be not the same. Thus the explanatory variable of interest may appear in only one or two of x, z and w, or in all of them. For a continuous variable x_k the marginal effect on the participation intention is in Eq(), which only relates to the explanatory variables in x, and is given by:

$$ME_{\Pr(R=0)} = \frac{\partial \Pr(R=0)}{\partial x_k} = \exp\ (-\ (\theta_1 + \theta_{12} + \theta_{13}))*\ (-\theta_1)\ *\beta_k$$

To derive the marginal effects on the overall probabilities for the sample-selected hurdle count data model with interdependence, we need to partition the explanatory variables and the associated coefficients as follows given the possible existence for only one or two of x, z and w:

$$x' = (u', \tilde{x}'),\ \beta' = ({\beta_u}', \widetilde{\beta}')$$

$$z' = (u', \tilde{z}'),\ \alpha' = ({\alpha_u}', \widetilde{\alpha}')$$

and $w' = (u', \widetilde{w}'),\ \gamma' = ({\gamma_u}', \widetilde{\gamma}')$

where u represents the common variables that appear in all the x, z, and w, with associated coefficients as β_u, α_u and γ_u for the participation intention, consumption intention, and consumption frequency equations, respectively. $\tilde{x}$ denotes the distinctive variables that only appear in the participation stage, with $\tilde{\beta}$ as the associated coefficients; similarly, $\tilde{z}$ and $\widetilde{w}$ denote the variables that only appear in the consumption decision stage and consumption frequency stage, with α' and γ'as the associated coefficients, respectively.

In order to express the marginal effects for the entire model, the unique explanatory variables are expressed as $x^{*\prime} = (u', \tilde{x}', \tilde{z}', \widetilde{w}')$ and the associated coefficient vectors set for the three stages are expressed as $\beta^{*\prime} = (\beta_u{}', \widetilde{\beta}', 0', 0')$, $\alpha^{*\prime} = (\alpha_u{}', 0', \widetilde{\alpha'}, 0')$, and $\gamma^{*\prime} = (\gamma_u{}', 0', 0', \widetilde{\gamma'})$.

The marginal effect of the explanatory variable vector x^* on the consumption probability is in Eq(), which relates to the explanatory variables in both x and z:

$$ME_{\Pr(D=0|R=1;x,z)} = \exp(-(\theta_2 + \theta_{12} + \theta_{23}))*(-\theta_2)*\alpha^* - \exp(-(\theta_1 + \theta_2 + \theta_{12} + \theta_{23} + \theta_{13}))*[-\theta_1 * \beta^* - \theta_2 * \alpha^*]$$

The marginal effect of the explanatory variables on the positive level of consumption y (y=1,2,…) is as follows in Eq()

$$\begin{aligned} &ME_{\Pr(Y=y|R=1,D=1;x,z,w)} \\ &= \exp\big(\big(-(\theta_3 + \theta_{13} + \theta_{23})\big) * (-\theta_3) * \gamma^* \\ &- \frac{\exp(-(\theta_1 + \theta_{12}) - (\theta_3 + \theta_{23}) - \theta_{13}) * (\theta_3 + \theta_{23})^y * y * (\theta_3 + \theta_{23})^{y-1} * \theta_3 * \gamma^*}{y!} \\ &- \frac{(\theta_3 + \theta_{23})^y * \exp(-(\theta_1 + \theta_{12}) - (\theta_3 + \theta_{23}) - \theta_{13}) * (-\theta_1 * \beta' - \theta_3 * \gamma^*)}{y!} \\ &- \frac{\exp(-(\theta_2 + \theta_{12}) - (\theta_3 + \theta_{23}) - \theta_{23}) * (\theta_3 + \theta_{13})^y * y * (\theta_3 + \theta_{13})^{y-1} * \theta_3 * \gamma^*}{y!} \\ &- \frac{(\theta_3 + \theta_{13})^y * \exp(-(\theta_2 + \theta_{12}) - (\theta_3 + \theta_{13}) - \theta_{23}) * (-\theta_2 * \alpha' - \theta_3 * \gamma^*)}{y!} \\ &+ \frac{\exp(-\theta_1 - \theta_2 - \theta_3 - \theta_{12} - \theta_{13} - \theta_{23}) * y * \theta_3{}^{y-1}}{y!} \\ &+ \frac{\exp(-\theta_1 - \theta_2 - \theta_3 - \theta_{12} - \theta_{13} - \theta_{23}) * \theta_3{}^{y} * (-\theta_1 * \beta' - \theta_3 * \gamma^* - \theta_2 * \alpha')}{y!} \end{aligned}$$

The marginal effects for the triple hurdle count data model with no interdependence are calculated as above but with $\theta_{13} = \theta_{23} = \theta_{12} = 0$.

The standard errors of the marginal effects could be calculated by the Delta Method or simulated asymptotic sampling techniques. Considering the complexity of the marginal effects, the sampling technique is used in this case. To be more specific, we randomly draw θ (where θ is the parameters in the Sample-selected Zero-inflated model) from MVN ($\hat{\theta}$, $\widehat{var}[\theta]$) 10,000 times, and for each draw we calculate the marginal effects based on equation (7) to equation (10), and then calculate the standard errors. These empirical standard deviations of the simulated marginal effects are the valid asymptotic estimates of the true marginal effects' standard errors.

Comparing Triple-hurdle Count Data model and the Double-Hurdle Models

One goal of this study is to discuss the difference in insights gained when differentiating potential consumers from non-consumers, and employing the triple hurdle model instead of the double-hurdle approach.

The double-hurdle alternative is similar to the model proposed by Shonkwiler and Shaw (1996), which assumes that the factors influencing consumption frequency decision are the same as the factors influencing the consumption intention.

In Shonkwiler and Shaw's model, the probability of observing an non-participant is:

$$\text{Prob}\,(D = 0|x) = \Pr(D^*=0|x) = \exp\,(-\,(\theta_1 + \theta_{12}))$$

and the probability of observing an individual who is a potential consumer is:

$$\Pr(D = 1,\ Y = 0|x, w) = \sum_{j=1}^{\infty} Pr\left(D^* = j,\ \ Y^* = 0\right)$$

$$= \exp(-(\theta_2 + \theta_{12})) - \exp(-\ \theta_1 - \theta_2 - \theta_{12})$$

The probability of observing positive consumption frequency is:

$$\Pr(D = 1, Y = y|x, w) = \sum_{j=1}^{\infty} Pr\left(D^* = j,\ \ Y^* = y\right)$$

$$= \frac{\exp(-\theta_2 - \theta_{12}))*((\theta_2 + \theta_{12})^y}{y!} - \frac{\exp(-\theta_1 - \theta_2 - \theta_{12}))*\theta_2{}^y}{y!}$$

Where $\theta_1 = \exp(x'\beta)$, β are the parameters on x in the market participation stage, and $\theta_2 = \exp(w'\gamma)$ with γ are the set of parameters on w in the consumption stage. In this case, there is an assumption that the probability of consumption intention and the probability of consumption frequency are related to the same explanatory factors (w) in similar ways.

Although this double-hurdle approach is non-nested with the triple-hurdle model, a generalized likelihood ratio (LR) statistic could be used, with degrees of freedom being given by the number of additional parameters estimated in the more general model. Additionally, in such a non-nested situation, information based model selection criteria such as AIC and BIC are appropriate for choosing between alternative models. These are given by AIC=-2ln(θ)+k, and BIC=-2ln(θ)+(lnN)*k, where k is the total number of parameters estimated and ln(θ) is the maximized log-likelihood function. The preferred model is that with smallest value.

Variables and Data

Data Set

In this paper, the sample selected hurdle count data model was fit using an online survey about consumers' consumption behavior and preferences forfresh blueberries. The survey was conducted with a random panel of respondents starting in September 2010 and lasted for 12 months, with approximately 350 participants recruited on a monthly basis. The target respondents are primary grocery shoppers in the Eastern States of the United States. Respondents answered a series of questions on how often and why (or why not) they purchase fresh blueberries.

Here, for modeling purposes, non-consumers and potential consumers were distinguished using survey design. Respondents were first asked whether they had ever purchased fresh blueberries and then asked whether they had purchased fresh blueberries in the past month. For those respondents who had purchased in the prior month, they were further asked to indicate how many times they purchased fresh blueberries in the past month. Purchase information was only asked for the past month to ensure accuracy of the data as it is difficult for people to recall purchases more than one month ago. By asking respondents whether they have purchased fresh blueberries before and whether they had purchased fresh blueberries last month in two questions, "non-consumers" and "potential consumers" can be differentiated according to the definition of the three types of consumers given above.

The questionnaire consisted of four parts. The first part included questions concerning consumers' frequency of consuming fresh blueberries. The second part focused on the reasons for consuming (or not consuming) fresh blueberries.The third part of the questionnaire focused on the consumers' awareness of health benefits of eating fresh blueberries and the last part includes socio-demographic variables, such as gender, age, educational level, employment status, family size, socioeconomic status, etc.

Variables

The key dependent variables are ANYPARTICIPATE, ANYCONSUME, and PURCHASEFREQ. The three variables are derived from the following questions, respectively: "Have you ever purchased fresh blueberries?" (where a binary "Yes/No" answer is required); "Have you purchased fresh blueberries in the LAST MONTH?"; and "In the last month, approximately how many times did you purchase fresh blueberries?". For the final question, respondents selected an answer from the categories 1 or 2 times; 3 or 4 times; 5 or 6 times; more than 6 times; and did not purchase (though they were not shown the question if they indicated no purchase, this was used as a consistency check). Thus, the use of these three dependent variables corresponds to examining a three-step decision made with respect to participation and consumption. As noted earlier, by asking the three questions in sequence, it allows identification of non-participants, potential consumers and consumers.

The covariates employed in the model are shown in Table 1, together with their means. In addition, descriptions of each variable, and whether the variable was employed in the participation stage (P), consumption intention stage (C) and consumption frequency stage (F) are also indicated in this table.

The individual characteristics include gender, education level, race, age and awareness of health benefits of blueberries. In this dataset, only 35.7% of the respondents are male, which was expected as only primary grocery shoppers for the household completed the survey. Education level is controlled for with a binary dummy variable indicating whether the respondents have a four-year college degree or not (40.6% of the participants have earned at least undergraduate degree). Consumers' awareness of health benefits of blueberries was controlled for by using a dummy variable, which allows us to test the effectiveness of knowledge of health benefits on consumption decisions. In this dataset, 51.9% of participants indicated that they were aware of specific health benefits of blueberries.

Together with individual characteristics, household characteristics are also controlled, including the number of people in the household, whether there are children living in the household, household income, and household food budget per week. Both household income and household food budget per week are included based on previous research that indicates income level works as a social class proxy for consumption participation, and food budget works more closely influencing consumption frequency.

The last set of variables is a ranking of how important the respondent finds different attributes of blueberries, including price and taste. Since the consumption of blueberries changes significantly over the year, we also include seasonal dummy variables.

Table 1. Variable Descriptions

Variables	Description		Value	Model
Male	Percent of sample male		35.7%	P/C/F
College	Percent of sample with at least four-year college degree		40.6%	P/C/F
Age	Age in years (continuous in analysis)	18-24 years	13.9%	P/C/F
		25-29 years	11.1%	
		30-34 years	10.8%	
		35-39 years	5.6%	
		40-44 years	8.2%	
		45-49 years	8.7%	
		50-54 years	11.5%	
		55-59 years	9.4%	
		60-64 years	9.2%	
		65 or above	11.6%	
Income	Estimated Household income	$14,999 or less	11.2%	P/C/F
		$15,000-$24,999	13.5%	
		$25,000-$34,999	14.7%	
		$35,000-$49,999	17.4%	
		$50,000-$74,999	21.0%	
		$75,000-$99,999	11.6%	
		$100,000 or above	10.6%	
Hispanic	Percent Hispanic		4.0%	P/C/F
Black	Percent Black/African American		10.1%	P/C/F
Asian	Percent Asian		3.2%	P/C/F
White	Percent White		82.3%	P/C/F
Otherrace	Percent other races		0.4%	P/C/F
Health_Aware	Percent who are aware of health benefits of blueberry		51.9%	P/C/F
Budget	Food budget per week	Less than $49	11.5%	P/C/F
		$50-99	36.1%	
		$100-149	28.9%	
		$150-199	13.4%	

		$200-$249	5.9%	
		$250+	4.2%	
WithChild	Percent who indicate have children live in the household		34.5%	P/C/F
Peop_number	People number in the house(continouse in the analysis)	1-2	55.0%	P/C/F
		3-4	34.6%	
		5-6	8.8%	
		7-8	1.3%	
		9 or above	0.3%	
Taste	Percent who indicate taste as a reason for eating/not eating blueberries		55.2%	C/F
Price	Percent who indicate price as a reason for eating/not eating blueberries		55.0%	C/F
Spring	Season Dummy for Spring		23.4%	C/F
Summer	Season Dummy for Summer		23.9%	C/F
Fall	Season Dummy for Fall		27.2%	C/F

Results

Summary statistics from both the Double Hurdle (DH) model and the Triple Hurdle (TH) model are presented in Table 2. The DH model is conditional only on X and W, which assumes that the same of explanatory factors influence consumption intention and consumption frequency. The TH model is conditional on X, Z, and W, which allows potential consumers to be differentiated from non-consumers. The likelihood ratio statistics from the models of fresh blueberry consumption clearly reject the DH model. Furthermore, both the AIC and BIC information criteria clearly suggest the superiority of the TH model over the DH. We therefore focus discussion on results from the triple hurdle count data model, but include discussion on the insights gained from using the triple hurdle apporach compared to the double hurdle approach.

Table 2. Fresh blueberry consumption: summary statistics from double hurdle approach and triple hurdle model

	Fresh mushroom Consumption	
	DH	TH
N	4038	4038
K	39	60
Loglikelihood	-5237.126	-5064.679
AIC	10513	**10189**
BIC	10798	**10627**
LR:TH versus DH	344.894***(df=21)	

(**) and (*) indicate statistical significance at 5% and 10% levels respectively. Preferred model with regard to each information criteria is indicated with bold.

Triple-Hurdle Count Data Model Results

Triple-hurdle count data model estimation results are displayed in Table 3 together with the estimation results from the Double-hurdle approach. Coefficient estimates for factors associated with the probability of having a positive market participation intention (Stage 1) are displayed in Column 1; coefficient estimates for predicting the probability of being observed with positive consumption intensity are displayed in Column 2; and coefficient estimates for factors associated with positive consumption levels given positive market participation intention and consumption intention are shown in Column 3. The marginal effects are shown in Table 4.

Table 3. Fresh blueberry consumption: regression results

	Triple-Hurdel Count Data Model			Double-Hurdle Count Data Model	
Explanatory Variables	Stage 1 Participatoin Intention	Stage 2 Consumption Intention	Stage 3 Consumption Intensity	Stage 1&Stage 2 Participation Intention	Stage 3 Consumption Intensity
Female	0.064	0.026	**0.430**	-0.063	**0.074**
	(0.057)	(0.028)	**(0.134)*****	(0.042)	**(0.030)*****
Caucasian	**0.340**	**-0.178**	-0.082	**0.212**	**-0.132**
	(0.161)**	**(0.083)****	(0.307)	**(0.104)****	**(0.068)****
Hispanic	**-1.062**	0.171	-0.681	**-0.241**	-0.112
	(0.304)**	(0.103)	(0.450)	**(0.118)****	(0.080)
Asian	-0.071	0.254	-0.700	0.080	0.115
	(0.194)	(0.133)	(0.504)	(0.150)	(0.088)
Black	0.141	0.033	-0.116	**-0.219**	0.104
	(0.166)	(0.089)	(0.341)	**(0.113)***	(0.073)
College	0.003	**-0.091**	0.165	**0.118**	-0.024
	(0.061)	**(0.026)***	(0.138)	**(0.043)*****	(0.030)
Health_Aware	**0.812**	0.021	**1.234**	**0.734**	**0.210**
	(0.065)***	(0.031)	**(0.282)*****	**(0.042)*****	**(0.031)*****
Age	**-0.055**	**-0.032**	**-0.123**	**-0.017**	**-0.036**
	(0.010)**	**(0.005)****	**(0.028)****	**(0.007)*****	**(0.005)*****
Income	0.020	0.004	0.035	**0.029**	0.008
	(0.011)	(0.005)	(0.032)	**(0.009)*****	(0.007)
Food budget	**0.126**	**0.090**	**0.198**	**0.050**	**0.104**
	(0.019)***	**(0.011)*****	**(0.028)*****	**(0.015)*****	**(0.009)*****
Peop_number	**-0.167**	-0.053	-0.110	-0.028	-0.010
	(0.051)**	(0.029)	(0.097)	(0.038)	(0.025)
With_child	**0.241**	-0.059	0.121	**0.142**	0.001
	(0.074)***	(0.038)	(0.151)	**(0.058)*****	(0.038)
Vegetarian	**0.503**	**-0.231**	-0.283	**0.244**	-0.019
	(0.181)***	**(0.042)****	(0.354)	**(0.125)***	(0.067)
Spring	-0.124	**-0.093**	-0.263	**-0.103**	0.055
	(0.085)	**(0.035)****	(0.255)	**(0.057)***	(0.045)
Summer	**0.399**	**0.422**	**0.583**	0.006	**0.459**

	(0.082)* **	**(0.039)*** **	**(0.187)*** **	(0.056)	**(0.040)*** **
Fall	**0.215**	**0.106**	0.026	-0.076	**0.236**
	(0.074)* **	**(0.036)** **	(0.231)	(0.054)	**(0.041)*** **
Price	-0.067	**-0.443**	**-0.190**	**-0.178**	**-0.293**
	(0.059)	**(0.030)** **	**(0.126)** **	**(0.040)*** **	**(0.028)*** **
Taste	-0.062	**0.589**	**1.155**	**0.131**	**0.589**
	(0.051)	**(0.023)*** **	**(0.534)*** **	**(0.041)*** **	**(0.036)*** **
Constant	**0.694**	**0.694**	**-3.029**	-0.032	-0.159
	(0.233)	**(0.107)*** **	**(0.753)** **	(0.143)	(0.105)
Rho(1,2)		**-0.794(0.055)*** **			
Rho(1,3)		**1.092(0.044)*** **		**-0.112(0.022)*** **	
Rho(2,3)		**0.140(0.033)*** **			
# of obs		4038		4038	
Log-Likelihood		5064.679		5237.126	

Of the individual characteristics, age is significantly negatively correlated with market participation intention, consumption intention, and also consumption frequency which indicates that younger people are more likely to consume fresh blueberries and also more likely to consume them more frequently. Similarly, females were more likely than males to purchase fresh blueberries, and also were more likely to purchase them with higher frequency. The variables representing race and ethnicity indicate that Caucasians are more likely to try the fresh blueberries than other races, yet there was no significant evidence indicating correlation with higher consumption frequency. What is more, results also indicate that Hispanics were more likely to be observed as non-consumers than other race (less likely to participate in the fresh blueberry market). However, it's not significantly related to consumption intention and consumption frequency. Interestingly, results suggest no relationship between household income level and fresh blueberry consumption in any of the three stages. However, weekly food budget is found to be significantly positively correlated with all of the three stages. People with higher weekly food budget would be more likely to participate in the fresh blueberry market, and also be more likely to have positive consumption intention with higher consumption frequency. The variable for education was not significantly related to market participation nor consumption frequency, yet it was found negatively related to the consumption intention. Considering consumers' food habits, vegetarian was not found to be significantly positive correlated with market participation, but negatively correlated with consumption intention. The estimated results indicated that consumers' awareness of health benefits of blueberries would significantly influence the decision of market participation and consumption frequency, but not consumption intention.

When looking at the household characteristics, consumers who indicate they have children living in the household are more likely to participate in the fresh blueberry market, yet there is no relationship with consumption intention or frequency. The estimated coefficients of number of people living in the household are significantly negatively correlated with market participation, but not significantly related to consumption intention and consumption frequency, indicating that households with larger family size would be less likely to participate in the market.

Participants that indicated taste and price are important when purchasing blueberries were significantly more likely to have positive consumption intention and higher consumption frequency. However, it was not related to the decision to participate in the market.

As expected, due to seasonality (the domestic blueberry season runs from April to late September), results indicate that consumers are more likely to purchase fresh blueberries and purchase at a higher consumption frequency in Summer compared to Winter. In Fall, although consumers are more likely to participate in the market, there is no significant effect detected on consumption frequency. In contrast, consumers are less likely to purchase and consume fresh blueberries in Spring compared to Winter.

Marginal effects of the Triple Hurdle Count Data Model

As previously mentioned, one of the advantages of the Triple Hurdle Count Data model is its capability to distinguish potential consumers from non-consumers, and explore the different generating processes of the three different types of consumers. This is most easily demonstrated by examining the marginal effects of different variables. One example in the case of the fresh blueberry consumption is the impact of the variable representing if a respondent feels price is an important factor in choosing blueberries (henceforth referred to as the price effect). When examining the price effect in the Triple Hurdle Count Data model, we see that the dominant effect is on the probability to be a potential consumer (by 0.178), and that there is no relationship between this variable and the probability to be a non-participant. Thus, we conclude that when price is identified as an important factor, the likelihood to be a potential consumer is higher, while the likelihood to be a non-consumer is unaffected. This is as expected as a high price might stop someone interested in purchasing from making that purchase decision, however a non-consumer is expected to be in their category because of more permanent reasons (such as allergies). Consumption frequency is ?? When we examine this variable in the double hurdle model.A similar effect is found for the taste factor of blueberries

Another example is the consumers' awareness of health benefits. From the marginal effect table, we see that consumers' awareness of health benefits is only significantly correlated with consumers' participation decision (by 0.18). This implies that being aware of the benefits of blueberries influences the likelihood to try blueberries, as well as the likelihood to consume more frequently, but does not impact the likelihood to be a potential consumer (once a consumer decides to participate, they are as likely to be a participant or not regardless of their awareness of health benefits, but if they do participate and consume, they are likely to consume more often.

Table 4. Marginal Effects for Triple Hurdle Count Data Model

Explanatory Variables	Pr（R=0）	Pr(D=0,R=1)	Pr(Y=1)	Pr（Y=2）	Pr（Y=3）	Pr (Y=4)

Female	-0.014	-0.009	**-0.030***	**-0.039***	**-0.041***	**0.146***
	(0.013)	(0.012)	**(0.014)**	**(0.013)**	**(0.013)**	**(0.015)**
Caucasian	**-0.075**	**0.083***	-0.020	-0.004	0.005	0.007
	(0.036)	**(0.035)**	(0.036)	(0.031)	(0.030)	(0.034)
Hispanic	**0.234***	**-0.104***	0.045	**0.070***	0.064	-0.328
	(0.064)	**(0.044)**	(0.046)	**(0.042)**	(0.043)	(0.047)
Asian	0.016	**-0.106***	**0.127**	**0.079***	0.066	-0.201
	(0.044)	**(0.054)**	**(0.056)**	**(0.047)**	(0.048)	(0.052)
Black	-0.031	-0.009	0.026	0.012	0.009	-0.009
	(0.037)	(0.037)	(0.040)	(0.034)	(0.033)	(0.038)
College	-0.001	**0.037***	**-0.039***	-0.021	-0.016	**0.045***
	(0.014)	**(0.011)**	**(0.014)**	(0.014)	(0.013)	**(0.015)**
Health_Aware	**-0.180***	0.018	**-0.063***	**-0.112***	**-0.116***	**0.492***
	(0.016)	(0.013)	**(0.022)**	**(0.023)**	**(0.024)**	**(0.027)**
Age	**0.012***	**0.011***	0.000	**0.010***	**0.012***	**-0.048***
	(0.002)	**(0.002)**	(0.002)	**(0.002)**	**(0.003)**	**(0.004)**
Income	-0.004	-0.001	-0.001	-0.003	-0.003	**0.014***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	**(0.003)**
Food budget	**-0.028***	**-0.033***	**0.012***	**0.013**	**-0.019***	**0.087***
	(0.004)	**(0.005)**	**(0.005)**	**(0.004)**	**(0.004)**	**(0.005)**
Peop_number	**0.037***	0.016	-0.013	0.007	0.010	**-0.061***
	(0.011)	(0.012)	(0.012)	(0.0102)	(0.010)	**(0.011)**
With_child	**-0.053***	**0.032***	-0.013	-0.014	-0.012	**0.064***
	(0.017)	**(0.016)**	(0.018)	(0.015)	(0.015)	**(0.017)**
Vegetarian	**-0.111***	**0.110***	-0.005	0.013	0.025	-0.043
	(0.041)	**(0.022)**	(0.035)	(0.034)	(0.034)	(0.039)
Spring	0.027	**0.034***	-0.008	0.018	0.024	**-0.103***
	(0.019)	**(0.015)**	(0.025)	(0.024)	(0.026)	**(0.028)**
Summer	**-0.088***	**-0.159***	**0.076***	-0.031	**-0.055**	**0.278***
	(0.017)	**(0.017)**	**(0.026)**	(0.024)	**(0.025)**	**(0.028)**
Fall	**-0.047***	**-0.036***	0.036	0.003	-0.004	**0.051**
	(0.017)	**(0.015)**	(0.024)	(0.023)	(0.024)	**(0.026)**
Price	0.015	**0.178***	**-0.101***	-0.009	0.016	**-0.107***
	(0.013)	**(0.012)**	**(0.014)**	(0.013)	(0.013)	**(0.015)**
Taste	0.014	**-0.241***	0.047	-0.068	**-0.103***	**0.385***
	(0.012)	**(0.012)**	(0.039)	(0.041)	**(0.040)**	**(0.045)**

Insights gained using the Triple-Hurdle Approach

A key benefit of the triple-hurdle model compared to the double-hurdle approach is its capability to isolate factors associated with the market participation decision from those correlated with the consumption intention. As mentioned in the model section, although the extended double-hurdle approach can allow two separate processes generating zero-consumption, it stills imposes the restriction that the underlying structural relationship with consumption intention is the same for each factor.

The marginal effects for the probability of observing non-consumers and potential-consumers from the two alternative models are compared in Table 5, which emphasizes the advantages of using the more flexible approach.

First the triple hurdle count data model found no significant relationships between the blueberries' price and taste and the market participation decision, yet it only estimates significant effects of the attributes on the consumption intention decision. When compared with the double-hurdle approach, the double-hurdle model found that both the price and taste are significantly influenced consumers market participation and consumption intention decisions, however, the magnitude of the estimated effects on consumption intention from the double-hurdle approach is much smaller than those found from the triple-hurdle model.

Second, for the race variables, the triple-hurdle approach find the race variables are significant for both market participation and consumption intention decisions, however, the double hurdle model only illicit significant effects on market participation decision.

As for the seasonal variables, the triple-hurdle model found significant seasonal effects for both market participation and consumption intention decision, but the double-hurdle model indicates no significant correlation between different seasons and market participation decision.

In summary, from the comparison, we found that the triple-hurdle model introduces more detailed information concerning the inferences about market participation and consumption decisions than the more restrictive alternative approach. The added information allows us to distinguish between factors associated primarily with market participation from those primarily associated with consumption decisions, and factors associated with both decisions.

Table 5. Comparison of the marginal effects for Triple Hurdle Count Data Model and Double Hurdle Count Model

	Pr(Non-consumer)		Pr(Potential-consumer)		Pr(Y=1)	
	Triple-hurdle	Double-hurdle	Triple-hurdle	Double-hurdle	Triple-hurdle	Double-hurdle
Female	-0.014	0.018	-0.009	**-0.025*** **	**-0.030*** **	**-0.010***
	(0.013)	(0.012)	(0.012)	**(0.009)**	**(0.014)**	**(0.006)**
Caucasian	**-0.075** **	**-0.061**	**0.083*** **	0.054	-0.020	**0.032*** **
	(0.036)	**(0.029)*** **	**(0.035)**	(0.021)	(0.036)	**(0.015)**
Hispanic	**0.234*** **	**0.069*** **	**-0.104*** **	0.010	0.045	**-0.035*** **
	(0.064)	**(0.034)**	**(0.044)**	(0.025)	(0.046)	**(0.017)**
Asian	0.016	-0.023	**-0.106***	-0.024	**0.127** **	0.010
	(0.044)	(0.043)	**(0.054)**	(0.028)	**(0.056)**	(0.022)
Black	-0.031	**0.063***	-0.009	**-0.047*** **	0.026	**-0.033*** **
	(0.037)	**(0.032)**	(0.037)	**(0.023)**	(0.040)	**(0.017)**
College	-0.001	**-0.034*** **	**0.037*** **	**0.016***	**-0.039*** **	**0.018*** **
	(0.014)	**(0.012)**	**(0.011)**	**(0.009)**	**(0.014)**	**(0.007)**

Health_Aware	**-0.180*** **	**-0.210*** **	0.018	0.005	**-0.063*** **	**0.106*** **
	(0.016)	**(0.010)**	(0.013)	(0.008)	**(0.022)**	**(0.005)**
Age	**0.012*** **	**0.005*** **	**0.011*** **	**0.008*** **	0.000	**-0.002** **
	(0.002)	**(0.002)**	**(0.002)**	**(0.001)**	(0.002)	**（0.001）**
Income	-0.004	**0.008*** **	-0.001	0.000	-0.001	**0.004*** **
	(0.002)	**(0.003)**	(0.002)	(0.002)	(0.003)	**(0.001)**
Food budget	**-0.028*** **	**-0.014*** **	**-0.033*** **	**-0.024*** **	**0.012*** **	**0.007*** **
	(0.004)	**(0.004)**	**(0.005)**	**(0.003)**	**(0.005)**	**（0.002）**
Peop_number	**0.037*** **	0.008	0.016	0.001	-0.013	-0.004
	(0.011)	(0.011)	(0.012)	(0.008)	(0.012)	(0.006)
With_child	**-0.053*** **	**-0.041*** **	**0.032* **	0.011	-0.013	**0.021*** **
	(0.017)	**(0.017)**	**(0.016)**	(0.012)	(0.018)	**(0.009)**
Vegetarian	**-0.111*** **	**-0.070* **	**0.110*** **	0.026	-0.005	0.036
	(0.041)	**(0.036)**	**(0.022)**	(0.021)	(0.035)	(0.019)
Spring	0.027	**0.030* **	**0.034*** **	**-0.024* **	-0.008	**-0.015* **
	(0.019)	**(0.016)**	**(0.015)**	**(0.013)**	(0.025)	**(0.008)**
Summer	**-0.088*** **	-0.002	**-0.159*** **	**-0.125*** **	**0.076*** **	-0.003
	(0.017)	(0.016)	**(0.017)**	**(0.012)**	**(0.026)**	(0.009)
Fall	**-0.047*** **	0.022	**-0.036*** **	**-0.071*** **	0.036	**-0.013* **
	(0.017)	(0.015)	**(0.015)**	**(0.012)**	(0.024)	**(0.008)**
Price	0.015	**0.051*** **	**0.178*** **	**0.065*** **	**-0.101*** **	**-0.024*** **
	(0.013)	**(0.011)**	**(0.012)**	**(0.009)**	**(0.014)**	**(0.006)**
Taste	0.014	**-0.037*** **	**-0.241*** **	**-0.149*** **	0.047	**0.014** **
	(0.012)	**(0.012)**	**(0.012)**	**(0.009)**	(0.039)	**(0.006)**

Conclusion

We argue that when consumers make consumption decisions, they need to make the following three decisions: whether to participate in the market, whether to consume during the current time period, and how much to consume. We therefore developed a triple hurdle model that accounts for all three stages. This differs from previous models by capturing the different generating processes of non-consumers from potential consumers. This approach facilitates improved inference because it accounts for the fact that market participation might be driven by a different structural process than consumption decisions. This triple hurdle approach should be useful in many other applications where consumers face similar decisions.

To compare to models commonly used to examine participation and consumption decisions in the market, we used the double hurdle approach. The likelihood ratio test, as well as model selection criteria (AIC and BIC) find that the triple-hurdle model is preferred statistically.

To demonstrate the differences in these models, we applied the double hurdle and triple hurdle models to the consumption of fresh blueberries. The application to the consumption of fresh blueberries highlights strong relationship between consumers' knowledge and awareness of the health benefits of blueberries and market participation and consumption frequency. The results suggest that the advertisements and claims of the health and nutrition benefits of fresh produce would be significantly important if policy makers intent to promote the fresh produce consumption.

These results show that there are different factors influencing non-consumers and potential consumers, thus emphasizing the contribution of using the triple hurdle count data model. The reasons behind non-consumers are mostly stable demographic variables like ethnicity, age, income level, family characteristics and consumers' perceptions towards fresh blueberries, which will not change quickly. The reasons behind potential consumers are more related to economic reasons like food expenditure. What is more, we also find that once consumers make their determination to participate, product characteristics like taste and price are important factors influencing consumption decisions. In this case, improvement in taste in the product will likely lead to increased consumption quantity, but not an increased quantity of consumers in the market.

Reference

Jiang, Y., House, L., Tejera, C., & Percival, S. S. (2015, January). Consumption of Mushrooms: A double-hurdle Approach. In *2015 Annual Meeting, January 31-February 3, 2015, Atlanta, Georgia* (No. 196902). Southern Agricultural Economics Association.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). *Discrete multivariate distributions* (Vol. 165). New York: Wiley.

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.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1-14.

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

Shonkwiler, J. S., & Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics*, 210-219.