# Double-hurdle Model with Bivariate Normal Errors: An Application to U.S. Rice Demand

**X. M. Gao, Eric J. Wailes and Gail L. Cramer***

*Abstract*

Per capita rice consumption in the U.S. has doubled over the past decade. The effects of social and demographic variables on the household's rice consumption decisions are analyzed along with income and price variables. A double-hurdle model is used to solve simultaneously the consumer decisions whether to purchase rice and how much. The joint decision hypothesis is tested and accepted. The non-normal distribution of error terms may be responsible for possible bias in the empirical test of the joint decision hypothesis. The hyperbolic sine transformation is used to correct the problem in this study prior to testing the joint decision hypothesis.

**Key Words: double hurdle, non-normal error distribution, rice demand**

## Introduction

Rice consumption in the United States has increased dramatically over the last decade. Total rice consumption in the United States has doubled, from 26.9 million cwt. in 1978/1979, to 54 million in 1990/1991. Annual average per capita consumption changed from 8.1 pounds for the 1978/1979 period to 15.5 pounds for the 1988/1990 period (Putnam and Allshouse). Increased rice consumption is part of an overall change in United States consumer's diets towards more complex carbohydrates, vegetables, and poultry, while reducing consumption of eggs, dairy products, and red meats (Senauer et al.). Consumption of grain products in general, including pasta, wheat, and rice products has increased from 143 pounds per capita in 1978/1980 to 171 pounds in 1988/1990. Between these two periods, the share of rice in U.S. grain product consumption increased from 5.6 percent to 9 percent. It is important to public and private policy makers to understand the reasons for this dramatic change, for which a new set of demand parameter estimates are needed.

Many factors appear to have led consumers to eat more rice. In addition to traditional demand determination factors such as changing relative prices and rising incomes, researchers have identified other factors such as demographic shifts, the growing popularity of ethnic foods, recommendations by health groups to increase consumption of complex carbohydrates, and aggressive advertising (Bunch and Wendland; Senauer et al.). If this observation is true, several factors point to continued expanding consumption of rice in the United States during the rest of the 1990's. These factors include: fast growing Asian-American and Hispanic-American populations improved health awareness among consumers, greater convenience in preparing rice, etc. However, there has not been conclusive statistical

*The authors are senior econometrician; TRS Risk Management, American Express, Phoenix, AZ 84021, professor and L.C. Carter professor respectively, Department of Agricultural Economics and Rural Sociology, University of Arkansas, Fayetteville, AR 72701.

*J. Agr. and Applied Econ.* 27 (2), December, 1995: 363-376

analysis, especially at the micro level, on the effects of these factors on rice consumption. This paper is intended to fill this gap of knowledge by estimating a new set of relationships affecting rice demand using the most recent U.S. household consumption survey data. This survey, for the first time, includes variables on consumer attitudes on food and health relationships and provides an unique opportunity to study the consumer health perception of rice, which has been the industry's major promotional theme over the last decade. Health concern is a cause for structural (taste) change. We intend to shed some light on the issue of taste change also in this research.

**Objectives**

The objectives of this article are twofold. The first is to present estimates of the demand structure for rice products. These estimates are derived from a microdata set which permits the measurement of demographic as well as price and income effects. Cross-sectional data are used to overcome the inherent problems of time-series data (Blundell et al.; Orcutt). We are especially interested in finding out if the price elasticity for rice is positive, as reported in a recent time series study (Huang, 1993). The second objective is to test a postulate that model misspecification, in particular the problem of non-normality, is responsible for rejection of the joint decision hypothesis in bivariate choice models.

When cross-sectional data are used for demand analysis, it is desirable to use limited dependent variable models to analyze censored data because some households may report zero purchases or demand during a short survey period. The Tobit model was widely used in early studies for this purpose, which treats all the zero observations as corner solutions and assumes all households consume the product. In more recent studies, various improvements of the Tobit model have been developed, modified, and applied for different problems (e.g. Cragg; Deaton and Irish; Blundell and Meghir; Jones, 1989, 1992). These bivariate decision models have also gained widespread applications in the food demand literature. A basic property of these models is that they model a consumer's zero value of purchase as a decision result. The Tobit assumption of equivalence between zero demand and a corner solution is relaxed. The double-hurdle model and the infrequency purchase model are the most frequently used models with this property.

When these bivariate models are applied in demand analysis, the decision to buy and the decisions of how much to buy depend on different sets of exogenous variables. These decisions can be modeled jointly, if consumers decide whether and how much to buy simultaneously. They can also be modeled sequentially, where the decision on whether to purchase will affect how much to purchase, but not vice versa (in some circumstances, the decision sequence can also be turned around when the second decision affects the first). Special forms of sequential models are often called dominance models when the sequential decisions are independent of each other (Jones, 1989). Of course, the two decisions can also be modeled separately in general terms when consumer decisions on whether and how much to purchase are independent of each other.

Food expenditures generally account for a small percentage of consumer budgets in the United States (11.5 percent of disposable per capita income, see Putnam and Allshouse). As argued by Ray, most food commodity demands, because of small budget shares, do not warrant careful consideration or perception formation by consumers before they purchase. This implies that consumers generally make their decisions on whether and how much to buy simultaneously. If Ray is correct, the joint decision market participation double-hurdle model should be the appropriate model to use in food demand analysis. However, a puzzling feature of previous food demand studies is that almost all reject the joint decision hypothesis and accept an independent decision model of some form (e.g. Blundell and Meghir; Jones, 1992; Gould).

In this paper, we propose a hypothesis to explain this empirical result. The reason for joint decision rejection may be the model specification, particularly, the assumed non-normal distribution of the error terms. Maximum likelihood (ML) estimation will give inconsistent estimates of the model parameters when the normality assumption is violated, and specification tests will not be reliable. An inverse hyperbolic sine transformation is used in this study to transform the error terms in a double hurdle model. Then a test of the joint decision hypothesis is made using rice demand as an example. The result supports our postulate that non-normality seriously affects the test power and biases the empirical test towards a separate decision hypothesis.

*Double-hurdle Model with Dependence*

The double-hurdle model is designed to deal with survey data which has many zero observations on a continuous dependent variable. It generalizes the Tobit model by allowing for a separate first hurdle which represents a consumer's decision to purchase, and a second hurdle which represents the consumer's decision about how much to purchase. A purchase is realized only after both hurdles are cleared. The two decisions can be modeled as dependent on or independent of each other. In fact, the two decisions also have been modeled as sequential (Lee and Maddala), but most studies treat the decisions as separate.

When $Y_h$ is the observed purchase and $Y_h^*$ represents the latent consumption level for household h, the double-hurdle model is represented as:

$$
\begin{aligned}
D_h &= Z_h\theta + \omega_h, \\
Y_h^* &= X_h\beta + \varepsilon_h, \\
Y_h &= Y_h^* \quad \textit{if } \{Z_h\theta + \omega_h > 0 \textit{ and } X_h\beta + \varepsilon_h > 0\}; \\
Y_h &= 0 \quad \textit{if } \{Z_h\theta + \omega_h \le 0 \textit{ and } X_h\beta + \varepsilon_h > 0\}, \\
&\quad \textit{or } \{Z_h\theta + \omega_h > 0 \textit{ and } X_h\beta + \varepsilon_h \le 0\}, \\
&\quad \textit{or } \{Z_h\theta + \omega_h \le 0 \textit{ and } X_h\beta + \varepsilon_h \le 0\};
\end{aligned}
\tag{1}
$$

where, $(\omega, e) \sim BVN(0,\Sigma)$, $\Sigma = \begin{bmatrix} 1 & \sigma\rho \\ \sigma\rho & \sigma^2 \end{bmatrix}$,
and $Z$ and $X$ are the sets of variables that enter the first and second hurdles. $D_h$ is a latent variable describing the first hurdle decision to purchase. Various modifications of this model have been used by Cragg, Blundell and Meghir, and Jones (1989, 1992) in a wide number of applications such as purchase frequency, cigarette demand, and labor supply. The sample likelihood function for this model can be specified as:

$$
L = \prod_0 1 - F(\omega > -Z\theta, e > -x\beta)
\prod_+ F(\omega > -z\theta, e > -x\beta) f(e \mid \omega > -z\theta, e > -x\beta),
\tag{2}
$$

where f(.) and F(.) represent density and cumulative distribution functions respectively. Since the error terms in the two hurdles have bivariate normal distributions, $F(Z_h\theta + \omega_h > 0, X_h\beta + e_h > 0)$, they can also be denoted as $\Phi(Z\theta, X\beta/\sigma, \rho)$. The marginal density in the last term of (2) can be simplified as:

$$
f(e \mid \omega > -Z\theta, e > -x\beta) = \frac{\int_{-Z\theta}^{\infty} f(e) f(\omega \mid e) d\omega}{\int_{-Z\theta}^{\infty}\int_{-X\beta}^{\infty} f(\omega, e) d\omega de} = \frac{f(e)\int_{-Z\theta}^{\infty} f(\omega \mid e) d\omega}{\Phi(Z\theta, X\beta/\sigma, \rho)}.
\tag{3}
$$

Using the well-known fact that the conditional distribution of $D_h$ given $Y_h = Y_h^*$ is normal with mean $Z_h\theta + \rho\sigma^{-1}(Y_h - X_h\beta)$ and variance $1-\rho^2$, (3) can be simplified as:

$$
f(e \mid \omega > -Z\theta, e > -X\beta) = \frac{\Phi\left((Z\theta + \frac{\rho}{\sigma}(Y - X\beta))/\sqrt{1-\rho^2}\right)\frac{1}{\sigma}\phi((Y - X\beta)/\sigma)}{\Phi(Z\theta, X\beta/\sigma, \rho)}.
\tag{4}
$$

Substituting (4) into (2), the sample likelihood function becomes:

$$
L = \prod_0 \left[1 - \Phi(Z\theta, X\beta/\sigma, \rho)\right] \prod_+
\left[\Phi\left((Z\theta + \frac{\rho}{\sigma}(Y - X\beta))/\sqrt{1-\rho^2}\right)\frac{1}{\sigma}\phi((Y - X\beta)/\sigma)\right]
\tag{5}
$$

It is straightforward to show that when $\rho=0$, equation (5) gives the likelihood function for the Cragg model; further when $\theta=\beta/\sigma$, we have the Tobit model. The Amemiya type II Tobit model is an approximation of equation (5). The optimization of (5) can be done using the GAUSS package, which provides a standard bivariate normal cumulative distribution function procedure.

*Correction for Non-normal Errors*

The hypothesis that whether and how much to purchase are independent decisions can be tested by a likelihood ratio statistic applied to equation (5). Blundell and Meghir also propose a Lagrange Multiplier (LM) scoring test statistic for this purpose, although the estimation algorithm they propose can not be applied to estimate joint decisions in non-custom statistical softwares. Numerous studies have used the Blundell and Meghir bivariate models for consumer demand analyses and tests, but almost all rejected the joint decision hypothesis. Rejection of the joint decision model did not require them to apply a bivariate joint decision model as presented in the previous section. Least-squares estimation of the linear demand model using all observations will generate consistent parameter estimates under independence (Keen). However, the LM scoring test of the joint decision hypothesis may not be robust, and in fact may be misleading. The LM scoring test is based on restricted maximum likelihood estimates, in which a normal distribution is an implicitly maintained hypothesis. The distribution assumption has a direct bearing on the test statistics[1]. If the latter scenario is true, future researchers should not circumvent the joint decision estimation problem by testing and rejecting the hypothesis using the LM scoring test.

If misspecification is the problem, we can immediately identify two potential errors. The first is the non-normal distribution. The derivation of equation (5) depends on the assumption of bivariate normal distributions. This may not be true for the error term $e_h$ since the observations are obviously truncated. There are two approaches to this problem: (1) transforming the data or (2) imposing a different distribution. Previous research has used a Box-Cox transformation to correct for the non-normality problem, where a logarithmic transformation is only a special case. Maddala has shown that this method of transformation is not correct because the Box-Cox transformation is not defined when the latent variable is not positive. In light of Maddala's findings, this paper uses a hyperbolic sine transformation, which is continuously defined over positive, negative, and zero values. The univariate inverse hyperbolic sine transformation has the form (Burbidge et al.):

$$g(y_h,\tau) = \tau^{-1} Ln(\tau y_h + (\tau^2 y_h^2 + 1)^{\frac{1}{2}}) \tag{6}$$

where $\tau$ is a parameter that controls kurtosis. The transformation function is defined over all values of $\tau$. The transformation is symmetric about 0 in $\tau$, and transforms large values of $y_h$ in a logarithmic fashion.

Replacing y by $g(y,\tau)$ in equation (5) and applying a Jacobian transformation, the likelihood function using the transformed dependent variables is defined as:

$$L = \prod_0 \left[1 - \Phi(Z\theta, X\beta/\sigma, \rho)\right] \prod_+ \left[\Phi\left((Z\theta + \frac{\rho}{\sigma}(g(y,\tau) - X\beta))/\sqrt{1-\rho^2}\right) \frac{1}{\sigma}\phi\big((g(y,\tau) - X\beta)/\sigma\big)(1+y^2)^{-1}\right]. \tag{7}$$

The second misspecification error is heteroskedasticity. The presence of a heteroskedastic error would lead to inconsistent covariance matrix estimates, but its effect on the model specification test using a likelihood ratio method, such as the test pertaining to joint decision hypothesis, is ambiguous. To isolate the specification error of non-normality, the heteroskedasticity problem is corrected. Since most studies have corrected heteroskedasticity in limited dependent variable models by specifying a relationship between the standard error and some causal variables, we adopt the same technique (Maddala):

$$\sigma_h = w_h \gamma , \tag{8}$$

where $w_h$ is a subset of exogenous variable $x_h$, and $\gamma$ is a parameter vector. In this study, we use both the log income and household size as the exogenous variables.

## Data and Quality-Price Adjustment

Data for this study are taken from the U.S. Nationwide Food Consumption Survey (NFCS) 1987-1988. The survey collected food purchase, consumption and dietary data from a random sample of 4,495 households[2]. This study used a subsample for analysis, classified as "housekeeping households", defined as "at least one member of the household had ten or more meals from the household food supply during the week of observation." This sample size included 4,273 households. Data on food purchased in each household were collected through interviews with the person recognized as the most responsible for food planning and preparation. These data were collected during a 7-day period in personal, in-house interviews.

About 33 percent of the households reported rice purchases during the survey week. Households which did not buy any rice products during the survey week might have purchased inventories of rice prior to the survey period, or might have desired to purchase rice but may have experienced impediments to purchasing rice, or might not have had any desire to purchase or consume rice at all. Since rice is not a commodity in which every household must always have positive consumption, the infrequency of purchase model may not be an ideal model to evaluate this data. An empirical test comparing the double-hurdle model and the infrequency of purchase model, using the framework by Blisard and Blaylock, is presented in the results section.

The household characteristic variables evaluated in this model include household composition, region of residence, urbanization, season, income, household head gender, having pregnant family members or nursing children, occupation, employment, education, food stamp participation, and information on health concerns. Many of these household characteristic variables have been used and discussed in Haines et al., Heien and Wessells, and Gould. Their results can be used as the maintained hypotheses. Factors affecting preferences are age, ethnicity, area of residence, household size and structure, employment status, occupation, and education levels of household heads. Seasonal variables are included to account for possible differences in consumption by time of year. Variables of food stamp program participation and having nursing children in household are also used, as suggested by Heien and Wessells; the former represents a poverty level delimitation and the latter represents physiological factors affecting eating patterns. The descriptive statistics of the data are presented in table 1.

One variable, in particular, is worth special notice. Unlike the previous surveys, a new variable which reflects consumer concern for health is included in this current data. Respondents were asked to report whether they had received information on diet/health issues during the previous year from doctors/nutritionists or dieticians. It is hypothesized that if these households received such information (or recalled receiving such information), these households are more health-conscious than others. It is of interest, therefore, to determine if these households have distinct demand patterns for rice demand.

One important aspect of the household level demand data is the heterogenous rice products having different quality and value-added features. Consumers buy "goods" (in the sense of Cramer), which, for instance, are different rice products, not the composite "commodity" itself. There are a total of 13 goods in the data set for the rice commodity, including rice cakes, precooked rice, enriched rice, and so on. This makes the data set interesting for analysis because it requires that the aggregation errors, which are basically consumer choices for quality, have to be filtered first. Consumers choose the quality, as well as the quantity of the purchase, and the calculated price reflects this choice and therefore should be adjusted before demand analysis (Cowling and Raynor; Deaton).

The quality-price adjustment is done by estimating a hedonic price equation. The two-step independent modeling of quality and quantity decisions is justified by assuming that the household first determines its demand for commodity quality (through the selection of component goods) and then determines its quantity of composite commodity (Cox and Wohlgenant). The quality adjusted price of a composite purchase is defined as the difference between the calculated price (unit value) and the expected price, given its specific quality characteristics. The expected price is calculated by a hedonic price function:

$$P_{ih} = \alpha_i + \Sigma_j \beta_j x_{ijh} + e_{ih} \tag{9}$$

**Table 1.** Data Statistics

| Variable | Minimum | Maximum | Mean | Std. Errors |
|---|---|---|---|---|
| Rice purchase | 0 | 125.0000 | 1.1022 | 1.1703 |
| Log (income) | 1.6094 | 12.7077 | 9.8782 | 0.9170 |
| Log (price) | -0.0045 | 1.4352 | 0.6751 | 0.5224 |
| Number of household members between ages of | | | | |
| 1-18 | 0 | 10.0000 | 0.8712 | 1.1839 |
| 19-49 | 0 | 6.0000 | 1.2016 | 0.9858 |
| 50+ | 0 | 4.0000 | 0.7107 | 0.8555 |
| Regions of domicile: | | | | |
| Midwest | 0 | 1.0000 | 0.2640 | 0.4408 |
| South | 0 | 1.0000 | 0.3438 | 0.4750 |
| West | 0 | 1.0000 | 0.1884 | 0.3911 |
| Seasons: | | | | |
| Winter | 0 | 1.0000 | 0.2865 | 0.4522 |
| Spring | 0 | 1.0000 | 0.4050 | 0.4909 |
| Summer | 0 | 1.0000 | 0.1495 | 0.3566 |
| Food planner: female | 0 | 1.0000 | 0.7880 | 0.4087 |
| Having nursing children | 0 | 1.0000 | 0.0141 | 0.1183 |
| Consulted doctor/dietician during last year on health/nutrition issues | 0 | 1.0000 | 0.3546 | 0.4784 |
| Male head of household characteristics: | | | | |
| Race: Black | 0 | 1.0000 | 0.0570 | 0.2319 |
| Other | 0 | 1.0000 | 0.0256 | 0.1582 |
| Employed | 0 | 1.0000 | 0.5386 | 0.4985 |
| Education: less than high school | 0 | 1.0000 | 0.1615 | 0.3684 |
| partial college | 0 | 1.0000 | 0.1407 | 0.3477 |
| college | 0 | 1 0000 | 0.1776 | 0.3822 |
| Female head of household charactenstics. | | | | |
| Race: Black | 0 | 1 0000 | 0.1054 | 0.3072 |
| Other | 0 | 1 0000 | 0.0308 | 0.1728 |
| Employed | 0 | 1.0000 | 0.4366 | 0.4960 |
| Proportion of away-from-home food budget | 0 | 0.9600 | 0.2560 | 0.2097 |

where $x_{ijh}$ are variables affecting the consumer's choice of qualities, such as income and household characteristics as proxies for household preferences for unobservable quality characteristics. Regional and seasonal dummy variables are not included because they reflect systematic supply variations. Their average effects are reflected by the intercept $\alpha_i$. The quality adjusted price is then defined by:

$$P_{ih}^{*} = P_{ih} - \Sigma_j \hat{\beta}_j x_{ijh}. \tag{10}$$

The quality-price adjustment is particularly important when the commodities under study have large quality and price variations. If the goods are more or less homogeneous, the adjustment would render small differences. In the case of this study, considerable quality variation exists for rice. Therefore, the quality-price adjustment specified above is applied.

One problem in estimating the above model is missing observations on prices for households that did not consume rice. These prices are estimated by regressing observed prices for purchasing households on dummy and household characteristic variables such as region, season, and income. The properties of such a method are discussed in Gourieroux and Monfort.

**Static and Stochastic Elasticities**

The elasticities can be calculated in two ways. The fixed effect elasticities can be obtained using only the households with observed purchase. In this case, the effect of a change in an exogenous variable is calculated conditional on the vector $X$ and the specific value of $e_h$ observed for household $h$. With a fixed effect interpretation, assuming the household has a strictly positive purchase, the impact of $X_i$ is given by $\beta_i$. For the households with recorded purchase, the income and price elasticities are calculated with the actual values of purchase. For households with zero purchases, the marginal effects of $X_i$ have no effect, and the income and price elasticities are zero.

The second case is where there is a random component and the relevant variable for the household is the mean spending on condition of $X$. The marginal effects (elasticities) of exogenous variables in the case of the censored data model cannot be calculated straight forwardly, just based on observations above limits. The expected value of Y in the model is:

$$EY = X\beta \cdot F(Z\theta, X\beta/\sigma) + \Sigma \cdot f(Z\theta, X\beta/\sigma), \tag{11}$$

where F and f are standard bivariate normal cumulative distribution and density functions. The expected value of Y for observations above the limit, here called $Y^{**}$, is

$$EY^{**} = E(Y/Y>0\ D>0) = X\beta + \Sigma \cdot f(Z\theta, X\beta/\sigma)/F(Z\theta, X\beta/\sigma). \tag{12}$$

Consequently, the basic relationship between the expected value of all observation, EY, the expected value conditional upon being above the limit, $EY^{**}$, and the probability of being above the limit, $F(Z\theta, X\beta/\sigma)$, is:

$$EY = F(Z\theta, X\beta/\sigma)EY^{**}, \tag{13}$$

and the effect of a change in X on EY is simply $F(Z\theta, X\beta/\sigma)\beta_i$ (McDonald and Moffit). Assuming the coefficients for the (log) price and income terms are $\beta_i$ (i=2,3), the stochastic elasticities are

$$\gamma_i = \frac{\beta_i F(Z\theta_i, X\beta_i/\sigma)}{X\beta_i \cdot F(Z\theta_i, X\beta_i/\sigma) + \Sigma \cdot f(Z\theta_i, X\beta_i/\sigma)}. \tag{14}$$

**Empirical Results**

*Model Specification Tests*

The double-hurdle model is first tested against two other common censored models, the Tobit and infrequency of purchase model. The Tobit specification (not present) is overwhelmingly rejected by a likelihood ratio test. The test statistic is 189.3 with a critical chi-square distribution value of 45 for 26 degrees of freedom. A non-nested test of the double-hurdle model and infrequency of purchase model (not present), using the framework outlined by Blisard and Blaylock, also shows the former is preferable. The test statistic is 2.45, which has a standard normal distribution, with the null hypothesis of equivalence of the two models. The test statistics are calculated using a simplified version of double hurdle and infrequency models in which preferences are assumed independent to reduce the computational burden. The conclusion is that the market participation double-hurdle model is a more appropriate model to perform demand analysis. This result is not surprising since most of the rice is consumed in different product forms, such as rice cereal, rice cake, etc., rather than in plain rice. The problem of infrequency of purchase may exist, but it is not significant enough to justify using the infrequency model. The double-hurdle model also accounts for the infrequency issue - but interprets it as a market participation issue where rice is not part of a regular diet.

The White information matrix test, which is a composite test for homoskedastic and normal disturbances (Chesher) is applied. The Chesher score test interpretation of the information matrix test is used to calculate the test statistics. The procedure involves computing, for a sample of $H$ observations, $H$ times $R^2$ from the least squares regression of a column of ones on a matrix whose elements are functions of first and second derivatives of the log likelihood function. The test statistic constructed similarly as Anderson and Shonkwiler is 180.2, which is distributed asymptotically as Chi-square with 54 degrees of freedom. The null hypothesis of correct model specification was soundly rejected.

The standard double-hurdle model, equation (5), with the heteroskedasticity correction from the equation (8), is estimated, and the information matrix test still rejects the null hypothesis (the test statistic is 121.2 for $\chi^2_{54}$). However, the heteroskedasticity correction fails to correct all the misspecification problems. Therefore, the non-normality misspecification is suspected and the LM test for normality based on Bera et al. is performed by calculating a simple $HR^2$ formulation following Chesher[3]. The null hypothesis of normal distributions of the probability and demand equation errors are all rejected, with calculated statistics of 45.7 and 23.8 for $\chi^2_2$, respectively.

The likelihood ratio test is used for the joint decision test. Equation (5) with the heteroskedasticity correction has a log likelihood value -6437.2, while the same model with separate decision restriction ($\rho$=0) has a log likelihood value -6438.1. The independence assumption is accepted without much difficulty.

The inverse hyperbolic sine transformed double-hurdle model, equation (7), is estimated with the heteroskedasticity correction from equation (8). A White misspecification test shows that it is free from misspecification errors. The test statistic which has 54 degrees of freedom was calculated as 35.4, and the null hypothesis of correct model specification fails to be rejected. Equation (7) with the heteroskedasticity correction has a log likelihood value of -6327.5, while the same model with a separate decision restriction has a log likelihood value of -6388.3. Thus, the independence assumption is clearly rejected. At this point, the evidence is clear that non-normality errors have biased the test for separate decisions. When the non-normality is corrected, the independence assumption is rejected by a large margin. Based on this observation, the results from the joint decision double-hurdle model are used in the subsequent rice demand analysis.

*Results and Interpretation*

The parameter estimates for the purchase and demand decision variables are presented in table 2. The dependent variables are the quantity of rice demanded. The independent variables are the price, income, and household characteristic variables[4]. We used the same explanatory variables for both purchase and demand decisions. The decision of whether to purchase is embodied in $\theta$, and $\beta$ embodies the second decision of how much to purchase.

The price of rice appears to be an important factor in both decisions of whether to purchase ($\theta$) and the decision of how much to purchase ($\beta$). The associated coefficients were both negative and significant. Household income has a significant positive effect on the decision to purchase rice, although it does not have a significant effect on the quantity purchased. The interpretation of price and income effects is best discussed in terms of elasticities. Given that we are using individual household data, price and income elasticities may be calculated for each household. To provide representative elasticities for the sample as a whole, the estimates are weighted averages of the individual household elasticities, where the weights are each household's food expenditure shares. This procedure has the effect of attaching a greater weight to the response of households with relatively larger expenditures. Summing these weighted elasticities yields a representative elasticity corresponding to the predicted aggregate response of the sample as a whole. The estimated weighted average price elasticity is -0.11, and the income elasticity is 0.14.

The stochastic price and income elasticities are calculated for each household. The distribution of the price and income elasticities is given in Figures 1 and 2. The price elasticities are concentrated in the range between -0.05 to -0.2 and income elasticities between 0.05 to 0.2. Compared with the distributions from static models (not present), the stochastic elasticity distributions are more concentrated. The estimated weighted average stochastic price and income elasticities for rice are -0.15 and 0.16, respectively.

Comparisons of the results to those found by other authors are difficult, and typically inconclusive, since the models and data used are not similar. However, it still can be shown that the estimates presented in this paper are in the reasonable range compared with other studies.

**Table 2.** U.S. Rice Demand: Double-hurdle Model with Dependence and Hyperbolic Sine Transformation

| | First Decision (θ) | | Second Decision (B) | |
|---|---|---|---|---|
| Variables | Param. | Std. Error | Param. | Std. Error |
| Intercept | -1.9429 | 0.3122 | -6.7622 | 2.4680 |
| Log (price) | -0.0408 | 0 0128 | -0.1411 | 0.0177 |
| Log (income) | 0 0964 | 0.0330 | 0.1244 | 0.2232 |
| Number of household members | | | | |
| 1-18 | 0 0414 | 0.0328 | 0.3770 | 0 1769 |
| 19-49 | 0.1959 | 0.0365 | 0.8999 | 0 2216 |
| 50+ | 0.1036 | 0 0392 | 0.3753 | 0.2772 |
| Regions: | | | | |
| Midwest | -0 1244 | 0.0622 | -1 0786 | 0.4900 |
| South | 0.0843 | 0.0623 | 1.6486 | 0.4616 |
| West | 0.1625 | 0 0690 | 1.2997 | 0 5020 |
| Seasons: | | | | |
| Winter | 0.2087 | 0.0681 | 0.4781 | 0.5704 |
| Spring | 0.0377 | 0.0650 | 0 7664 | 0.5581 |
| Summer | 0 0274 | 0.0779 | 0.5109 | 0.7098 |
| Food planner: female | 0.1097 | 0.0558 | -0.5859 | 0 4399 |
| Having nursing children | 0.1092 | 0.1730 | -0 1068 | 1.1516 |
| Consulted doctor/dietician during last year | 0.0403 | 0.0058 | 0 3140 | 0.3373 |
| Male head of household characteristics | | | | |
| Race: Black | 0.3660 | 0.1159 | -0.0451 | 0.8029 |
| Other minorities | 0.6489 | 0.1713 | 2.1814 | 0.7045 |
| Employed | -0.0905 | 0.0571 | 0 0125 | 0.4331 |
| Education: less than high school | -0 2197 | 0 0699 | 1.0508 | 0.7705 |
| partial college | 0.0372 | 0 0689 | 0.9835 | 0 5649 |
| college | 0 2594 | 0 0649 | 0 9127 | 0 4675 |
| Female head of household characteristics | | | | |
| Race: Black | 0.4640 | 0.0897 | 1 3773 | 0.6987 |
| Other | 0 5950 | 0.1558 | 2.4948 | 0 6900 |
| Employed | -0 0273 | 0 0477 | -0 0259 | 0 3485 |
| Proportion of away-from-home food consumption | -0.6496 | 0 1203 | -0.8524 | 1.0425 |
| Error term std. Error ($\sigma_0$) | | | 1.0415 | 0 4795 |
| Correlation coefficient ($\rho$) | | | 0.2391 | 0 0837 |
| Kurtosis parameter ($\tau$) | | | 0.0147 | 0 0037 |
| Heteroskedastic transformation parameters: | | | | |
| Income ($\gamma_1$) | | | 0.0043 | 0 0017 |
| Household size ($\gamma_2$) | | | 0.0054 | 0 0015 |

Huang (1993), using aggregate time series data from 1953 to 1990 for 39 food commodities, estimated the own price elasticities for rice to be 0.06. Huang also found the expenditure elasticity for rice to be 0.15 and income elasticity to be 0.04. Using the same model, Huang's earlier estimates (1985) are -0.14 for price elasticity and -0.36 for expenditure elasticity. The sign reversal of the elasticity estimates by adding only a few more years of data reveals a problem with the use of time series data. The microeconometric evidence presented by Blundell et al. suggests that aggregate studies are likely to display both bias and instability in measuring price and income responses. Micro-analysis may also yield substantially greater precision in the estimation of the parameters than estimates based on aggregate data (Orcutt et al.). A still earlier study by George and King, using 1965 survey data, found the own price elasticity for rice is -0.32. Their price data was not adjusted for quality effects. Wailes et al. found the price elasticity for converted rice equivalent to be -0.58 and expenditure elasticity to be 0.98 (income elasticity is 0.12 if using income elasticity of grain estimates of 0.12 from Huang, 1993) using also the 1987-88 USDA NFCS data. The difference between the price elasticities is mainly from different quantity definitions.

Household characteristic variables are also significant in affecting rice demand. Since most of the household characteristic variables are binary

**Figure 1.** Stochastic Price Elasticity Estimates

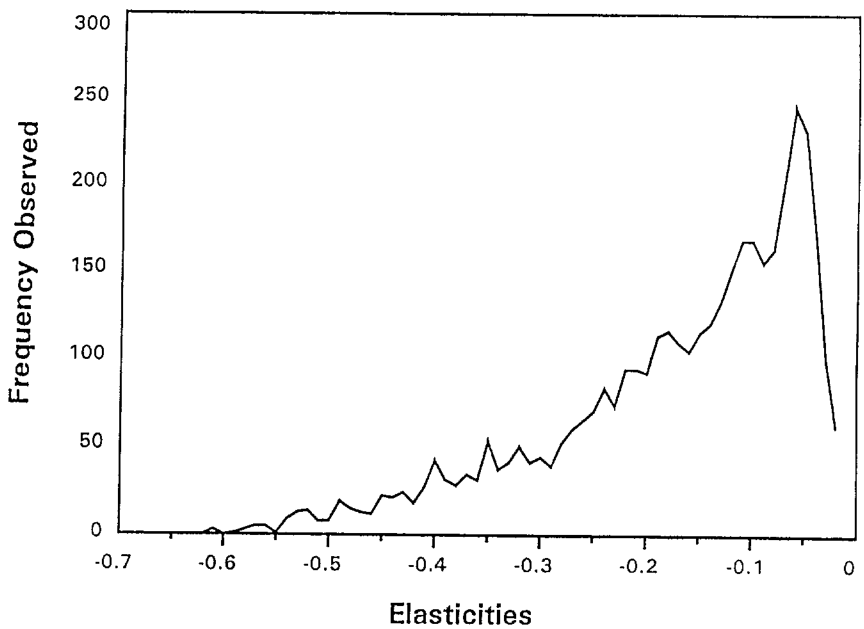
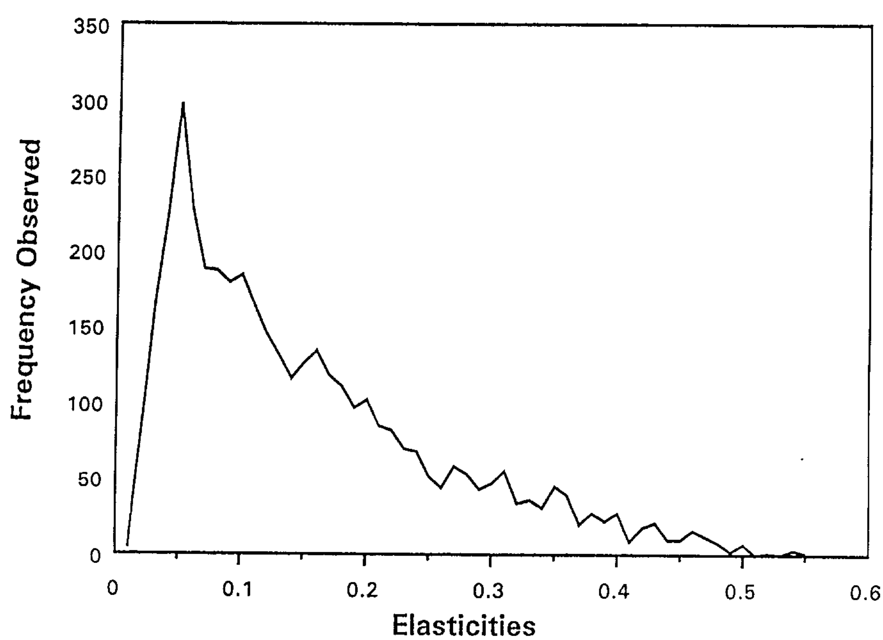



**Figure 2.** Stochastic Income Elasticity Estimates



variables, their elasticities are not calculated[5]. The number of household members between the age of 19 and 49 has a positive impact on the decisions of whether to purchase and how much to purchase. For the number of household members with ages over 50, the impact on the decision of whether to purchase is still positive, but the impact on the amount of purchase is not significant.

Consumers living in the south and west are more likely to purchase rice and tend to purchase more than residents in the east. Midwestern residents are less likely to purchase rice, and when they do, they purchase less.

The results also indicate that families who have consulted with doctors or dieticians about nutritional matters during the last year are significantly more likely to purchase rice. However, once the decision is made to purchase, the amount to purchase was not affected by this factor. The gender of the household meal planner has a significantly positive impact on whether to purchase but no impact on how much rice to purchase. With regard to race, families with black male household heads are more likely to decide to purchase rice. However, once the decision is made, they do not seem to purchase more than households with white household heads. When the male household head is an "other minority" (a category which mainly includes Asians and Hispanics), the decision to purchase and the quantity to purchase are both significantly positive. A similar observation is also found for female headed households; nonwhites are more likely to purchase rice, and when they do, they purchase more rice than whites.

Households headed by male college graduates are more likely to purchase rice, and when they do, they tend to purchase more than households headed by males with only a high school degree. People who have a higher percentage expenditure on away-from-home food consumption are less likely to purchase rice for home consumption. Once the decision is made to purchase rice, however, they do not tend to purchase less than households that spend less on away-from-home food consumption.

Among the household characteristic variables, season of year, having nursing children, and employment are not found to be significant in affecting consumer decisions. The impacts of the household characteristic variables are largely consistent with the results of Wailes et al. The causes of increasing rice consumption are attributed to income, social-demographic changes, and perhaps consumer taste shifts.

## Summary and Conclusions

A double-hurdle model is used to analyze the U.S. rice consumers' decisions on whether to purchase and how much. Households which consumed any rice during the survey week were studied in comparison with all surveyed households. The effects of social and demographic variables on the household's decision to purchase and the amount to purchase are analyzed with income and price variables. The decision to purchase and the amount to purchase are modeled as separate functions of prices, income, and household characteristic variables, however, the two decisions are tested to be jointly determined. The joint purchase decision hypothesis was tested and accepted in this study. The non-normal distribution may be responsible in previous studies for the rejection of the joint decision hypothesis. The inverse hyperbolic sine transformation is used to correct the problem of non-normal distribution of error terms. After the correction, the correlation of the two decisions becomes significant.

The household characteristic variables which are found to be significant in affecting consumer rice demand decisions are age structure, region of residence, health consciousness, and race. The result supports the argument by Bunch and Wendland that the changing population composition and rising health concerns are important influences on food demand. Unlike the positive price elasticity estimates from the time-series study by Huang, we still found the price effect to be negative but very inelastic.

The positive price elasticity for rice using time series data reported by Huang is because both the demand and relative price of rice have been increasing over the last few years. Since changes in socio-demographic effects are ignored in time series studies, the positive price elasticity estimate is believed to be a result of misspecification. If our estimate of price elasticity is correct, the main factors associated with increasing consumption may be those reflecting changes in taste. This finding has important implications for demand augmentation policy as advertising may be more effectively directed at the households which account for increasing rice demand in the United States. Although there is strong evidence of structural change in consumer demand for rice, the structural change hypothesis could not be formally tested from the cross-sectional data used in this study. A follow-up test on the taste change hypothesis is needed in a well-specified time series framework.

**References**


Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.

Anderson, R. and J. S. Shonkwiler. "Testing and Correcting for Distributional Misspecifications in the Tobit Model: An Application of the Information Matrix Test." *Empirical Econ.*, 16(1991):313-323.

Blaylock, J. R. and W. N. Blisard. "U.S. Cigarette Consumption: The Case of Low-Income Women." *Amer. J. Agr. Econ.*, 74(1992):698-705.

Blisard, W. N. and J. R. Blaylock. "Distinguishing Between Marketing Participation and Infrequency of Purchase Models of Butter Demand." *Amer. J. Agr. Econ.*, 75(1993):314-320.

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

Blundell, R., P. Pashardes, and G. Weber. "What Do We Learn About Consumer Demand Patterns from Micro Data?" *Amer. Econ. Rev.*, 83(1993): 570-597.

Bunch, K. and B. Wendland. "Grain Products Regain Popularity." *National Food Review.* Econ. Re. Ser., U.S. Dept Agr. NFR-34(1986):1-7.

Burbidge, J. B., L. Magee. and A.L. Robb. "Alternative Transformation to Handle Extreme Values of the Dependent Variable." *J. Amer. Stat. Assoc.*, 83(1983):123-127.

Chesher, A. "The Information Matrix Test." *Econ. Letters.* 13(1983):45-48.

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

Cramer, J. S. "Interaction of Income and Price in Consumer Demand." *Int. Econ. Rev.*, 14(1973):351-363.

Cox, T. L. and M. K. Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *Amer. J. Agr. Econ.*, 68(1986):908-919.

Cowling, R.G.D. and A. J. Raynor. "Price, Quality, and Market Shares." *J. Pol. Econ.*, 78(1970):1292-1309.

Deaton, A. "Quality, Quantity, and Spatial Variation of Price." *Amer. Econ. Rev.*, 78(1989):418-30.

Deaton A. and M. Irish. "Statistical Models for Zero Expenditures in Household Budgets." *J. Pub. Econ.*, 23(1984):59-80.

Gao, X. M., Eric J. Wailes and Gail L. Cramer. "A Synthetic Demand System: An Application to U.S. Consumer Demand for Rice and Selected Rice Substitutes." *Rev. Agr. Econ.* 16(1994):13-24.

Gao, X. M., Eric J. Wailes and Gail L. Cramer. "A Microeconometric Model Analysis of U.S. Consumer Demand for Alcoholic Beverages." *Applied Econ.* 27(1995):59-69.

General Accounting Office. *Nutrition Monitoring, Mismanagement of Nutrition Survey Has Resulted in Questionable Data.* RCED-91-117, July, 1991.

George, P. S. and G. A. King. *Consumer Demand for Food Commodities In the United States With Projections for 1990.* Giannini Foundation Monograph No. 26, Univ. of Cal. Davis, 1971.

Gourieroux, C. and Monfort. "On the Problem of Missing Data in Linear Models." *Rev. Econ. Stud.*, 48(1981):579-586.

Gould, B. W. "At Home Consumption of Cheese: A Purchase Infrequency Model." *Amer. J. Agr. Econ.*, 74(1992):453-459.

Haines, P., D. Guilkey and B. Popkin. "Modeling Food Consumption as a Two-Step Decision." *Amer. J. Agr. Econ.*, 70(1988):543-552.

Heien, D. and C.R. Wessells. "Demand System Estimation With Microdata: A Censored Regression Approach." *J. Bus. and Econ. Stat.*, 8(1990):365-371.

Huang, K. S. U.S. *Demand for Food: A Complete System of Price and Income Effects.* USDA technical Bulletin 1714, Washington D.C. 1985.

Huang, K. S. *A Complete System of U.S. Demand for Food.* USDA technical bulletin 1821, Washington D.C. 1993.

Jones, A. M. "A Double Hurdle Model of Cigarette Consumption." *J. App. Econometrics*, 4(1989):23-39.

Jones, A. M. "A Note on Computation of the Double Hurdle Model with Dependence with An Application to Tobacco Expenditure" *Bull. Econ. Res.*, 44(1992):67-73.

Keen, M. J. "Zero Expenditures and the estimation od Engel Curves." *J. App. Econometrics*, 1(1986):277-286.

Lee, L.F. and G.S. Maddala. "Sequential Selection Rules and Selectivity in Discrete Choice Econometric Models." mimeo, University of Florida, 1985.

Lutz, S. M., D. M. Smallwood, J. R. Blaylock, and M. Y. Hama. *Changes in Food Consumption and Expenditures in American Households During the 1980's.* USDA technical Bulletin 849, Washington D.C. 1992.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press, 1983.

McDonald, J. F. and R. A. Moffit. "The Uses of Tobit Analysis." *Rev. Econ. and Stat.*, 60(1980):318-321.

Orcutt, G. H., H. W. Watts and J. B. Edwards. "Data Aggregation and Information Loss." *Amer. Econ. Rev.*, 58(1968): 773-787.

Putnam, J.J. and J.E. Allshouse. *Food Consumption, Prices, and Expenditures, 1968-89.* Commodity Econ.

Div., Econ. Res. Ser., U.S. Dept. Agr. Statistical Bulletin No. 825, May, 1991.

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.

Senauer, B., E. Asp, and J. Kinsey. *Food Trends and the Changing Consumer.* Eagan Press, St. Paul, MN. 1991.

USDA (U.S. Department of Agriculture), Human Nutrition Information Service, "Evaluation of Nonresponse in the Nationwide Food Consumption Survey 1987-88." Washington: USDA, NFCS REP. NO. 87-M-2 (forthcoming).

Wailes, E., X. M. Gao and G. Cramer "U.S. Consumer Demand for Complex Carbohydrate Products: The Microdata Evidence." presented at the AAEA meetings, Orlando, Aug. 1993.

White, H. "A Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50(1982):1-25.

**Endnotes**

1. As observed by Blundell and Meghir, "unlike the infrequency of purchase model, the asymptotic properties of the LM estimator (of double hurdle models) require the error distribution to be correctly specified (p.186)."

2. A weakness of this data set is the relatively low survey response rate of 38 percent. Under-representation also occurred for a number of demographic groups including, central-city households, higher-income households and households with a female head. Greater detail on the survey design is found in two reports by the USDA and General Accounting Office.

3. The test is greatly simplified by basing it on an independence decision model, since the bivariate Pearson distribution is difficult to define. The result, however, is not affected by this simplification because if individual variables are not distributed univariate normal, they cannot be jointly bivariate normal.

4. The prices of other food and non-food commodities are assumed to be constant during the survey period, and their effects are filtered into the intercept term.

5. Household characteristic variables are discrete whereas elasticity is a continuous variable concept. It is not an easy exercise to calculate correct elasticities in this case because the underlying latent variable distributions of dummy variables have to be estimated. The coefficients for household characteristic variables can adequately represent their effects. Signs and significances of dummy variable coefficients are more important than magnitude in interpreting their effects.