AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

RESEARCH PAPER: 1992-3

MODELING CONSUMER'S PERCEPTION OF ORANGE JUICE

BY

Xiaoming Gao, Fellow -- UF

Jonq-Ying Lee

Senior Research Economist

FLORIDA DEPARTMENT OF CITRUS

Economic and Market Research Department

P.O. Box 110249

Gainesville, Florida 32611-2049 USA

Phone: 352-392-1874

Fax: 352-392-8634

Email: jonqying@ufl.edu

www.floridajuice.com

Modeling Consumer's Perception of Orange Juice

Xiaoming Gao

and

Jonq-Ying Lee*

* Xiaoming Gao was a McKethan-Matherly Post-Doctoral Fellow at the Graduate School of Business Administration, University of Florida. Jonq-Ying Lee is a Research Economist, Florida Department of Citrus and an Adjunct Professor, Department of Food and Resource Economics, University of Florida. This research is partially supported by McKethan-Matherly Eminent Scholar Chair of the University of Florida.

The authors thank A. Reynolds, J. S. Shonkwiler, and H. Theil for comments and suggestions.

Modeling Consumer's Perception of Orange Juice

Abstract

A multiple indicator and multiple cause model with dichotomous indicators was used to study consumer's perception toward orange juice. Results indicate that recalls of orange juice advertising messages by the respondent had a positive impact on his/her perception toward orange juice. Results also suggest that selected socioeconomic variables were important determinants of consumer perception.

Key Words: Perception, Latent, Dichotomous.

Modeling Consumer's Perception of Orange Juice

Most econometric studies analyzing the effectiveness of advertising have taken the advertising message as given, and interests have focused on quantifying the impact of advertising on sales. However, it is widely accepted (see Kinnucan and Venkateswaran, 1990) that advertising influences consumption by changing consumers' perception (fancied or real) of the attributes of the good in question. Therefore, research concerned with characterizing consumers' perception and obtaining a better understanding of the relationship existing between advertising and consumers' perception can provide useful information for guiding the design of advertising messages in a manner that would maximize the effectiveness of such messages.

One possible reason that economists have not investigated consumer perception when analyzing the effectiveness of advertising is that consumer perception is not observable. Moreover, even when data for consumer perception in the form of proxies or indicators are available, the use of proxies with the standard regression approach results in biased and inconsistent parameter estimates (Fuller, 1987). Therefore, alternative models need to be considered. The latent variable structural equation model is widely used for studies having unobservable variables, but the estimation is further complicated when (as is often the case) the proxy variables for perception (or indicators) are dichotomous (0 and 1 variables) in nature.

Recognizing that there are linkages existing among advertising, consumer perception and consumer demand, this study employs a structural latent variable model with Multiple Indicators and Multiple Cause variables (MIMIC) to analyze the impact of household socioeconomic factors on consumer perception of orange juice and to determine the effect of orange juice advertising messages on consumer attitudes toward orange juice. The MIMIC model used in this study takes the dichotomous nature of the proxies or indicators used to represent the unobservable perception variable. This model may have wide application to analyzing unobserved economic phenomena when indicators and/or cause variables are available in a dichotomous form.

Advertising is a prominent feature of the U.S. food industry. Each year, citrus producers and processors spend millions of dollars promoting their products through generic and brand advertising. For example, the orange-juice industry and the Florida Department of Citrus (FDOC) spent more than $30 million a year during the last decade on the advertising of orange juice. The FDOC's advertising messages have been changed constantly. During the late 1960s, the FDOC advertised that orange juice is good for breakfast. Later, the emphasis changed from being good for breakfast to being good for any time of day. In the 1980s, the advertising themes changed to new lifestyle themes, such as orange juice is refreshing, invigorating and is most natural. However, past studies of the economic impacts of orange-juice advertising have not examined the relationship between advertising themes and consumers' perception of orange juice.

The MIMIC Model With Dichotomous Indicators

In this study, it is assumed that the consumer's perception of orange juice (the latent variable which will be discussed in this section) is influenced by his/her socioeconomic background (these are cause variables), e.g., education, household size, race, and the location of consumer's residency. In addition, it is assumed that the consumer's positive and/or negative perception of orange juice can be expressed by his/her choice of orange juice as the answers for certain selected image questions (these are indicator variables for consumer's perception), such as what beverage is good for breakfast, what beverage is good for your health, etc.

The MIMIC model (Jöreskog and Goldberger, 1975) considers a latent variable structural model where the latent variables are associated with a set of indicators, and a set of cause variables. The model is defined by two parts, cause equations and indicator equations:

(1) $\eta = \Gamma x + \zeta$

(2) $y = \Lambda_y \eta + \epsilon$

where η is considered for our purpose to be a scalar latent variable for consumer's perception of orange juice; x ($n \times 1$) and y ($p \times 1$) are vectors of observed variables (they are the cause or socioeconomic variables and indicator (answers to image questions) variables, respectively); Γ and Λ_y are coefficient matrices compatibly defined; ϵ ($p \times 1$) are the errors of measurement for y; and ζ is the disturbance. The measurement errors, ϵ, and the disturbance ζ have expected values of zero and covariance matrices Ψ and Θ_ϵ, respectively; $Cov(\zeta,\epsilon)=0$; and ϵ and ζ are uncorrelated with the latent variable, η.

The estimation of the structural equation model differs from the traditional econometric approach in that, instead of minimizing the functions of observed and predicted individual dependent values, the procedure used here minimizes the difference between the sample covariances and the covariances predicted by the model. The fundamental hypothesis for these structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters, which can be expressed as:

(3) $\Sigma = \Sigma(\theta)$

where Σ is the population covariance matrix of observed variables; θ is a vector that contains the model parameters, Γ, Λ_y and Θ_ϵ; and $\Sigma(\theta)$ is the implied structural covariance matrix written as a function of θ.

The sample covariance matrix of the observed variables, S=COV(y,x), can be used as an approximation of Σ. The implied structural covariance matrix is specified as

(4) $$\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix}$$

where $\Sigma_{yy}(\theta) = E(yy')$

$$= E[(\Lambda_y \eta + \epsilon)(\Lambda_y \eta + \epsilon)']$$

$$= \Lambda_y E(\eta\eta')\Lambda'_y + \Theta_\epsilon$$

$$= \Lambda_Y^{-1}(\Gamma S_{xx}\Gamma' + \Psi)\Lambda'_y + \Theta_\epsilon$$

Other elements of the implied covariance matrix can be defined accordingly. The complete expression of the implied covariance matrix is specified as (Bollen, 1989):

(5) $$\Sigma(\theta) = \begin{bmatrix} \Lambda_Y(\Gamma S_{xx}\Gamma' + \Psi)\Lambda'_y + \Theta_\epsilon & \Lambda_y \Gamma S_{xx} \\ S_{xx}\Gamma'\Lambda'y & S_{xx} \end{bmatrix}$$

in which S_{xx} is defined as covariance matrix for the cause variables x. The Maximum Likelihood algorithm can be used to estimate the parameters ($\hat{\theta}$) by minimizing the fitting function which equals the difference between S and $\Sigma(\hat{\theta})$. The model is identified when $n \geq 1$ and $p \geq 2$ (Bollen, pp.331).

The above estimator is only valid when both of the observed indicator y and cause variable x are continuous. When the indicators y are dichotomous, equation (2) becomes

(2') $y^* = \Lambda_y \eta + \epsilon.$

One does not observe y; instead, one observes y^*. Where

$$(6) \qquad y^* = 1 \quad \text{if } y \geq \kappa$$
$$= 0 \quad \text{otherwise}$$

and κ (p×1) are threshold variables; y are the underlying continuous latent variables; and y^* are indicators. We now have a case where the indicators (y) of a latent variable (η) are latent variables themselves.

The consequences of observing dichotomous indicators make direct application of the Jöreskog and Goldberger estimator inappropriate because $y^* \neq y$. In general, Σ^*, the population covariance matrix of dichotomous y^* and x, does not equal Σ, the population covariance matrix of the underlying continuous latent variable y. Thus, $\Sigma^* \neq \Sigma(\theta)$. The fundamental hypothesis, therefore, holds for the continuous indicators but not for the dichotomously observed indicators. If S^* is a consistent estimator of Σ^*, the parameter estimator based on S^* would give an inconsistent estimator of θ.

Maddala and Trost (1980, 1981) proposed a structural Probit (and Logit) estimator for the above problem by substituting equation (1) into equation (2') and estimating a set of nonlinear Probit (Logit) equations using the ML method. Such a model is easy to specify but computationally demanding (Gao and Lee, 1991). In this paper, a general model developed by Muthén et al. (1981, 1985) for the categorical indicator model is used for estimation.

The method developed by Muthén et al. corrects the problems of the presence of dichotomous indicators by estimating underlying continuous latent variables y and the corresponding covariance matrix between y and x, Σ. To estimate Σ from the dichotomous variables, one must assume a distribution for the underlying latent continuous indicators. It is commonly assumed that y follows a multivariate normal distribution. Given a multivariate normal distribution, estimated correlation coefficients can be obtained for every pair of y and x. For dichotomous indicators y^*, the correlation between the underlying continuous indicators is called a tetrachoric correlation; while the correlation between y^* and the continuous variable x is referred to as a biserial correlation. The correlation matrix, S, so obtained, is a consistent estimate of Σ.

Let a, b, c, and d be the elements of a fourfold (2×2 contingency) table, where a is the number of pairs

(y^*_{it} and y^*_{jt}), with $t=1, 2, ..., T$ with $y^*_{it} = y^*_{jt} = 0$.[1] The tetrachoric correlation between y_i and y_j in a standard bivariate normal distribution $\Phi(y_i, y_j; r_t)$ is approximately estimated (using information about y^*_{it} and y^*_{jt}) by Kotz et al. (1981, vol. 9: pp.223).

(7) $\quad r_t = \sin[(\pi/2)((ad)^{-1/2} - (bc)^{-1/2})/((ad)^{-1/2} + (bc)^{-1/2})]$.

The biserial correlation coefficient between y_i and x_j is calculated, based on the sample from (y^*,x) by assuming (y,x) has a bivariate normal distribution, $\Phi(y_i, x_j; r_b)$. Thus

(8) $\quad r_b = pq(x_1 - x_0)/(s_x u)$,

where

$$u = (2\pi)^{-1/2}\exp(-h^2/2)$$

with h defined as

$$pr[z \geq h] = p$$

where z is a standard normal variate, x_1 and x_0 are the mean x-values of observations having $y_i=1$ and $y_i=0$, respectively; s^2_x is the sample variance of x; and p is the proportion of the y^*-sample with $y^*_i = 1$ $(q = 1-p)$.

The general fitting function to be minimized (Muthén, 1984) can be expressed as

(9) $\quad F= [\hat{\rho} - \sigma(\hat{\theta})]'W^{-1}[\hat{\rho} - \sigma(\hat{\theta})]$

where $\hat{\rho}$ is a $(p+n)(p+n+1)(1/2) \times 1$ vector containing tetrachoric, biserial and Pearson correlation coefficients for the non-redundant correlations between all pairs of y and x variables (upper triangle of S), $\sigma(\hat{\theta})$ is the corresponding vector for the implied covariance matrix; and W is a consistent estimator of the asymptotic covariance matrix of $\hat{\rho}$ (Muthén, 1983 and 1984).

---

[1] A fourfold (2x2 contingency) table with frequency is given by

	$y^*_i=0$	$y^*_i=1$
$y^*_j=0$	a	b
$y^*_j=1$	c	d

Data Source and Estimation

Data generated from a survey sponsored by the FDOC, covering 1434 respondents nationwide during November 1990, were used for the study. Respondents were asked to indicate which beverage is good for 14 different occasions; e.g, what beverage is good for breakfast, good for your health, a good value, etc. The choice of orange juice as the answer signifies if consumers have positive perceptions toward orange juice. Each question asked had two categories, aided and unaided. In addition, respondents were asked to provide certain socioeconomic information. Sample means and standard deviations of selected indicator and cause variables are shown in Table 1.

Previous research on consumer perception of products revealed that the consumer image of a product can only be adequately understood by techniques that tap subconscious feelings, such as top-of-mind responses to certain questions (Zeitlin and Westwood, 1986). Therefore, only the unaided or spontaneous responses were used in this study[2]. Preliminary analysis showed that some of the indicators had no influence on consumer perception of orange juice. Therefore, although the FDOC collected data on 14 perception indicators, only seven were used in this study.

It is assumed in this study that the consumer's positive and negative perception of orange juice can be expressed by his/her choice of orange juice as the answer for the image questions. These relationships are expressed in the indicator equations (equation 2). Note that when the respondent chose orange juice as the top-of-mind answer to an image question (e.g., what beverage is good for breakfast?), it signified a positive perception toward orange juice and it is represented by "1" in a dichotomous variable. As shown in Table 1, more than half of the respondents considered orange juice was good for breakfast, about a third of them considered orange juice was the most natural beverage and good for their health, and less than six percent of the respondents considered orange juice as their top-of-mind overall choice beverage.

As discussed previously, the consumer's perception of orange juice (η) is influenced by his/her socioeconomic background (see equation (1)). Seven socioeconomic variables (presence of children under 18

[2] The response to the aided question "Do you think orange juice is good for your health? " represents the consumer's general knowledge of orange juice instead of his/her perception of orange juice. Using the same model and aided-question data gave very confusing results.

years old, respondent had college education, respondent was employed, respondent's sex and race, respondent resided in metropolitan areas, and household size) were used in the analysis. In addition, two dummy variables representing whether respondents' recall seeing an orange juice advertisement and had bought orange juice during the last 30 days prior to the survey were also included in the cause variables for perception. Statistics in Table 1 show that about 80 percent of the respondents were white females who had bought orange juice during the last 30 days before the survey. Other statistics indicated that about half of the respondents were employed and had a college education.

The x and y vectors in (1) and (2) have dimensions of (9×1) and (7×1), respectively. Since η is not measurable, a normalization rule has to be used to make it identifiable. In this case $\Lambda_1 = 1$ is used, i.e., the slope coefficient for the image question of "orange juice is refreshing" is set to one.

Equations (7) and (8) were used to obtain estimates of the tetrachoric and biserial correlation coefficients (S or $\hat{\rho}$) and their covariance matrix (M). The problem with this particular data set is that most of the cause variables are dichotomous in nature. Some of them, such as "college education," "employed," "metropolitan residents" and "bought orange juice," can be argued to have underlying continuous latent variables. These continuous latent variables can be defined as "propensity for education," "propensity for employment," etc. This makes tetrachoric correlation a more suitable estimate of association among elements of the indicators. On the other hand, cause variables such as "sex" and "race," although dichotomous in nature, have no underlying continuous latent variables. Therefore, they are treated like continuous variables, and biserial correlation coefficients are estimated. The fitting function (9) is then minimized by a numeric algorithm with respect to the parameters $\hat{\theta}$.

Results

The results of the MIMIC model are presented in Table 2. In general, most estimates are significantly different from zero. The coefficient estimates for the cause equation (1) indicate that respondents who were white and had a college education were more likely to have a less favorable perception toward orange juice (i.e., indicated by mentioning beverages other than orange juice to the seven perception questions). Results also indicate that respondents who were female, have children under 18, resided in metropolitan areas, recalled orange juice advertisements, and had purchased orange juice during the last 30 days before the survey would likely have a positive perception toward orange juice. In addition, the positive perception toward orange juice is positively related to household size. Results also indicate that respondents who recalled seeing an orange juice advertisement in the last 30 days prior to the survey were more likely to have a positive perception.

The coefficient estimates for equation (2) indicate that, if a consumer had a positive perception toward orange juice, he/she would likely to choose orange juice as the choice for "good for anytime," "most natural," "a good value," "good for your health," and "good for breakfast". The negative slope coefficient estimate for the indicator variable "comes to top of mind" is unexpected (it is not significantly different from zero). A close examination reveals that this unexpected result may be caused by the low number of positive responses to "comes to top of mind" (only 5.79% of the respondents chose orange juice). This view is corroborated by the tetrachoric correlation coefficient estimates for all indicator variables in Table 3, where the tetrachoric correlation between the indicator "comes to top of mind" and the other indicators are positive, but of least magnitude. This result suggests that this variable is a weak but not contradictory indicator.

Since the coefficient for the indicator "refreshing" (λ_1) is set to one for identification purposes, the results presented in Table 2 do not provide any information about whether "refreshing" is associated with the consumer's positive perception toward orange juice, or whether it is a good (significant) indicator for the latent perception variable.

Concluding Remarks

The MIMIC specification of consumers' perception of orange juice provides insight into the characteristics of perception. This model shows what variables are good indicators of respondent's perception toward orange juice and which demographic and socioeconomic variables are important determinants of a positive perception toward orange juice.

The results of this study indicate that a respondent's recall of advertising had a positive impact on his/her perception of the advertised goods. Results also suggest that advertising themes describing orange juice as "a good value," "good for breakfast" and "a healthy" beverage are important to the consumer's positive perception toward orange juice.

The results of this study also suggest that white, higher educated respondents are less likely to have a favorable perception toward orange juice. If the purpose of future orange-juice advertising is to expand the number of orange juice consumers, it should focus on this group of consumers, and advertising messages should be adjusted to have more appeal to them.

Table 1. Sample means and standard deviations of data.

Variable	Mean	Standard Deviation
	Cause Variables	
Children under 18	.4247	.4945
College Education	.5607	.4965
Employed	.5690	.4954
Sex (Female = 1)	.7803	.4142
Race (White = 1)	.8013	.3992
Metropolitan Resident	.4205	.4938
Purchase of Orange Juice	.8075	.3944
Household Size	2.8400	1.4800
Recalled Seeing OJ Ad.[a]	.3152	.4648
	Indicator Variables	
Refreshing	.0990	.2988
Top of Mind	.0579	.2336
Any Time of Day	.1318	.3384
Most Natural	.3180	.4659
Food for Health	.2706	.4444
A Good Value	.1806	.3847
Good for Breakfast	.5823	.4934

[a]The definition of "recall" is different from that used in marketing literature. It merely represents the awareness of having previously experienced the advertising stimuli. It is more like the term "recognition" in that part of the literature.

Table 2. MIMIC model with dichotomous indicators (transformed).

Variable	Estimate	Standard Deviation
	Cause Variable Coefficients (Γ)	
Children under 18	.2921	.1938
College Education	-.7690	.0374
Employed	.0458	.0438
Sex (Female = 1)	.6639	.4117
Race (White = 1)	-.5809	.4301
Metropolitan Resident	.2868	.1190
Purchase of Orange Juice	1.6248	.7274
Household Size	.4237	.2712
Recalled OJ Ads.	.5648	.2349
	Indicator Variable Coefficients (Λ)	
Refreshing	1.0000	
Top of Mind	-1.5717	1.7746
Any Time of Day	.8518	.5531
Most Natural	.9219	.7145
Good for Health	.9231	.4539
A Good Value	.9311	.3716
Good for Breakfast	.9240	.4481
	Variables of Indicator Equation Residuals (Θ^{ϵ})	
Refreshing	1.1086	.3981
Top of Mind	.7494	.4002
Any Time of Day	.4139	.1769
Most Natural	.7308	.3451
Good for Health	.4579	.2938
A Good Value	.0458	.0212
Good for Breakfast	.2336	.1218
Ψ	26.5222	12.4384

Table 3. Tetrachoric correlations of indicators

Variables	Refreshing	Top of Mind	Any Time of Day	Most Natural	Good for Health	A Good Value	Good for Breakfast
Refreshing	1.0000	.7723	.8049	.7432	.9948	.9800	.9364
Top of Mind		1.0000	.7310	.8890	.6641	.5688	.4626
Any Time of Day			1.0000	.9220	.9103	.9592	.8372
Most Natural				1.0000	.9914	.8634	.7699
Good for Health					1.0000	.9990	.9911
A Good Value						1.0000	.9865
Good for Breakfast							1.0000

References

Bogart, L. "Consumer and Advertising Research," in *Handbook of Communication*, I. de Sola Pool et al. (eds.), Rand McNally College Publishing Company, Chicago, 1973.

Bollen, K. *Structural Equations With Latent Variables*. John Wiley & Sons, New York, 1989.

Fuller, W. *Measurement Error Models*. John Wiley & Sons, New York, 1987.

Gao, X. and J. Y. Lee. "Consumer Perceptions About Orange Juice" Mimeo. University of Florida. 1991.

Jöreskog, K. G. and A. S. Goldberger. "Estimation of a Model With Multiple Indicators and Multiple Causes of a Single Latent Variable." *Journal of American Statistical Association*. 79(1975):631-9.

Kinnucan, H. and M. Venkateswaran. "Effects of Generic Advertising on Perceptions and Behavior: The Case of Catfish." *Southern Journal of Agricultural Economics*, 23(1990):137-51.

Kotz, S. and N. Johnson (Chief Editors). *Encyclopedia of Statistical Sciences*. John Wiley & Sons, New York, 1981.

Maddala, G. S. and R. P. Trost. "Some Extensions of the Nerlove-Press Model," *Proceedings of the American Statistical Association (Business and Economic Statistics Section)*. 1980: 481-5.

Maddala, G. S. and R. P. Trost. "Alternative Formulation of Nerlove-Press Models." *Journal of Econometrics*. 49(1981):35-49.

Muthén, B. "A General Structural Equation Model With Dichotomous, Ordered Categoric and Continuous Latent Variable Indicators." *Psychometrika* 49(1984):115-32.

Muthén, B. "Latent Variable Structural Equation Modeling With Categorical Data." *Journal of Econometrics*, 22(1983):43-65.

Muthén, B. and A. Christoffersson. "Simultaneous Factor Analysis of Dichotomous Variables in Several Groups." *Psychometrika*. 46(1981):407-19.

Muthén, B. and G. Speckart. "Latent Variable Probit ANCOVA: Treatment Effects in the California Civil Addict Program." *British Journal of Mathematical and Statistical Psychology*. 38(1985):161-70.

Ray, M. L. "Marketing Communication and the Hierarchy of Effects," in *New Models for Mass Communication Research*, p. Clark (ed.), Sage Publications, Beverly Hill, 1973.

Zeitlin, D. M. and R. A. Westwood. "Measuring Emotional Response," *Journal of Advertising Research*, 26(5): 34-6.