Chinese Preference for Online Grocery Shopping: Shopping for Convenience, Quality or Price?

Junhong Chen
Department of Agricultural Economics, Purdue University
chen1422@purdue.edu

Qiujie Zheng
College of Business and Public Policy, University of Alaska Anchorage
qzheng3@alaska.edu

Robin W. Zhang
Department of Economics, Cornell University
rwz8@cornell.edu

Holly Wang
Department of Agricultural Economics, Purdue University
wanghong@purdue.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

Please do not cite the paper without permission of the authors.

Chinese Preference for Online Grocery Shopping: Shopping for Convenience, Quality or Price?

Introduction

Online shopping has become increasingly popular in recent years in China. The accelerated use of smartphones and availability of online payment platforms have made shopping online an easy and convenient experience. According to 2015 China Internet Network Information Centers (CNNIC) annual report, the number of online shoppers reached 413 million by the end of 2015, about one-third of the entire population, which increased by 14.3% from 2014. In the early 2000s online shopping was not fully accepted by Chinese consumers due to the lack of computer access, credit card systems and online payment platforms (Bin et al, 2003). With the emergence of affordable mobile devices and many online payment and escrow service tools such as AliPay (launched by the e-commence tycoon Alibaba) and Wechat Pay (launched by the giant social media app Wechat developed by Tencent), online shopping has quickly become popular across the country, reaching a total transaction value of 620 billion USD in 2015. The Chinese e-commence retailing is represented by two leading companies, Alibaba and JD.com Inc., which account for more than 90% of market shares (CNNIC, 2016). Alibaba's Taobao alone reaches 485 billion USD in its 2016 fiscal year, more than that of Amazon (107 billion USD) and eBay (8.6 billion USD) combined in that year.

Consumers purchase various types of products and services online. The top five most popular categories purchased online in 2015 are clothes and shoes, articles of daily use, books and audio and video disks, computer and digital products and home appliances. If we count the top five categories purchased through mobile e-commerce, foods and supplements replace the books and audio and video disks and rank within the top five categories (CNNIC, 2016). While shopping for non-food products online has gained more popularity among consumers, shopping for food especially fresh (i.e. perishable) food online is still at an early stage in China, largely due to quality and safety concerns. Nowadays, people switch from the traditional in-store shopping to online shopping in order to save time and cost, especially for those living in the big cities. While online shopping provides the advantages of convenience and lower price, the touch-and-feel experience to select food products is not available in this form of shopping (Pechtl, 2003; Ramus and Asger Nielsen, 2005; Chu et al, 2010; Gong et al, 2013;). Thus, the information asymmetry and lacking of trust especially against the background of food safety problems in Chinese market are critical issues for online grocery shopping.

The online shopping environment is fundamentally different from that of a conventional physical retail environment and thus consumer's purchase behavior is different between these two formats (Shankar et al, 2003; Liu et al, 2008). A key difference between online shopping and conventional retail store is the ability of online consumers to obtain more information about both price and non-price attributes (Alba et al, 1997; Degeratu et al, 2000). Choosing to shop online, a customer is expecting a more competitive price resulting from a reduction in operational costs and internalizing activities traditionally performed by intermediaries (Anckar et al, 2002; Pechtl, 2003). Meanwhile, the increased availability of comparative price information online may make consumers more price-sensitive (Degeratu et al, 2000). Online shopping is convenient in that there are no travel involved, no shopping hour restrictions and the products can be arranged to get shipped and delivered to a destination directly, with a fee (Huang and Oppewal, 2006; Chu et al, 2010). However, consumers loose the touch and feel experience possibility and need to rely on the information provided by e-vendors to make purchase decisions. Consumers learn the product attributes through the product brand, descriptions, pictures and packages. The trustworthiness of the e-vendors is an important factor that consumers usually evaluate through information provided by the website and ratings and reviews of former consumers (Gong et al, 2013; Clemes et al, 2014; Kindra et al, 2014).

Many existing research on online shopping apply consumer behavior theory in a qualitative analysis, however, quantitative studies are limited. Especially, few studies have explicitly examined the impact of various factors on consumer preference and behavior on online shopping for fresh food. In this paper, we investigate Chinese consumer's perception of online shopping for fresh food and examine factors that influence their online shopping behavior. E-vendor characteristics, product attributes, and consumer's characteristics will be studied as they may influence the online shopping decisions. The results can help us gain a better understanding on consumer's online shopping choice and make comparison with the existing knowledge on consumer's preference for conventional in-store shopping. With online shopping continues to play an increasing role in people's life, e-vendors need to learn the factors that consumers care and consider as important and effective in order to better improve the quality of their products and services.

Data

A consumer survey was designed and implemented in Beijing, Guangzhou and Shanghai in the summer of 2015. Graduate Students from Renmin University, Shanghai Jiaotong University and Guangdong University of Foreign Studies were recruited and trained to conduct face-to-face interviews. In each city, we selected nine representative supermarkets within the urban area of the city. Two students formed a team and they randomly invited shoppers to participate in the survey. We interviewed about 340 grocery shoppers in each city and the total sample size is 1028. Because grocery store is still the primary shopping outlet in Chinese topline cities, randomly selecting shoppers there can obtain a representative sample of overall shoppers.

The survey includes questions on consumer's frequency of online shopping for food/drink and for fresh food in particular and intention to try fresh food shopping online in the future if they never did it before. We listed 11 factors that may affect consumer's choice of online shopping for fresh food including five product attribute factors (i.e. food safety and quality, price, place of origin, package and brand) and six e-vendor characteristic factors (i.e. e-vendor's reputation, product sales volume which usually shown on e-vendor's website, consumer's ratings and reviews, delivery cost, delivery speed, and product display picture on the website). We asked consumers to first select the factors that they perceive as important ones affecting their purchase decisions and then rank the relative importance of these selected factors. The rank value of one indicates the most important factor, two means the second most important factor and so on.

In order to compare consumer's perception of online shopping for fresh food with that of conventional in-store shopping, we asked respondents to compare the two shopping formats in terms of convenience, product quality, price and product freshness/taste. For each perspective, they chose one of the five options which are online shopping is much better than, or better than, or the same as, or worse than, or much worse than conventional in-store shopping. Additionally, socio-demographic information was collected to capture the heterogeneity among consumers. This includes family size (numbers of adults and children in a household), gender, age, income, education level and whether they migrated to a big city from rural areas in the past two years.

Table 1 summarizes descriptive statistics of demographic variables. Overall, 65% of the total respondents is female and the average age is 37 years old. In the survey, we asked respondents to provide their education levels by selecting one out five categories. To create a continuous variable of years of education, we assign value of 12 if the respondent has high school or lower education, 14 if vocation school or associate, 16 if bachelor's degree and 19 if graduate degree. The average years of education is 15 years. Similarly, we create a continuous variable of income by assigning values of 50,000 RMB, 60,000 RMB, 85,000 RMB, 125,000 RMB and 150,000 RMB to respondents whose annual household income is less than 50,000 RMB, between 50,000 RMB and 70,000 RMB, between 70,000 RMB and 100,000 RMB, between 100,000 RMB and 150,000 RMB and above 150,000 RMB, respectively. The average household income pear year is 113,600 RMB. On average, each household has three adults and 0.66 children. Around a quarter of the respondents stated that they moved to urban areas from rural areas or countryside over the last two years. Consumers shop food and drink online more frequently than shopping for fresh food products. About 23% of the total consumers have bought fresh food online before and around 65% have shopped for food/drink from the Internet. As for the main reasons that hold consumers from conducting online shopping, the top two choices are inconvenient Internet access and untrustworthy online vendors.

Table 2 illustrates summary statistics of variables used in the models.

Methods

Under random utility framework, we employ ordered logit models to measure the effect of various factors on consumer's frequency of online shopping for food/drink and use Heckman two-stage model to examine the effect on frequency of online shopping for fresh food.

In the food/drink model, the frequency of online shopping for food and drink over the last three months is an ordinal ranking variable with five values. We assigned values of zero through four to indicate purchase frequency over the past three months of never, once or twice, once per month on average, once per week on average and multiple times per week, respectively. The ordered logit model is specified following Green (1993) and Wang and Kockelman (2005).

$$y^* = X\beta + \epsilon \tag{1}$$

$$y = \begin{cases} 0(\textit{never bought}), & y * \leq \mu_1 \\ 1(\textit{once or twice}), & \mu_1 < y * \leq \mu_2 \\ 2(\textit{once per month on average}), & \mu_2 < y * \leq \mu_3 \\ 3(\textit{once per week on average}), & \mu_3 < y * \leq \mu_4 \\ 4(\textit{multiple times per week}), & y * > \mu_5 \end{cases}$$

where y is the observed frequency of online food/drink shopping over the last three months, y^* is the latent utility level, and $\mu_i (i = 1,2,3,4,5)$ represents the utility thresholds of online shopping for food/drink. X is the vector of independent variables including product attributes, e-vendor characteristics, consumer's perception on comparison between online and offline shopping, and socio-demographic variables, β is the corresponding coefficient vector, and ϵ is the random error term.

The probability of observing shopping frequency of *i* for an individual consumer is

$$Prob(y = i) = \mathrm{Prob}(\mu_i < y * \leq \mu_{i+1}) = G(X\beta - \mu_i) - G(X\beta - \mu_{i+1}),\ i = 0, 1, \ldots, 4 \tag{2}$$

where G(.) is the logistic cumulative distribution function in the form of $G(z) = \frac{\exp(z)}{1+\exp(z)}$, and we assume $\mu_0 = -\infty$ and $\mu_5 = +\infty$.

Log likelihood function is obtained by multiplying the probabilities across all respondents. Maximizing the log likelihood function provides the estimates of coefficients and cutoff levels. In the ordered logit model, the underlying assumption of proportional odds indicates the odds are cumulative odds of belonging to a certain category or higher versus of belonging to the lower categories, which is represented as

$$Odds\ Ratio = \frac{Prob(y \geq i)}{Prob(y < i)} = \exp(X\beta - \mu_i) \tag{3}$$

where i = 0, 1,…, 4. The log odds ratio is then $X\beta - \mu_i$. The β coefficients are the marginal log odds ratio for corresponding independent variables. Marginal odds ratio is defined as the exponential of each β. Marginal effect of a factor change on the probability of each category, i.e. $\frac{\partial Prob(y=i)}{\partial X_j}$ is also commonly reported.

The fact that online shopping for fresh food is still at early stage is reflected in the survey data in that only 23% of the respondents bought fresh food online over the last three months while 77% never bought it online. The dominance of respondents without fresh food online shopping experience raises concerns because their answers to the perception questions are based on hypothesis instead of actual experience. As a result, the selection bias may exist in the sample and the effect of different factors on consumer's decision to shop fresh food online may differ from the effect on their purchase frequency.

To examine effect of different factors on consumer's frequency of online shopping for fresh food, we adopt the Heckman two-stage model (Heckman, 1976) which treats selection bias as an omitted variable problem. Khodaverdizadeh et al (2009) stated that the assumption of Heckman two-step method allows that different series of variables enter the model in each stage. Instead of using a choice model in the first stage and a simple regression model in the second stage, Degeratu et al (2000) adopted a binary probit model in the first stage and a multinomial logit model in the second stage to analyze consumer's choice of shopping online and offline.

Following Degeratu et al (2000), we use a probit model in the first stage to estimate consumer's online purchase decision and an ordered logit model in the second stage to estimate effect on consumer's online shopping frequency. Different explanatory variables are used in the two stages. Since most respondents never bought fresh food online before, we assume their answers on whether the listed factors are important to make purchase decisions are accurate while their rankings on the selected important factors are not reliable due to lack of experience. In the first stage, the product attributes and e-vendor characteristics enter the model in dummy variable forms, i.e. the variable takes value of one if the respondent considered it as an important factor and zero otherwise. In the second stage, we select only observations of respondents who shopped online for fresh food over the last three months and apply ordered logit model. Due to consumer's previous experience, we consider their ranking of the selected important factors reliable and use the ranks as explanatory variables in the model. In particular, the variable is assigned value one if the factor was ranked as one of the top three most important factors by the respondent and value zero if not.

The first stage probit model is specified as

$$Z^* = X\beta + \epsilon$$

$$Z = \begin{cases} 1, & Z^* > 0 \\ 0, & Z^* \leq 0 \end{cases} \tag{4}$$

where Z^* is the unobservable latent variable representing utility level, Z is observed online shopping frequency for fresh food. X is a vector of explanatory variables including product attributes, e-vendor characteristics, consumer perceptions and socio-demographic variables. β is the vector of corresponding coefficients and ϵ is the random error term.

Following Zheng et al (2010), predicted probability is calculated in the first stage probit regression and used to create the Inverse Mills Ratio (λ).

$$P_i = Prob(Z = 1) = Prob(\epsilon > -X\beta) = 1 - \Phi(-X\beta) = \Phi(X\beta) \tag{5}$$

We can obtain Inverse Mills Ratio (λ) as follows:

$$\lambda = \frac{\phi[-(X\hat{\beta})/\sigma]}{\Phi[-(X\hat{\beta})/\sigma]} \tag{6}$$

where $\phi(.)$ is the probability density function of the standard normal distribution and $\Phi(.)$ is the corresponding cumulative distribution function. σ represents the standard deviation of ϵ in equation (4), which is one in this case.

In the second stage, we employ an ordered logit model. The dependent variable of online shopping frequency for fresh food takes values of one through four given we only use observations of respondents who purchased fresh food online over the last three months. The Inverse Mills Ratio (λ) is incorporated in the second stage to correct the selection bias, which is specified as

$$y^* = \alpha\lambda + X\beta + u \tag{7}$$

$$y = \begin{cases} 1(once\ or\ twice), & y* \leq \mu_1 \\ 2(once\ per\ month\ on\ average), & \mu_1 < y* \leq \mu_2 \\ 3(once\ per\ week\ on\ average), & \mu_2 < y* \leq \mu_3 \\ 4(multiple\ times\ per\ week), & y* > \mu_4 \end{cases}$$

where X is a vector of explanatory variables including product attributes, e-vendor characteristics, consumer perceptions and socio-demographic variables. β is the vector of corresponding coefficients and ϵ is the random error term. α is the coefficient on Inverse Mills Ratio. If α is statistically significant, we can confirm there is sample selection bias in the model and Inverse Mills Ratio is needed for correction.

Results

Maximum log-likelihood method is used for the model estimation and is implemented in Stata. Table 3 illustrates the estimation results of food/drink model and fresh food model side-by-side. In the food/drink model, we find that the indicator variable of high ranking of product's place of origin (i.e. respondent ranks the place of origin as one of the top three important factors affecting their online purchase decisions. For simplicity, we mention it as high ranking of the factor thereafter.) is positive and significant at the 1% level. This indicates consumers ranking product's place of origin as a top important factor tend to shop online for food/drink more frequently. The marginal odds ratio of 1.732 means the odds of consumers who consider place of origin as a top important factor shopping online for food/drink more frequently are 1.732 times the odds of consumers who do not consider it as a top factor. This shows that consumers who care about place of origin prefers online shopping for food/drink more than their counterparts since the online platforms carry a bigger variety of products imported from different areas, domestically and internationally.

The variables indicating high ranking of e-vendor's reputation, delivery cost and delivery speed are all positive and significant at the 5% level which shows consumers who view these three e-vendor related factors as important ones tend to shop online for food/drink more frequently. Consumers are not able to touch and feel the products or communicate with sellers face-to-face during online shopping. Thus the e-vendor's reputation becomes an important factor for consumers to evaluate the seller's trustworthy and make purchase decision. Positive experience on e-vendor's reputation motivates consumers to shop online more frequently. Being able to know the reputation of e-vendor motivates consumer to shop more online, because the e-vendor reputation labeled online is a result of previous customers' satisfaction rating which is not available in offline grocery stores. Convenience is one of the motives for consumers to shop online since the products will be delivered to destination directly so that it saves time and efforts of traveling and carrying products back. However, the convenience advantage is realized and online shopping is chosen only if consumers view the delivery cost is reasonable and delivery speed is fast enough.

Consumer's perception on comparison between online and offline shopping in terms of price and product freshness/taste are significant at the 1% level. Consumers who consider the online shopping price is better than that in conventional stores tend to shop online for food/drink more frequently. Consumers who view the online shopping products are less fresh and tastey than conventional stores tend to shop online less frequently. While the more competitive price for online products attract consumers, the freshness and taste of products are also critical. Consumers may not choose to shop food/drink online if the vendors cannot effectively preserve the freshness and taste of the products.

Demographic variables also affect consumer's online shopping decisions. Consumers having more years of education are likely to shop online for food/drink more frequently, probably due to familiarity with electronic devices, acceptability of technological trends and a better understanding of potential risks. Consumers who moved to big cities from rural areas over the last two years are less likely to shop online frequently. This group of consumers may be more used to conventional in-store shopping. The older the consumers, the less likely they shop online frequently. Comparing to young consumer group, the older consumers may not have the equipment or tools (such as electronic devices, credit cards, online transaction platform accounts, etc) for online shopping.

Table 4 shows marginal effects of changes in regressors on predicted probabilities at each level of the dependent variable for food/drink model. For example, if the value of *d_price* changes from 0 to 1, the probabilities of y=0 and y=1 increase by 5.6% and 0.4%, respectively. The probabilities of y=2, y=3 and y=4 decrease by 3.5%, 1.8% and 0.8%, respectively. The interpretation for other variables are similar.

The right side panel in Table 3 shows estimation results for the fresh food model. In the first stage model, the dummy variables of place of origin and consumer's reviews are positive and significant at the 1% level. This indicates consumers who consider place of origin and reviews as important factors are more likely to shop fresh food online. The dummy variable of e-vendor's reputation is negative and significant at the 5% level showing consumers viewing reputation as an important factor is less likely to shop online for fresh food. Thus e-vendor's reputation is a concern for consumers to make decisions on purchasing fresh food online. Notice this sample is dominated by shoppers who have no online grocery shopping experience, those who think e-vender reputation is important they may assume they don't have good reputations in general and don't go online shopping.

Some of the consumer perception factors are also significant at the 5% or 1% levels. Consumers who perceive online shopping is inferior to in-store shopping in terms of convenience and product freshness are less likely to choose to shop fresh food online. While consumers who view online shopping is superior to in-store shopping in terms of price is more likely to shop online for fresh food. This confirms the competitive price motivates consumers to choose online shopping. However, if they perceive the fresh food purchased online is less fresh than the in-store purchase, they do not shop online.

People who recently moved to the big cities from countryside over the last two years are less likely to shop online for fresh foods. This is consistent with food/drink model in that migrants may be more used to conventional in-store shopping and are still adjusting themselves to the catch up the pace of city life.

Based on the first stage prediction, we calculated the Inverse Mills Ratio and applied it to the estimation in the second stage. See Figure 1 for the predicted probability density function in Heckman second stage. The coefficient of λ is 1.278 and is not significant at the 10% level, indicating a weak selection bias. The high ranking indicator variable of price and quality are negative and significant at the 5% level indicating consumers ranking those factors as top importance are likely to shop fresh food less frequently online. Consumers perceiving online shopping inferior to in-store shopping in freshness is likely to shop fresh food online less frequently. Thus for fresh food shopping, price becomes a concern instead of a motive for purchase decision. Due to the preservation needs of fresh food, the price may not be competitive with purchase in supermarkets or farmer markets. In addition, product quality and freshness are also concerns due to the perishable feature of products. If the consumers are not satisfied with the quality and freshness of online purchased fresh food, they may choose not to shop them online even there are advantages in convenience or price.

Table 5 shows marginal effects of changes in regressors on predicted probabilities at each level of the dependent variable for fresh food model. For example, if the value of *r_price* changes from 0 to 1, the probability of y=1 increases by 25.6%. The probabilities of y=2, y=3 and y=4 decrease by 18.3%, 6.2% and 1.1%, respectively. The interpretation for other variables are similar.

Conclusion

With the innovation in online payment platforms and development in delivery systems, online grocery market is developing rapidly during recent years in China. In addition to the traditionally popular product categories provided in the online "stores", more and more e-commerce started to add "grocery stores" as part of their business, such as Tmall market, the business-to-consumer market developed by Alibaba. However, online grocery store is still very new to Chinese consumers and more regulations need to be established. In our survey dataset, there are more than 90% respondents stated that they had online shopping experience, while only 15% shopped online for fresh food. Consumers are still learning and establishing their online shopping preference. Due to the high standards of Chinese consumers on the freshness and quality of grocery products, a timely and efficient storage and delivery system with reasonable cost is a challenge for e-commerce to expand their online grocery business.

In this paper, we investigate consumer's online grocery shopping preference and behavior by estimating the effects of various factors on consumer's online shopping frequency for food/drink and for fresh food, respectively. A rank ordered logit model is used to examine consumer's online shopping frequency for food/drink. Because only a small portion of consumers shopped online for fresh food before, a Heckman two stage model is used to analyze consumer's online shopping decision and frequency for fresh food.

In both models, we find that consumers considering product's place of origin as an important factor are more likely to shop online. This is consistent with the trends that Chinese consumers, especially the rising middle class, have emerging preference for high quality food, such as imported food from reputational origin. Consumer's perception of price and freshness of products affect their frequency of online shopping. One of the motives of online shopping is that the price is more competitive than conventional stores due to lower operational and intermediate cost. Consumers who consider online shopping has a better price than in-store shopping are more likely to shop online for food/drink and fresh food. Freshness and thus quality are important factors for food/drink especially fresh food purchase. Consumers who view the freshness of online products as inferior to in-store products are less likely to shop online. Thus consumers need to balance among price, quality and convenience to make online purchase decisions.

For food/drink online consumption, consumers who consider e-vendor's reputation, delivery cost and delivery speed as top important factors are more likely to shop online. This indicates that trustworthiness of e-vendors and effective delivery system are critical factors to motivate consumers to shop online. For fresh food online consumption, consumers who consider e-vendor reputation as an important factor is less likely to shop online. This indicates that reputation is a concern for consumers. Positive perception and experience of e-vendor's reputation encourages consumers to purchase online while uncertainty and concerns on the reputation and trustworthiness prevent consumers from choosing the online shopping form. Consumers who consider ratings and reviews by previous consumers as an important factor are more likely to purchase fresh food online which indicates consumers also rely on their peers' comments to investigate e-vendor's trustworthiness so as to make purchase decisions.

This paper has shed light on factors that impact consumers' online shopping frequency for food/drink and fresh food. The results are useful for policy makers on online grocery shopping regulations and for online grocery vendors on developing the e-commerce industry. The results on product characteristics and e-vendor characteristics can help e-retailers to expand their business. Moreover, the analysis on socio-demographic variables can assist e-vendors on targeting and segmenting potential consumers.

Reference

Anckar, B., Walden, P., & Jelassi, T. (2002). Creating customer value in online grocery shopping. International Journal of Retail & Distribution Management, 30(4), 211-220.

Bin, Q., Chen, S. J., and Sun, S. Q. (2003). Cultural Differences in E-commerce: A Comparison Between the US and China. *Journal of Global Information Management*, 11(2), 48.

Chu, J., Arce-Urriza, M., Cebollada-Calvo, J. J., and Chintagunta, P. K. (2010). An Empirical Analysis of Shopping Behavior Across Online and Offline Channels for Grocery Products: the Moderating Effects of Household and Product Characteristics. *Journal of Interactive Marketing*, 24(4), 251-268.

Clemes, M. D., Gan, C., and Zhang, J. (2014). An Empirical Analysis of Online Shopping Adoption in Beijing, China. *Journal of Retailing and Consumer Services*, 21(3), 364-375.

CNNIC (2016). 2015 Statistical Report on Internet Development in China. Retrieved from http://www.cnnic.net.cn/

Degeratu, A. M., Rangaswamy, A., and Wu, J. (2000). Consumer Choice Behavior in Online and Traditional Supermarkets: the Effects of Brand Name, Price, and Other Search Attributes. *International Journal of Research in Marketing*, 17(1), 55-78.

Gong, W., Stump, R. L., and Maddox, L. M. (2013). Factors Influencing Consumers' Online Shopping in China. *Journal of Asia Business Studies*, 7(3), 214-230.

Greene, W. H. (1993). Econometric Analysis. *Engelwood Cliffs*.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Aample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement, Volume 5, number 4* (pp. 475-492).

Huang, Y., and Oppewal, H. (2006). Why Consumers Hesitate to Shop Online: An Experimental Choice Analysis of Grocery Shopping and the Role of Delivery Fees. *International Journal of Retail & Distribution Management*, 34(4/5), 334-353.

Khodaverdizadeh, M., Kelashemi, M. K., Hayati, B., and Molaei, M. (2009). Estimation of Recreation Value and Determining the Factors Effective in Visitors' WTP for Saint Stepanus Church Using the Heckman Two-stage and CV Methods. *World Applied Sciences Journal*, 7(4):543-551.

Kindra, G., Ahmed, S. A., & Shawli, N. (2014). Entrepreneurship in Saudi Arabia: An Empirical Investigation of Online Grocery Shopping Behaviour. *Journal of Management Information System & E-commerce.* 1(2), 01-24.

Liu, X., He, M., Gao, F., & Xie, P. (2008). An empirical study of online shopping customer satisfaction in China: a holistic perspective. *International Journal of Retail & Distribution Management*, *36*(11), 919-940.

Pechtl, H. (2003). Adoption of Online Shopping by German Grocery Shoppers. *The International Review of Retail, Distribution and Consumer Research,* 13(2), 145-159.

Ramus, K., and Asger Nielsen, N. (2005). Online Grocery Retailing: What Do Consumers Think? *Internet Research,* 15(3), 335-352.

Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. International journal of research in marketing, 20(2), 153-175.

Wang, X., and Kockelman, K. (2005). Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury: Distinguishing Effects of Vehicle Weight and Type. *Transportation Research Record: Journal of the Transportation Research Board*, (1908), 195-204.

Zheng, S., Wang, Z., Wang, H., and Song, S. (2010). Do Nutrition and Health Affect Migrant Workers' Incomes? Some Evidence from Beijing, China. *China & World Economy*, *18*(5), 105-124.

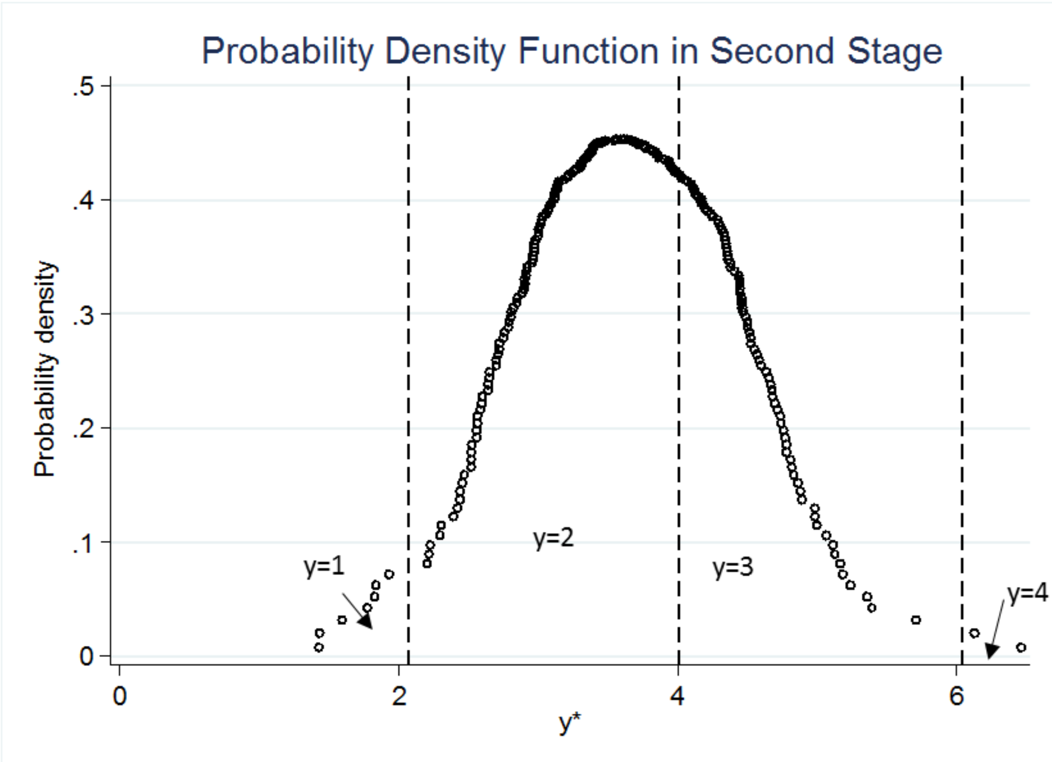

Figure 1. Probability Density Function in Heckman Second Stage

Table 1: Summary Statistics of Demographic Variables

Variable	Description	Mean	Std. Dev.	Min	Max
female	=1 if female; 0 otherwise	0.65	0.48	0	1
eduyear	=12 if high school or lower; =14 if vocation school or associate; =16 if bachelor's degree; =19 if graduate degree	14.97	2.35	12	24
move	=1 if migrate to city over last two years; =0 otherwise	0.24	0.43	0	1
income (in 10,000 RMB)	=5 if income is below 50,000 RMB; =6 if ≥50,000 RMB and <70,000 RMB; =8.5 if ≥70,000 RMB and <100,000 RMB; =12.5 if ≥100,000 RMB and <150,000 RMB; =15 if ≥150,000 RMB	11.36	3.69	5	15
n_Adult	number of adults in household	2.97	1.20	0	12
n_Child	number of children in household	0.66	0.82	0	9
age	in years	37.16	13.54	10	85

Table 2: Summary Statistics of Variables Used in the Models

	Variable	Description	Values	Mean	Std. Dev.
Dependent Variable	freq_fresh	fresh food online shopping frequency	(0,1,2,3,4)	0.36	0.74
	freq_fodr	food/drink online shopping frequency	(0,1,2,3,4)	1.22	1.14
Independent Variable:					
product attributes	*d_price*	Price selected as an important factor	(0,1)	0.66	0.47
	d_package	Package selected as an important factor	(0,1)	0.64	0.48
	d_quality	Quality selected as an important factor	(0,1)	0.45	0.50
	d_origin	Place of origin selected as an important factor	(0,1)	0.88	0.33
	d_brand	Brand selected as an important factor	(0,1)	0.61	0.49
	r_price	Rank price as top 3 important factors	(0,1)	0.29	0.45
	r_package	Rank package as top 3	(0,1)	0.08	0.26
	r_quality	Rank quality as top 3	(0,1)	0.63	0.48
	r_origin	Rank place of origin as top 3	(0,1)	0.24	0.43
	r_brand	Rank brand as top 3	(0,1)	0.27	0.45
e-vendor characteristics	*d_reputation*	Reputation selected as an important factor	(0,1)	0.82	0.38
	d_salesvolume	Sales volume selected as an important factor	(0,1)	0.50	0.50
	d_review	Reviews selected as an important factor	(0,1)	0.79	0.41
	d_deliverycost	Delivery cost selected as an important factor	(0,1)	0.38	0.48
	d_deliveryspeed	Delivery speed selected as an important factor	(0,1)	0.54	0.50
	d_displaypicture	Display picture selected as an important factor	(0,1)	0.39	0.49
	r_reputation	Rank reputation as top 3	(0,1)	0.58	0.50
	r_salesvolume	Rank sales volume as top 3	(0,1)	0.20	0.40
	r_review	Rank reviews as top 3	(0,1)	0.47	0.50
	r_deliverycost	Rank delivery cost as top 3	(0,1)	0.04	0.19
	r_deliveryspeed	Rank delivery speed as top 3	(0,1)	0.11	0.31
	r_displaypicture	Rank display pictures as top 3	(0,1)	0.05	0.22
convenience	*con_onlineneg*	Online less convenient than offline	(0,1)	0.20	0.40
	con_onlinepos	Online more convenient than offline	(0,1)	0.68	0.47
quality	*qua_onlineneg*	Online less reliable than offline	(0,1)	0.48	0.50
	qua_onlinepos	Online more reliable than offline	(0,1)	0.12	0.33
price	*pri_onlineneg*	Online higher price than offline	(0,1)	0.07	0.26
	pri_onlinepos	Online lower price than offline	(0,1)	0.74	0.44
freshness/taste	*fre_onlineneg*	Online less fresh than offline	(0,1)	0.53	0.50
	fre_onlinepos	Online more fresh than offline	(0,1)	0.12	0.32

Table 3. Maximum-Likelihood Coefficient Estimates and Marginal Odds Ratios in Two Models

	Variables		Food/drink Ordered Logit Model				Fresh Food Heckman Two Stage Model						
							1st stage (y=0,1)			2nd stage (y=1,2,3,4)			
			Coef.	Std. Err.	Marginal Odds Ratio	Sig.	Coef.	Std Err.	Sig.	Coef.	Std Err.	Marginal Odds Ratio	Sig.
Product Attributes	dummy	*d_price*	-0.260	0.158	0.773	*	-0.097	0.110					
		d_package	-0.190	0.167	0.827		-0.003	0.118					
		d_quality	-0.275	0.224	0.758		-0.184	0.143					
		d_origin	0.288	0.157	1.335	*	0.386	0.109	***				
		d_brand	0.143	0.159	1.155		0.045	0.106					
	rank	*r_price*	0.255	0.215	1.289					-1.208	0.537	0.299	**
		r_package	0.511	0.288	1.666	*				-0.155	0.768	0.857	
		r_quality	-0.080	0.212	0.923					-1.063	0.458	0.346	**
		r_origin	0.552	0.233	1.732	***				0.587	0.600	1.799	
		r_brand	0.096	0.217	1.099					-0.200	0.504	0.818	
E-vendor Characteristics	dummy	*d_reputation*	-0.278	0.201	0.755		-0.263	0.128	**				
		d_salesvolume	-0.048	0.164	0.952		-0.070	0.108					
		d_review	0.229	0.189	1.262		0.219	0.131	***				
		d_deliverycost	0.089	0.167	1.090		-0.049	0.127					
		d_deliveryspeed	-0.023	0.153	0.978		0.030	0.110					
		d_displaypicture	0.278	0.172	1.322	*	0.095	0.122					
	rank	*r_reputation*	0.449	0.214	1.569	**				-0.778	0.477	0.459	*
		r_salesvolume	0.435	0.231	1.546	*				0.091	0.503	1.095	
		r_review	0.345	0.213	1.410	*				-0.161	0.532	0.852	
		r_deliverycost	0.680	0.332	1.979	**				0.561	0.884	1.752	
		r_deliveryspeed	0.585	0.264	1.794	**				-0.419	0.510	0.658	
		r_displaypicture	0.108	0.321	1.113					-0.524	1.185	0.592	
Perception of Online vs Offline Shopping	Convenience	*con_onlineneg*	-0.412	0.227	0.662	*	-0.393	0.180	**	-0.190	0.733	0.827	
		con_onlinepos	0.069	0.190	1.071		-0.046	0.145		-0.168	0.480	0.846	
	Quality	*qua_onlineneg*	-0.255	0.140	0.776	*	-0.112	0.110		-0.106	0.425	0.899	
		qua_onlinepos	-0.268	0.243	0.765		0.216	0.170		-0.331	0.475	0.718	
	Price	*pri_onlineneg*	0.210	0.279	1.235		0.301	0.222		1.084	0.862	2.955	
		pri_onlinepos	0.659	0.160	1.933	***	0.466	0.132	***	0.416	0.626	1.515	
	Freshness/ Taste	*fre_onlineneg*	-0.427	0.139	0.653	***	-0.347	0.109	***	-0.966	0.401	0.380	***
		fre_onlinepos	0.207	0.255	1.232		0.116	0.175		0.689	0.450	1.991	
demographics		*female*	0.232	0.126	1.262	*	-0.010	0.096		-0.049	0.322	0.952	
		eduyear	0.099	0.030	1.103	***	0.010	0.023		0.060	0.071	1.062	
		move	-0.282	0.143	0.755	**	-0.257	0.113	**	0.094	0.454	1.099	
		income	0.000	0.050	1.003		-0.005	0.013		-0.048	0.042	0.953	
		n_Adult	0.032	0.052	1.032		-0.025	0.041		0.231	0.168	1.260	
		n_Child	0.064	0.076	1.066		0.043	0.060		-0.028	0.262	0.973	
		age	-0.025	0.005	1.025	***	-0.007	0.004	*	0.014	0.012	1.014	
		Inverse Mills								1.278	1.047		

Note: significant level: ***---1%, **---5%, *---10%.

Table 4: Marginal Effect in Food/drink model at different dependent variable level

Variables			Frequency				
			0	1	2	3	4
Product Attributes	dummy	*d_price*	0.056	0.004	-0.035	-0.018	-0.008
		d_package	0.042	0.002	-0.026	-0.013	-0.006
		d_quality	0.059	0.007	-0.037	-0.020	-0.009
		d_origin	-0.065	-0.002	0.039	0.019	0.009
		d_brand	-0.032	-0.001	0.019	0.009	0.004
	rank	*r_price*	-0.055	-0.005	0.034	0.017	0.008
		r_package	-0.104	-0.020	0.065	0.039	0.019
		r_quality	0.018	0.001	-0.011	-0.005	-0.003
		r_origin	-0.115	-0.016	0.071	0.040	0.020
		r_brand	-0.021	-0.001	0.013	0.006	0.003
E-vendor Characteristics	dummy	*d_reputation*	0.060	0.007	-0.037	-0.020	-0.009
		d_salesvolume	0.011	0.001	-0.007	-0.003	-0.002
		d_review	-0.053	0.000	0.032	0.015	0.007
		d_deliverycost	-0.019	-0.001	0.012	0.006	0.003
		d_deliveryspeed	0.005	0.000	-0.003	-0.001	-0.001
		d_displaypicture	-0.061	-0.004	0.037	0.019	0.009
	rank	*r_reputation*	-0.100	-0.003	0.061	0.029	0.014
		r_salesvolume	-0.092	-0.012	0.057	0.032	0.015
		r_review	-0.076	-0.004	0.046	0.023	0.011
		r_deliverycost	-0.132	-0.034	0.082	0.056	0.029
		r_deliveryspeed	-0.118	-0.024	0.074	0.045	0.023
		r_displaypicture	-0.023	-0.002	0.014	0.007	0.003
Perception of Online vs Offline Shopping	Convenience	*con_onlineneg*	0.095	-0.002	-0.056	-0.025	-0.011
		con_onlinepos	-0.015	-0.001	0.009	0.004	0.002
	Quality	*qua_onlineneg*	0.056	0.003	-0.034	-0.017	-0.008
		qua_onlinepos	0.061	-0.001	-0.036	-0.017	-0.008
	Price	*pri_onlineneg*	-0.045	-0.005	0.028	0.015	0.007
		pri_onlinepos	-0.152	0.007	0.088	0.039	0.018
	Freshness/ Taste	*fre_onlineneg*	0.094	0.005	-0.057	-0.029	-0.013
		fre_onlinepos	-0.045	-0.005	0.028	0.015	0.007
Demographics		*female*	-0.052	-0.001	0.031	0.015	0.007
		eduyear	-0.022	-0.001	0.013	0.007	0.003
		move	0.064	0.000	-0.038	-0.018	-0.008
		income	-0.001	0.000	0.000	0.000	0.000
		n_Adult	-0.007	0.000	0.004	0.002	0.001
		n_Child	-0.014	-0.001	0.009	0.004	0.002
		age	-0.006	0.000	0.003	0.002	0.001

Table 5: Marginal Effect in Fresh Food model at different dependent variable level

Variables			Frequency			
			1	2	3	4
Product Attributes	rank	*r_price*	0.256	-0.183	-0.062	-0.011
		r_package	0.036	-0.025	-0.009	-0.002
		r_quality	0.250	-0.164	-0.072	-0.014
		r_origin	-0.141	0.093	0.041	0.008
		r_brand	0.047	-0.032	-0.012	-0.002
E-vendor Characteristics	rank	*r_reputation*	0.183	-0.123	-0.051	-0.009
		r_salesvolume	-0.021	0.015	0.006	0.001
		r_review	0.038	-0.026	-0.010	-0.002
		r_deliverycost	-0.138	0.086	0.043	0.008
		r_deliveryspeed	0.094	-0.067	-0.023	-0.004
		r_displaypicture	0.115	-0.083	-0.027	-0.005
Perception of Online vs Offline Shopping	Convenience	*con_onlineneg*	0.044	-0.031	-0.011	-0.002
		con_onlinepos	0.040	-0.027	-0.011	-0.002
	Quality	*qua_onlineneg*	0.025	-0.017	-0.007	-0.001
		qua_onlinepos	0.076	-0.053	-0.019	-0.003
	Price	*pri_onlineneg*	-0.264	0.145	0.099	0.020
		pri_onlinepos	-0.094	0.066	0.023	0.004
	Freshness/ Taste	*fre_onlineneg*	0.217	-0.151	-0.056	-0.010
		fre_onlinepos	-0.167	0.106	0.051	0.010
Demographics		*female*	0.012	-0.008	-0.003	-0.001
		eduyear	-0.014	0.010	0.004	0.001
		move	-0.022	0.015	0.006	0.001
		income	0.011	-0.008	-0.003	-0.001
		n_Adult	-0.054	0.037	0.014	0.003
		n_Child	0.006	-0.004	-0.002	0.000
		age	-0.003	0.002	0.001	0.000